

THE TRIANGULATION COMPLEXITY OF FIBRED 3-MANIFOLDS

MARC LACKENBY AND JESSICA S. PURCELL

ABSTRACT. The *triangulation complexity* of a closed orientable 3-manifold is the minimal number of tetrahedra in any triangulation of M . The main theorem of the paper gives upper and lower bounds on the triangulation complexity of any closed orientable hyperbolic 3-manifold that fibres over the circle. We show that the triangulation complexity of the manifold is equal to the translation length of the monodromy action on the mapping class group of the fibre S , up to a bounded factor, where the bound depends only on the genus of S .

1. INTRODUCTION

The *triangulation complexity* $\Delta(M)$ of a closed orientable 3-manifold M is the minimal number of tetrahedra in any triangulation of M . Despite its naive definition, it has some attractive properties. An obvious but important property is that only finitely many 3-manifolds have triangulation complexity less than a given number. It is also useful in normal surface theory, where it is a natural measure of complexity for M . However, like other invariants of manifolds that are defined as the minimum of some quantity, it is neither easy to compute nor of obvious topological significance. Indeed, precise values for $\Delta(M)$ are known for a few relatively small examples (see for example [19], [21]) and for a few infinite families [23, 25, 10, 13, 14, 11, 12].

It is the main goal of this paper to establish that triangulation complexity is a good invariant of 3-manifolds that relates to many other key topological and geometric quantities. Our focus will be on 3-manifolds M that fibre over the circle. We will relate $\Delta(M)$ to the geometry of the mapping class group of the fibre and to its Teichmüller space. As a consequence, the modern theory of mapping class groups can be applied to compute $\Delta(M)$ to within a bounded factor, where the bound depends on the topology of the fibre.

1.1. Translation length and analogous results. There is an obvious analogy between the triangulation complexity of a hyperbolic 3-manifold M and its volume. Indeed, there is a well-known inequality due to Gromov and Thurston [28], which states that the hyperbolic volume of M is at most $v_3\Delta(M)$, where v_3 is the volume of a regular hyperbolic ideal 3-simplex. A beautiful and important theorem of Brock [6] relates the hyperbolic geometry of a fibred hyperbolic 3-manifold with the Weil-Petersson geometry of Teichmüller space. Specifically, suppose that M fibres over the circle with fibre S and monodromy $\phi: S \rightarrow S$. The monodromy induces an action on the Teichmüller space of S that is an isometry with respect to both the Weil-Petersson and Teichmüller metrics.

Whenever one has an isometry h of a metric space (X, d) , one can consider its *translation length* $\ell_X(h)$, which is defined to be

$$\ell_X(h) = \inf\{d(h(x), x) : x \in X\}.$$

One can also define its *stable translation length* $\bar{\ell}_X(h)$ by

$$\bar{\ell}_X(h) = \inf\{d(x, h^N(x))/N : N \in \mathbb{Z}_{>0}\},$$

where $x \in X$ is chosen arbitrarily. This is independent of $x \in X$, the infimum is in fact a limit as $N \rightarrow \infty$, and it is at most the translation length; see [4, II.6.6]. We denote the translation length of the action of ϕ on the Teichmüller space of S with the Weil-Petersson metric by $\ell_{\text{WP}(S)}(\phi)$.

There are also many simplicial complexes associated with the surface S . Brock considers the *pants complex* $\mathcal{P}(S)$. This has a vertex for each collection of disjoint simple closed curves on S that

divide it into a union of pairs of pants, and two vertices are joined by an edge when the associated collections of curves are related by a simple type of move. One can assign a path metric to this complex by declaring that each edge has length one, and then the monodromy ϕ acts on it by an isometry. One can again therefore define its translation length $\ell_{\mathcal{P}(S)}(\phi)$. The following is Brock's theorem [6, Theorem 1.1].

Theorem 1.1 (Brock). *Let S be a compact orientable surface. Then the following quantities are all within bounded ratios of each other, where the bounds only depend on the Euler characteristic of S , for a pseudo-anosov homeomorphism ϕ of S :*

- (1) *the hyperbolic volume of the fibred manifold $(S \times I)/\phi$;*
- (2) *the Weil-Petersson translation length $\ell_{\text{WP}(S)}(\phi)$;*
- (3) *the translation length $\ell_{\mathcal{P}(S)}(\phi)$ in the pants complex of S ;*
- (4) *the stable translation length $\bar{\ell}_{\mathcal{P}(S)}(\phi)$.*

There are many other interesting invariants of hyperbolic 3-manifolds and many other complexes associated with a compact orientable surface. One is naturally led to ask whether any of these other quantities are related as in Brock's theorem. A theorem of Futer and Schleimer asserts that there is another relationship of this form [7]. This involves the *arc complex* $\mathcal{A}(S)$, which has a vertex for each isotopy class of properly embedded essential arc in S and where two vertices are joined by an edge if their associated arcs can be isotoped to be disjoint.

Theorem 1.2 (Futer–Schleimer). *Let S be a compact orientable surface with non-empty boundary. Then the following quantities are all within bounded ratios of each other, where the bounds only depend on the Euler characteristic of S , for a pseudo-anosov homeomorphism ϕ of S :*

- (1) *the volume of a maximal collection of cusps for the hyperbolic manifold $(S \times I)/\phi$;*
- (2) *the stable translation length $\bar{\ell}_{\mathcal{A}(S)}(\phi)$ in the arc complex of S .*

1.2. Main results. In this paper, we continue this theme by relating the triangulation complexity $\Delta(M)$ with the discrete geometry of the mapping class group of S and with the Teichmüller space of S . The following is our main theorem.

Theorem 1.3. *Let S be a closed orientable surface of genus at least two, and let $\phi: S \rightarrow S$ be a pseudo-anosov homeomorphism. Then the following quantities are within bounded ratios of each other, where the bounds only depend on the genus of S :*

- (1) *the triangulation complexity of $(S \times I)/\phi$;*
- (2) *the translation length of ϕ in the thick part of the Teichmüller space of S ;*
- (3) *the translation length of ϕ in the mapping class group of S ;*
- (4) *the stable translation length of ϕ in the mapping class group of S .*

We now explain these terms in a bit more detail. The *mapping class group* of S , denoted $\text{MCG}(S)$, is well known to be a finitely generated group. Once one picks a finite set of generators, it inherits a word metric. Different choices of finite generating sets give different metrics, but any two such metrics remain within a bounded ratio of each other. For the sake of being definite, we can pick a standard generating set, for example as in [18], which then determines the metric on $\text{MCG}(S)$.

The Teichmüller space of S can be viewed as the space of marked hyperbolic structures on S . Its *thick part* is the subset consisting of hyperbolic structures where every geodesic has length at least some $\epsilon > 0$. A suitable choice of ϵ must be made. One normally uses a version of the Margulis constant, so that the union of geodesics with length at most ϵ is a union of disjoint simple closed curves. We will choose ϵ to have the additional property that the thick part of Teichmüller space is path-connected. When Teichmüller space is given one of its usual metrics, say the Weil-Petersson metric or the Teichmüller metric, the thick part inherits a path metric. We use either of these metrics.

The mapping class group of S acts properly discontinuously and cocompactly on the thick part. Thus it is a standard consequence of the Švarc-Milnor lemma that $\text{MCG}(S)$ with its word metric and the thick part of Teichmüller space are quasi-isometric; see [4, Proposition 8.19]. Hence, for

any orientation-preserving homeomorphism ϕ of S , the translation lengths of ϕ on $\text{MCG}(S)$ and on the thick part of Teichmüller space are within a bounded ratio of each other. So, the relationship between the second and third quantities in Theorem 1.3 is easy and well known. The relationship between the third and fourth quantities is also probably well known; we give a proof in Section 3.

It is the relationship with the triangulation complexity of the fibred manifold $(S \times I)/\phi$ that is new. One of the main consequences of Theorem 1.3 is that it is now possible to compute the triangulation complexity of fibred 3-manifolds up to a multiplicative error. This is because the translation length in $\text{MCG}(S)$ is computable up to a bounded factor, by a theorem of Masur, Mosher and Schleimer [20]; see Section 3 for more details.

Analogous to his theorem about the volume of fibred 3-manifolds, Brock [5] also proved a result about geometrically finite hyperbolic structures on $S \times [0, 1]$. He showed that the volume of the convex core of such a 3-manifold is bounded above and below by linear functions of the Weil-Petersson distance between the points in Teichmüller space associated to $S \times \{0\}$ and $S \times \{1\}$. We also have a result in this spirit.

Theorem 1.4. *Let S be a closed orientable surface of genus at least two and let \mathcal{T}_0 and \mathcal{T}_1 be non-isotopic 1-vertex triangulations of S . Then the following are within a bounded ratio of each other, the bound only depending on the genus of S :*

- (1) *the minimal number of tetrahedra in any triangulation \mathcal{T} of $S \times [0, 1]$ such that the restriction of \mathcal{T} to $S \times \{0\}$ equals \mathcal{T}_0 , and the restriction of \mathcal{T} to $S \times \{1\}$ equals \mathcal{T}_1 ;*
- (2) *the minimal number of 2-2 Pachner moves relating \mathcal{T}_0 and \mathcal{T}_1 .*

In fact, a version of this theorem is the main technical result of the paper, and forms the core of the proof of Theorem 1.3.

1.3. Applications. Implicit in the statement of Theorem 1.3 is that for a pseudo-anosov homeomorphism $\phi: S \rightarrow S$, its translation length $\ell_{\text{MCG}(S)}(\phi)$ and its stable translation length $\bar{\ell}_{\text{MCG}(S)}(\phi)$ lie within a bounded ratio of each other, where the bound depends only on the genus of S . This is in fact a rapid consequence of the theorem of Masur, Mosher and Schleimer mentioned above [20], which interprets translation length in terms of splitting sequences for train tracks. However, a consequence of Theorem 1.3 is that the triangulation complexity of a fibred 3-manifold and its ‘stable’ triangulation complexity (when this term is defined in the obvious way using powers of the monodromy) lie within a bounded ratio of each other. Specifically, we have the following result.

Corollary 1.5. *Let S be a closed orientable surface. Then there is a constant $k > 0$ (depending only on the genus of S) such that, for any pseudo-anosov homeomorphism ϕ of S and any positive integer N ,*

$$k \Delta((S \times I)/\phi) \leq \frac{\Delta((S \times I)/\phi^N)}{N} \leq \Delta((S \times I)/\phi).$$

Our methods can also determine the triangulation complexity of lens spaces. These have also been extensively studied by Jaco, Rubinstein and Tillmann [10]. They conjectured that the triangulation complexity of the lens space $L(p, q)$ can be computed in terms of the continued fraction expansion of p/q . Using Theorem 1.4, we confirm that their conjecture is true, up to a bounded multiplicative error. This will be appear in a forthcoming paper [17].

Theorem 1.6. *Let $L(p, q)$ be a lens space, where p and q are coprime integers satisfying $0 < q < p$. Let $[a_1, \dots, a_n]$ be the continued fraction expansions of p/q where each $a_i > 0$. Then there is a universal constant $k > 0$ such that*

$$k \sum a_i \leq \Delta(L(p, q)) \leq \sum a_i.$$

1.4. Further work. We conjecture that the main result of this paper, Theorem 1.3, should also be true for compact orientable fibred 3-manifolds with non-empty boundary. It seems likely that the methods introduced in this paper might lead to a proof of this. However, the generalisation does not seem to be immediate. One might be tempted to double the bounded manifold to obtain a closed one, but the resulting monodromy is not pseudo-anosov. Alternatively, one might attach solid tori to the boundary to extend the manifold to a closed one, but again the resulting monodromy need not be pseudo-anosov and even if it is, then its translation length in the mapping class group of the closed surface may be much less than the original translation length in the bounded surface. One might alternatively try to adapt the proof of the main theorem. However, various aspects of the argument require the fibre to be closed. Nevertheless, this seems to be a promising area for further research, which would have some attractive applications.

We also believe that the main theorem should hold for arbitrary orientation-preserving homeomorphisms $\phi: S \rightarrow S$ (other than those that are isotopic to the identity), not just pseudo-anosov ones. One might attempt to prove such a result by cutting the surface along some disjoint essential simple closed curves C into pieces, where on each piece ϕ is either the identity or pseudo-anosov. This is possible after possibly passing to a power of the monodromy. These curves in S then determine a collection of disjoint incompressible tori in the fibred manifold M , and if we cut along these tori, then we obtain a lower bound on the triangulation complexity of each of the pieces, using the conjectured version of our main theorem in the case with non-empty boundary. However, this is not enough to obtain the correct lower bound on the triangulation complexity of the original manifold M . This is because one loses track of possible Dehn twists along the curves C in S . Thus, to prove the main theorem for general homeomorphisms ϕ , one either needs to work with a version of the mapping class group for bounded manifolds, where isotopies of the boundary are not allowed, or one needs to adapt the techniques of this paper to deal with homeomorphisms that may have invariant multi-curves.

Another useful direction that one might take is to consider manifolds M that are not given as a fibration, but that are given using a Heegaard splitting. Specifically, one can fix a Heegaard surface S that separates M into two handlebodies V and W . In this paper, we consider the *spine graph* $\text{Sp}(S)$, which has a vertex for each spine of S up to isotopy and where two vertices are joined by an edge if and only if the corresponding spines are related by the contraction or expansion of an edge; see Definitions 2.1 and 2.5 for more details. Associated to each of the handlebodies V and W , there are subsets \mathcal{D}_V and \mathcal{D}_W of $\text{Sp}(S)$, which we call *disc subsets* defined as follows. We say that \mathcal{D}_V consists of those spines Γ in S with the property that some subset of Γ forms the boundary of a union of disjoint properly embedded discs D in V , such that $V \setminus D$ is a ball. We define \mathcal{D}_W similarly.

It is reasonable to make the following conjecture.

Conjecture 1.7. *Let M be a closed orientable manifold with a Heegaard surface S that divides M into handlebodies V and W . Then the complexity $\Delta(M)$ is bounded above and below by linear functions of the distance in $\text{Sp}(S)$ between the disc subsets \mathcal{D}_V and \mathcal{D}_W .*

Our result about lens spaces, Theorem 1.6, is a confirmation of this conjecture in the case where S is a torus. Given that the techniques of this paper rely so heavily on almost normal surface theory, which was specifically designed to deal with Heegaard splittings, it is tempting to think that this conjecture may be within reach.

1.5. Outline of proof. The first step is to transfer the problem from considering distances in the mapping class group to considering spines of surfaces, which are more geometric and easier to work with. In Section 2, we introduce the *spine graph* $\text{Sp}(S)$ for the surface S mentioned above. We show that the spine graph $\text{Sp}(S)$ is quasi-isometric to $\text{MCG}(S)$. So, instead of considering $\ell_{\text{MCG}(S)}(\phi)$ and $\bar{\ell}_{\text{MCG}(S)}(\phi)$ in Theorem 1.3, we can consider $\ell_{\text{Sp}(S)}(\phi)$ and $\bar{\ell}_{\text{Sp}(S)}(\phi)$. The central part of the proof is to show that these quantities are within a bounded factor of $\Delta(M)$, where $M = (S \times [0, 1])/\phi$.

It is fairly straightforward to bound $\Delta(M)$ linearly above in terms of $\ell_{\text{Sp}(S)}(\phi)$. We achieve this in Section 2 by building a triangulation of M from a path in $\text{Sp}(S)$ joining a spine to its image under ϕ . The main work is in showing that $\ell_{\text{Sp}(S)}(\phi)$ is bounded above by a linear function of $\Delta(M)$.

The broad idea is as follows. Start with a triangulation \mathcal{T} for M where the number of tetrahedra $\Delta(\mathcal{T})$ equals $\Delta(M)$. Put a fibre S in normal form, taking one with a minimal *weight*, defined in Section 4. When we cut along S we obtain a copy of $S \times [0, 1]$ that inherits a handle structure \mathcal{H} from the triangulation \mathcal{T} . The surfaces $S \times \{0\}$ and $S \times \{1\}$ inherit handle structures, where each handle is a component of intersection between S and a handle of \mathcal{H} . These handle structures determine cell structures on $S \times \{0\}$ and $S \times \{1\}$ by declaring that each handle is a 2-cell. The gluing map ϕ from $S \times \{1\}$ to $S \times \{0\}$ preserves these handle structures and so is cell-preserving. Pick a spine Γ in $S \times \{0\}$ that is a subcomplex of the 1-skeleton of the cell structure; such a spine is called *cellular*.

The main part of the proof is to isotope Γ across $S \times [0, 1]$, making modifications to the spine as we go, until it is a subcomplex Γ' of the cell structure on $S \times \{1\}$. The key claim is that the number of steps in $\text{Sp}(S)$ between Γ and Γ' is at most a linear function of $\Delta(\mathcal{T})$. Of course, the image of Γ' under the gluing map ϕ will probably not be equal to Γ . But both it and Γ are subcomplexes of the same cell structure on $S \times \{0\}$, and this allows us to relate them by a bounded number of steps in $\text{Sp}(S)$, done in Section 2. By the end of this process, we have found a path in $\text{Sp}(S)$ relating Γ and $\phi(\Gamma)$, with length that is at most a linear function of $\Delta(\mathcal{T})$. This will prove the main theorem.

We now explain how to find the sequence of steps in $\text{Sp}(S)$ taking Γ to Γ' . Using the machinery of normal and almost normal surface theory, between $S \times \{0\}$ and $S \times \{1\}$ lies a collection of normal surfaces, almost normal surfaces, and surfaces interpolating between the two. To form this collection, start with a collection of disjoint non-parallel normal fibres that satisfy a maximality property. Between each of these, there is an almost normal fibre. Start with any almost normal fibre and apply weight-reducing isotopies, all in the same transverse direction. We define these isotopies in Section 5, having introduced the basic definitions and results about normal and almost normal surfaces in Section 4. If we apply these isotopies to an almost normal surface in our collection, then we show that we end up at a normal surface. By the maximality of our initial collection of normal surfaces, this normal surface must be parallel to one in our collection. In this way, we get a collection of fibres interpolating between $S \times \{0\}$ and $S \times \{1\}$. Some of these are normal, some are almost normal and, between these, we have a type of surface that we call *nearly normal*.

Using usual isotopy moves to simplify surfaces, it will not be the case that we obtain an upper bound on the number of surfaces in our collection. Thus, even if we were able to bound the number of steps to transfer a spine between any pair using these usual moves, it would still not be the case that we could obtain a linear upper bound on the distance in $\text{Sp}(S)$ between Γ and Γ' .

In order to deal with this problem, we introduce isotopies, known as generalised isotopy moves, that have a more drastic effect on the fibres. These are defined in terms of the *parallelity bundles* of the normal and almost normal surfaces. This terminology is recalled in Section 6. A large portion of the argument of this paper concerns parallelity bundles, and analysing their topology.

For a normal or almost normal surface S , the space between two parallel normal discs of S is an I -bundle. These I -bundles patch together to form the parallelity bundle for the exterior of S . Its horizontal boundary is a subsurface of the boundary of the exterior of S , and its vertical boundary is a collection of properly embedded annuli. It was shown in [16] that this I -bundle may be extended to a possibly larger I -bundle \mathcal{B} with incompressible horizontal boundary, called a *generalised parallelity bundle*. Each component of \mathcal{B} either is an I -bundle over a disc or has incompressible vertical boundary. In particular, when a component of \mathcal{B} has its entire horizontal boundary in the same side of the same fibre, then it is an I -bundle over a disc or an annulus. This is a consequence of the way that the vertical annuli can be embedded in $S \times [0, 1]$. In our sequence of isotopies, when a fibre starts to enter such a component of \mathcal{B} , we can perform a generalised isotopy move, which moves it across this component in a single step. By allowing these larger moves, we can ensure that the number of moves is at most a linear function of $\Delta(\mathcal{T})$. Properties of such moves are given in Section 7.

We need to control the number of modifications to the spine when a generalised isotopy move is performed. This is done for all but one case in Section 8. The difficult part is the move involving an I -bundle over an annulus. This is called an *annular simplification*. What is important here is the

width of this annulus, which is the distance between its boundary components, as measured using the cell structure. Annular simplifications are analysed in Section 9.

In Section 10, we combine these results to give a proof of Theorem 1.4, using a delicate inductive argument to pick normal fibres carefully.

We give the proof of the main theorem in Section 11. This uses bounds on modifications as we pass a spine Γ from $S \times \{0\}$ to a spine Γ' in $S \times \{1\}$. We then need to convert this to $\phi(\Gamma)$ in $\text{Sp}(S)$. Using the results in Section 2, the number of moves is bounded above by the number of 1-cells in Γ . But unfortunately we do not have a good upper bound on this quantity. In order to circumvent this difficulty, we consider not the fibred manifold with monodromy ϕ , but rather some finite cover with monodromy some high power ϕ^n . In order to compare the translation length $\ell_{\text{Sp}(S)}(\phi)$ and $\ell_{\text{Sp}(S)}(\phi^n)$, we relate them both to the stable translation length $\bar{\ell}_{\text{Sp}(S)}(\phi)$. We do this in Section 3, using the machinery of train tracks.

1.6. Acknowledgements. J. Purcell was partially supported by the Australian Research Council. M. Lackenby was partially supported by the EPSRC.

2. THE SPINE GRAPH AND TRIANGULATION GRAPHS

In this section, we show that we can transfer the problem of proving Theorem 1.3 from considering distances in the mapping class group to considering spines of surfaces, which are more geometric and easier to work with. This is done in Proposition 2.7. Using this, we bound the triangulation complexity $\Delta((S \times I)/\phi)$ from above by a linear function of the translation length of ϕ in the mapping class group, in Proposition 2.8.

2.1. The spine graph. Throughout this paper, our graphs may have multiple edges between vertices and may have edge loops.

Definition 2.1. A *spine* for a closed surface S is an embedded graph Γ such that $S \setminus \Gamma$ is a disc, and Γ has no vertices with valence 0, 1 or 2. It is a *trivalent spine* if every vertex has valence 3.

When an embedded graph satisfies the first of the above conditions, but has some vertices of valence 1 or 2, then there are some easy modifications that one can apply to turn it into a spine. If there is a valence 1 vertex, then we remove it and its incident edge. We continue until there are no valence 1 vertices. We can then remove each valence 2 vertex by amalgamating its incident edges into a single edge. In some of the arguments below, the changes that we make to a spine may create vertices with valence 2, but in this case, we may immediately perform the above modification to turn it back into a spine.

Dual to a spine is a cell structure for the surface with a single vertex, in which every 2-cell has at least 3 sides. When the spine is trivalent, then its dual is a 1-vertex triangulation of the surface.

Lemma 2.2 (Bound on vertices and edges, spine). *Let Γ be a spine for a closed orientable surface S . Then Γ has at most $4g(S) - 2$ vertices and at most $6g(S) - 3$ edges.*

Proof. Let $d(v)$ denote the degree of a vertex v . Let V and E denote the number of vertices and edges of the spine. Then, because $S \setminus \Gamma$ is a single disc, we deduce that

$$\chi(S) = V - E + 1 = 1 + \sum_v \left(1 - \frac{d(v)}{2}\right) \leq 1 - \frac{V}{2},$$

where the sum ranges over all vertices v . Hence, $V \leq 2 - 2\chi(S) = 4g(S) - 2$. So

$$E = V + 1 - \chi(S) \leq 3 - 3\chi(S) = 6g(S) - 3. \quad \square$$

This implies that the sphere does not have a spine, since it cannot have a negative number of vertices. In a similar spirit, we have the following.

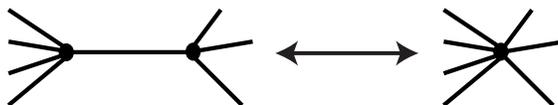


FIGURE 1. An edge contraction/expansion



FIGURE 2. 2-2 and 1-3 Pachner moves

Lemma 2.3 (Bound on triangles and edges, triangulation). *Let \mathcal{T} be a triangulation of a closed orientable surface with V vertices. Then the number of triangles is $2V + 4g(S) - 4$ and the number of edges is $3V + 6g(S) - 6$.*

Proof. Let E and F be the number of edges and triangles. Then $3F = 2E$, and so $\chi(S) = V - E + F = V - (F/2)$. Rearranging gives the required formulae. \square

We now describe some modifications to spines.

Definition 2.4. In an *edge contraction* on a spine Γ , one collapses an edge that joins distinct vertices, thereby amalgamating these vertices into a single vertex. An *edge expansion* is the reverse of this operation. See Figure 1.

Definition 2.5. The *spine graph* $\text{Sp}(S)$ for a closed orientable surface S other than a 2-sphere is a graph defined as follows. It has a vertex for each spine of S , up to isotopy of S . Two vertices are joined by an edge if and only if their spines differ by an edge contraction or expansion.

Closely related is the following concept.

Definition 2.6. The *triangulation graph* $\text{Tr}(S)$ for a closed orientable surface S is a graph defined as follows. It has a vertex for each 1-vertex triangulation of S , up to isotopy of S . Two vertices are joined by an edge if they differ by a 2-2 Pachner move.

Recall that a *2-2 Pachner move* on a triangulation removes an edge with distinct triangles on each side of it, thereby forming a quad, and then it introduces the edge that is the other diagonal of this quad. Similarly, a *1-3 Pachner move* subdivides one triangle into three; see Figure 2. A *3-1 Pachner move* is the reverse of a 1-3 Pachner move.

More generally, for any positive integer v , define the graph $\text{Tr}(S; v)$ as follows. It has a vertex for each triangulation of S with at most v vertices, up to ambient isotopy. Two vertices of $\text{Tr}(S; v)$ are joined by an edge if the corresponding triangulations differ by a 1-3, 2-2 or 3-1 Pachner move.

A well known application of the Švarc-Milnor lemma gives the following result.

Proposition 2.7 (Quasi-isometric to $\text{MCG}(S)$). *For a closed orientable surface S other than a 2-sphere and a positive integer v , its spine graph $\text{Sp}(S)$ and triangulation graphs $\text{Tr}(S)$ and $\text{Tr}(S; v)$ are all quasi-isometric to the mapping class group of S .*

Proof. This is an application of the Švarc-Milnor lemma; see [4, Proposition 8.19]. We just need to verify that the hypotheses of this lemma hold.

The triangulation graph $\text{Tr}(S)$ is well-known to be connected. For example, given two 1-vertex triangulations of S , we can remove their vertices to obtain ideal triangulations of the surface S' that has a single puncture. Two such ideal triangulations differ by a sequence of 2-2 Pachner moves, by [15, Lemma 6]. Hence, the original 1-vertex triangulations differ by a sequence of 2-2 Pachner moves.

A similar argument gives that, for any positive integer v , the triangulation graph $\text{Tr}(S; v)$ is connected. Suppose that we are given two triangulations of S , each with at most v vertices. If they have different numbers of vertices, then we can apply 1-3 Pachner moves to one of them until they have the same number of vertices. We can then remove their vertices, creating ideal triangulations of the same punctured surface S' . As above, these are related by a sequence of 2-2 Pachner moves.

The spine graph is connected, for a similar reason. Given two spines for S , we may apply edge expansions to each of them until each of their vertices is 3-valent. The resulting spines are dual to 1-vertex triangulations of S . These are related by a sequence of 2-2 Pachner moves. Dual to a 2-2 Pachner move is an operation on a spine that can be achieved by an edge contraction followed by an edge expansion.

Thus the spine graph and triangulation graphs are geodesic metric spaces. The mapping class group acts on them properly, using the fact that if a homeomorphism of a surface fixes a spine or triangulation pointwise, then it is isotopic to the identity. Hence, the subgroup of the mapping class group that sends a spine or triangulation back to itself, up to isotopy, is finite.

Finally, the actions of the mapping class group on the spine graph and triangulation graphs are cocompact. That is, up to orientation-preserving homeomorphism of S , there are only finitely many spines for S and only finitely many triangulations with at most v vertices. This follows from Lemmas 2.2 and 2.3.

Thus, all the requirements of the Švarc-Milnor are verified, and so the mapping class group of S is quasi-isometric to the spine graph and to the triangulation graphs. \square

We are now in a position to prove one direction in Theorem 1.3, which is that the triangulation complexity $\Delta(M)$ of a fibred 3-manifold M with monodromy ϕ is bounded above by a linear function of the translation length of ϕ in the mapping class group of S .

Proposition 2.8 (Upper bound on complexity). *Let S be a closed orientable surface, ϕ an orientation preserving homeomorphism of S , and let M denote the mapping torus $(S \times I)/\phi$. There exist constants A and B , depending only on S , such that the triangulation complexity $\Delta(M)$ is bounded above by*

$$\Delta(M) \leq A \cdot \ell_{\text{MCG}(S)}(\phi) + B.$$

Proof. It suffices to focus on the case where S is not a 2-sphere, since any orientation-preserving homeomorphism of the 2-sphere is isotopic to the identity. Thus, we may assume that S has a 1-vertex triangulation. Since $\text{MCG}(S)$ is quasi-isometric to $\text{Tr}(S)$, it suffices to bound $\Delta(M)$ above in terms of the translation length in $\text{Tr}(S)$. Let \mathcal{T} be a 1-vertex triangulation of S such that $d(\phi(\mathcal{T}), \mathcal{T})$ equals the translation length $\ell_{\text{Tr}(S)}(\phi)$. First build a triangulation of $S \times [0, 1]$ so that the induced triangulations of $S \times \{0\}$ and $S \times \{1\}$ are both isotopic to \mathcal{T} . For example, for each triangle T of \mathcal{T} , take the prism $T \times [0, 1]$, then triangulate it by coning to a central vertex. This can be achieved so that the number of tetrahedra is bounded above by a linear function of the genus of S .

A shortest path in $\text{Tr}(S)$ joining \mathcal{T} to $\phi(\mathcal{T})$ specifies a sequence of 1-vertex triangulations of S , each obtained from its predecessor by a 2-2 Pachner move. At each such move, attach a tetrahedron onto $S \times \{1\}$ so that in the new triangulation of $S \times [0, 1]$ the induced triangulation of $S \times \{1\}$ inherits the new triangulation. In this way, we build a triangulation of $S \times [0, 1]$ where the bottom $S \times \{0\}$ is triangulated using \mathcal{T} and the top $S \times \{1\}$ is triangulated using $\phi(\mathcal{T})$ and where the number of tetrahedra is bounded above by the translation length of ϕ in $\text{Tr}(S)$ plus a constant. Glue bottom to top via ϕ and we thereby construct our required triangulation of M . \square

2.2. Edge contractions and expansions on the same surface. Proposition 2.8 gives an upper bound on triangulation complexity in terms of translation length in the mapping class group. We need a lower bound to finish the proof of Theorem 1.3. We will obtain the lower bound by transferring spines of surfaces through the triangulation of the manifold $(S \times I)/\phi$, bounding the number of edge contractions and expansions along the way. There are two steps for this transfer. First, we will transfer a spine Γ from $S \times \{0\}$ to a spine Γ' in $S \times \{1\}$. This step is difficult, and will require most of the rest of the work in the paper. The second step is to bound the number of edge contractions

and expansions to transfer the spine Γ' in $S \times \{1\}$ to $\phi(\Gamma)$. This second step does not require much additional work, and so we give a bound in this subsection.

We begin with a lemma that bounds the number of edge expansions and contractions taking a given spine on a surface with boundary to the dual of a fixed ideal triangulation. This is not exactly the setting we need, since our surfaces are closed, but we can easily convert our problem to this one, in Lemma 2.10. The ideas of the proof are taken from [15].

Recall that a *spine* for a compact orientable surface S with non-empty boundary is a graph Γ embedded in the interior of S with no vertices of valence 0, 1 or 2, and such that $S \setminus \Gamma$ is a regular neighbourhood of ∂S .

Lemma 2.9 (Replacing spine with the dual of an ideal triangulation). *Let S be a compact orientable surface with non-empty boundary. Let Γ be a spine for S and let T be an ideal triangulation with edges T^1 . Then there is a sequence of at most $4|\chi(S)| \cdot (|\Gamma \cap T^1| + 1)$ edge expansions and contractions taking Γ to the dual of T .*

Proof. In [15], it was shown how to perform edge expansions and contractions to transform Γ to the dual of T . If one follows this proof, one readily obtains the bound in the lemma. We briefly sketch the argument. We first convert Γ to a spine with only trivalent vertices using edge expansions, without changing its intersection with T^1 . Each edge expansion increases the number of edges by 1. By an argument analogous to that in Lemma 2.2, the number of edges of a spine of S with trivalent vertices is $3|\chi(S)|$. So, the number of edge expansions is less than $3|\chi(S)|$.

The surface $S \setminus \Gamma$ is a collection of annuli, one for each component of ∂S . The 1-skeleton of T intersects these annuli in arcs. We may assume that these arcs miss the vertices of Γ . No arc can run from ∂S back to ∂S (without meeting Γ), since every edge of an ideal triangulation is essential in S . If every arc runs from ∂S to Γ , then T is dual to Γ . On the other hand, if some arc runs from Γ to Γ , then an outermost one in $S \setminus \Gamma$ separates off a disc. This disc lies in a triangle of T , and has both endpoints on the same edge. Using this disc, we may apply at most $4|\chi(S)|$ edge contractions and expansions to Γ and reduce $|T^1 \cap \Gamma|$. So after at most $4|\chi(S)| \cdot |\Gamma \cap T^1|$ edge expansions and contractions, we end with the spine dual to T . \square

The following lemma will be used in the proof of Theorem 1.3, to transfer a spine coming from $\phi(S \times \{1\})$ to a spine on $S \times \{0\}$.

Lemma 2.10 (Spines in the same cell structure). *Let S be a closed orientable surface with a cell structure \mathcal{C} . Let Γ and Γ' be spines for S that are both subcomplexes of \mathcal{C} . Let n be the number of 1-cells in Γ . Then Γ and Γ' differ by a sequence of edge expansions and contractions with length at most $48g(S)^2n$.*

Proof. Form a 1-vertex triangulation T dual to Γ' , as follows. Pick a point p in the interior of a 2-cell to be the vertex. For each edge of Γ' , pick a point in a 1-cell lying in that edge. Construct arcs coming out from this point, from each side of the edge, and ending at p . The union of these two arcs will form edges of T . Since the vertices of Γ' need not be trivalent, we may need to add further edges to T to make it a triangulation. We can ensure that the edges of T are transverse to the 1-cells of \mathcal{C} . We can further ensure that each edge of T intersects each 1-cell of e at most once. The number of edges of T is $6g(S) - 3$ by Lemma 2.3. So the number of intersections between the the edges of T and Γ is at most $n(6g(S) - 3)$.

Now apply Lemma 2.9 to the ideal triangulation obtained from T by removing p . We deduce that there is a sequence of at most $4|\chi(S \setminus N(p))| (|\Gamma \cap T^1| + 1) \leq 48g(S)^2n$ edge expansions and contractions taking Γ to Γ' . \square

3. STABLE TRANSLATION LENGTH IN THE MAPPING CLASS GROUP

In this section, we will collate some useful facts about the mapping class group $\text{MCG}(S)$. We will discuss results of Agol [1], Hamenstadt [8] and Masur, Mosher, and Schleimer [20], which will

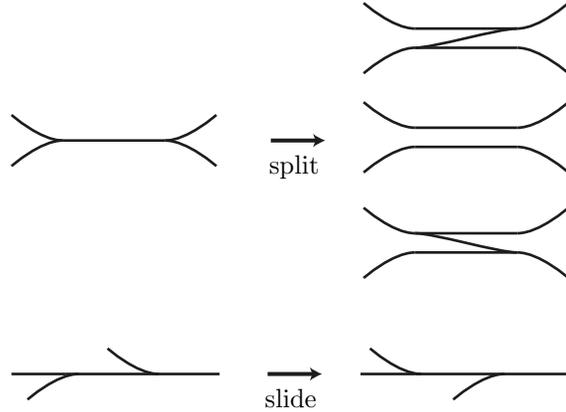


FIGURE 3. Top: The three types of split; Bottom: A slide

demonstrate that the translation length in $\text{MCG}(S)$ of a pseudo-anosov ϕ is readily calculable up to a bounded factor, where the bound only depends on the genus of S . Hence, as a consequence of our main theorem, the triangulation complexity of $(S \times I)/\phi$ is also calculable up to such a factor.

3.1. Train tracks. We recall some terminology related to train tracks. A *pre-track* τ is a graph smoothly embedded in the closed orientable surface S such that at each vertex of the graph, all three edges coming into that vertex share the same tangent line, with one edge coming along the line in one direction, and the other two edges entering from the other direction. The edges of the graph are called *branches* and the vertices are called *switches*. See Figure 3 for some pictures of pre-tracks.

Now let τ be a pre-track, and consider a component F of $S \setminus \tau$. Its boundary consists of a union of arcs, where each arc is identified with a branch of τ . When two such arcs meet in ∂F , it might or might not be possible to join them to form a single smooth arc. If they cannot be joined in this way at a point p in ∂F , we say that p is a *cuspl* of F . Define the *index* $I(F)$ of F to be $\chi(F)$ minus half the number of cusps in ∂F . The pre-track τ in the closed orientable surface S is said to be a *train track* if each component of $S \setminus \tau$ has negative index. It is *filling* if each component of $S \setminus \tau$ is, in addition, a disc.

Lemma 3.1 (Complement of train track). *A train track τ in a closed orientable surface S has at most $4g(S) - 4$ complementary regions.*

Proof. The sum of indices $I(F)$ over all components F of $S \setminus \tau$ is equal to $\chi(S)$. Because τ is a train track, each component has index at most $-1/2$. Thus the number of components of $S \setminus \tau$ is at most $-2\chi(S)$. The lemma follows. \square

3.2. Splits and slides. In this subsection, we present some well known modifications that one can make to a train track τ .

If we pick a point p in the interior of a branch b of τ , the closure of each component of $b - p$ is a *half-branch*. Make such a choice of point for each branch, thereby expressing every branch as the union of two half-branches. We say that a half-branch is *large* if at the switch to which it is incident, there is no other half-branch coming in from the same direction. Otherwise, the half-branch is *small*.

A branch is *large* if both its half-branches are large. It is *small* if both its half-branches are small. Otherwise, it is *mixed*.

A *split* or a *slide* is one of the modifications to a train track τ shown in Figure 3. There are three possible splits that can take place at a large branch; the middle case is called a *central split*.

If we forget the tangential structure at each switch of a train track τ , we get a trivalent graph embedded in S . When τ is filling, this graph is dual to a triangulation of S . By Lemma 3.1, the dual triangulation has at most $k = 4g(S) - 4$ vertices. When two filling train tracks differ by a slide, the

associated triangulations differ by a 2-2 Pachner move. The same is true of non-central splits. When two filling train tracks differ by a central split, their dual triangulations also differ by a bounded number of Pachner moves, where the bound only depends on the genus of S . However, we will not need to compute this bound. Hence, a sequence of slides and splits can be viewed as determining a sequence of points in $\text{Tr}(S; k)$. The following important result is [20, Theorem 6.2].

Theorem 3.2 (Masur–Mosher–Schleimer). *There is a constant $Q > 0$, depending only on the genus of S , with the following property. Let τ_i be a sequence of filling train tracks in S , each obtained from its predecessor from a slide or a split. Suppose that, in all subsequences consisting only of slides, no two train tracks in the subsequence are isotopic. Then this sequence of train tracks determines a Q -quasigeodesic in $\text{Tr}(S; k)$, where $k = 4g(S) - 4$.*

3.3. Pseudo-anosov homeomorphisms. Let ϕ be a pseudo-anosov homeomorphism of the closed orientable surface S . Recall that there are two associated measured singular foliations (\mathcal{F}_s, μ_s) and (\mathcal{F}_u, μ_u) and a real number $\lambda > 1$ such that $\phi(\mathcal{F}_s, \mu_s) = (\mathcal{F}_s, \lambda\mu_s)$ and $\phi(\mathcal{F}_u, \mu_u) = (\mathcal{F}_u, \lambda^{-1}\mu_u)$. The measured singular foliations are known as the *stable* and *unstable* measured singular foliations for ϕ , and λ is the *dilatation*.

We only focus on the stable foliation \mathcal{F}_s . One may split its leaves to form a measured lamination (\mathcal{L}, μ) , which is then carried by a train track τ . The transverse measure of each branch is a positive real number. These numbers satisfy simple linear constraints, which assert that at each switch, the weight of the large half-branch coming into it is equal to the sum of the weights of two small half-branches. Such an assignment of positive real numbers to the branches of τ makes it a *measured train track* (τ, μ) . This data can be computed, via the following result of Bestvina and Handel [3].

Theorem 3.3 (Bestvina–Handel). *Let ϕ be an element of the mapping class group of S , given as a product of standard generators. If ϕ has a pseudo-anosov representative, then there is an algorithm to compute a measured train track (τ, μ) for the stable measured foliation of ϕ and to compute the dilatation of ϕ , as an exact algebraic number.*

Indeed, it is worth emphasizing that τ and μ are readily computable in practice. For example, in the closely related case of a pseudo-anosov on a punctured surface, the program *flipper* [2] is a helpful practical method of doing this.

3.4. Eventually periodic train track sequences. The stable measured lamination satisfies the condition $\phi(\mathcal{L}, \mu) = (\mathcal{L}, \lambda\mu)$. Hence, the measured train tracks (τ, μ) and $(\phi(\tau), \lambda^{-1}\mu)$ both carry the same measured lamination. When two measured train tracks carry the same measured lamination, then they differ by a sequence of measured splits, slides and their inverses, up to isotopy [24, Theorem 2.8.5]. This terminology is defined as follows. Suppose that (τ, μ) is a measured train track and that b is a mixed branch. Then one may slide along b , as defined in Section 3.2, and the new train track τ' inherits a measure. If b is a large branch, then one may split τ at b , but in order that the new train track naturally inherits a measure, only one of the three ways of doing this is permitted; this is determined by the weights of the incoming small half-branches. We call this a *measured split or slide*.

A *maximal split* of a measured train track (τ, μ) is the operation of simultaneously performing a measured split along all the branches of maximal weight. Note that the branches of maximal weight are necessarily large branches. As any two large branches cannot share any switches, there is no concern about performing all these measured splits simultaneously.

It is important to note that there is one and only one possible maximal split that can be applied to a measured train track.

The following was proved by Agol [1, Theorem 3.5], based on work of Hamenstadt [8].

Theorem 3.4 (Agol). *Let ϕ be a pseudo-anosov homeomorphism of a closed orientable surface S . Let (τ, μ) be a measured train track that carries the stable measured lamination of ϕ . Then there exists a sequence of measured train tracks $(\tau, \mu) = (\tau_0, \mu_0), (\tau_1, \mu_1), \dots$, each obtained from the predecessor by a maximal split, such that the sequence is eventually periodic. That is, there exist $n \geq 0, m > 0$ such that $\tau_{n+m} = \phi(\tau_n)$ and $\mu_{n+m} = \lambda^{-1}\mu_n$, with λ the dilatation of ϕ .*

3.5. Stable translation length. Using the results in the previous subsections, we will prove the following result, which is presumably well known.

Theorem 3.5 (Translation length and stable translation length). *Let S be a closed orientable surface. Then, there is a constant $k > 0$, depending only on the genus of S , such that for any pseudo-anosov homeomorphism ϕ of S ,*

$$k \ell_{\text{MCG}(S)}(\phi) \leq \bar{\ell}_{\text{MCG}(S)}(\phi) \leq \ell_{\text{MCG}(S)}(\phi).$$

As a consequence, the translation length in the spine graph $\ell_{\text{Sp}(S)}(\phi)$ and the stable translation length in the spine graph $\bar{\ell}_{\text{Sp}(S)}(\phi)$ also lie within a bounded ratio of each other.

Proof. The stable translation length of an isometry is always at most the translation length; see [4, II.6.6]. So we focus on the first inequality.

By Proposition 2.7, for each positive integer v , $\text{Tr}(S; v)$ and $\text{Sp}(S)$ are quasi-isometric to $\text{MCG}(S)$. It therefore suffices to prove that $\bar{\ell}_{\text{Tr}(S; v)}(\phi) \geq k' \ell_{\text{Tr}(S; v)}(\phi)$ for some positive integer v and some $k' > 0$ depending only on the genus of S .

Let (τ, μ) be a measured train track carrying the stable measured lamination for ϕ . Let $(\tau, \mu) = (\tau_0, \mu_0), (\tau_1, \mu_1), \dots$ be the sequence of measured train tracks, each obtained from its predecessor by a maximal split. By Theorem 3.4, there are integers $n \geq 0$ and $m > 0$ such that $\tau_{n+m} = \phi(\tau_n)$ and $\mu_{n+m} = \lambda^{-1} \mu_n$. By forgetting the smoothing information at each vertex, and taking the dual of the resulting graph, each train track determines an element of $\text{Tr}(S; v)$ for $v = 4g - 4$, by Lemma 3.1. Thus, we obtain a sequence x_0, \dots, x_{n+m} in $\text{Tr}(S; v)$ where $x_{n+m} = \phi(x_n)$. Each train track is obtained from its predecessor by a maximal split, which is a composition of splits, each of which occurs at a branch, dual to an edge. By Lemma 2.3, there are at most $18g(S) - 18$ edges in the triangulation. Hence this is a composition of at most $18g(S) - 18$ splits.

From τ_n and beyond, no central splits are performed, because any central split reduces the number of complementary regions by 1, whereas the equality $\tau_{n+m} = \phi(\tau_n)$ implies that the number of complementary regions is unchanged. Hence, for $i \geq n$, x_{i+1} is obtained from x_i by at most $18g(S) - 18$ Pachner moves. If we continue this sequence by setting $x_{m+i} = \phi(x_i)$, we can obtain a sequence of elements of $\text{Tr}(S; v)$ of length Nm , say, joining x_n to $\phi^N(x_n)$. By Theorem 3.2, this is a Q -quasi-geodesic, for a constant Q that depends only on the genus of S . So, $d(x_n, \phi^N(x_n)) \geq (Nm/Q) - Q$. So, letting N tend to infinity, we deduce that the stable translation length satisfies

$$\bar{\ell}_{\text{Tr}(S; v)}(\phi) \geq m/Q.$$

But $\ell_{\text{Tr}(S; v)}(\phi) \leq (18g(S) - 18)m$, and so setting $k' = (18g(S) - 18)^{-1}Q^{-1}$, we have proved

$$\bar{\ell}_{\text{Tr}(S; v)}(\phi) \geq k' \ell_{\text{Tr}(S; v)}(\phi). \quad \square$$

Note that the integer m in the previous proof is calculable, as follows. By Theorem 3.3, (τ, μ) may be constructed from an expression of ϕ as a product of standard generators of $\text{MCG}(S)$. We start with the train track $(\tau, \mu) = (\tau_0, \mu_0)$ and then start to perform maximal splits, creating a sequence of measured train tracks (τ_i, μ_i) . Each time we create a new measured train track (τ_i, μ_i) , we search for all possible homeomorphisms $h: S \rightarrow S$ taking τ_i to some earlier τ_j . This is clearly algorithmically possible, because τ_i and τ_j induce cell structures on S , and the search for h is the search for homeomorphisms that preserves this cell structure and that respects the smoothing information at each switch. Hence, up to cell-preserving isotopy, there are only finitely many such h . For each one, we check whether h is isotopic to our given homeomorphism ϕ and whether its action on the measure μ is just multiplication by λ^{-1} . According to Theorem 3.4, for some train tracks (τ_{n+m}, μ_{n+m}) and (τ_n, μ_n) in our sequence, such a homeomorphism h with these properties will eventually be found. Since m is therefore calculable, $\bar{\ell}_{\text{Tr}(S; v)}(\phi)$ and $\ell_{\text{Tr}(S; v)}(\phi)$ are also calculable to within a bounded factor. As $\text{Tr}(S; v)$ and $\text{MCG}(S)$ are quasi-isometric, this also implies that $\bar{\ell}_{\text{MCG}(S)}(\phi)$ and $\ell_{\text{MCG}(S)}(\phi)$ are also calculable to within a bounded factor. Unfortunately, this factor itself is not known to be calculable.

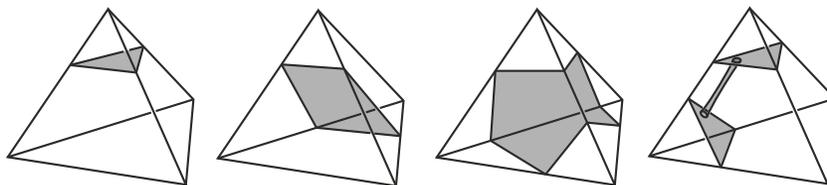


FIGURE 4. Normal and almost normal pieces, left to right: triangle, square, octagon, tubed piece.

4. NORMAL AND ALMOST NORMAL SURFACES

Road map: For M the mapping torus of ϕ , we are trying to prove that there exist constants A, B depending only on the genus of S such that $\Delta(M) \geq A \ell_{\text{Sp}(S)}(\phi) + B$. The idea of the proof will be to cut M along a (nice) fibre S into $S \times [0, 1]$, then transfer a spine Γ from $S \times \{0\}$ to a spine Γ' from $S \times \{1\}$, and then use Lemma 2.10 to bound the distance in $\text{Sp}(S)$ between $\phi(\Gamma)$ and Γ' . Of course, $S \times [0, 1]$ is a product and so we could simply isotope Γ onto Γ' , but we will retain greater control over these spines by ensuring that they are subcomplexes of suitable cell structures on S . We will transfer the spine in $S \times \{0\}$ to the spine in $S \times \{1\}$ by passing it along a sequence of normal surfaces, almost normal surfaces, and surfaces interpolating between them. This section reviews definitions and results from normal and almost normal surface theory.

Throughout, M will be a compact orientable 3-manifold with a triangulation \mathcal{T} . To simplify later arguments, we will also introduce a handle structure on M dual to \mathcal{T} . We will therefore extend results on normal and almost normal surfaces in triangulations to those in a handle structure \mathcal{H} .

Definition 4.1 (Normal surface, triangulation). An arc properly embedded in a 2-simplex is *normal* if it is disjoint from the vertices and has endpoints on distinct edges.

A disc properly embedded in a tetrahedron is a *triangle* if its boundary consists of three normal arcs. It is a *square* if its boundary consists of four normal arcs. See Figure 4, left.

A surface properly embedded in M is *normal* if its intersection with each tetrahedron of \mathcal{T} is a union of disjoint triangles and squares.

Definition 4.2 (Almost normal surface, triangulation). An *almost normal piece* properly embedded in a tetrahedron is either a disc with boundary equal to eight normal arcs (known as an *octagon*) or an annulus that is obtained from two disjoint normal triangles or squares by attaching a tube that runs parallel to an edge of the tetrahedron (known as a *tubed piece*). See Figure 4.

A surface properly embedded in M is *almost normal* if it intersects each tetrahedron in a collection of triangles and squares, except in precisely one tetrahedron, where it is a collection of triangles and squares and exactly one almost normal piece.

For many of our results it is actually easier to consider a handle structure arising from a triangulation \mathcal{T} rather than the triangulation itself.

Definition 4.3 (Handle structure). In this paper, a *handle structure* on a 3-manifold or a 2-manifold will always satisfy the following conditions:

- (1) each i -handle $D^i \times D^{k-i}$ ($k = 2$ or 3) intersects the handles of lower index in $\partial D^i \times D^{k-i}$;
- (2) any two i -handles are disjoint;
- (3) in the case of a 3-manifold, the intersection of any 1-handle $D^1 \times D^2$ with any 2-handle $D^2 \times D^1$ is of the form $D^1 \times \alpha$ in $D^1 \times D^2$ where α is a collection of arcs in ∂D^2 , and of the form $\beta \times D^1$ in $D^2 \times D^1$ where β is a collection of arcs in ∂D^2 ;
- (4) any 2-handle runs over at least one 1-handle.

Definition 4.4 (Associated handle structure on a 0-handle, or ∂M). Let \mathcal{H} be a handle structure of a 3-manifold M . Suppose each 0-handle has connected intersection with the union of the 1-handles and the 2-handles.

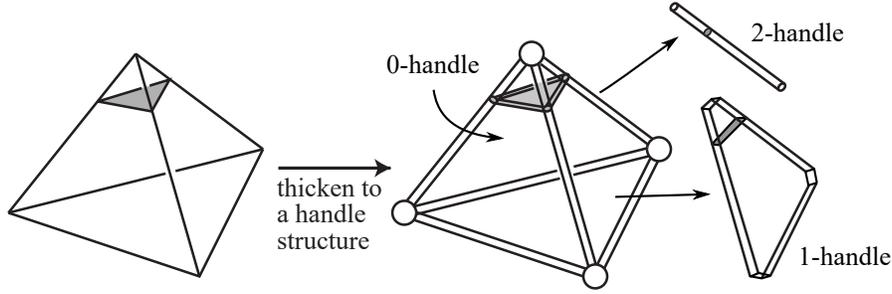


FIGURE 5. A handle structure arising from a triangulation. Shown also is a shaded surface that respects the handle structure.

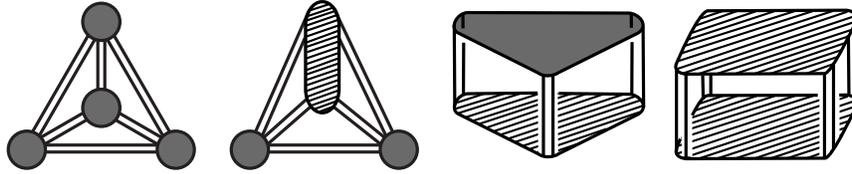


FIGURE 6. Pre-tetrahedral handle structures. Left to right: tetrahedral, semi-tetrahedral, product handle of length three, parallelity handle of length four. The shaded handles arise either from 3-handles or ∂M . The striped arise from ∂M only.

- (1) The boundary of each 0-handle inherits an *associated handle structure*. The 0-handles of this structure are the components of intersection between the 0-handles and the 1-handles. The 1-handles of the associated handle structure are the components of intersection between the 0-handles and the 2-handles. The 2-handles of the associated handle structure are the remainder of the boundary of the 0-handles, which are discs by assumption.
- (2) The boundary of the 3-manifold also inherits a handle structure, where the i -handles are the components of intersection between ∂M and the i -handles of \mathcal{H} .

Definition 4.5 (Dual handle structure to triangulation). Given a triangulation \mathcal{T} of a closed 3-manifold, there is a handle structure, called the *dual handle structure*, obtained as follows. Each tetrahedron gives rise to a 0-handle, which is obtained from the tetrahedron by removing a thin open regular neighbourhood of its boundary. Each face of \mathcal{T} gives rise to 1-handle, which we take to be a thin regular neighbourhood of the face with a thin open regular neighbourhood of the edges of \mathcal{T} removed. Each edge of \mathcal{T} forms a 2-handle, by taking a regular neighbourhood of the edge and removing a small open regular neighbourhood of the endpoints of the edge. Finally, regular neighbourhoods of the vertices correspond to 3-handles. See Figure 5.

Consider a normal surface in a triangulation, intersecting the tetrahedra in triangles and squares. When we cut along the normal surface, tetrahedra are not necessarily split into tetrahedra; the cut also creates pieces with quadrilateral faces, or parallel triangle faces. However, normal surfaces can be defined in handle structures, and cutting along a normal surface in a handle structure will still give rise to a handle structure. The boundaries of the 0-handles will inherit handle structures, such those shown in Figure 6. We now define these more formally.

Definition 4.6 (Pre-tetrahedral handle structure). Let M be a compact 3-manifold, possibly with boundary, with a handle structure \mathcal{H} . We will define various types of 0-handle of \mathcal{H} . Each 0-handle will have connected intersection with the union of the 1-handles and 2-handles and so its boundary will have an associated handle structure. Let H_0 be a 0-handle of \mathcal{H} .

- (1) The handle H_0 is *tetrahedral* if ∂H_0 inherits a handle structure as shown on Figure 6 left; that is, it inherits the same handle structure as the boundary of a 0-handle coming from a tetrahedron. The 2-handles in ∂H_0 , which are shaded in the figure, may arise either as intersections of ∂H_0 with 3-handles of \mathcal{H} , or as intersections of ∂H_0 with ∂M .
- (2) H_0 is *semi-tetrahedral* if the intersection of ∂H_0 with 1-handles, 2-handles, 3-handles, and ∂M is as shown in Figure 6 second from left; that is, it has the handle structure of the boundary of a thickened tetrahedron cut along a single square. Two of the 2-handles, shaded in the figure, may arise either as intersections of ∂H_0 with 3-handles or with ∂M . The third 2-handle, which meets four 1-handles, arises only as a component of intersection between ∂H_0 and ∂M .
- (3) H_0 is a *product handle of length 3*, if it is as shown in the middle right of Figure 6. That is, ∂H_0 meets exactly three 1-handles of \mathcal{H} and exactly three 2-handles, connected in a cycle. Furthermore, it is a *parallellity handle of length 3* if it meets ∂M only and no 3-handles.
- (4) H_0 is a *parallellity handle of length 4* if it is as shown in Figure 6 right; that is it intersects four 1-handles and 2-handles of \mathcal{H} in a cycle, and meets ∂M on either side of the cycle.

Parallellity handles arise, for example, when a 0-handle dual to a tetrahedron is cut along parallel normal triangles or squares.

We say that the handle structure \mathcal{H} is *tetrahedral* if each of its 0-handles is tetrahedral. We say it is *pre-tetrahedral* if each of its 0-handles is tetrahedral, semi-tetrahedral, a product handle or a parallellity handle, as above.

The following definition gives an analogue to Definition 4.1.

Definition 4.7 (Surface respects a handle structure). A closed surface S embedded in M *respects* a handle structure \mathcal{H} of M if

- (1) its intersection with each 1-handle $D^1 \times D^2$ is of the form $D^1 \times \alpha$, where α is a properly embedded 1-manifold in D^2 ;
- (2) it intersects each 2-handle $D^2 \times D^1$ in a collection of discs of the form $D^2 \times p$, where p is a point in the interior of D^1 ;
- (3) it is disjoint from the 3-handles.

Any surface properly embedded in M that is in general position with respect to a triangulation \mathcal{T} gives rise to a surface that respects the dual handle structure \mathcal{H} and vice versa.

Definition 4.8 (Normal, handle structure). If a closed surface S embedded in M respects a handle structure \mathcal{H} and, in addition, it intersects each 0-handle and each 1-handle in a collection of properly embedded discs, it is *standard*. If, furthermore, S has the property that for each 0-handle H_0 of \mathcal{H} , each component of $S \cap H_0$ runs over any component of intersection between H_0 and the 2-handles in at most one arc, then S is called *normal with respect to \mathcal{H}* .

Figure 5 shows a surface that is normal in a handle structure.

We also define a notion of standard for 1-manifolds in surfaces with handle structures, as follows.

Definition 4.9 (Standard 1-manifold). Let S be a compact surface with a handle structure \mathcal{H} . Then a 1-manifold properly embedded in S is *standard* if

- (1) it intersects each 0-handle in a collection of properly embedded arcs;
- (2) it intersects each 1-handle in a collection of arcs, each of which runs parallel to the core of the 1-handle and respects its product structure; and
- (3) it is disjoint from the 2-handles.

Definition 4.10 (Normal 1-manifold). Let S be a compact surface with a handle structure \mathcal{H} . An arc properly embedded in a 0-handle H_0 is *normal* if its endpoints lie in distinct components of intersection between H_0 and the 1-handles of \mathcal{H} . A closed 1-manifold properly embedded in S is *normal* if it is standard and each arc of intersection with each 0-handle is normal.

Definition 4.11 (Triangles and squares in a 0-handle). Let \mathcal{H} be a handle structure of a 3-manifold M . Let S be a closed normal surface properly embedded in M . Then a component F of intersection between a 0-handle H_0 of \mathcal{H} and S is a *triangle* (respectively, *square*) if it is a disc and ∂F is a normal curve that runs over the 1-handles of ∂H_0 exactly three (respectively, four) times.

Recall that the 1-handles on the boundary of a 0-handle H_0 arise from 2-handles of \mathcal{H} . When \mathcal{H} is dual to a triangulation, its 2-handles can be viewed as thickened edges of this triangulation. Thus Definition 4.11 is completely analogous to Definition 4.1. The following result is easily checked.

Lemma 4.12 (Normal surface meets 0-handles in triangles and squares). *Let \mathcal{H} be a pre-tetrahedral handle structure of a 3-manifold M . Let S be a closed normal surface properly embedded in M . Then any component of intersection between S and a 0-handle of \mathcal{H} is a triangle or square.* \square

We may define the complexity of a handle structure by analogy with that of a triangulation.

Definition 4.13 (Tetrahedral complexity, handle structure). Let \mathcal{H} be a pre-tetrahedral handle structure of a 3-manifold M , and let H_0 be a 0-handle of \mathcal{H} . Let α denote the number of intersections of H_0 with the 3-handles of \mathcal{H} . Define the *complexity* of H_0 to be $(\alpha/8) + \beta$, where

- $\beta = 1/2$ if H_0 is tetrahedral,
- $\beta = 1/4$ if H_0 is semi-tetrahedral, and
- $\beta = 0$ if H_0 is a product handle or parallelity handle.

Define the *tetrahedral complexity* $\Delta(\mathcal{H})$ to be the sum of the complexities of its 0-handles.

When \mathcal{H} is dual to a triangulation \mathcal{T} of a closed 3-manifold M , then $\Delta(\mathcal{H}) = \Delta(\mathcal{T})$. This is because every 0-handle of \mathcal{H} is tetrahedral and has 4 components of intersection with the 3-handles.

Lemma 4.14 (Complexity unchanged under cutting along normal surface). *Let \mathcal{H} be a pre-tetrahedral handle structure of a 3-manifold. Let S be a closed normal surface properly embedded in M . Then $M \setminus S$ inherits a pre-tetrahedral handle structure \mathcal{H}' with $\Delta(\mathcal{H}') = \Delta(\mathcal{H})$.*

Proof. Each tetrahedral 0-handle of \mathcal{H} is decomposed into one of the following:

- (1) a single tetrahedral 0-handle plus possibly some parallelity and product handles;
- (2) two semi-tetrahedral 0-handles plus possibly some parallelity and product handles.

A semi-tetrahedral 0-handle is decomposed into exactly one semi-tetrahedral 0-handle plus possibly some parallelity and product handles. A product or parallelity 0-handle is decomposed into product and parallelity 0-handles. In every case, the components of intersection between the 0-handle and the 3-handles are shared out among the resulting 0-handles of \mathcal{H}' . \square

Analogous to Definition 4.2, we also define almost normal surfaces in pre-tetrahedral handle structures.

Definition 4.15 (Almost normal surface, handle structure). Let \mathcal{H} be a pre-tetrahedral handle structure of a 3-manifold.

- (1) An *octagon* is a disc properly embedded in a 0-handle H_0 of \mathcal{H} with boundary that is a normal curve that runs over eight 1-handles and eight 0-handles of ∂H_0 .
- (2) A *tubed piece* is an annulus in a 0-handle H_0 obtained by tubing together two disjoint triangles or squares. The 0-handle is required to be tetrahedral or semi-tetrahedral. The tube runs parallel to an arc in ∂H_0 that is a co-core of a 1-handle of ∂H_0 .

An *almost normal piece* is an octagon or tubed piece.

A closed surface properly embedded in a 3-manifold M with a pre-tetrahedral handle structure \mathcal{H} is *almost normal* if it respects the handle structure, it intersects each 0-handle in a collection of triangles and squares, except in precisely one 0-handle, where it is a collection of triangles and squares and exactly one almost normal piece.

Definition 4.16 (Handle structure, cutting along almost normal). Suppose S is almost normal. Then in the case where S has an octagonal piece, $M \setminus S$ inherits a handle structure, where each i -handle is a component of intersection between $M \setminus S$ and an i -handle of M . In the case where S has a tubed piece, this does not quite work because one component of intersection between $M \setminus S$ and a 0-handle of M is a solid torus V . This solid torus is naturally the union of a 0-handle and a 1-handle as follows. There is a disc D in the 0-handle H_0 containing the tubed piece such that $D \cap S$ is an arc in ∂D running over the tubed piece, and the remainder of ∂D is part of the co-core of a 1-handle of ∂H_0 . Then a regular neighbourhood of D is the 1-handle of V , and the remainder of V is the 0-handle. The remaining handles of the handle structure are defined in the same way as previously, where each i -handle is a component of intersection between $M \setminus S$ and an i -handle of M . We say that this handle structure is the one that $M \setminus S$ inherits from M .

We now recall a few definitions and results from normal surface theory, and their analogues for handle structures.

Definition 4.17 (Weight). If S is a surface properly embedded in M , in general position with respect to a triangulation \mathcal{T} , then its *weight* is defined to be the number of intersections between S and the edges of \mathcal{T} .

If S is a surface properly embedded in M that respects a handle structure \mathcal{H} , define its *weight* to be the number of components of intersection between S and the 2-handles of \mathcal{H} .

Definition 4.18 (Topologically parallel surfaces). Two disjoint closed surfaces S_0 and S_1 properly embedded in M are *topologically parallel* if there is an embedding of $S \times [0, 1]$ in M such that $S_0 = S \times \{0\}$ and $S_1 = S \times \{1\}$.

When we are dealing with surfaces that respect a triangulation or a handle structure of M , there is a stronger notion of parallel surfaces, as follows.

Definition 4.19 (Normally parallel surfaces, tetrahedra). For a triangulation \mathcal{T} of M , a *normal isotopy* of M is an isotopy $F: M \times [0, 1] \rightarrow M$ such that, for each $t \in [0, 1]$, the map $F_t: M \times \{t\} \rightarrow M$ is a homeomorphism that preserves each simplex of \mathcal{T} and such that F_0 is the identity on M . We say that two surfaces D_0 and D_1 properly embedded in a tetrahedron of \mathcal{T} have the same *type* if there is a normal isotopy taking one to the other. More specifically, there is a normal isotopy $F: M \times [0, 1] \rightarrow M$ such that $F_1(D_0) = D_1$. We say that D_0 and D_1 are *normally parallel* if there is a normal isotopy $F: M \times [0, 1] \rightarrow M$ taking D_0 to D_1 such that the restriction of F to $D_0 \times [0, 1]$ is an embedding into M . Similarly, disjoint surfaces S_0 and S_1 properly embedded in M are *normally parallel* if there is a normal isotopy $F: M \times [0, 1] \rightarrow M$ taking S_0 to S_1 , such that the restriction of F to $S_0 \times [0, 1]$ is an embedding into M .

In Figure 7, left, two discs properly embedded within a tetrahedron are shown. They are of the same type but are not normally parallel.

Definition 4.20 (Normally parallel surfaces, handle structure). Let \mathcal{H} be a handle structure of a 3-manifold M . Let S be a properly embedded surface that respects \mathcal{H} . An *elementary piece* of S is a component of intersection between S and some handle of \mathcal{H} .

Two elementary pieces D_0 and D_1 are *normally parallel* if there is an isotopy $F: M \times [0, 1] \rightarrow M$ that preserves all the handles of \mathcal{H} and sends D_0 to D_1 , and such that the restriction of F to $D_0 \times [0, 1]$ is an embedding into M .

Two disjoint surfaces S_0 and S_1 properly embedded in M are *normally parallel* if there is an isotopy $F: M \times [0, 1] \rightarrow M$ taking S_0 to S_1 , preserving all the handles of \mathcal{H} , such that the restriction of F to $S_0 \times [0, 1]$ is an embedding into M .

Between normally parallel elementary pieces lie parallelity regions, defined as follows.

Definition 4.21 (Parallelity regions). Let S be a surface properly embedded in M that respects the handle structure \mathcal{H} . For any handle H of \mathcal{H} , a component of $H \setminus S$ is a *parallelity region* if it

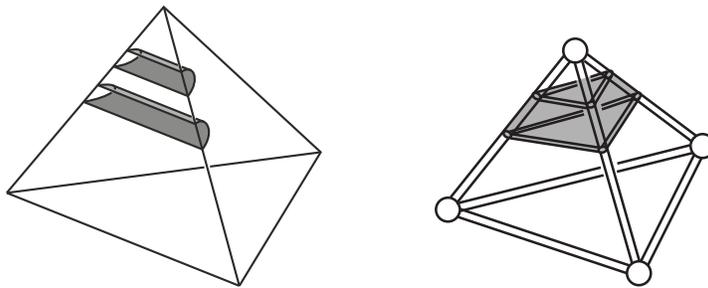


FIGURE 7. Left: Discs that are of the same type but that are not normally parallel. Right: Normally parallel disks in a handle structure. The region between is the union of parallelity regions in 0, 1, and 2-handles.

lies between two normally parallel elementary pieces D_0 and D_1 of $S \cap H$. There is an identification between this region and $D_0 \times [0, 1]$, where $D_0 = D_0 \times \{0\}$ and $D_1 = D_0 \times \{1\}$. The *projection map* for this region is the map $D_0 \times [0, 1] \rightarrow D_0$ onto the first factor. See Figure 7, right.

Theorem 4.22 (Stocking’s theorem). *Let M be a compact orientable 3-manifold equipped with a pre-tetrahedral handle structure \mathcal{H} . Let S_0 and S_1 be disjoint normal closed connected surfaces embedded in M that are not 2-spheres and that are topologically parallel but not normally parallel. Then there is an almost normal surface between them that is topologically parallel to each of them.*

The proof of Theorem 4.22 is essentially contained within the arguments of [26]. There, Stocking considers a closed triangulated orientable 3-manifold, and shows that any strongly irreducible Heegaard surface for the manifold may be isotoped into almost normal form. The argument also applies to compact orientable 3-manifolds with boundary. In that situation, the right structure to use is more general than a triangulation. We call it a generalised triangulation, and define it as follows.

If a tetrahedron is cut along a union of disjoint triangles and squares, each resulting component is called a *sub-tetrahedron*. It inherits a cell structure, with a single 3-cell, and where the 2-cells arise as components of intersection with the faces of the original tetrahedron and as copies of the triangles and squares. We call the latter type of 2-cells *∂ -faces*.

A *generalised triangulation* for a compact 3-manifold M is a cell structure for M with the following properties. Each 3-cell is a copy of a sub-tetrahedron. The ∂ -faces of these sub-tetrahedra form the 2-cells in ∂M . The other faces of these sub-tetrahedra are identified in pairs so that they form 2-cells with interior lying in the interior of M . Generalised triangulations are exactly the dual concept to a pre-tetrahedral handle structure.

Many of the familiar concepts for normal surfaces in triangulated 3-manifolds apply equally well to 3-manifolds with a generalised triangulation. One can speak of a triangle, square, octagon or tubed piece in a sub-tetrahedron, as long as one requires that these are disjoint from the ∂ -faces. One can also therefore speak of a closed normal or almost normal surface. Furthermore, a closed surface in a generalised triangulation is normal if and only if it is normal in the dual pre-tetrahedral handle structure. The same is true of almost normal surfaces, except that in the case of a handle structure, we require that a tubed piece does not lie in a product or parallelity handle.

Proof of Theorem 4.22. Stocking’s argument for closed triangulated 3-manifolds applies just as well to compact orientable 3-manifolds with a generalised triangulation (see [26, Section 6]). In fact, in her argument, one cuts along closed normal surfaces and considers the complementary pieces. These inherit a generalised triangulation, which is what Stocking works with.

We are given a pre-tetrahedral handle structure \mathcal{H} . Cutting this along S_0 and S_1 , we obtain pre-tetrahedral handle structures on each of the pieces. In particular, the copy of $S \times [0, 1]$ between S_0 and S_1 inherits such a handle structure \mathcal{H}' . Let \mathcal{T} be the generalised triangulation dual to \mathcal{H}' . Let \mathcal{T}^1 be the union of the 1-cells of \mathcal{T} not lying in $S \times [0, 1]$. Now $S \times [0, 1]$ admits a Heegaard

splitting where $S \times \{1/2\}$ is the Heegaard surface, dividing the manifold into two I -bundles. Thus, we can apply Stocking's methods to this splitting.

Stocking's argument started by finding within $S \times [0, 1]$ a maximal collection \mathcal{S} of disjoint properly embedded normal 2-spheres, no two of which are normally parallel. The proof divides into the case where \mathcal{S} is empty or non-empty.

If \mathcal{S} is non-empty, then one component M' of $(S \times [0, 1]) \setminus \mathcal{S}$ is homeomorphic to $S \times [0, 1]$ with a single open 3-ball removed ([27, Lemma 2]). No vertices of \mathcal{T} lie in M' , and hence the intersection between \mathcal{T}^1 and M' is a disjoint union of properly embedded arcs. Since one obtains a cell structure on M' by starting with $\partial M' \cup (M' \cap \mathcal{T}^1)$ and adding only 2-cells and 3-cells, we deduce that there must be some arc of $\mathcal{T}^1 \cap M'$ that joins $S \times \{0\}$ to the boundary of the ball. If we attach a tube that follows this arc, we obtain a tubed almost normal surface, as required.

So suppose that \mathcal{S} is empty. One can view the product projection map $S \times [0, 1] \rightarrow [0, 1]$ as a height function. One can then place \mathcal{T}^1 into thin position with respect to this height function. This is defined as follows. For each level $S \times \{t\}$, where $t \in [0, 1]$, one considers the number of intersections between \mathcal{T}^1 and $S \times \{t\}$, and this gives a function $[0, 1] \rightarrow \mathbb{N}$. Since $S \times \{0\}$ and $S \times \{1\}$ are normal, $\{0\}$ and $\{1\}$ are local minima for this function. Since $S \times \{0\}$ and $S \times \{1\}$ are not normally parallel, the function is not constant. Hence, there is some local maximum realised by a surface $S \times \{t\}$ that is in general position with respect to the 1-skeleton. It is shown (see [26, Lemma 5]) that one may isotope and compress $S \times \{t\}$, keeping its intersection with \mathcal{T}^1 unchanged, taking it to an octagonal almost normal surface. However, since $S \times \{t\}$ is, in our case, incompressible, and $S \times [0, 1]$ is irreducible, this implies that no compressions are required. Thus, this almost normal surface is topologically parallel to S_0 and S_1 .

We now convert this almost normal surface in \mathcal{T} to one in \mathcal{H}' . It will then be almost normal in \mathcal{H} . The only extra condition that needs to be ensured is that if the almost normal surface contains a tubed piece, then this must miss the product and parallelity handles in \mathcal{H}' . Now, the union of the product and parallelity handles and any incident 1-handles and 2-handles is an I -bundle. The tube is vertical in this I -bundle. When the tube is compressed, the resulting normal surfaces are horizontal in the I -bundle. Note that this I -bundle is not all of $S \times [0, 1]$ because then \mathcal{H}' would consist entirely of product and parallelity handles and so S_0 and S_1 would be normally parallel. Thus, we can slide the tube, pulling it away from the I -bundle into an adjacent 0-handle that is not a product or parallelity handle. The resulting surface is then almost normal in \mathcal{H} . \square

A useful feature of pre-tetrahedral handle structures that is not always true of triangulated 3-manifolds is the following. Its proof is immediate.

Lemma 4.23 (Boundary surface is normal). *Let \mathcal{H} be a pre-tetrahedral handle structure of a compact 3-manifold M . When ∂M is pushed a little into the interior of M , it becomes a normal surface.* \square

As a consequence of the above lemma and of Theorem 4.22, we have the following theorem.

Theorem 4.24 (Almost normal surface exists). *Let \mathcal{H} be a pre-tetrahedral handle structure of $S \times [0, 1]$, where S is a closed orientable surface, such that $\Delta(\mathcal{H}) > 0$. Then \mathcal{H} contains an almost normal surface that is isotopic to $S \times \{1/2\}$.*

Proof. By Lemma 4.23, when $S \times \{0\}$ and $S \times \{1\}$ are pushed a little into the interior of the manifold, they become normal surfaces. They are not normally parallel, because then every handle of \mathcal{H} would be a parallelity handle and hence $\Delta(\mathcal{H})$ would be zero. Hence, by Theorem 4.22, there is an almost normal surface between them that is topologically parallel to each. \square

5. NORMALISING ALMOST NORMAL SURFACES

Road map: We are still trying to build a sequence of surfaces between $S \times \{0\}$ and $S \times \{1\}$ in a manifold $M \setminus S \cong S \times [0, 1]$. So far, we have defined normal and almost normal surfaces. But we do not yet know enough to transfer spines between such surfaces in a way that will allow us to

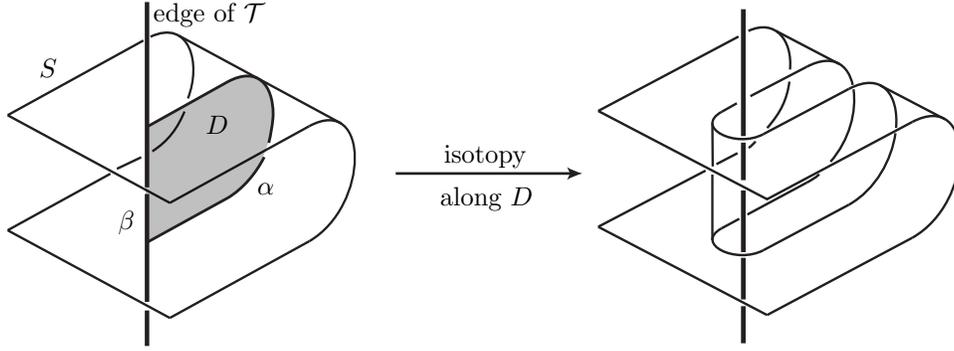


FIGURE 8. Isotopy along an edge compression disc.

bound from above the number edge expansions and contractions. In this section, we introduce nearly normal surfaces and moves between them. These will give us a first sequence of surfaces on which we can transfer spines.

Throughout this section M is a compact orientable irreducible 3-manifold and S is a closed incompressible surface properly embedded in M . The moves we describe are easiest to visualise when M is equipped with a triangulation \mathcal{T} . However, we will need to apply them in the case that M admits a pre-tetrahedral handle structure \mathcal{H} . Thus we will define the moves first in the triangulation setting, and then explain how to define them for a handle structure.

Definition 5.1 (Edge compression disc, triangulation). An *edge compression disc* for S is a disc D lying in a tetrahedron Δ of \mathcal{T} such that ∂D is the union of two arcs $\alpha \subset S$ and $\beta \subset \partial\Delta$, where $\alpha = D \cap S$, $\beta = D \cap \partial\Delta$ is a sub-arc of an edge of Δ , and $\alpha \cap \beta = \partial\alpha = \partial\beta$.

Definition 5.2 (Face compression disc, triangulation). A *face compression disc* for S is a disc D lying in a 2-simplex F such that ∂D is the union of two arcs $\alpha \subset S \cap F$ and $\beta \subset \partial F$, where $\alpha = D \cap S$, $\beta = D \cap \partial F$ is a sub-arc of an edge of F , and $\alpha \cap \beta = \partial\alpha = \partial\beta$.

Figure 8 left shows an edge or face compression disc.

Definition 5.3 (Isotopy ball, triangulation). Given an edge or face compression disc D for a properly embedded surface S , its associated *isotopy ball* B is a small regular neighbourhood of D in $M \setminus S$. Thus, $\partial B \cap S$ is a regular neighbourhood of $D \cap S$ in S . The remainder of ∂B is obtained by taking two parallel copies of D and attaching a band that goes around the edge of \mathcal{T} incident to D .

An *isotopy along D* is the isotopy that moves S across this ball. When S is transversely oriented, we require that this transverse orientation points towards D and hence that the isotopy moves S in that direction. See Figure 8.

Definitions 5.1, 5.2, and 5.3 arise when applying normalisation moves to a surface in a triangulation \mathcal{T} . While they are not as natural in a handle structure, they can be easily moved to this setting as in the following definitions, which assume that S respects a handle structure \mathcal{H} .

Definition 5.4 (Edge compression disc, handle structure). Let S be a closed surface S embedded in M that respects a pre-tetrahedral handle structure \mathcal{H} on M . An *edge compression disc* for S in \mathcal{H} is a disc D lying in a 0-handle H_0 of \mathcal{H} such that ∂D is the union of two arcs $\alpha \subset S$ and $\beta \subset \partial H_0$, where $\alpha = D \cap S$, $\beta = D \cap \partial H_0$ is an arc on the boundary of a 2-handle $H_2 \subset \mathcal{H}$ that is vertical in its product structure, and $\alpha \cap \beta = \partial\alpha = \partial\beta$.

Note that because S respects the handle structure, its intersection with the 2-handle $H_2 = D^2 \times D^1$ contains the discs $D^2 \times \partial\alpha$. Thus when an edge compression disc exists, we actually have the product $D^2 \times \beta$ lying in the adjacent 2-handle $H_2 = D^2 \times D^1$, with $(D^2 \times \beta) \cap S = D^2 \times \partial\beta = D^2 \times \partial\alpha$.

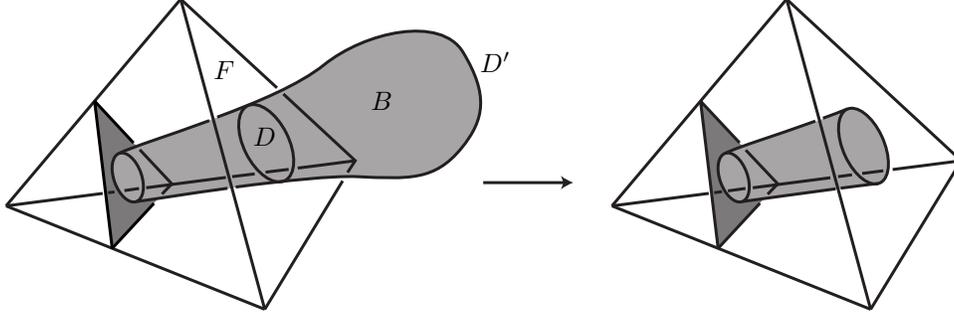


FIGURE 9. A compression isotopy.

Definition 5.5 (Face compression disc, handle structure). A *face compression disc* for S in a pre-tetrahedral handle structure \mathcal{H} is a disc D lying in the intersection of a 1-handle H_1 with a 0-handle H_0 of \mathcal{H} , such that ∂D is the union of two arcs $\alpha \subset S$ and $\beta \subset \partial(H_1 \cap H_0)$, where $\alpha = D \cap S$, and $\beta = D \cap \partial(H_1 \cap H_0)$ is an arc on the boundary of a 2-handle $H_2 \subset \mathcal{H}$, and $\alpha \cap \beta = \partial\alpha = \partial\beta$.

As in the case of an edge compression disc, because S respects the handle structure \mathcal{H} , its intersection with the 1-handle $H_1 = D^1 \times D^2$ contains $D^1 \times \alpha$, and its intersection with the 2-handle $H_2 = D^2 \times D^1$ contains $D^2 \times \partial\alpha$. Thus when a face compression disc exists, in fact there is a product $D^1 \times D \subset H_1$ with $D^1 \times \alpha \subset S \cap H_1$, and $D^2 \times \beta \subset H_2$ with $D^2 \times \partial\beta \subset S \cap H_2$.

Definition 5.6 (Isotopy along edge or face compression disc). Let D be an edge or face compression disc for a properly embedded surface S that respects a pre-tetrahedral handle structure \mathcal{H} . Thus ∂D is the union of arcs α and β , as in Definition 5.4 or 5.5. Let $H_1 = D^1 \times D^2$ be the incident 1-handle in the case of a face compression disc.

When D is an edge compression disc, its associated *isotopy ball* B is a small regular neighbourhood in $M \setminus S$ of the union of D and the incident 2-handle of $M \setminus S$. When D is a face compression disc, its associated *isotopy ball* B is a small regular neighbourhood in $M \setminus S$ of the union of D , the incident 2-handle H_2 of $M \setminus S$ and $D^1 \times D \subset D^1 \times D^2 = H_1$. Then as in the triangulation case, $\partial B \cap S$ is a regular neighbourhood of $D \cap S$ in S that contains $D^1 \times \alpha$ in $S \cap H_1$ and contains $D^2 \times \partial\beta$ in $S \cap H_2$, and the remainder of ∂B is obtained by taking two parallel copies of D (that are disjoint from H_1 in the face compression case) and attaching a band that goes around the 2-handle H_2 incident to D .

An *isotopy along D* is the isotopy that moves S across this ball. When S is transversely oriented, we require that this transverse orientation points towards D and hence that the isotopy moves S in that direction.

Another isotopy that is applied to normalise surfaces in triangulations is a compression isotopy, defined below.

Definition 5.7 (Compression isotopy, triangulation). Let S be a transversely oriented surface properly embedded in M , in general position with respect to a triangulation \mathcal{T} . Let D be a disc embedded in the interior of a face F of \mathcal{T} , such that $D \cap S = \partial D$ and so that S points into D . Suppose that ∂D bounds a disc D' in S , and that $D \cup D'$ bounds a ball B with interior disjoint from S . Then the *isotopy induced by D* is the isotopy that moves S across B and then a little further so that the curve ∂D is removed from $S \cap F$. We call this a *compression isotopy* in \mathcal{T} . See Figure 9.

As before, we adapt Definition 5.7 to the handle structure case.

Definition 5.8 (Compression isotopy, handle structure). Let S be a transversely oriented surface properly embedded in M , that respects a pre-tetrahedral handle structure \mathcal{H} on M . Suppose a disc D lies in the interior of the intersection $F = \partial H_0 \cap H_1$, for a 0-handle H_0 and a 1-handle H_1 , with $D \cap S = \partial D$ and S pointing into D . Suppose ∂D bounds a disc D' in S , and that $D \cup D'$ bounds

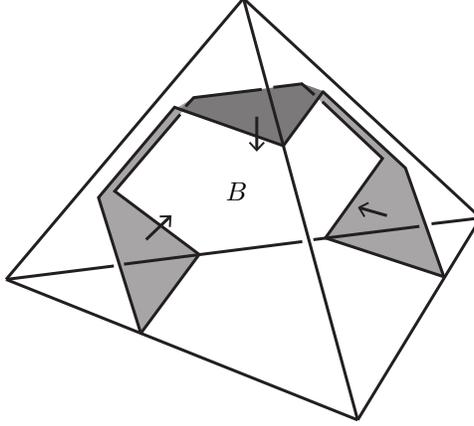


FIGURE 10. A nearly normal disc. This might have been created from three normal triangles by performing isotopies along face compression discs in nearby faces.

a ball B with interior disjoint from S . Then the *isotopy induced by D* is the isotopy that moves S across B and then (if necessary) a little further so that the annulus $D^1 \times \partial D$ is removed from $S \cap H_1$. We call this a *compression isotopy* in \mathcal{H} .

When we move between almost normal and normal surfaces, isotoping along edge and face compressing discs, we obtain surfaces that are not necessarily normal. However, they still have a nice form, described by the following definition.

Definition 5.9 (Nearly normal, triangulation). Let F be a connected transversely oriented surface properly embedded in a tetrahedron Δ in general position with respect to the 1-skeleton of Δ . Then F divides Δ into two components. Let B be the component into which F points. We say that F is a *nearly normal piece* if the all the following conditions hold:

- (1) B is a 3-ball;
- (2) $\Delta \setminus B$ forms a product region between F and a subsurface of $\partial \Delta$;
- (3) the intersection between B and any face of Δ is either empty or a disc;
- (4) B intersects each edge of Δ in at most one component.

We say that a surface S properly embedded in M is *nearly normal* if it has a transverse orientation which makes each component of $S \cap \Delta$ a nearly normal piece for each tetrahedron Δ of \mathcal{T} .

Triangles and squares form nearly normal pieces. However, neither an octagon nor a tubed piece is nearly normal.

Notice as well that nearly normal surfaces cannot have an edge compression disc by item (4).

The analogous definition for handle structures is the following.

Definition 5.10 (Nearly normal, handle structure). Let \mathcal{H} be a pre-tetrahedral handle structure for a compact 3-manifold M , and let S be a transversely oriented surface properly embedded in M that respects \mathcal{H} . Let H_0 be a 0-handle of \mathcal{H} and let F be a component of $S \cap H_0$. Then F divides H_0 into two components; let B denote the component into which F points. We say that F is a *nearly normal piece* if the following hold:

- (1) F is disjoint from $\partial M \cap \partial H_0$,
- (2) B is a 3-ball,
- (3) $H_0 \setminus B$ forms a product region between F and a subsurface of ∂H_0 ,
- (4) for any component of intersection between H_0 and a 1-handle H_1 , its intersection with B is either empty or a disc,
- (5) for any component E of intersection between H_0 and a 2-handle, $B \cap E$ is either empty or connected.

The surface S is *nearly normal* if each component of $S \cap H_0$ is a nearly normal piece for each 0-handle H_0 of \mathcal{H} .

Lemma 5.11 (Bound on nearly normal pieces). *A tetrahedral or semi-tetrahedral 0-handle contains only finitely many nearly normal pieces up to normal isotopy. Moreover, each nearly normal piece intersects the union of the 1-handles in at most 12 arcs or curves.*

Proof. A nearly normal piece in a 0-handle H_0 is parallel to a subsurface of ∂H_0 . This surface F is determined by its boundary curves ∂F , up to two choices. For any 1-handle H_1 , the intersection between a component of $H_0 \cap H_1$ and ∂F is at most three arcs or is a simple closed curve, because otherwise the 3-ball B in the definition of a nearly normal surface would intersect this component of $H_0 \cap H_1$ in more than one component or in something other than a disc. Hence, there are only finitely many possibilities for ∂F up to normal isotopy. Since H_0 intersects the 1-handles in at most 4 discs, we deduce that the number of components of intersection between ∂F and these 1-handles is at most 12. \square

We will see that the isotopies we have introduced either take an almost normal surface to a nearly normal one, or take a nearly normal surface to a nearly normal one. We prove this here for the case of the compression isotopy. We will save the proof for isotopies along face and edge compression discs for a slightly more general setting in the next section.

Lemma 5.12 (Compression isotopy preserves nearly normal). *Let S be a nearly normal surface, and let S' be the result of applying a compression isotopy along a disc D that lies in the interior of the intersection of a 1-handle H_1 with a 0-handle H_0 . Then S' is nearly normal.*

Proof. Nearly normal pieces are removed in their entirety if they were a subset of the disc in S bounded by ∂D , but removing pieces does not affect whether a surface is nearly normal. One of the two nearly normal pieces that are incident to $(D^1 \times D) \cap H_1$ is removed, the other is modified. Let F be the nearly normal piece of S that is changed, by adding a parallel copy of D . This piece divides the 0-handle H_0 that contains it into two components, one of which is the 3-ball B . The topology of B remains a ball after this modification. The other component $H_0 \setminus B$ remains a product. The intersection between B and any 1-handle or 2-handle remains unchanged, except for the component of $H_0 \cap H_1$ containing D , where the intersection becomes empty. These are the only changes that are made to the surface, and so the surface S' is nearly normal. \square

Recall from Definition 4.17 that the weight of a surface is its number of components of intersection with the 2-handles.

Lemma 5.13 (Effect on weight, compression isotopy). *Suppose S' is obtained from S by a compression isotopy. Then the weight of S' is at most that of S .*

Proof. For a compression isotopy, a disc D parallel to the compression disc replaces a disc D' on the surface S , and otherwise S' agrees with S . Thus no new intersections with 2-handles are introduced. If D' meets 2-handles, the components of intersection with these 2-handles are removed. In any case, the weight does not increase. \square

There is one more move on almost normal surfaces that we need to introduce, as follows.

Definition 5.14 (Tube compression). Suppose S is an almost normal incompressible surface with a tubed piece. Suppose S is transversely oriented in the direction that points into the tube. Due to the incompressibility of S and the irreducibility of M , compressing this tube yields a 2-sphere plus an isotopic copy of S , which we denote by S' . The isotopy that we perform moves S across to S' . We call this a *tube compression*.

Lemma 5.15 (Tube compression gives normal). *A tube compression takes an almost normal surface to a normal surface, hence a nearly normal surface.*

Proof. Compressing a tubed piece gives two normal discs, one of which belongs to the surface after tube compression. Since the remaining pieces are unchanged, the result is normal. \square

We generally assume that if S contains a tubed piece, then its orientation points out of the tube, because otherwise we immediately perform a tube compression and no longer have such a piece.

Proposition 5.16 (Unsimplifiable implies normal). *Let S be a closed incompressible surface properly embedded in a compact orientable irreducible 3-manifold M equipped with a pre-tetrahedral handle structure \mathcal{H} . Suppose that S is a transversely oriented nearly normal surface that admits no face compression disc in the specified transverse orientation and admits no compression isotopies. Suppose also that S has no component that is a 2-sphere lying in a 0-handle. Then S is normal.*

Proof. We check the conditions of Definition 4.8. First, by definition of nearly normal, S respects the handle structure \mathcal{H} .

Consider a 0-handle H_0 . We first show that for any 1-handle H_1 incident to H_0 , S intersects $H_1 \cap H_0$ in normal arcs. Suppose not. Then at least one component of intersection with $H_1 \cap H_0$ is a simple closed curve or an arc with endpoints on the same 2-handle meeting H_1 . Consider an innermost simple closed curve in $H_1 \cap H_0$. This bounds a disc D . The component of $S \cap H_0$ containing ∂D is a nearly normal piece. This divides H_0 into two components, one of which is the 3-ball B in the definition of nearly normal. The intersection between B and the face is a disc, and hence it must be D . So, S points into D and so we can perform a compression isotopy, which is contrary to assumption.

So now consider an arc of intersection between S and $H_1 \cap H_0$ that is outermost in $H_1 \cap H_0$ with endpoints on the same 2-handle H_2 . This separates off a face compression disc D . This arc lies in a nearly normal piece, which again divides H_0 into two components, one of which is the 3-ball B . Since B has connected intersection with each component of $H_2 \cap H_0$, we deduce that D lies in B . Hence, we have a face compression disc in the specified transverse direction, which again is a contradiction.

Now consider any piece F of $S \cap H_0$ for some 0-handle H_0 . We claim that F is a disc. Since F is nearly normal, it is parallel to a subsurface of ∂H_0 and hence it is planar. We are assuming that F is not a 2-sphere. So if F is not a disc, then it has at least two boundary components. These are normal curves, and hence we may find a component σ of intersection between H_0 and a 1-handle that intersects two different boundary components of F . We may find an embedded arc α in σ with interior disjoint from F and with endpoints lying on two different components of ∂F . Let B be the 3-ball in the definition of nearly normal. Then α lies in B because otherwise $B \cap \sigma$ is disconnected. We may find an arc β properly embedded in F joining the two endpoints of α . Then $\alpha \cup \beta$ is a simple closed curve in ∂B that intersects the two boundary components of F in one point each. This is a contradiction, because any simple closed curve in a sphere is separating.

We now show that S is normal. We have shown that each component F of $S \cap H_0$ is a disc and that its boundary curve is a union of normal arcs. If F runs over any of the 2-handles meeting H_0 in more than one arc, then its boundary has length at least 8 ([26, Section 2.2]), and for each component of $H_0 \setminus F$, we can then find a 1-handle meeting H_0 that intersects this component in at least two discs. This contradicts the definition of being nearly normal. \square

5.1. Canonical Handle Structures. We have defined isotopy moves of surfaces that take almost normal surfaces through nearly normal surfaces to normal surfaces. Each of these surfaces intersects the handle structure of the larger 3-manifold in well-understood ways. Recall that eventually, we wish to count edge contractions and expansions on spines of surfaces as we pass from $S \times \{0\}$ to $S \times \{1\}$. Some of the normal, almost normal, and nearly normal surfaces we have encountered under the above isotopy moves will be the surfaces we analyse. In order to transfer spines, however, we need to determine how a cell structure on the surface itself changes. We start the process in this subsection.

Definition 5.17 (Canonical handle structure, embedded surface). Let \mathcal{H} be a handle structure for a 3-manifold M , and let S be an incompressible surface properly embedded in M that respects \mathcal{H} and

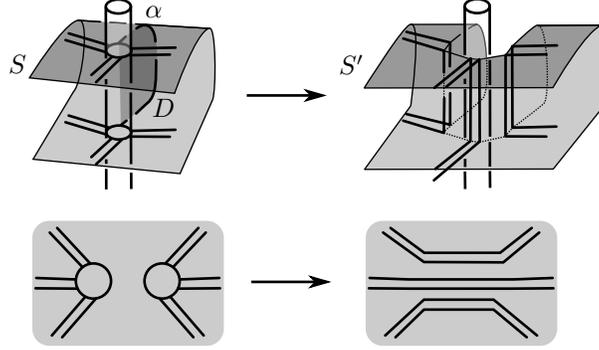


FIGURE 11. On the left is the surface S admitting an edge compression disc. At the top, it is shown in M , below is shown its inherited canonical handle structure. On the right is the surface S' after isotopy along the edge compression disc, with its canonical handle structure.

that has no 2-sphere components. Let F be the intersection between the 0-handles of \mathcal{H} and S . If any component of F is not a disc, then its boundary curves all bound discs in S , since S is incompressible. These discs may be nested. Under these circumstances, enlarge F by including these discs, forming a surface F^+ . The *canonical handle structure* on S has 0-handles equal to the components of F^+ . The 1-handles are the components of intersection between the 1-handles of \mathcal{H} and $S \setminus F^+$. The 2-handles are the components of intersection between the 2-handles of \mathcal{H} and $S \setminus F^+$.

Examples of canonical handle structures before and after an isotopy along an edge compression disc are shown in Figure 11. Note that, by Lemma 5.11, each 0-handle of the canonical handle structure on a nearly normal surface has at most 12 components of intersection with the 1-handles and at most 12 components of intersection with the 2-handles.

Lemma 5.18 (Compression isotopy and canonical handle structure). *Let S be a nearly normal surface, and let S' be obtained from it by a compression isotopy. Then the compression isotopy takes the canonical handle structure on S to the canonical handle structure on S' .*

Proof. Let D and D' be the discs in Definition 5.8. Let H_0 be the 0-handle of \mathcal{H} containing D . The disc D forms a compression disc for a component of $H_0 \cap S$, and hence D' lies in the interior of a 0-handle of the canonical handle structure on S . The isotopy moves D' across to D and then pushes it a little further off the 1-handle containing D . Thus after this isotopy, the image of D' lies within a 0-handle of the canonical handle structure on S' . The remainder of the canonical handle structures on S and S' are equal. \square

Lemma 5.19 (Edge or face compression and canonical handle structure). *Suppose S is an almost normal or nearly normal surface, and S' is a nearly normal surface obtained from S by isotopy along an edge or face compression disc D , with $\partial D = \alpha \cup \beta$ as in Definitions 5.4 and 5.5. Suppose that α does not lie entirely in the interior of a 0-handle of the canonical handle structure of S , in other words, that α does not lie in the interior of a component of F_+ as in Definition 5.17. Suppose also that we are not performing an edge compression to a tubed piece. Then the canonical handle structure of S' is obtained from that of S as follows:*

- (1) Remove the 2-handles at the end of α .
- (2) If α lies in the boundary of a 0-handle of S (i.e. D is a face compression disc), remove the 1-handle containing α .
- (3) Join each 1-handle at one end of α to a 1-handle at the other end, pairing those that lie in the same 1-handle of \mathcal{H} .
- (4) The spaces between these 1-handles become part of 0-handles in the new handle structure.

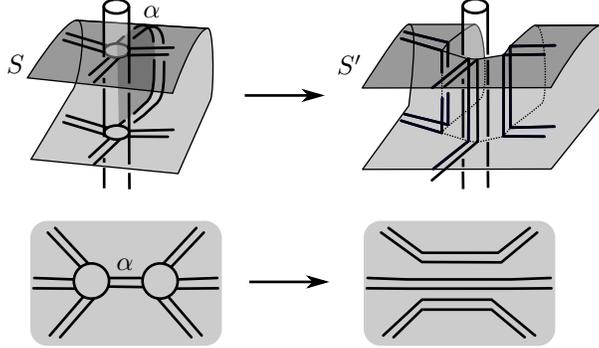


FIGURE 12. How the canonical handle structure on S is affected by an isotopy along a face compression disc.

- (5) However, if this results in any regions that are not discs, these are replaced by discs that each form a single 0-handle.

Proof. The proof follows from the definition of isotopy along an edge or face compression disc, Definition 5.6. See Figures 11 and 12. \square

We now consider the cases excluded in the above lemma.

Lemma 5.20 (Edge or face compression in 0-handle). *Suppose S is an almost normal or nearly normal surface, and S' is a nearly normal surface obtained from S by isotopy along an edge or face compression disc D , with $\partial D = \alpha \cup \beta$ as in Definitions 5.4 and 5.5. If α lies entirely in the interior of a 0-handle of the canonical handle structure of S , then the isotopy taking S to S' transfers the canonical handle structure of S to the canonical handle structure of S' .*

Proof. The assumption of the lemma implies that α lies in a disc in S with boundary lying in a 0-handle of \mathcal{H} . The isotopy from S to S' takes this disc to a corresponding disc in S' . Hence, the 0-handle of the canonical handle structure containing α is sent to a 0-handle of S' , and all the other handles remain unchanged. \square

Lemma 5.21 (Edge compression in tubed piece). *Suppose S is an incompressible almost normal surface with a tubed piece, and S' is a nearly normal surface obtained from S by isotopy along an edge compression disc D , with $\partial D = \alpha \cup \beta$ as in Definitions 5.4 and 5.5. Then the canonical handle structure on S' is obtained from that of S as follows. Let E be a disc in S whose boundary ∂E is a simple closed curve forming a core curve of the tube. Components of intersection between E and the 1-handles and 2-handles of \mathcal{H} have been removed to form the canonical handle structure on S . These are reinstated as 1-handles and 2-handles. Then steps (1) to (5) of Lemma 5.19 are applied.*

Proof. Definition 5.17 ensures that $E \subset S$ lies in a 0-handle of the canonical handle structure of S , with all 1-handles and 2-handles of E removed. Applying the isotopy along the edge compression disc may adjust this structure. However, after reinstating the 1- and 2-handles of E , the adjustment will then be identical to that of Lemma 5.19. \square

6. PARALLELITY BUNDLES AND GENERALISED ISOTOPY MOVES

Road map: We are working towards a bound on the number of edge expansions and contractions required to transfer a spine on $S \times \{0\}$ to $S \times \{1\}$. We will be transferring the spines along a sequence of surfaces that consist of normal, almost normal, and nearly normal surfaces as defined in the previous section, and we will bound the number of edge contractions and expansions at each step. A consequence of Proposition 5.16 is that compression isotopies and isotopies along edge and face compression discs take an almost normal surface to a normal one. So this will give us a complete

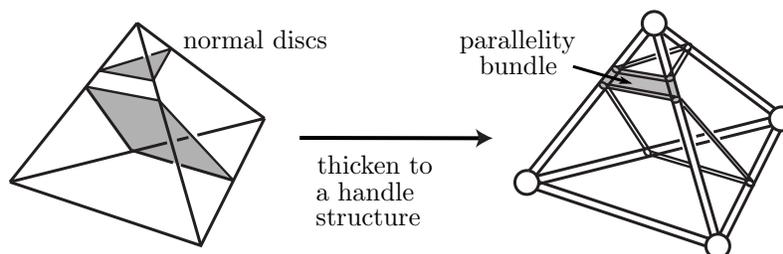


FIGURE 13. A portion of the parallelity bundle

sequence of surfaces to work with. Unfortunately, using these moves alone gives far too many surfaces. There is no good way to bound the number of such surfaces in a sequence; without a bound on the number of surfaces, having a bound on edge expansions and contractions at each step will not lead to the bound we need. Therefore, we need to introduce more drastic moves. This section introduces the moves and the necessary terminology.

6.1. Parallelity bundles. Consider an isotopy across a face compression disc, as in Definition 5.6. This pushes a surface S past a 2-handle. If on the other side of that 2-handle, S cuts the handles of M into parallelity handles, then this isotopy move will give rise to new face compression discs, and the move must be repeated to slide S . There is no good bound on the number of parallelity handles that $M \setminus S$ may contain, and it could be the case that each one leads to a required isotopy across a face compression disc. Instead of performing these isotopies one by one, we want to perform them all in a single step that depends on the parallelity handles adjacent to the face compression disc.

In this subsection, we define parallelity bundles and their generalisations, which will allow us to define the more general isotopy moves that we need. These bundles first appeared in [16]. The idea is as follows. Normal surfaces are made up of triangles and squares in tetrahedra. We cannot bound the number of normal triangles and squares in the intersection of a normal surface S with a given tetrahedron. But if S intersects a single tetrahedron many times, then it must do so in more and more parallel triangles and squares. These cut the tetrahedron, and hence the 3-manifold $M \setminus S$, into I -bundles.

Throughout this section, let \mathcal{H} be a handle structure of a 3-manifold M . Let S be a surface properly embedded in M that respects the handle structure.

Definition 6.1 (Parallelity bundle for S). The *parallelity bundle* \mathcal{B} for S is the submanifold of $M \setminus S$ that is the union of the parallelity regions, as in Definition 4.21.

When \mathcal{H} is dual to a triangulation \mathcal{T} of M , then it is possible to visualise the parallelity bundle for a surface S in \mathcal{T} without specifically having to pass to the dual handle structure. If we have two adjacent normally parallel pieces of S in a tetrahedron, the space between them forms a parallelity region in $M \setminus S$. Similarly, if the intersection between S and a face of \mathcal{T} contains two adjacent parallel arcs, the space between them, thickened a little, forms a parallelity region. When S intersects an edge of \mathcal{T} , it divides the edge into arcs, and all but the outermost two arcs produce parallelity regions in the dual 2-handle of \mathcal{H} . See Figure 13.

We also define parallelity handles in more general handle structures \mathcal{H}' for a 3-manifold M' that is not necessarily of the form $M \setminus S$, and where a specified subsurface F of $\partial M'$ is given.

Definition 6.2 (Parallelity bundle, handle structure). Let \mathcal{H}' be a handle structure of a compact orientable 3-manifold M' , and let F be a subsurface of $\partial M'$ such that ∂F is standard. A handle H of \mathcal{H}' is a *parallelity handle* if it admits a product structure $D^2 \times I$ such that

- (1) $D^2 \times \partial I = H \cap F = H \cap \partial M'$;
- (2) each component of intersection between H and another handle is $\beta \times I$ for a subset β of ∂D^2 .

The *parallelity bundle* \mathcal{B} for \mathcal{H}' is the union of the parallelity handles.

It is shown in [16, Lemma 5.3] that the product structures on the parallelity handles can be chosen so they make the parallelity bundle into an I -bundle over a surface E .

Definition 6.3 (Horizontal and vertical boundary). The *horizontal boundary* $\partial_h \mathcal{B}$ is the (∂I) -bundle over E . It is a subsurface of F . The *vertical boundary* $\partial_v \mathcal{B}$ is the I -bundle over ∂E . It is an orientable I -bundle properly embedded in M' , and hence it is a collection of annuli.

We will want to expand the parallelity bundle to include not just the parallelity handles, but also other simple I -bundles whose fibring matches that of the parallelity bundle. This is done by the following definition.

Definition 6.4 (Generalised parallelity bundle). Let M' be a compact orientable 3-manifold with a handle structure \mathcal{H}' . Let F be a subsurface of $\partial M'$ such that ∂F is standard. A *generalised parallelity bundle* \mathcal{B} is a 3-dimensional submanifold of M' such that

- (1) \mathcal{B} is an I -bundle over a compact surface;
- (2) the horizontal boundary of \mathcal{B} is the intersection between \mathcal{B} and F ;
- (3) \mathcal{B} is a union of handles of \mathcal{H}' ;
- (4) any handle of \mathcal{B} that intersects $\partial_v \mathcal{B}$ is a parallelity handle, where the I -bundle structure on the parallelity handle agrees with the I -bundle structure of \mathcal{B} ;
- (5) whenever a handle of \mathcal{H}' lies in \mathcal{B} then so do all incident handles of \mathcal{H}' with higher index;
- (6) the intersection between $\partial_h \mathcal{B}$ and the non-parallelity handles lies in a union of disjoint discs in the interior of $\partial_h \mathcal{B}$.

We also say that \mathcal{B} is a *generalised parallelity bundle for* (M', F) , when we wish to emphasize the manifold and surface. When we do not specify F , we take F to be all of $\partial M'$.

A generalised parallelity bundle is *maximal* if it is not a proper subset of another generalised parallelity bundle.

Of course, the parallelity bundle for \mathcal{H}' is a generalised parallelity bundle. But the more general concept is useful. For example, suppose that a vertical boundary component of the parallelity bundle \mathcal{B} separates off a component of $M' \setminus \mathcal{B}$ that is an I -bundle over disc, with the (∂I) -bundle lying in F , and that this I -bundle structure extends that on \mathcal{B} . Then it is natural to enlarge \mathcal{B} by including the I -bundle over the disc. The result is a generalised parallelity bundle.

Note that condition (6) in the above definition is new to this paper; it does not appear in [16, Definition 5.2].

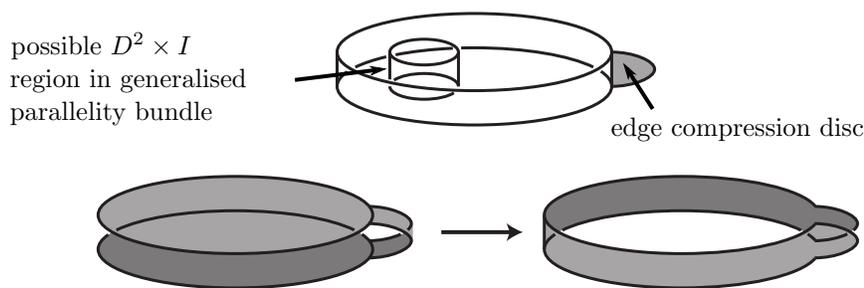
6.2. Isotopy moves across I -bundles over discs. We now generalise Definition 5.6. Let \mathcal{H} be a pre-tetrahedral handle structure of $M = S \times [0, 1]$. Let S be a normal or almost normal fibre. Let \mathcal{B} be a maximal generalised parallelity bundle for $M \setminus S$.

We suppose that we have performed some isotopies to S , all in the same transverse direction, taking it to a nearly normal surface S' . We also allow the possibility that no isotopies have been performed, and hence that $S' = S$, which is normal or almost normal.

Definition 6.5 (Generalised edge or face compression). Suppose S' is nearly normal and D is a face compression disc for S' , or suppose S' is almost normal and D is an edge compression disc for S' . Suppose that the interior of D is disjoint from \mathcal{B} . Its intersection with a 2-handle is an arc (denoted β in Definition 5.5) in the vertical boundary of \mathcal{B} . Let W be the isotopy ball for D . Let B be the component of \mathcal{B} incident to D . Suppose that B is an I -bundle over a disc and that its horizontal boundary lies in S' . Then the *generalised isotopy move across B* moves S' across the ball $W \cup B$. Thus, it removes the horizontal boundary of B from S' , together with $S' \cap W$, and it replaces it with $\partial_v B \setminus W$ together with two parallel copies of D . See Figure 14.

We also call this move a *generalised isotopy along D* , where D is an edge compression disc or face compression disc. We also occasionally refer to it for short as a *generalised edge or face compression*.

Note that the 2-handle meeting the arc β is a parallelity handle. If this handle is not adjacent to any other parallelity handles, then the usual isotopy along an edge or face compression disc, as in


 FIGURE 14. An generalised isotope move across an I -bundle over a disc

Definition 5.6, is an example of a generalised isotope along D , where the component B consists just of the parallelity 2-handle meeting β .

The effect on the canonical handle structure on the surface is nearly identical to that in the usual isotope along D , as follows.

Lemma 6.6 (Effect of generalised isotope on canonical handle structure). *Suppose S' is a nearly normal or almost normal surface, and S'' is a nearly normal surface obtained from S' by a generalised isotope along an edge or face compression disc D , with $\partial D = \alpha \cup \beta$ as in Definitions 5.4 and 5.5. Then the canonical handle structure of S'' is obtained from that of S' as follows:*

- *If α lies entirely in the interior of a 0-handle of the canonical handle structure of S' , then the handle structure of S'' is unchanged.*
- *If the generalised isotope is not along an edge compression disc running over a tubed piece, then the canonical handle structure of S'' is obtained completely analogously to steps (1) to (5) of Lemma 5.19:*
 - (1) *Remove from S' all handles in the horizontal boundary $\partial_h B$ of B in S' .*
 - (2) *If D is a face compression disc, remove the 1-handle containing α .*
 - (3) *Each 1-handle adjacent to one component of $\partial_h B$ is paired with a 1-handle adjacent to the other component of $\partial_h B$, where the two handles lie in the same handle of \mathcal{H} , both adjacent to the same parallelity handle on ∂B .*
 - (4) *The spaces between these 1-handles become part of 0-handles in the new handle structure.*
 - (5) *However, if this results in any regions that are not discs, these are replaced by discs that each form a single 0-handle.*
- *If the generalised isotope is along an edge compression disc running over a tubed piece, then as in Lemma 5.20, reinstate 1-handles and 2-handles in a disc on S' with boundary a core curve of the tube, and then apply steps (1) through (5) above.*

Proof. This follows by definition. See Figure 15. □

We now show that these moves result in a nearly normal surface.

Lemma 6.7 (Generalised face compression preserves nearly normal). *Let S' be a nearly normal surface, and let D be a face compression disc for S' . Let S'' be the result of a generalised isotope of S' along D . Then S'' is nearly normal.*

Proof. Let H_1 denote the 1-handle containing D , and H_2 the 2-handle meeting an arc β of ∂D . The surface S'' is obtained from S' by modifying it in the two 0-handles incident to H_1 , and in all handles in a component W of a generalised parallelity bundle that is an I -bundle over a disc, and in handles adjacent to $\partial_v W$.

Within H_2 and W , all intersections are removed in all handles. Within H_1 , the component of $H_1 \cap S'$ incident to D is removed. So there is nothing to consider in these handles.

Let us first consider a 0-handle H_0 that intersects $D^1 \times D \subset H_1$. Let F be the nearly normal piece of S' incident to D . Then, by the definition of nearly normal, F divides H_0 into two components,

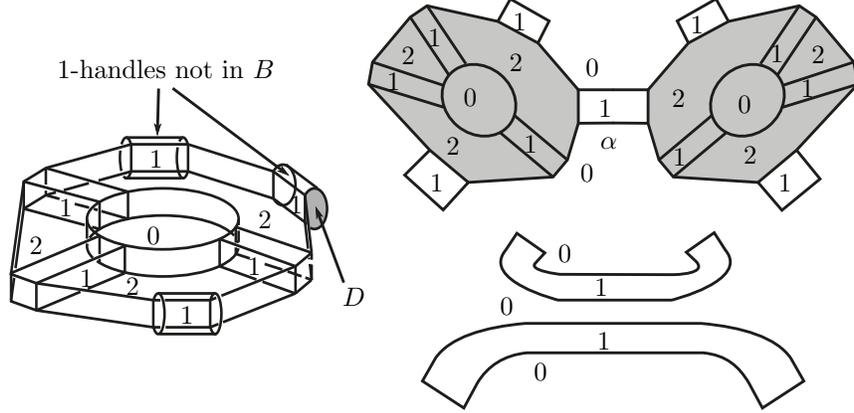


FIGURE 15. Left: A component of the generalised parallelity bundle that is an I -bundle over a disc. Right: The effect on the canonical handle structure of a generalised isotopy move across this component. The numbers show the index of the handles. The shaded region is the horizontal boundary of the bundle.

one of which is a 3-ball B containing D . When we perform the isotopy, B is reduced by removing a thin regular neighbourhood of $(D^1 \times D) \cap \partial H_0$, and of $(D^2 \times \beta) \cap \partial H_0$, but it remains a 3-ball B' say. The intersection between B' and the 1-handle H_1 meeting D is now empty. There will be another 1-handle adjacent to H_2 on ∂H_0 , and the intersection of B' with this 1-handle has shrunk, by removing a small regular neighbourhood of $D^2 \times \beta \subset H_2$, but it remains a disc.

The region $H_0 \setminus B'$ is obtained from $H_0 \setminus B$ by attaching a thin regular neighbourhood of $(D^1 \times D) \cap \partial H_0$ and $(D^2 \times \beta) \cap \partial H_0$, and hence it remains a product region. A region of B meeting $D^2 \times \beta \subset H_2$ is removed in this operation, but the remaining components of intersection with 2-handles are unchanged. Thus, the resulting piece of S'' is still nearly normal.

Let us now consider a 0-handle H_0 that is incident to the I -bundle over a disc W , but does not meet $D^1 \times D$. Consider any nearly normal piece F of S' in H_0 that is modified. It is altered by banding it onto another nearly normal piece F' in H_0 , or possibly banding it to itself. Then F (respectively F') divides H_0 into two regions, one of which is a ball B (respectively, B') and the other of which is a product region. The transverse orientations on F and F' point into B and B' , by the definition of a nearly normal piece. So the band is disjoint from the product regions, and hence when the band is attached to form a piece F'' , one component of $H_0 \setminus F''$ is again a product region between F'' and a subsurface of ∂H_0 . Hence, the other component is homeomorphic to H_0 and therefore a ball B'' . The intersection between B'' and any component σ of intersection between H_0 and a 1-handle must be empty or connected, for the following reason. The intersection between B and σ is empty or a disc, by hypothesis, as is $B' \cap \sigma$. Thus $B' \cap B \cap \sigma$ is empty or a disc. The intersection $B'' \cap \sigma$ either is equal to $B' \cap B \cap \sigma$ or is obtained from $B' \cap B \cap \sigma$ by removing a thin strip running along $\sigma \cap \partial_v W$, and hence remains empty or a disc. The intersection between B and any other 2-handle H'_2 meeting ∂H_0 is empty or a product $[0, 1] \times a$, where a is an arc. The same is true for $B' \cap H'_2$. Hence, $B' \cap B \cap H'_2$ is empty or a product $[0, 1] \times (a \cap a')$, and therefore so is $B'' \cap H'_2$.

In the above argument, we assumed that the intersection of $((D^1 \times D) \cup (D^2 \times \beta))$ with H_0 was either equal to a disc in a single 1-handle $H_1 \cap H_0$ and an adjacent 2-handle $H_2 \cap H_0$, or to a subset of a single 2-handle, $(D^2 \times \beta) \cap \partial H_0$. However, the boundary of the 0-handle H_0 might actually meet the same 1-handle H_1 twice, or the same 2-handle H_2 multiple times. Thus $(D^1 \times D) \cup (D^2 \times \beta)$ may actually have several components of intersection on ∂H_0 of this form. But the conclusion of the argument remains unchanged, because we can view the modification from $S' \cap H_0$ to $S'' \cap H_0$ as being achieved in several steps, each of the form discussed above. \square

Lemma 6.8 (Generalised edge compression gives nearly normal). *Let S be a transversely oriented almost normal surface. If S contains a tubed piece, suppose that its orientation points out of the tube. Let S' be obtained from S by a generalised isotopy along an edge compression disc. Then S' is nearly normal.*

Proof. The edge compression disc D must be incident to the almost normal piece of S . The effect on this almost normal piece yields a nearly normal piece. Other pieces of S are also affected, if they are incident to the I -bundle over a disc that is adjacent to D . A band is added to these pieces. As argued above in the proof of Lemma 6.7, attaching a band in this way does not alter the fact that the pieces are nearly normal. \square

Lemma 6.9 (Generalised isotopy along a disc reduces weight). *Suppose S'' is obtained from S' by applying a generalised isotopy along an edge or face compression disc. Then the weight of S'' is strictly less than that of S' .*

Proof. Let B denote the I -bundle over a disc in Definition 6.5. It is a component of a generalised parallelity bundle \mathcal{B} , and by definition whenever a handle lies in \mathcal{B} , so do all incident handles of higher index. Since 2-handles are the highest index parallelity handle, B must meet a 2-handle. Then S' , which runs over the horizontal boundary of B , must meet the same 2-handle. When we apply the isotopy move, the new surface S'' becomes disjoint from all 2-handles in B , and does not meet new 2-handles. Thus the weight strictly decreases. \square

6.3. Parallelisation isotopy. Again let \mathcal{H} be a pre-tetrahedral handle structure of $M = S \times [0, 1]$. Let S be an almost normal fibre.

Definition 6.10 (Parallelisation isotopy). Let S' be a normal fibre that is topologically parallel to S . Suppose that there is a copy of $D^2 \times [0, 1]$ in $M \setminus (S \cup S')$, with $D^2 \times \{0\} \subset S$, $D^2 \times \{1\} \subset S'$ and $\partial D^2 \times [0, 1]$ a vertical boundary component of the parallelity bundle of $M \setminus (S \cup S')$. Then a *parallelisation isotopy* moves S across $D^2 \times [0, 1]$, taking it to S' . When S has transverse orientation, we require that this points into $D^2 \times [0, 1]$ so that the isotopy moves S in this direction.

We do not require that a parallelisation isotopy reduces the weight of S , and so there is no analogue of Lemma 6.9 in this setting. By assumption, S' is normal, and so the conclusion of Lemma 6.8 trivially holds.

The motivation for introducing parallelisation isotopies comes from the following.

Lemma 6.11 (Almost normal piece in bundle). *Let \mathcal{H} be a pre-tetrahedral handle structure for $M = S \times [0, 1]$. Let S be an almost normal fibre, transversely oriented in some way. Let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points. Suppose that the almost normal piece of S intersects the interior of $\partial_h \mathcal{B}$. Then S admits a parallelisation isotopy.*

Proof. The almost normal surface S has only one almost normal piece P . Hence, P cannot be part of the parallelity bundle of $M \setminus S$. However, we are assuming that it intersects the interior of the horizontal boundary of the maximal generalised parallelity bundle \mathcal{B} . By (6) in the definition of a generalised parallelity bundle, Definition 6.4, the surface that is obtained by removing the parallelity handles from $\partial_h \mathcal{B}$ lies in a union of disjoint discs in the interior of $\partial_h \mathcal{B}$. The almost normal piece P must lie in one of these discs. The I -bundle structure on \mathcal{B} gives a copy of $D^2 \times [0, 1]$ in \mathcal{B} , where $D^2 \times \{0\}$ is the disc containing P , $D^2 \times \{1\}$ also lies in $\partial_h \mathcal{B}$ and $\partial D^2 \times [0, 1]$ is a vertical boundary component of the parallelity bundle. Thus, we can perform the parallelisation isotopy that moves S across $D^2 \times [0, 1]$. \square

6.4. Annular simplification. In addition to isotopy moves across I -bundles over a disc and parallelisation isotopies, we also need moves that can be applied to generalised parallelity bundles that are I -bundles over annuli. The simplification will be applied both to 3-manifolds and to surfaces; both definitions are below.

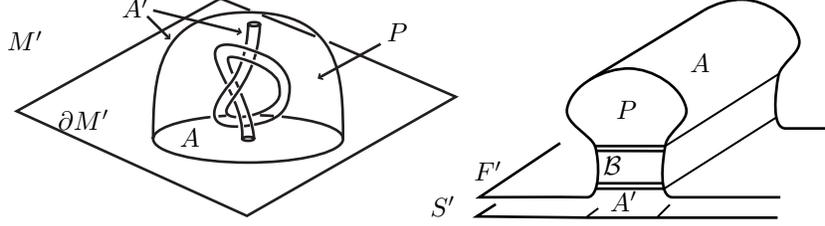


FIGURE 16. An annular simplification removes P and replaces A with A' . Left: a trivial annular simplification. Right: a cross-section of an essential annular simplification.

Definition 6.12 (Annular simplification). Let M' be a compact orientable irreducible 3-manifold with a handle structure \mathcal{H}' , and let F be an incompressible surface of $\partial M'$ such that ∂F is standard and F is not a 2-sphere.

Suppose M' contains the following:

- an annulus A' that is a vertical boundary component of a generalised parallelity bundle \mathcal{B} ;
- an annulus A contained in F such that $\partial A = \partial A'$;
- a 3-manifold P with $\partial P = A \cup A'$ such that P either lies in a 3-ball or is a product region between A and A' .

Suppose also that P is a union of handles of \mathcal{H}' , that whenever a handle of \mathcal{H}' lies in P , so do all incident handles with higher index, and that any parallelity handle of \mathcal{H}' that intersects P lies in P . Finally, suppose that apart from the component of the generalised parallelity bundle incident to A' , all other components of \mathcal{B} in P are I -bundles over discs.

An *annular simplification* of the 3-manifold M' is the manifold obtained by removing the interiors of P and A from M' . When P lies in a 3-ball, it is a *trivial annular simplification*. On the other hand, when A is an essential subsurface of $\partial M'$, it is an *essential annular simplification*. Note that the resulting 3-manifold is homeomorphic to M' , even though P may be homeomorphic to the exterior of a non-trivial knot. See Figure 16.

Suppose that the component F' of F containing A is transversely oriented, with orientation pointing into M' . An *annular simplification* of the surface F' is obtained by replacing A in F' by A' , and then isotoping slightly past the handles containing A' , in the direction of the orientation, obtaining a surface S' .

Lemma 6.13 (Annular simplification and canonical handle structure). *Let S' be obtained from a nearly normal surface S by an annular simplification. This replaces an annulus A in S with an annulus A' in S' . Then the effect on the canonical handle structure of S is as follows:*

- (1) Remove all handles in A .
- (2) Adjacent to each component of ∂A , in $S \setminus A$, there are 0-handles and 1-handles in an alternating fashion. Each such 0-handle or 1-handle adjacent to one component of ∂A corresponds to a 0-handle or 1-handle adjacent to the other component of ∂A , where the correspondence arises from the I -bundle structure on A' . Combine each pair of 0-handles into a single 0-handle of S' , and do the same for each pair of 1-handles.
- (3) If this results in any regions that are not discs, these are replaced by discs that each form a single 0-handle F^+ , as in Definition 5.17.

Proof. This follows from the definition. □

Analogous to other isotopy moves, an annular simplification preserves nearly normal surfaces and decreases weight.

Lemma 6.14 (Annular simplification preserves nearly normal). *Let S be a nearly normal surface, and let S' be obtained from S by applying an annular simplification. Then S' is nearly normal.*

Proof. We use the terminology of Definition 6.12: P is a submanifold of $M \setminus S$, lying between an annulus A in S and an annulus A' properly embedded in $M \setminus S$. The modification to S is the removal of A and the addition of A' . This annulus A' is a vertical boundary component of the parallelity bundle of $M \setminus S$, and the component of the parallelity bundle that contains A' lies in P . The effect of the removal of A and P is to remove some nearly normal pieces of S . We must consider what happens when we add A' .

Let H_0 be a 0-handle of \mathcal{H} that has non-empty intersection with A' . The intersection between A' and H_0 is a union of fibres in the I -bundle structure on A' . These therefore form bands that are added to nearly normal pieces of $S \cap H_0$. In the definition of an annular simplification, the transverse orientation on the new surface S' is such that it points out of P along A' . Thus, just as in the case of an isotopy along a face compression disc, Lemma 6.7, the new surface S' remains nearly normal. \square

Lemma 6.15 (Annular simplification reduces weight). *Suppose S' is obtained from S by an annular simplification. Then the weight of S' is strictly less than that of S .*

Proof. Let A' be an annulus as in the definition of an annular simplification, which replaces the annulus A in S . If A' meets any 2-handles, then since $\partial A = \partial A' \subset S$, the surface S meets the same 2-handles. When we replace A with A' , and then isotope past the handles containing A' , we will remove those intersections with 2-handles without adding any additional intersections with 2-handles. Thus if we can show that A' runs through some 2-handle, we will have shown that the weight decreases strictly. But A' lies in the vertical boundary of a generalised parallelity bundle \mathcal{B} , and by definition whenever a handle lies in \mathcal{B} , then so do all incident handles of higher index. Hence, as one travels around A' , the handles of \mathcal{B} that are incident to A' are alternately 1-handles and 2-handles. In particular, A' meets at least one 2-handle. \square

The left side of Figure 16 shows a special type of parallelity bundle, which we now define.

Definition 6.16 (Boundary trivial). Let M' and F be as in Definition 6.4, and let \mathcal{B} be a generalised parallelity bundle for (M', F) . A component P of \mathcal{B} is *boundary-trivial* if the following all hold.

- (1) P lies within a 3-ball B such that $B \cap \partial M'$ is a single disc lying in F ,
- (2) the disc $\partial B \setminus \partial M'$ is disjoint from any component of \mathcal{B} that is not an I -bundle over a disc,
- (3) and P is not an I -bundle over a disc.

Lemma 6.17 (Boundary trivial admits annular simplification). *Let M' and F be as in Definition 6.12. Let \mathcal{B} be a maximal generalised parallelity bundle for (M', F) . Suppose that a component of \mathcal{B} is boundary-trivial. Then M' admits an annular simplification.*

Proof. Let P be a component of a maximal generalised parallelity bundle \mathcal{B} that is boundary-trivial. By definition, P lies within a 3-ball B such that $B \cap \partial M'$ is a single disc lying in F , and $\partial B \setminus \partial M'$ is disjoint from any component of \mathcal{B} that is not an I -bundle over a disc. Choose P so that it is furthest from $\partial B \setminus \partial M'$, in the sense that if P' is any other component of \mathcal{B} in B that is not an I -bundle over a disc, then P does not separate P' from $\partial B \setminus \partial M'$.

Let A' be a component of the vertical boundary of P that is closer to $\partial B \setminus \partial M'$ than any other vertical boundary component of P . Then A' is properly embedded in M' with both boundary components in the disc $B \cap \partial M'$. If the components of $\partial A'$ are parallel within $B \cap \partial M'$, then they cobound an annulus A , and all the conditions on A and A' of Definition 6.12 are satisfied. Thus it admits an annular simplification.

If the components of ∂A are not parallel within $B \cap \partial M'$, then they bound disjoint discs. But then the two discs must form the horizontal boundary of an I -bundle over a disc with vertical boundary A' . Thus because \mathcal{B} is maximal, this I -bundle makes up P , and P is not boundary-trivial. \square

6.5. Generalised isotopy moves. At this point we have described four isotopies that involve drastic changes along parallelity bundles, namely generalised isotopies along edge or face compression discs, parallelisation isotopies and annular simplifications. In this section, we show that these isotopies

are, in some sense, exactly the right isotopies to move an almost normal surface to a normal surface in the presence of large parallelity bundles. We do this by first classifying all components of the maximal generalised parallelity bundle in Theorem 6.18. We then investigate properties of parallelity bundles that admit a simplification under one of our isotopy moves. We show that any such parallelity bundle either is boundary-trivial, or an I -bundle over a disc or annulus, in Proposition 6.20, which means that if the surface admits a simplification through a parallelity bundle, it will occur under one of the isotopies we have defined.

The following is a version of [16, Corollary 5.7].

Theorem 6.18 (Classification of generalised parallelity bundles). *Let M' be a compact orientable irreducible 3-manifold with a handle structure \mathcal{H}' , and let F be an incompressible surface of $\partial M'$ such that ∂F is standard and F is not a 2-sphere. Let \mathcal{B} be a generalised parallelity bundle for (M', F) that is maximal, in the sense that it is not a proper subset of another generalised parallelity bundle. Then \mathcal{B} contains every parallelity handle of \mathcal{H}' . Moreover, every component of \mathcal{B} is either*

- an I -bundle over a disc, or
- boundary-trivial, or
- has incompressible vertical and horizontal boundary.

Proof. Note first that \mathcal{B} must contain every parallelity handle of \mathcal{H}' , as otherwise we may take any component of the parallelity bundle of (M', F) that does not lie wholly in \mathcal{B} and add it to \mathcal{B} . It is easy to check that the result is still a generalised parallelity bundle, which contradicts the hypothesis that \mathcal{B} is maximal.

We divide \mathcal{B} into three subsets, each of which is a union of components of \mathcal{B} :

- (1) the union of the I -bundles over discs, denoted \mathcal{B}_D ;
- (2) the union of the boundary-trivial components, denoted \mathcal{B}_∂ ;
- (3) the union of the remaining components, denoted \mathcal{B}_I .

We will show that \mathcal{B}_I has incompressible vertical boundary and incompressible horizontal boundary.

Our first task is show that we may find a collection of disjoint properly embedded discs that separate \mathcal{B}_∂ from \mathcal{B}_I . These discs may intersect \mathcal{B}_D . The boundary of the discs will lie in F .

Let D be a collection of disjoint compression discs for $F - \partial_h(\mathcal{B}_\partial \cup \mathcal{B}_I)$ in the complement of $\mathcal{B}_\partial \cup \mathcal{B}_I$, which is maximal in the sense that any other compression disc that is disjoint from D is parallel to a component of D . Since F is incompressible and M' is irreducible, each component of D is parallel to a disc in F , via a 3-ball. Any two of these 3-balls must be disjoint or nested, because otherwise F is a 2-sphere. Hence the union of these 3-balls is a collection of balls B , each of which intersects F in a single disc. Note that $\partial B \setminus F$ is disjoint from $\mathcal{B}_\partial \cup \mathcal{B}_I$.

Any component of \mathcal{B} lying within B that is not an I -bundle over a disc must be boundary-trivial by definition. We now show that B contains \mathcal{B}_∂ . Suppose that, on the contrary, there is a component of \mathcal{B}_∂ that is disjoint from B . By definition, this lies in a ball B' such that $\partial B' \setminus \partial M'$ is a disc disjoint from $\mathcal{B}_\partial \cup \mathcal{B}_I$. By cutting and pasting, we may modify B' so that $\partial B' \setminus \partial M'$ is disjoint from D . By maximality of D , $\partial B' \setminus \partial M'$ must be parallel into D , via a product region disjoint from $\mathcal{B}_\partial \cup \mathcal{B}_I$. But then B' can be isotoped into B , without moving $\mathcal{B}_\partial \cup \mathcal{B}_I$, which contradicts our assumption that this component of \mathcal{B}_∂ is disjoint from B .

Suppose that the vertical boundary of \mathcal{B}_I is compressible. Let D' be a compression disc, with boundary lying in an annular component A of $\partial_v \mathcal{B}_I$. The interior of D' is disjoint from \mathcal{B}_I because no component of \mathcal{B}_I is I -bundle over a disc. By cutting and pasting, we may assume that D' is disjoint from the discs $\partial B \setminus F$. Then A compresses to two discs D_1 and D_2 . These are disjoint from $\mathcal{B}_\partial \cup \mathcal{B}_I$. They are parallel to discs D'_1 and D'_2 in F , and for $i = 1$ and 2 , $D_i \cup D'_i$ bounds a 3-ball B_i . The balls B_1 and B_2 are either disjoint or nested. We consider these two cases separately. The argument is illustrated in Figure 17.

If B_1 and B_2 are disjoint, then the union of B_1 , B_2 and a regular neighbourhood of D' forms a copy of $D^2 \times I$, with interior disjoint from \mathcal{B}_I and with $(D^2 \times I) \cap F = D^2 \times \partial I$. Therefore, we

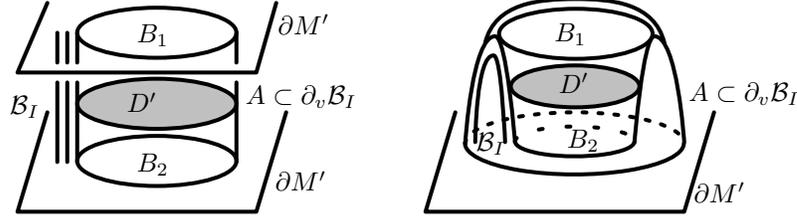


FIGURE 17. Left: if B_1 and B_2 are disjoint. Right: if $B_2 \subset B_1$.

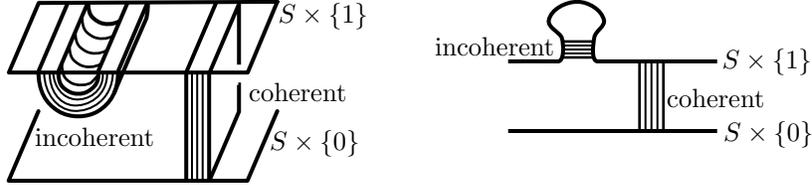


FIGURE 18. Schematic pictures of coherent and incoherent I -bundles

may enlarge \mathcal{B} by attaching $D^2 \times I$, forming a larger generalised parallelity bundle than \mathcal{B} . This contradicts the assumption that \mathcal{B} is maximal.

Suppose now B_1 and B_2 are nested. Say that B_2 lies in B_1 . Then because D' is disjoint from the interior of \mathcal{B}_I , we deduce that the component of \mathcal{B}_I containing A lies in B_1 . This implies that this component is boundary-trivial, which is a contradiction.

Thus, we have shown that every component of \mathcal{B} either is an I -bundle over a disc, is boundary-trivial, or has incompressible vertical boundary.

We will show that the horizontal boundary of \mathcal{B}_I is incompressible. We can assume that a compression disc D' is disjoint from the discs $D = \partial B \setminus F$. It must necessarily intersect the vertical boundary of \mathcal{B}_I . Then consider the curves of intersection between D' and $\partial_v \mathcal{B}_I$. Any curve can either be removed by an isotopy, or it bounds a disc in D' that forms a compression disc for $\partial_v \mathcal{B}_I$. But we have just shown that $\partial_v \mathcal{B}_I$ is incompressible. So $\partial_h \mathcal{B}_I$ is incompressible. \square

One of our goals is to show that generalised isotopy moves are powerful enough to isotope an almost normal surface to a normal one. In order to show this, we need to understand the way that generalised parallelity bundles can lie in $S \times [0, 1]$. We are therefore led to the following definition.

Definition 6.19 (Coherent, incoherent I -bundle). Let B be a connected I -bundle embedded in $S \times [0, 1]$ with $\partial_h B = B \cap (S \times \{0, 1\})$. Then B is *coherent* if it intersects both $S \times \{0\}$ and $S \times \{1\}$. Otherwise, it is *incoherent*. See Figure 18.

Note that isotopy moves along edge and face compression discs move a surface through an incoherent I -bundle. Similarly, an annular simplification occurs in an incoherent I -bundle. The following proposition shows that these are essentially the only moves needed to reduce a surface.

Proposition 6.20 (Incoherent is trivial or I -bundle over disc or annulus). Let \mathcal{H} be a handle structure for $S \times [0, 1]$ where S is a closed orientable surface. Let \mathcal{B} be a maximal generalised parallelity bundle for $S \times [0, 1]$. Let B be an incoherent component of \mathcal{B} . Then B is either boundary-trivial or an I -bundle over a disc or an annulus.

Proof. Suppose that B is neither boundary-trivial nor an I -bundle over a disc. By Theorem 6.18, the horizontal and vertical boundaries of B are incompressible. Any vertical boundary component A of B is an incompressible annulus properly embedded in $S \times [0, 1]$ with ∂A lying in $S \times \{0\}$, say. Hence, it is parallel to an annulus in $S \times \{0\}$. It cannot be the case that, for every vertical boundary component of B , the product region between it and the corresponding annulus in $S \times \{0\}$ has interior

disjoint from B . Hence, there is some vertical boundary component A with the property that the product region between A and the annulus in $S \times \{0\}$ contains B . Therefore, $\partial_h B$ lies in this annulus in $S \times \{0\}$. By the incompressibility of $\partial_h B$, we deduce that each component of $\partial_h B$ is an annulus. Therefore, B is an I -bundle over an annulus or Möbius band. In fact, the latter case cannot arise because a Möbius band does not properly embed in a copy of $S \times [0, 1]$. \square

Definition 6.21 (Essential annular I -bundle). Let B be a connected I -bundle embedded in $S \times [0, 1]$ with $\partial_h B = B \cap (S \times \{0, 1\})$. Then B is *annular* if it is an I -bundle over an annulus. It is *essential* if its vertical boundary is incompressible.

Note that by Proposition 6.20, any incoherent component of the maximal generalised parallelity bundle that is neither boundary-trivial nor an I -bundle over a disc must be both annular and essential.

We now list all the isotopy moves that we will use in this paper.

Definition 6.22 (Generalised isotopy moves). Let S be a normal or almost normal fibre in $M = S \times [0, 1]$. Give it a transverse orientation. Let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points. Let S' be either S or a nearly normal surface in M , on the side of S towards which S points. A *generalised isotopy move* is one of the following moves applied to S' :

- (1) a generalised isotopy along an edge compression disc, provided S' is almost normal (Definition 5.4 and Definition 6.5);
- (2) a parallelisation isotopy, provided S' is almost normal (Definition 6.10);
- (3) a compression isotopy, provided S' is nearly normal (Definition 5.8);
- (4) an annular simplification, on the side into which S' points (Definition 6.12);
- (5) a generalised isotopy along a face compression disc (Definition 6.5);
- (6) a tube compression, if S' is a tubed almost normal surface that is oriented into the tube.

Remark 6.23. Observe that if S is an almost normal or nearly normal surface with a transverse orientation, then performing any of the generalised isotopy moves above to S , and then isotoping slightly in the direction in which S points, yields a surface that:

- is disjoint from S ,
- lies in the component of $M \setminus S$ to which S points, and
- admits a transverse orientation, pointing away from S .

Thus repeatedly applying generalised isotopy moves yields a sequence of pairwise disjoint, transversely oriented surfaces.

Remark 6.24. Many of the generalised isotopy moves are only applicable to almost normal surfaces. So if S' is nearly normal, then the only possible moves that might be applied to S' are compression isotopies, annular simplifications and generalised isotopy along a face compression disc. Thus, if S' is normal, the only possible move that might be applicable is an annular simplification.

Proposition 6.25 (Nearly normal preserved). *Let S be a transversely oriented almost normal surface. Let S' be obtained from S by an isotopy along an edge compression disc, a generalised isotopy along an edge compression disc, a tube compression or a parallelisation isotopy, followed by a sequence of generalised isotopy moves. Then S' is nearly normal.*

Proof. By Lemmas 6.8 and 5.15, a first isotopy along an edge compression disc or tube compression yields a nearly normal surface. By definition, a parallelisation isotopy takes S to a normal surface, which is therefore nearly normal. By Lemmas 5.12, 6.7 and 6.14, any compression isotopies, generalised isotopies along face compression discs, and annular simplifications leave the surface nearly normal. So the result is nearly normal. \square

6.6. Avoiding the interior of the generalised parallelity bundle. We have defined generalised isotopy moves. However, there is an important property that we would like them to satisfy. If S is the initial normal or almost normal surface, and \mathcal{B} is a maximal generalised parallelity bundle for $M \setminus S$, then we would like to ensure that for the sequence of surfaces arising from isotopy moves, each surface is disjoint from the interior of \mathcal{B} . The following lemma nearly achieves this. It also places the moves into a convenient order.

Lemma 6.26 (Almost normal to normal, mostly disjoint \mathcal{B}). *Let S be an almost normal surface that is transversely oriented. Let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points. Then there is a sequence of generalised isotopy moves, all in the direction specified by the transverse orientation of S , taking S to a normal surface, with the following properties:*

- (1) *If a generalised isotopy along an edge compression disc is performed, this is the first move of the sequence.*
- (2) *If a tube compression or a parallelisation isotopy is performed, this is the only move of the sequence.*
- (3) *If any compression isotopies are performed, these all take place at the end of the sequence.*
- (4) *For each surface in the sequence before the compression isotopies, the surface is disjoint from the interior of each component of \mathcal{B} , except possibly in a regular neighbourhood of the edge compression disc, where it is allowed to go a little into an annular component of \mathcal{B} .*
- (5) *From the second surface onwards, all surfaces in the sequence are nearly normal.*

Proof. Suppose first that the initial surface S has a tubed piece, and the transverse orientation points into the tube. Then we perform a tube compression and end with a normal surface. Similarly, if S admits a parallelisation isotopy, then we perform this and end with a normal surface.

Suppose now that if S contains a tubed piece, then the transverse orientation points out of the tube. Suppose also that S does not admit a parallelisation isotopy. The initial almost normal surface S has an edge compression disc D on the side into which S points. By Lemma 6.11, the almost normal piece is disjoint from the interior of $\partial_h \mathcal{B}$, and hence D is disjoint from the interior of \mathcal{B} . This disc D is incident to a 2-handle, which is part of \mathcal{B} . If the 2-handle is an entire component of \mathcal{B} , then perform an edge compression along D . If this component of \mathcal{B} is a larger I -bundle over a disc, then perform a generalised isotopy along the edge compression disc D . If this component of \mathcal{B} is an I -bundle over an annulus, then we just perform the edge compression along D and allow the resulting surface to go slightly into this component of \mathcal{B} .

After this process, the result is a nearly normal surface. By Proposition 6.25, any later surfaces in the sequence will be nearly normal. Let S' be the surface that has been obtained so far.

Suppose that there is an incoherent component B of \mathcal{B} on the side of S' into which it points, such that B is not an I -bundle over a disc. Then B is boundary-trivial or an I -bundle over an essential annulus, by Proposition 6.20. If there is a boundary-trivial component, then S' admits an annular simplification by Lemma 6.17. If there is an essential incoherent annular I -bundle, then we may perform an annular simplification along an extrememost one in $M \setminus S'$.

So, we may assume that every incoherent component of \mathcal{B} on the side of S' into which it points is an I -bundle over a disc. If S' has a face compression disc, pick one, called D . Let β be the arc as in Definition 5.5. This lies in a 2-handle $D^2 \times D^1$ with $D^2 \times \partial D^1$ in S . Hence, this 2-handle is a parallelity 2-handle, and is therefore part of a component B of \mathcal{B} . By assumption, B is an I -bundle over a disc, and so we can perform a generalised isotopy move across B .

Each of these moves reduces the weight of S' , by Lemmas 6.9 and 6.15. Thus, this process eventually terminates at a surface S' . This is nearly normal but does not have a face compression disc. If S' is not normal, then by Proposition 5.16, S' admits a compression isotopy. Perform this compression isotopy. This does not create any new face compression disc. Hence, we may repeat, performing compression isotopies at each stage. Each compression isotopy reduces the number of

annular components of intersection between S' and the 1-handles. Hence, this process eventually terminates with a surface that must be normal. \square

In the sequence of surfaces taking S to a normal surface, we want to ensure that the surfaces are disjoint from the interior of \mathcal{B} . The reason is that this will enable us to provide upper bounds on the number of edge swaps taking a spine in one surface to a spine in the other. These bounds will mostly be in terms of the length of the vertical boundary of \mathcal{B} . A formal definition of this length will be given in Definition 8.11 but it is roughly the number of handles of \mathcal{H} that $\partial_v \mathcal{B}$ runs through.

Unfortunately, Lemma 6.26 is not quite sufficient for our purposes. One reason for this is that in (4), the surface is allowed to go a little into the interior of \mathcal{B} . We now develop an extension to Lemma 6.26 that will allow to us to improve (4). However, first we need a definition.

Definition 6.27 (Clean annular simplification). Let M be a compact orientable 3-manifold with a pre-tetrahedral handle structure. Let S be a transversely oriented normal or almost normal surface properly embedded in M . Let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points. Let S' be obtained from S by a sequence of generalised isotopy moves, all in the given transverse direction. Suppose that S' is disjoint from the interior of \mathcal{B} . Suppose that S'' is obtained from S' by an annular simplification that moves S' across a boundary-trivial or essential annular component B of \mathcal{B} . This annular simplification isotopes an annulus A' in S' across to a vertical boundary component A'' of B . Now S is isotopic to S' , and before the annular simplification, this has not moved $S \cap \partial_h B$. Hence, there is a corresponding annulus A in S with boundary equal to $\partial A''$. We say that this annular simplification is *clean* if the isotopies taking A to A' are all annular simplifications.

Lemma 6.28 (Isotopy to normal, extra properties). *In the sequence of generalised isotopy moves in Lemma 6.26, we may in addition ensure that the following properties hold:*

- (1) *The annular simplifications are all clean.*
- (2) *For each surface in the sequence before the compression isotopies, the surface is disjoint from the interior of each component of \mathcal{B} , unless the first isotopy is a parallelisation isotopy, in which case it is the only generalised isotopy move that is performed.*

Proof. We suppose that a parallelisation isotopy is not performed, as in this case, it is the only move in the sequence and the lemma is automatically true.

We now explain how to make all the annular simplifications clean. Suppose that in the sequence of Lemma 6.26, there is some annular simplification that is not clean. Consider the first such annular simplification. This involves an annular component B of \mathcal{B} . It moves an annulus A' in S' across to a vertical boundary component A'' of B . Let S'' be the resulting surface. By definition of an annular simplification, $A' \cup A''$ bounds a 3-manifold P' that lies in a 3-ball or is a product region between A' and A'' . Now, S is isotopic to S' , and before the annular simplification, this has not moved $S \cap \partial_h B$. Hence, there is a corresponding annulus A in S and a corresponding 3-manifold P bounded by $A'' \cup A$. As we are considering the first annular simplification that is not clean, all previous ones were clean. Hence, we can perform these, taking S to some surface \tilde{S}' . For any generalised isotopies that were applied to S outside of P , we apply the same isotopies to \tilde{S}' , giving a surface \hat{S}' . We can then perform the annular simplification across B taking \hat{S}' to S'' . This involves exactly the same number of annular simplification as before, but it involves no other generalised isotopies within P . Thus, it is clean. Hence, we have created a shorter sequence of generalised isotopy moves satisfying the conclusions of Lemma 6.26. So in a shortest such sequence, all annular simplifications are clean.

This actually implies that before the compression isotopies, the surfaces are disjoint from the interior of each component of \mathcal{B} . For the only way that this can be violated is at the first move, when an isotopy is performed along an edge compression disc, taking the surface a little into an annular component of \mathcal{B} . At some stage, the surface must be isotoped across this annular I -bundle, necessarily by an annular simplification. But this annular simplification is then not clean, because the isotopy along the edge compression disc precedes it. \square

Remark 6.29. A consequence of the above proof is that the sequence of generalised isotopy moves does not necessarily start with an isotopy along an edge compression disc. Instead, we may start with annular simplifications which end up by removing the almost normal piece of S . Alternatively, we may perform a parallelisation isotopy or a tube compression.

The following result is similar to Lemma 6.26 but where we start with a normal surface and apply an annular simplification.

Lemma 6.30 (Normal to normal, disjoint \mathcal{B}). *Let S be a normal surface that is transversely oriented. Let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points. Suppose that S admits an annular simplification in this direction. Then there is a sequence of generalised isotopy moves, all in the direction specified by the transverse orientation of S , taking S to a normal surface, with the following properties:*

- (1) *The annular simplifications are all clean, and at least one annular simplification is performed.*
- (2) *If any compression isotopies are performed, these all take place at the end of the sequence.*
- (3) *For each surface in the sequence before the compression isotopies, the surface is disjoint from the interior of each component of \mathcal{B} .*
- (4) *All surfaces in the sequence are nearly normal.*
- (5) *Each incoherent component of \mathcal{B} incident to S that is not an I -bundle over a disc is isotoped across in this sequence.*

Proof. This is very similar to the proof of Lemma 6.26. Consider all the incoherent components of \mathcal{B} that are not I -bundles over discs and that are incident to S . Since S admits an annular simplification, there is at least one such component of \mathcal{B} . If there is a boundary-trivial component, then the surface admits an annular simplification by Lemma 6.17. If there is an essential incoherent annular I -bundle, then we may perform an annular simplification along an extrememost one. Since these annular simplifications occur before any other generalised isotopy moves, they are clean. We keep doing this until we end with a surface S' such that every incoherent component of \mathcal{B} on the side of S' into which it points is an I -bundle over a disc. If this surface has a face compression disc, we apply a generalised isotopy move across it. We keep doing this until the resulting surface has no face compression discs. We then apply compression isotopies as many times as possible. The result is the final normal surface. \square

The following lemma is an application of Lemmas 6.26 and 6.30 multiple times until no further generalised isotopy moves are possible.

Lemma 6.31 (Iterating generalised isotopy moves). *Let S be an almost normal or normal surface in that is transversely oriented. Then there is a sequence of disjoint transversely oriented surfaces $S = S_0, \dots, S_n$ with the following properties:*

- (1) *Each S_i points towards S_{i+1} , and the final surface S_n points away from the others.*
- (2) *For all $i > 0$, S_i is normal.*
- (3) *The final surface S_n admits no generalised isotopy moves.*

For each i , let \mathcal{B}_i be a maximal generalised parallelity bundle for the manifold between S_i and S_{i+1} . Then there is a sequence of generalised isotopy moves taking S_i to S_{i+1} with the following properties:

- (1) *If a generalised isotopy along an edge compression disc is performed, this is the first move of the sequence.*
- (2) *If a tube compression or a parallelisation isotopy is performed, this is the only move of the sequence.*
- (3) *If any compression isotopies are performed, these all take place at the end of the sequence.*
- (4) *For each surface in the sequence before the compression isotopies, the surface is disjoint from the interior of each component of \mathcal{B}_i , unless the first isotopy is a parallelisation isotopy, in which case it is the only generalised isotopy move that is performed.*
- (5) *From the second surface onwards, all surfaces in the sequence are nearly normal.*

- (6) *The annular simplifications are all clean.*
 (7) *When S_i is normal, for every incoherent component B of \mathcal{B}_i incident to S_i such that B is not an I -bundle over a disc, we isotope across B in this sequence.*

Proof. We start with $S = S_0$. If S_0 is normal and admits no annular simplifications, then we set $k = 0$ and stop. Otherwise, we may apply Lemmas 6.26 and 6.28 or Lemma 6.30 to get a sequence of generalised isotopy moves taking S_0 to a normal surface S_1 . If this admits an annular simplification, we apply Lemma 6.30 again. We repeat until we reach a normal surface S_n that admits no annular simplifications. This must exist because at each stage when going from S_i to S_{i+1} (except possibly in the parallelisation isotopy which happens only once), the weight of the surface strictly decreases. \square

7. SURFACES AND GENERALISED ISOTOPY MOVES

Road map: The generalised isotopy moves defined in Section 6 will give us a sequence of surfaces in $S \times [0, 1]$. Eventually, we want to show that there is a bounded number of moves required to transfer a spine on $S \times \{0\}$ to $S \times \{1\}$. These spines will step through a sequence of surfaces produced by generalised isotopy moves. In this section, we gather results that determine properties of such surfaces. These will be used in the proof of the main theorem.

Definition 7.1 (Normally cylindrical/acylindrical). A properly embedded normal or almost normal 2-sided surface S in M is *normally cylindrical* if there exists an annulus A embedded in M with the following properties:

- (1) $A \cap S = \partial A$;
- (2) each curve of ∂A is essential in S ;
- (3) A lies in the parallelity bundle of $M \setminus S$ and is vertical in it;
- (4) near both components of ∂A , A emanates from the same side of S .

We also say that S is normally cylindrical on the side containing A . If no such annulus exists as above, then we say that S is *normally acylindrical*. A surface can also be normally acylindrical on one side.

Definition 7.2 (Coherently bundled). Let \mathcal{H} be a handle structure for $S \times [0, 1]$. Then \mathcal{H} is *coherently bundled* if no vertical boundary component of its maximal generalised parallelity bundle is an incompressible annulus with both boundary components in $S \times \{0\}$ or both boundary components in $S \times \{1\}$.

The following lemmas are immediate consequences of the definitions and Theorem 6.18.

Lemma 7.3 (Coherently bundled alternative). *Let \mathcal{H} be a handle structure for $S \times [0, 1]$. Then \mathcal{H} is coherently bundled if and only if it admits a maximal generalised parallelity bundle with the property that every incoherent component either is an I -bundle over a disc or is boundary-trivial.* \square

Lemma 7.4 (Acylindrical fibre gives coherently bundled). *Let \mathcal{H} be a handle structure for a closed orientable 3-manifold M that fibres over the circle. Let S be a normal fibre. Then the handle structure that $M \setminus S$ inherits is coherently bundled if and only if S is normally acylindrical.* \square

Lemma 7.5 (Acylindrical fibre in product gives coherently bundled). *Let \mathcal{H} be a coherently bundled handle structure for $M = S \times [0, 1]$. Let S' be a normal fibre. Then the handle structure that $M \setminus S'$ inherits is coherently bundled if and only if S' is normally acylindrical.* \square

Lemma 7.6 (Acylindrical one side gives coherent one side). *Let \mathcal{H} be a coherently bundled handle structure for $M = S \times [0, 1]$. Let S' be a normal fibre that is normally acylindrical on one side. Let \mathcal{H}' be the handle structure for the component of $M \setminus S'$ on that side. Then \mathcal{H}' is coherently bundled.* \square

Suppose \mathcal{H} is a pre-tetrahedral handle structure of $M = S \times [0, 1]$. Then the maximal generalised parallelity bundle for \mathcal{H} , which we will denote by $\mathcal{B}(\mathcal{H})$, has components that can be partitioned into three subsets, by Theorem 6.18:

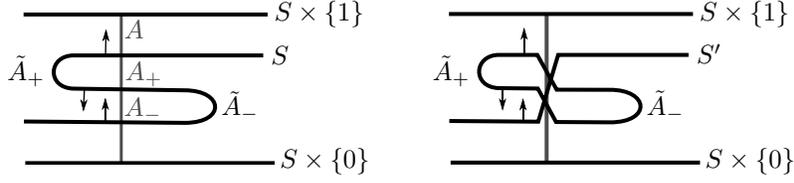


FIGURE 19. Left: Schematic picture of annuli A in $S \times [0, 1]$, A_- and A_+ in A , and \tilde{A}_- and \tilde{A}_+ in S . Right: S is the normal sum of S' and the torus $\tilde{A}_- \cup \tilde{A}_+$.

- $\mathcal{B}_D(\mathcal{H})$, the I -bundles over discs;
- $\mathcal{B}_\partial(\mathcal{H})$, the boundary-trivial components;
- $\mathcal{B}_I(\mathcal{H})$, the remaining components, which have incompressible horizontal boundary and incompressible vertical boundary.

Lemma 7.7 (Incoherent components for $M \setminus S$ in coherent bundle for M). *Let \mathcal{H} be a pre-tetrahedral handle structure of $M = S \times [0, 1]$, with components of its maximal generalised parallelity bundle partitioned into $\mathcal{B}_D(\mathcal{H})$, $\mathcal{B}_\partial(\mathcal{H})$, and $\mathcal{B}_I(\mathcal{H})$ as above. Let $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$ denote the coherent components of $\mathcal{B}_I(\mathcal{H})$. Suppose that $\Delta(\mathcal{H}) > 0$.*

- *If \mathcal{H} contains a normal fibre that is not normally parallel to a boundary component, then let S be one such fibre with least weight.*
- *If every normal fibre in \mathcal{H} is normally parallel to a boundary component, then let S be an almost normal fibre with least weight.*

Let \mathcal{B} be a maximal generalised parallelity bundle for the handle structure obtained by cutting along S . Let \mathcal{B}_A be the union of the incoherent essential annular components of \mathcal{B} that are incident to S . Then \mathcal{B}_A is disjoint from $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$.

Proof. We first show that no component of \mathcal{B}_A lies in the interior of $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$. For suppose instead that a component B of \mathcal{B}_A lies completely in the interior of $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$. The component of $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$ containing B consists of a component P of the parallelity bundle for \mathcal{H} , possibly with some $D^2 \times I$ pieces attached to it. Now B cannot intersect P , since S would have to be horizontal within P and hence B would intersect $\partial_v \mathcal{B}_I^{\text{coh}}(\mathcal{H})$. Hence, B lies within the $D^2 \times I$ pieces of this component of $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$, which is a collection of 3-balls. But this implies that the horizontal boundary of this component of \mathcal{B}_A is inessential in S , which contradicts the definition of \mathcal{B}_A .

We now show that \mathcal{B}_A is disjoint from $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$. To do this, it suffices to show that it is disjoint from $\partial_v \mathcal{B}_I^{\text{coh}}(\mathcal{H})$, because, in the previous paragraph, we showed that no component of \mathcal{B}_A lies in the interior of $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$. Suppose that \mathcal{B}_A intersects a component A of $\partial_v \mathcal{B}_I^{\text{coh}}(\mathcal{H})$. Then $S \cap A$ would be at least two parallel core curves of A . In fact, because $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$ is coherent and because S separates $S \times \{0\}$ from $S \times \{1\}$, we deduce that any essential arc in A must intersect S an odd number of times. Hence $S \cap A$ is at least three core curves of A . Pick three such core curves that are adjacent in A , and let A_- and A_+ be the annuli between them in A . Since S is separating, we may choose a transverse orientation on S so that, near ∂A_- , it points into A_- , and near ∂A_+ , it points out of A_+ . Because A is incompressible, ∂A_- bounds an annulus \tilde{A}_- in S , and similarly ∂A_+ bounds an annulus \tilde{A}_+ in S . See Figure 19, left.

Suppose first that \tilde{A}_- and \tilde{A}_+ emanate from the curve $A_- \cap A_+$ in opposite directions in S . Then the interiors of \tilde{A}_- and \tilde{A}_+ are disjoint, unless S is a torus. So suppose for the moment that S is not a torus. Then we may remove $\tilde{A}_- \cup \tilde{A}_+$ from S and attach the boundary components of the resulting surface, to obtain a normal or almost normal fibre S' . We claim that S' is not normally parallel to a boundary component of M . One way to see this is to note that S is the normal sum of S' and a torus. This torus is obtained from \tilde{A}_- and \tilde{A}_+ by perturbing using a small normal isotopy and then gluing their boundary components together in pairs, as in Figure 19, right. (The normal sum is obtained by cutting along the parallel normal curves and reconnecting into the surface S ; see,

for example [9] for information on normal sums.) If S' is normally parallel to a boundary component, then this normal sum would just be a disjoint union, which is a contradiction. Now, S' is a fibre and it has smaller weight than S . If S was normal, then so is S' . If S is almost normal, then S' may be normal or almost normal. But in all cases, this contradicts our hypotheses.

Suppose now that \tilde{A}_- and \tilde{A}_+ emanate from the curve $A_- \cap A_+$ in opposite directions in S , but that the interiors of \tilde{A}_- and \tilde{A}_+ overlap. Then S is a torus and the three curves $\partial\tilde{A}_- \cup \partial\tilde{A}_+$ are parallel curves on S . Hence, we may consider $S \setminus \setminus \tilde{A}_-$ instead of \tilde{A}_- and consider $S \setminus \setminus \tilde{A}_+$ instead of \tilde{A}_+ . These annuli then have disjoint interiors, and we may then argue as above to deduce that S is the normal sum of a fibre S' and a torus, and thereby reach a contradiction.

Suppose now that \tilde{A}_- and \tilde{A}_+ emanate from $A_- \cap A_+$ in the same direction in S . Because the other curves of ∂A_- and ∂A_+ are disjoint, we deduce that \tilde{A}_- and \tilde{A}_+ are nested. Say that $\tilde{A}_- \subset \tilde{A}_+$. Note that near its two boundary curves, \tilde{A}_- emanates from the same side of A . For otherwise, $\tilde{A}_- \cup A_-$ is the union of two annuli embedded in $S \times [0, 1]$ with matching boundary components. This union is a Klein bottle, because the transverse orientations on S at the two components of ∂A_- are oppositely oriented. But a Klein bottle does not embed in $S \times [0, 1]$, which is a contradiction. The same argument proves that \tilde{A}_+ emanates from the same side of A . But then the annulus $\tilde{A}_+ \setminus \setminus \tilde{A}_-$ emanates from opposite sides of A near its boundary. We can reglue the boundary components of $\tilde{A}_+ \setminus \setminus \tilde{A}_-$ to form an embedded torus T . The surface S is the normal sum of T and another normal or almost normal fibre S' . As argued above, this has smaller weight than S and is not normally parallel to a boundary component of M . This is again a contradiction. \square

Recall the weight of a surface that respects a handle structure is the number of components of intersection between the surface and the 2-handles, Definition 4.17.

Lemma 7.8 (Least weight surface is acylindrical). *Let M be a closed orientable 3-manifold that fibres over the circle. Then, for any pre-tetrahedral handle structure \mathcal{H} of M , there is a normal fibre that is normally acylindrical. Indeed, any normal fibre that has least weight in its isotopy class is normally acylindrical and admits no annular simplifications.*

Proof. Let S be a fibre that is standard with respect to \mathcal{H} and that has least weight among all such fibres. Then we may isotope S to a normal surface without increasing the weight: the fact that such an isotopy exists is well-known. See, for example [22, Theorem 3.4.7] or [16, Proposition 4.4].

Suppose that S admits an annular simplification. Then this reduces the weight of S by Lemma 6.15, which is a contradiction. Suppose that S is normally cylindrical. Then by Lemma 7.4, $M \setminus \setminus S$ is not coherently bundled. So, by Lemma 7.3, it has an essential incoherent annular component of its maximal generalised parallelity bundle. So S admits an annular simplification, which we have established not to be the case. \square

The next lemma ensures that a normal surface in $M \setminus \setminus S$ is also normal in M .

Lemma 7.9 (Normal in $M \setminus \setminus S$ is normal in M). *Let \mathcal{H} be a handle structure of a compact orientable 3-manifold M . Let S be a closed normal surface in M . Let \mathcal{H}' be the handle structure that $M \setminus \setminus S$ inherits. Let S' be a closed surface in $M \setminus \setminus S$ that is normal with respect to \mathcal{H}' . Then S' is normal with respect to \mathcal{H} in M .*

Proof. Let H_0 be a 0-handle of \mathcal{H} and let D be a component of $S' \cap H_0$. Suppose that D is not normal. Then it runs over a component E of intersection between H_0 and the 2-handles more than once. This component E is divided up by S . However, since S is normal in \mathcal{H} , each component of intersection between S and H_0 runs over E at most once. Hence, any 0-handle of $H_0 \cap \mathcal{H}'$ intersects E in either the empty set or a single disc. Therefore, if D runs over E more than once, then S' is not normal in \mathcal{H}' , which is a contradiction. \square

Similarly, if a surface is normally acylindrical in $M \setminus \setminus S$, then it is normally acylindrical in M , under the following assumptions.

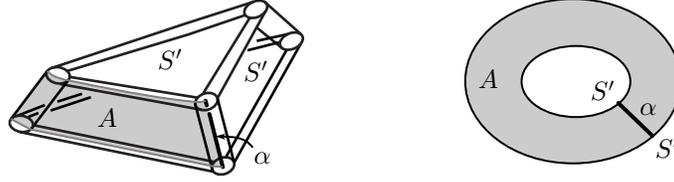


FIGURE 20. The annulus A may be taken to run through the co-core of a 2-handle, intersecting it in an arc α that is essential in A , with both endpoints on S' .

Lemma 7.10 (Normally acylindrical in $M \setminus S$ and M). *Let \mathcal{H} be a handle structure of a compact orientable 3-manifold M . Let S be a closed separating incompressible normal surface in M that is normally acylindrical. Let \mathcal{H}' be the handle structure that $M \setminus S$ inherits. Let S' be a closed incompressible connected normal surface in \mathcal{H}' that is normally acylindrical with respect to \mathcal{H}' . Then S' is a normal surface in \mathcal{H} that is normally acylindrical with respect to \mathcal{H} .*

Proof. By Lemma 7.9, S' is normal in \mathcal{H} . Suppose that S' is normally cylindrical in \mathcal{H} , via an annulus A . If this annulus has empty intersection with S , then S' is normally cylindrical with respect to \mathcal{H}' , which is contrary to assumption. Therefore, A is divided up by curves of $S \cap A$.

If there is more than one circle of $A \cap S$, then two adjacent circles bound an annulus in A with interior disjoint from S and which emanate from the same side of S . Since S is normally acylindrical, we deduce that these curves must be inessential in S . But then the curves $A \cap S'$ bound discs in the complement of S' , which contradicts the assumption that S' is incompressible.

So suppose that there is just one curve of $A \cap S$. Then an essential arc in A intersects this curve just once, and using the fact that S' is connected, we may join the endpoints of this arc by a curve parallel to S' . This implies that S is non-separating, which is a contradiction. \square

Proposition 7.11 (Acylindrical one side stays acylindrical one side). *Let \mathcal{H} be a pre-tetrahedral handle structure for $M = S \times [0, 1]$. Let S be a normal fibre that is normally cylindrical on one side only. Transversely orient S in this direction. Let S' be a normal fibre that is obtained from S by performing generalised isotopy moves, all in this transverse direction. Then S' is normally acylindrical on the side into which S' does not point.*

Proof. Each time that we perform an annular simplification, a generalised isotopy along an I -bundle over a disc, an isotopy along a face compression disc, or a compression isotopy, we remove points of intersection with 2-handles of \mathcal{H} without adding more such intersections. Thus, we can view each component of intersection between one of the surfaces and the 2-handles of \mathcal{H} as corresponding to a component of intersection with the previous surface and the 2-handles of \mathcal{H} . In order to ensure that all these surfaces in this isotopy are disjoint, each time we perform an isotopy, we push the entire surface a little in its specified transverse orientation (as in Remark 6.23).

Suppose that S' is cylindrical on the side into which it does not point, via an annulus A . The annulus A is vertical in the parallelity bundle of $M \setminus S'$. Thus by isotoping A to the boundary of some parallelity 0-handle through which it runs, we may ensure that A contains the co-core of a 2-handle of $M \setminus S'$. This co-core will be an essential arc α in A , with both endpoints of α on S' . See Figure 20.

If S were to intersect A , it would do so in a collection of parallel core curves of A , because S is normal. Since each component of intersection of S' with 2-handles corresponds to a component of intersection of S with 2-handles, the two intersections between α and S' at the endpoints of α also give points of intersection with S . These must lie in the interior of α , because the transverse orientation on S' points away from α . Hence $S \cap A$ contains at least two core curves, one parallel to one component of ∂A and the other parallel to the other component of ∂A . These core curves point towards the relevant component of ∂A . We deduce that there are two adjacent curves of $S \cap A$ that

co-bound an annulus A' in A , where S points out of A' near $\partial A'$. Hence, S is normally cylindrical on the side into which it does not point, contrary to assumption. \square

Corollary 7.12 (Maximal isotopies, cylindrical one side gives normal). *Suppose \mathcal{H} is a pre-tetrahedral handle structure for $M = S \times [0, 1]$. Suppose that M contains no normal fibre that is normally acylindrical, other than those that are normally parallel to $S \times \{0\}$ or $S \times \{1\}$. Let S be a transversely oriented normal fibre that is cylindrical on one side only, the side into which it points. Let S' be a surface that is obtained from S by performing generalised isotopy moves, all in the specified transverse direction, until no further generalised isotopy moves are possible.*

Proof. By Proposition 5.16, S' is normal. By Proposition 7.11, S' is acylindrical on the side into which it does not point. It cannot be cylindrical on the side into which it points, as otherwise an annular simplification could be performed upon it. Hence, it is acylindrical and therefore, by hypothesis, it is boundary parallel. \square

Proposition 7.13 (Maximal isotopies, everything cylindrical). *Suppose \mathcal{H} is a pre-tetrahedral handle structure for $M = S \times [0, 1]$. Suppose that every normal fibre is cylindrical on the $S \times \{1\}$ side at least, with the exception of those that are normally parallel to $S \times \{0\}$ or $S \times \{1\}$. Let S be a normal or almost normal fibre that is not boundary parallel. Let S' be a surface that is obtained from S by a maximal sequence of generalised isotopy moves, all in the $S \times \{1\}$ direction. Then S' is normally parallel to $S \times \{1\}$.*

Proof. The surface S' is normal, by Proposition 5.16. If it is not parallel to $S \times \{1\}$ it admits an annular simplification in the $S \times \{1\}$ direction, by hypothesis. Because S' admits no such simplification, it must therefore be normally parallel to $S \times \{1\}$. \square

Proposition 7.14 (Isotopies from a least weight fibre). *Let $M = S \times [0, 1]$, and suppose \mathcal{H} is a pre-tetrahedral handle structure for M .*

- *If there is a normal fibre that is not normally parallel to $S \times \{0\}$ or $S \times \{1\}$, let S be one with least weight. In this case, suppose that S is normally cylindrical on the $S \times \{1\}$ side.*
- *If the only normal fibres are those that are normally parallel to $S \times \{0\}$ or $S \times \{1\}$, let S be an almost normal fibre with least weight.*

Then there is a sequence of generalised isotopies satisfying the conclusions of Lemma 6.26 and 6.28 (in the case where S is almost normal) or 6.30 (in the case where S is normal) taking S to $S \times \{1\}$.

Proof. Orient S towards $S \times \{1\}$. By Lemmas 6.26 and 6.28, or Lemma 6.30, there is a sequence of generalised isotopy moves as in those lemmas, taking S to a normal surface S' . When S is almost normal, then by assumption, the only normal fibres are parallel to $S \times \{0\}$ or $S \times \{1\}$ and so S' must be parallel to $S \times \{1\}$. When S is normal, then S' has smaller weight than S and hence again it must be parallel to $S \times \{1\}$. \square

Definition 7.15 (Innermost). Let \mathcal{H} be a handle structure of $M = S \times [0, 1]$. Let S be a normal fibre that is cylindrical on exactly one side. Then S is *innermost* in M if, for every normal fibre S' that is disjoint from S , that lies on the normally acylindrical side of S and that is normally cylindrical in M on the side pointing towards S and normally acylindrical on the other side, S' is normally parallel to S . See Figure 21.

Lemma 7.16 (Cutting along innermost). *Let \mathcal{H} be a handle structure of $M = S \times [0, 1]$. Suppose that every normal fibre in M is normally cylindrical, with the exception of those that are boundary parallel. Let S be a normal fibre that is normally cylindrical on exactly one side. Suppose that S is innermost. Then, in the component of $M \setminus S$ on the normally acylindrical side of S , every normal fibre is either normally cylindrical in $M \setminus S$ on the side away from S , or boundary parallel.*

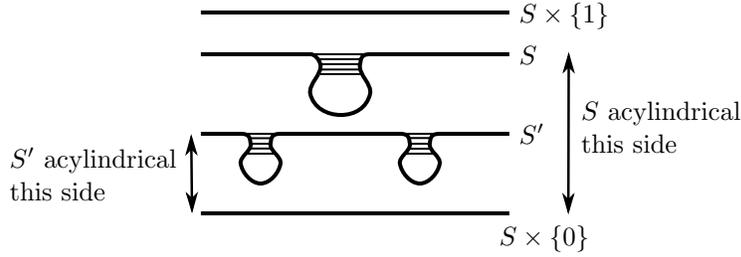


FIGURE 21. If S as shown is innermost, then S' must be normally parallel to S .

Proof. Suppose that, on the contrary, there is a normal fibre S' in M on the normally acylindrical side of S , such that S' not boundary parallel, but S' is normally acylindrical in $M \setminus S$ on the side away from S . This fibre S' is normal in M by Lemma 7.9. It is normally cylindrical in M by hypothesis, but because S is innermost, S' cannot be normally cylindrical in M only on the side that points toward S . Note that if S' is normally cylindrical in $M \setminus S$ only on the side that points toward S , then it will still be normally cylindrical on this side in M , since the incoherent parallelity bundle for S' in $M \setminus S$ will be an incoherent parallelity bundle for S' in M . Thus S' must be normally cylindrical in M on the side away from S . But then the incoherent parallelity bundle for S' on this side is disjoint from S , and thus remains an incoherent parallelity bundle for $M \setminus S$, and so S' is normally cylindrical in $M \setminus S$ on the side away from S . This contradicts our assumption. \square

8. SPINES ON SURFACES

Road map: We have built nearly normal surfaces and described generalised isotopy moves that interpolate between almost normal, nearly normal, and normal surfaces. By tracing through the effect of these moves on a surface $S \subset S \times [0, 1]$, we will be able to transfer a spine from one nearly normal or almost normal surface to another. This section describes how the spine transfer is done.

Finally, recall that to prove the main theorem, we need to bound the number of edge contractions and expansions required when we transfer spines in $S \times [0, 1]$. This section also give some initial upper bounds on the number of edge contractions and expansions under generalised isotopy moves.

8.1. Edge swaps in the spine graph. Given a sequence of nearly normal surfaces obtained from generalised isotopy moves, we know how to transfer the canonical handle structure across the surfaces. However, we actually want to be transferring spines across surfaces, and bounding the number of edge contractions and expansions required in order to do so. Given canonical handle structures, we now describe how to transfer spines.

Our main tool will be a modification to a spine called an edge swap.

Definition 8.1 (Edge swap). Let Γ be a spine for a closed surface S . Let e_1 be an arc properly embedded in the disc $S \setminus \Gamma$. Let e_2 be an edge of the graph $\Gamma \cup e_1$ that has distinct components of $S \setminus (\Gamma \cup e_1)$ on either side of it. Then the result of removing e_2 from Γ and adding e_1 is a new spine Γ' for S . We say that Γ and Γ' are related by an *edge swap*.

The following is an important example.

Lemma 8.2 (Edge swap and Dehn twist). *Let Γ be a spine for a closed orientable surface. Let C be a simple closed curve intersecting Γ transversely in a single point in the interior of an edge of Γ . Let Γ' be obtained from Γ by Dehn twisting about C . Then Γ' is obtained from Γ by a single edge swap.*

Proof. Let e_2 be the edge of Γ that intersects C . Let e_1 be an arc in $S \setminus \Gamma$ joining the endpoints of e_2 and emanating from opposite sides of e_2 . There are two possible choices for e_1 . The spine Γ' is obtained from Γ by removing e_2 and adding one choice for e_1 . \square

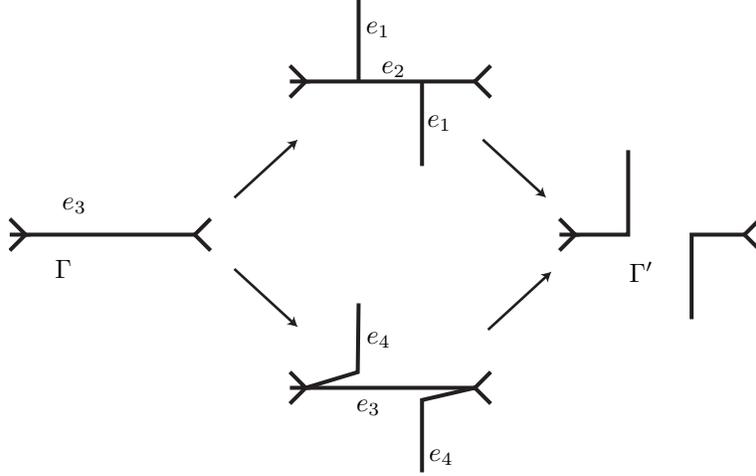


FIGURE 22. Two ways of performing an edge swap

The next lemma shows that a single edge swap is realised by a bounded number of edge contractions and expansions, so we will be able to bound distances in the spine graph by bounding edge swaps.

Lemma 8.3 (Edge swap and contraction/expansion). *Let S be a closed orientable surface. Let Γ be a spine for S . Then an edge swap can be realised by a sequence of at most $24g(S)$ edge expansions and contractions.*

Proof. Let e_1 and e_2 be as in Definition 8.1. Let Γ' be obtained from Γ by adding e_1 and removing e_2 . We first reduce to the case where e_2 is an edge of Γ . Suppose that it is not. Then e_2 is a sub-arc of an edge e_3 of Γ . Let e_4 be the arc $e_1 \cup \text{cl}(e_3 - e_2)$. A small isotopy makes it disjoint from e_3 . Then Γ' is obtained from Γ by removing e_3 and adding e_4 . See Figure 22.

Thus, we may assume that e_2 is an edge of Γ . Let Γ_- be the result of removing e_2 from Γ . Then $S \setminus \Gamma_-$ is an annulus A . The two edges e_1 and e_2 are both properly embedded in A . They are both essential in A , since if A is cut along either e_1 or e_2 , the result in both cases is a disc. Hence, there is an isotopy of A taking e_2 to e_1 . We will show that this isotopy can be achieved using a sequence of edge expansions and contractions, starting with Γ and ending with Γ' .

We perform the isotopy in three stages: first we move one endpoint of e_2 to an endpoint of e_1 , then we move the other endpoint of e_2 to the other endpoint of e_1 , in such a way that the interiors of e_1 and e_2 are then isotopic, then we move the interior of e_2 to the interior of e_1 . The latter step is achieved by an isotopy of the spine and so requires no expansions or contractions. So we need only consider the first two stages.

If an endpoint of e_2 has valence more than 3, then we first perform an edge expansion so that the valence at this end of e_2 becomes exactly 3. Then we start to perform the isotopy. When we move the endpoint of e_2 , it will, at various moments in time, move past a vertex of Γ_- or move past the other endpoint of e_2 . Each such move can be achieved by an edge contraction followed by an edge expansion. So, the total number of contractions and expansions to move one endpoint is at most twice the number of vertices in one component of the boundary of A . To move the other requires at most twice the number of vertices in the other component of ∂A . Thus we consider the number of vertices of both components of ∂A . This may be more than the number of vertices of Γ_- because when identifications are made on the boundary of A to form S , distinct vertices of A may become identified. But the number of vertices in ∂A is equal to the number of edges in ∂A and this is twice the number of edges in Γ_- . Lemma 2.2 implies there are at most $6g(S) - 4$ edges in Γ_- . Hence there are less than $24g(S)$ edge expansions and contractions required. \square

Lemma 8.4 (Edge of a spine has essential dual curve). *Let Γ be a spine for a closed orientable surface S . Let e be an edge of Γ , and let Γ_- be the graph obtained from Γ by removing e . Then the core curve of the annulus $S \setminus \Gamma_-$ is essential in S .*

Proof. Suppose that the core curve C is inessential in S . Then it bounds a disc D . We may assume that C intersects e transversely at a single point. The intersection between Γ and D is a graph G in D that intersects ∂D at one point. Since $D \setminus G$ contains a single face, the one incident to ∂D , we deduce that G is a tree. Hence it contains a vertex with valence 1, other than the one on ∂D . This is a vertex with valence 1 in Γ , contradicting the hypothesis that Γ is a spine for S , Definition 2.1. \square

Definition 8.5 (Associated cell structure from surface handle structure). Given a handle structure \mathcal{H} for a closed surface S , its *associated cell structure* is defined as follows. Each 2-cell is a handle of \mathcal{H} . Thus, its 1-cells are the components of intersection between pairs of handles. Its 0-cells are components of intersection between triples of handles. We say that a graph embedded in S is *cellular* if it is a subcomplex of this cell structure.

Definition 8.6 (Cellular spine). Let \mathcal{C} be a cell structure for S . A spine embedded in S is *cellular* if it is a subcomplex of \mathcal{C} .

Lemma 8.7 (Moving spine off a disc). *Let S be a closed orientable surface with a cell structure \mathcal{C} , and with a cellular spine Γ . Let D be a cellular subset of S that is an embedded disc, and let ℓ be the length of ∂D , where each 1-cell of \mathcal{C} is declared to have length 1. Then there is a sequence of at most $6g(S) + 2\ell$ edge swaps taking Γ to a cellular spine Γ' that is disjoint from the interior of D . Moreover, $\Gamma' - D$ is contained in $\Gamma - D$.*

Proof. We will perform the modifications to Γ in two stages. In the first stage, we will remove edges that lie entirely in the interior of D . The number of such edges is at most the number of edges of Γ , which is at most $6g(S)$ by Lemma 2.2. Let e be an edge of Γ that lies entirely in the interior of D , and let x be a point in the interior of e . We will perform an edge swap that removes e , but we need to find a suitable arc e' to replace it. Removing e from Γ gives a graph Γ_- , the exterior of which is an annulus A . The arc e is essential in A . Hence, there is an embedded core curve C for A that intersects e once at exactly the point x . The boundary ∂D intersects the interior of A in a collection of properly embedded arcs. By choosing C suitably, we may assume that C intersects each of these arcs at most once. It has to intersect at least one of these arcs, by Lemma 8.4. So, let e' be one such arc. Then the edge swap removes e and inserts e' .

After this procedure, Γ intersects the interior of D in a union of embedded arcs, each of which has at least one endpoint in ∂D . The closures of these arcs may intersect on ∂D . But these arcs may also intersect each other at their endpoints in the interior of D . So each component of $\Gamma \cap \text{int}(D)$ is either an arc or a star-shaped graph. These arcs and graphs divide D into discs, and each of these discs contains at least one 1-cell lying in ∂D , for otherwise cutting along Γ would separate S into at least two components, one entirely contained in the interior of D , contradicting the definition of a spine. The number of these discs is more than half the number of arcs of $\Gamma \cap \text{int}(D)$; this can be shown by induction on the number of components of intersection, using the fact that there are no vertices of valence one in a spine. Hence, the number of arcs is less than twice the length ℓ of ∂D . Each modification that we make will reduce the number of arcs of $\text{int}(D) \cap \Gamma$, and so the number of edge swaps needed in this stage will be at most 2ℓ . Let e be an arc of $\text{int}(D) \cap \Gamma$, and let x be a point in $\text{int}(e) \cap \text{int}(D)$. Again, there is a closed embedded curve C intersecting Γ exactly at the point x . We may assume that C intersects each arc of $\partial D \setminus \Gamma$ at most once. Let e_1 be an arc of $\partial D \setminus \Gamma$ that meets C exactly once. Let e_2 be the edge of $\Gamma \cup e_1$ containing x . Then $\Gamma \cup e_1$ divides S into two discs, which lie on either side of e_2 . Hence, we may perform an edge swap that adds e_1 and removes e_2 . \square

Corollary 8.8 (Moving spine off multiple discs). *Let S be a closed orientable surface with a cell structure \mathcal{C} , and with a cellular spine Γ . Let D_1, \dots, D_k be cellular subsets of S , each of which is an*

embedded disc, and with disjoint interiors. Let ℓ be the sum of the lengths of $\partial D_1, \dots, \partial D_k$. Then there is a sequence of at most $6kg(S) + 2\ell$ edge swaps taking Γ to a cellular spine Γ' that is disjoint from the interior of D_1, \dots, D_k . Moreover, $\Gamma' - (D_1 \cup \dots \cup D_k)$ is contained in $\Gamma - (D_1 \cup \dots \cup D_k)$.

Proof. Use Lemma 8.7 to move the spine off the interior of D_1 , then the interior of D_2 , and so on. At the i th stage, we do not add to the spine away from D_i . Hence, when we remove it from the interior of D_i , the new spine does not go into the interior of an earlier D_j . Hence, after this process, the spine is disjoint from the interior of D_1, \dots, D_k . Moreover, $\Gamma' - (D_1 \cup \dots \cup D_k)$ is contained in $\Gamma - (D_1 \cup \dots \cup D_k)$. \square

8.2. Moving spines on nearly normal surfaces. In this subsection, surfaces will have a handle structure, and we will work with the associated cell structure of Definition 8.5. We give concrete bounds on the number of edge swaps required to transfer a spine in the case of isotopy along edge and face compression discs and generalised isotopy across an I -bundle over a disc.

Remark 8.9. Note that since the canonical handle structure is unchanged under a compression isotopy, the spine on one transfers to the other immediately.

Lemma 8.10 (Edge swap bound, edge and face compression discs). *Suppose S is almost normal or nearly normal, and S' is the nearly normal surface obtained from S by an isotopy along an edge or face compression disc D . Let v be the valence of the 2-handle meeting the endpoints of $D \cap S$, i.e. the number of components of intersection between this 2-handle and the 1-handles. Give S and S' their canonical handle structures.*

- If S is equipped with a cellular spine Γ , then we may build a cellular spine on S' from Γ by a sequence of at most $\max\{0, 6(2v - 3)g(S) + 88v - 152\}$ edge swaps in the face compression case, and by at most $6(2v - 1)g(S) + 88v - 64$ edge swaps in the edge compression case.
- If instead S' is equipped with a cellular spine Γ' , then we may build a cellular spine on S from Γ' by a sequence of at most $6g(S) + 8v$ edge swaps in the face compression case, and at most $6g(S) + 8v + 40$ edge swaps in the edge compression case.

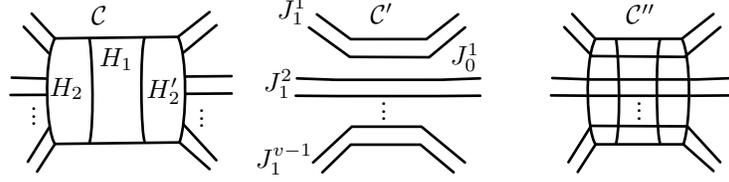
Proof. Consider the arc $\alpha = D \cap S$ of Definitions 5.4 and 5.5. We may suppose that α does not lie entirely in the interior of a 0-handle of S , else Lemma 5.20 implies Γ or Γ' transfers unchanged, and the bound is trivial. In the face compression case, this implies that $v \geq 1$.

Suppose first that the isotopy is not across an edge compression disc incident to a tubed piece of S . Then Lemma 5.19 applies. Let H_2 and H'_2 be the 2-handles of S at the endpoints of α . If α lies on a 1-handle, i.e. in the face compression disc case, let H_1 denote that 1-handle. If not, α runs through the interior of a 0-handle; let H_0 denote this 0-handle.

The effect of the isotopy along D is given in Lemma 5.19; see also Figures 11 and 12. The 2-handles H_2 and H'_2 are removed, along with H_1 in the face compression case, and handles incident to H_2 and H'_2 are combined. Finally, complementary regions that are not discs are replaced by discs. In the edge compression case, this produces (at most) v 1-handles in S' , which we denote by J_1^1, \dots, J_1^v , and (at most) $(v - 1)$ 0-handles between them, which we denote by J_0^1, \dots, J_0^{v-1} . In the face compression case, we only join $(v - 1)$ 1-handles, since H_1 was removed, and so we produce (at most) $(v - 1)$ 1-handles, denoted J_1^1, \dots, J_1^{v-1} and (at most) $(v - 2)$ 0-handles, denoted J_0^1, \dots, J_0^{v-2} .

Let \mathcal{C} and \mathcal{C}' be the cell structures associated to the handle structures on S and S' . We form a common refinement \mathcal{C}'' , as follows. Starting from \mathcal{C} , overlay the cells of \mathcal{C}' to form \mathcal{C}'' . In the face compression case, this divides the 1-handle H_1 into $(2v - 3)$ 2-cells by introducing 1-cells that run along it, and divides the 2-handles H_2 and H'_2 similarly, each becoming $(2v - 3)$ 2-cells. See Figure 23. The edge compression case is similar, with the only difference being that additional 1-cells at the top and bottom are added.

Suppose that we start with a spine Γ that is cellular in \mathcal{C} . This is therefore cellular in \mathcal{C}'' . We will move the spine through \mathcal{C}'' to a spine that is cellular in \mathcal{C}' . To do so, we move the spine to the boundary of any 1-handle J_1^j and 0-handle J_0^k . These determine cellular discs in \mathcal{C}'' . The length of


 FIGURE 23. \mathcal{C}'' is a common refinement of \mathcal{C} and \mathcal{C}' .

∂J_1^j in \mathcal{C}' is four: a nearly normal surface meets 1-handles in product regions. Because we divide the 1-handle to form the refinement, the length of ∂J_1^j in \mathcal{C}'' is 12. (We are assuming here that $J_1^j \neq J_1^k$ for $j \neq k$. In other words, we are assuming that no 1-handle J_1^j winds several times through the disc region shown in Figure 23. If some J_1^j does enter this disc region several times, then our estimates for the total length of the discs $J_1^1, \dots, J_1^v, J_0^1, \dots, J_0^{v-1}$ will still hold.) The length of ∂J_0^k in \mathcal{C}' is at most 24: 12 intersections with 1-handles by Lemma 5.11 and 12 intersections with 2-handles between 1-handles. Therefore the length of ∂J_0^k in \mathcal{C}'' is at most 32. Thus the total length of all boundaries is at most:

$$\begin{aligned} 12(v-1) + 32(v-2) &= 44v - 76 && \text{face compression case} \\ 12v + 32(v-1) &= 44v - 32 && \text{edge compression case} \end{aligned}$$

Then Corollary 8.8 implies that Γ can be transferred to Γ' by a sequence of edge swaps, with total number bounded by:

$$\begin{aligned} 6(2v-3)g(S) + (88v-152) &&& \text{face compression case} \\ 6(2v-1)g(S) + (88v-64) &&& \text{edge compression case} \end{aligned}$$

Suppose now that we have a spine Γ' that is cellular in \mathcal{C}' . It is then cellular with respect to \mathcal{C}'' . We move the spine off the interior of the discs H_2 and H_2' , and in the face compression case off the interior of the disc H_1 . In the edge compression case, we move it off the interior of the disc H_0 .

In the case of a face compression disc, in \mathcal{C}'' , the union of the discs H_1, H_2, H_2' has boundary of length $2((v-1) + (v-2)) + 6 = 4v$.

In the case of an edge compression disc, we will consider the disc that is the union of discs H_2, H_2' , and H_0 . The portion of the boundary of this disc that meets ∂H_2 has length $v + (v-1)$, and similarly for the portion of the boundary that meets $\partial H_2'$. As for H_0 , if it is a nearly normal piece in S , then its boundary has length at most 24 in \mathcal{C} by Lemma 5.11. Note that two of these 24 curves come from intersections with H_2 and H_2' . Thus the length of ∂H_0 away from H_2 and H_2' in \mathcal{C}'' is at most 22 in this case, and so the total length of the boundary in \mathcal{C}'' is at most $2(2v-1) + 22 = 4v + 20$.

If H_0 is not a nearly normal piece, then it comes from an octagon or tubed piece in S . An octagon is a disc whose boundary has length 16 in \mathcal{C} . Again two of these curves come from intersections with H_2 and H_2' , so the length of the portion of the boundary of H_0 that is disjoint from $\partial H_2 \cup \partial H_2'$ is 14. It follows that the total length of the boundary of $H_0 \cup H_2 \cup H_2'$ in \mathcal{C}'' is at most $2(2v-1) + 14$ in this case.

In all cases other than the tubed case, Lemma 8.7 implies that there is a sequence of edge swaps in \mathcal{C}'' taking Γ' to a spine Γ , which is cellular in \mathcal{C} , and that sequence has length at most:

$$\begin{aligned} 6g(S) + 8v &&& \text{face compression case} \\ 6g(S) + 8v + 40 &&& \text{edge compression case.} \end{aligned}$$

Finally suppose H_0 came from a tubed piece. Lemma 5.21 describes the way that the canonical handle structure on S' is obtained from that of S . As above, \mathcal{C} and \mathcal{C}' are the cell structures associated to the handle structures on S and S' . We define their common refinement \mathcal{C}'' again by superimposing them. Thus, the 2-handle H_2 in S at one endpoint of α is divided into (at most) $(2v-1)$ 2-cells.

At most v 1-handles of S' and at most $(v - 1)$ 0-handles of S' run through H_2 . Denote these by J_1^1, \dots, J_1^v and J_0^1, \dots, J_0^{v-1} .

Suppose first that we have a spine that is cellular in \mathcal{C}' . Then it is also cellular in \mathcal{C}'' . Both boundary components of the tube bound discs in S . The larger of these two discs contains either H_2 or H_2' , say H_2' . Let W be the union of H_2 and the larger of these two discs. Its boundary has length in \mathcal{C}'' at most $2v + 6$. So we can move the spine in \mathcal{C}'' off the interior of W using at most $6g(S) + 4v + 12$ edge swaps, by Lemma 8.7. It is then cellular in \mathcal{C} .

A spine that is cellular in \mathcal{C} is cellular in \mathcal{C}'' . We can then move it off the interior of J_1^1, \dots, J_1^v and J_0^1, \dots, J_0^{v-1} using Corollary 8.8. The length of ∂J_1^j in \mathcal{C}'' is 8. Then length of ∂J_0^j in \mathcal{C}'' is at most 28. So, the total length of their boundaries is at most

$$8v + 28(v - 1) = 36v - 28.$$

The number of edge swaps required to take the spine off these handles is therefore at most

$$6(2v - 1)g(S) + (72v - 56). \quad \square$$

We want to extend Lemma 8.10 to generalised isotopies across edge or face compression discs. To do so, we need a notion of the length of the vertical boundary of an I -bundle.

Definition 8.11 (Length of vertical boundary). Let \mathcal{B} be the parallelity bundle for a compact orientable 3-manifold M . Then its vertical boundary $\partial_v \mathcal{B}$ is a union of annuli, which inherit a cell structure, as follows. Each 2-cell is a component of intersection between a parallelity handle and a non-parallelity handle. Thus, each such 2-cell is a union of fibres in \mathcal{B} . Define the *length* of $\partial_v \mathcal{B}$ to be the number of such 2-cells. Similarly, if \mathcal{B} is a maximal generalised parallelity bundle, then $\partial_v \mathcal{B}$ is a union of vertical boundary components of the parallelity bundle, and so its length is also defined.

Lemma 8.12 (Edge swap bound, isotopy across I -bundle over disc). *Let S' be obtained from S by a generalised isotopy move across an I -bundle over a disc, where the disc has boundary of length ℓ . Let \mathcal{C} and \mathcal{C}' be the cell structures associated with the canonical handle structures of S and S' . Let Γ be a spine for S that is cellular in \mathcal{C} . Then there is a spine for S' that is cellular in \mathcal{C}' and that is obtained from Γ by performing at most $6(2\ell + 1)g(S) + 92\ell - 64$ edge swaps. Similarly, if Γ' is a cellular spine for S' , then there is a cellular spine for S obtained by performing at most $6g(S) + 8\ell + 40$ edge swaps.*

Proof. We may assume that α does not lie wholly in the interior of a 0-handle of S , as in this case, the isotopy from S to S' takes canonical handle structure to canonical handle structure. Let us also assume, for the moment, that if S is almost normal, then the edge compression disc does not run over a tubed piece.

Let \mathcal{C}_- be the cell structure obtained from \mathcal{C} by replacing each component of $\partial_h B$ by a 2-cell. Let \mathcal{C}'' be the cell structure that arises by superimposing \mathcal{C}_- and \mathcal{C}' , as in the proof of Lemma 8.10.

Consider first the spine Γ that is cellular in \mathcal{C} . Then using Corollary 8.8, we can move Γ off the interior of $\partial_h B$ using at most $12g(S) + 4\ell$ edge swaps. It is then cellular in \mathcal{C}_- and hence cellular in \mathcal{C}'' . As in the proof of Lemma 8.10, there is then a sequence of at most $6(2\ell - 1)g(S) + 88\ell - 64$ edge swaps taking to a spine that is cellular in \mathcal{C}' .

Consider now the spine Γ' that is cellular in \mathcal{C}' . As in the proof of Lemma 8.10, there is a sequence of at most $6g(S) + 8\ell + 40$ edge swaps taking to a spine that is cellular in \mathcal{C}_- . It is then cellular in \mathcal{C} .

Finally, suppose that S is almost normal and that the edge compression disc runs over a tubed piece. By Lemma 6.6, in this case we reinstate 1-handles and 2-handles of \mathcal{H} , replace each component of $\partial_h B$ by a 2-cell, and apply steps (1) to (5) of Lemma 6.6 to obtain the canonical handle structure of S' .

Suppose that we are given a spine Γ' that is cellular in \mathcal{C}' . Then the argument of Lemma 8.10 applies. Specifically, each boundary component of the tube bounds a disc in S ; let W be the union of $\partial_h B$ and the larger of these two discs. Its boundary in \mathcal{C}'' has length at most $2\ell + 6$. Then we make the spine disjoint from the interior of W using at most $6g(S) + 4\ell + 12$ edge swaps. It is then cellular in \mathcal{C} .

To convert a spine Γ in S to a cellular spine in S' requires at most $6(2\ell - 1)g(S) + (72\ell - 56)$ edge swaps, as in the proof of Lemma 8.10. \square

We can also deal with parallelisation isotopies at this stage.

Lemma 8.13 (Edge swap bound, parallelisation isotopy). *Let S be an almost normal surface, and let S' be a normal surface that is obtained from S by a parallelisation isotopy. This moves S across $D^2 \times [0, 1]$, where $\partial D^2 \times [0, 1]$ is a vertical boundary component of the parallelity bundle for $M \setminus (S \cup S')$. Give S and S' their canonical handle structures. Consider a spine that is cellular in one of these surfaces. Then it may be converted to a spine that is cellular in the other surface using at most $6g(S) + 2\ell$ edge swaps, where ℓ is the length of $\partial D^2 \times [0, 1]$.*

Proof. The canonical handle structures on S and S' agree away from $D^2 \times \{0, 1\}$. Thus, we make the spine disjoint from the interior of one of these discs using Lemma 8.7. It is then cellular in the handle structure on the other surface. \square

To apply Lemmas 8.12 and 8.13 we need a bound on the length of vertical boundary components of parallelity bundles. This is obtained by the following lemmas.

Lemma 8.14 (Length bound). *Let \mathcal{H} be a pre-tetrahedral handle structure for a compact orientable 3-manifold M . Let \mathcal{B} be a maximal generalised parallelity bundle, or a union of components of the parallelity bundle. Then the length of $\partial_v \mathcal{B}$ is at most $56\Delta(\mathcal{H})$.*

Proof. In both a maximal generalised parallelity bundle, and simply a parallelity bundle, each 2-cell in $\partial_v \mathcal{B}$ is a component of intersection between a parallelity handle and a non-parallelity handle. The non-parallelity handle is a 0-handle or a 1-handle. As one travels around $\partial_v \mathcal{B}$, one meets these 0-handles and 1-handles alternately. The number of 2-cells in $\partial_v \mathcal{B}$ lying in the non-parallelity 1-handles is at most the number lying in the non-parallelity 0-handles. So it suffices to bound the latter quantity. A tetrahedral 0-handle has at most six components of intersection with the parallelity handles. By Definition 4.13, its complexity is at least $1/2$, so it meets a number of parallelity handles equal to at most 12 times its complexity. A semi-tetrahedral 0-handle has at most seven components of intersection with the parallelity handles (five 2-handles and two 1-handles). Again by Definition 4.13, its complexity is at least $1/4$, so it meets a number of parallelity handles equal to at most 28 times its complexity. A product non-parallelity 0-handle is disjoint from the parallelity handles. So, the number of 2-cells in $\partial_v \mathcal{B}$ is at most $56\Delta(\mathcal{H})$. \square

Lemma 8.15 (Length bound, almost normal case). *Let \mathcal{H} be a pre-tetrahedral handle structure for a compact orientable 3-manifold M . Let S be an almost normal surface in M , and let \mathcal{H}' be the handle structure on $M \setminus S$ described in Definition 4.16. Let \mathcal{B} be a maximal generalised parallelity bundle for $M \setminus S$, or a union of components of the parallelity bundle. Then the length of $\partial_v \mathcal{B}$ is at most $88\Delta(\mathcal{H})$.*

Proof. As above, it suffices to bound the number of components of intersection between the non-parallelity 0-handles of \mathcal{H}' and the parallelity handles. When a 0-handle H_0 of \mathcal{H} is divided along normal discs into 0-handles of \mathcal{H}' , the sum of the complexities of these 0-handles is equal to the complexity of H_0 . The number of intersections between each such 0-handle and the parallelity handles is at most 28 times its complexity, as argued in the proof of Lemma 8.14.

Thus, it suffices to consider when a 0-handle H_0 of \mathcal{H} contains an almost normal piece, plus possibly some triangles and squares. If the almost normal piece is tubed, the number of intersections of the resulting 0-handles with the parallelity handles is the same as in the situation when the tube is compressed. So again the number of intersections is at most 28 times the complexity of H_0 .

So, we now consider the situation with H_0 contains an octagon plus possibly some triangles and squares. Then H_0 must be tetrahedral. One can check that the resulting 0-handles of \mathcal{H}' have at most 14 intersections with the parallelity 2-handles and at most 8 intersections with the parallelity 1-handles. Again, since the complexity of a tetrahedral 0-handle is at least $1/2$, it follows that the

number of intersections between these 0-handles and the parallelity handles is at most 44 times the complexity of H_0 .

We must then double this to get an upper bound on the length of $\partial_v \mathcal{B}$, because we must also take account of the intersections between non-parallelity 1-handles of \mathcal{H}' and the parallelity handles. But, as argued in the proof of Lemma 8.14, the number of these is at most the number of intersections between the non-parallelity 0-handles of \mathcal{H}' and the parallelity handles. \square

9. EDGE SWAP BOUNDS FOR ANNULAR SIMPLIFICATION

Road map: Using generalised isotopy moves, we have a sequence of surfaces interpolating between almost normal, nearly normal, and normal surfaces. For compression isotopies, tube compressions, parallelisation isotopies and generalised isotopy along an edge or face compression disc, we know how to transfer spines along those surfaces, and we can bound the number of edge swaps to do so. What we are missing is a similar bound for annular simplification. This case is somewhat more difficult than the others, and so we put it in its own section.

The main goal of this section is to bound the number of edge swaps required to transfer a cellular spine from one surface to another under an annular simplification. Our bound on edge swaps will be in terms of the width of an annulus, defined as follows.

Definition 9.1 (Width). Let M' be a compact orientable irreducible 3-manifold with a handle structure \mathcal{H}' . Let F be an incompressible subsurface of $\partial M'$ such that ∂F is standard and F is not a 2-sphere. Let \mathcal{B} be a maximal generalised parallelity bundle for (M', F) , and let B be an essential annular component of \mathcal{B} . Then the *width* of B is defined as follows.

Pick a component A of $\partial_h B$ and let P be its intersection with the parallelity bundle. By (6) in the definition of a generalised parallelity bundle, Definition 6.4, $A \setminus \setminus P$ lies in a union of disjoint discs in the interior of A . Hence, one component A_- of P is obtained from A by removing some discs from its interior. Define the width of B to be length of the shortest cellular path in A_- joining the two components of ∂A , where the length of a path is the number of edges traversed.

Note that width is independent of the choice of component A of $\partial_h B$, since the parallelity bundle sets up a handle preserving homeomorphism from A_- to the corresponding surface in the other component of $\partial_h B$.

9.1. Edge swaps in annuli. We consider again cellular spines on surfaces, as in Subsection 8.1. In that subsection, we discussed edge swaps within discs. We now need to build similar tools within annuli.

Lemma 9.2 (Arranging a spine and an annulus). *Let S be a closed orientable surface with a cell structure \mathcal{C} , and Γ a spine for S that is cellular with respect to \mathcal{C} . Let A be an essential annulus embedded in S that is a union of cells, such that ∂A has length ℓ . Finally, let α be an essential properly embedded arc in A that is cellular and has length d . Then there is a sequence of at most $6g(S) + 2\ell + 4d$ edge swaps taking Γ to a cellular spine Γ' such that $\Gamma' \cap \text{int}(A) = \text{int}(\alpha)$.*

Proof. The proof is very similar to that of Lemma 8.7. Let D be the disc $A \setminus \setminus \alpha$. We will modify Γ so that it misses the interior of D .

Let e be an edge of Γ with interior that lies entirely in D , and let x be a point in the interior of e . Removing e from Γ gives a graph Γ_- , the exterior of which is an annulus A' . The arc e is essential in A' . Hence, there is an embedded core curve C for A' that intersects e once at exactly the point x . The boundary ∂A intersects the interior of A' in a (possibly empty) collection of properly embedded arcs. By choosing C suitably, we may assume that C intersects each of these arcs at most once. If C does intersect one of these arcs, e' say, then we perform the edge swap that removes e and inserts e' . On the other hand, if C is disjoint from ∂A , then it lies entirely in A . It must then intersect α by Lemma 8.4, and we can assume that it intersects each arc of $\alpha \setminus \setminus \Gamma_-$ at most once. Let e' be one such arc. If necessary, add sub-arcs of ∂A to e' so that it becomes properly embedded in A' . Perform an

edge swap that removes e and replaces it by this enlarged e' . By Lemma 2.2, at most $6g(S)$ edges lie in the interior of D , so at most $6g(S)$ edge swaps take Γ to a spine without edges in $\text{int}(D)$.

After this procedure, Γ intersects the interior of D in a collection of embedded arcs and star-shaped graphs. These divide D into discs, and each of these discs contains at least one 1-cell lying in ∂D . Hence, the number of arcs is less than twice the length $\ell + 2d$ of ∂D . Let e be an arc of $\text{int}(D) \cap \Gamma$, and let x be a point in $\text{int}(e) \cap \text{int}(D)$. Again, there is a closed embedded curve C intersecting Γ exactly at the point x . We may assume that C intersects each arc of $\partial A \setminus \Gamma$ at most once. If it does intersect some arc of $\partial A \setminus \Gamma$, then perform the edge swap that adds this arc plus possibly a sub-arc of $\alpha \setminus \Gamma$, forming a graph Γ_+ , and then removes the edge of Γ_+ that contains x . On the other hand, if C is disjoint from ∂A , then it intersects some arc of $\alpha \setminus \Gamma$, and we may perform an edge swap that adds this arc, plus possibly some sub-arcs of ∂A and removes the arc containing x . We do such an edge swap for each arc intersecting the interior of D , and there are at most $2\ell + 4d$ of these.

After at most $6g(S) + 2\ell + 4d$ edge swaps, we end with a spine Γ' such that $\Gamma' \cap \text{int}(A) \subseteq \text{int}(\alpha)$. In fact, $\Gamma' \cap \text{int}(A) = \text{int}(\alpha)$ since otherwise a core curve of A lies in the disc $S \setminus \Gamma'$, contradicting the assumption that A is essential in S . \square

Lemma 9.3 (Replacing an annulus). *Let S be a closed orientable surface with a cell structure \mathcal{C} , and let Γ be a spine that is cellular with respect to \mathcal{C} . Let A be an essential annulus embedded in S that is a union of cells, with ℓ the length of ∂A . Suppose \mathcal{C}' is obtained from \mathcal{C} by removing the interior of A and inserting some other cell structure. Then the number of edge swaps required to take Γ to a spine that is cellular with respect to \mathcal{C}' is at most*

$$6g(S) + 2\ell + 4d + |\alpha \cap \alpha'| + 2,$$

where α is any essential properly embedded arc in A that is cellular with respect to \mathcal{C} , d is the length of α , and α' is an embedded arc in A with the same endpoints as α that is cellular with respect to \mathcal{C}' .

Proof. First, by Lemma 9.2, we may apply at most $6g(S) + 2\ell + 4d$ edge swaps to take Γ to a spine Γ_1 that is cellular in \mathcal{C} and such that $\Gamma_1 \cap \text{int}(A) = \text{int}(\alpha)$. Now, we are assuming that α' is an embedded arc in A , but it is not necessarily properly embedded. So it might intersect Γ_1 and ∂A in its interior. But it contains a sub-arc α'' that is embedded and essential in A such that $\alpha'' \cap \partial A \cap \Gamma_1 = \partial \alpha''$. Also, let α''' be the arc obtained from α' by pushing its interior a little into the interior of A . By Lemma 8.2, we may perform $|\alpha \cap \alpha'|$ edge swaps to Γ_1 , that leaves $\Gamma_1 - A$ unchanged and that takes α to α''' . Then, at most two further edge swaps adjust the endpoints of α''' so that they are equal to those of α'' . After this process, the spine is cellular with respect to \mathcal{C}' . \square

9.2. Essential annular simplifications. Recall from Definition 6.12 that annular simplifications are either trivial or essential. Our next goal is to find bounds on edge swaps when an essential annular simplification is performed. To do so, we first set up some notation.

Let M be a compact orientable 3-manifold with a pre-tetrahedral handle structure \mathcal{H} . Let S be a closed normal or almost normal surface with a given transverse orientation, and let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points.

Inductively, assume we have performed a (possibly empty) sequence of annular simplifications in the direction S points, obtaining a surface S' disjoint from the interior of \mathcal{B} . Suppose that S' admits a further essential annular simplification. Then the setup is as follows, illustrated in Figure 24.

Recall from Definition 6.12 that we will replace an annulus $A \subset S'$ with an annulus A' that lies on the vertical boundary of a component of \mathcal{B} . Because this is an essential annular simplification, there is a component B of the generalised parallelity bundle of the form $A_B \times I$, where A_B is an annulus, and $(A_B \times I) \cap S' = A_B \times \partial I$. The annuli $A_B \times \{0\}$ and $A_B \times \{1\}$ form sub-annuli of $A \subset S'$; see Figure 24. Let A'' denote the annulus $A \setminus (A_B \times \partial I)$.

The vertical boundary $\partial A_B \times I$ is incompressible by assumption. It has two components, one of which is A' and will replace A after the annular simplification. Denote the other by V . Then ∂V and $\partial A''$ agree.

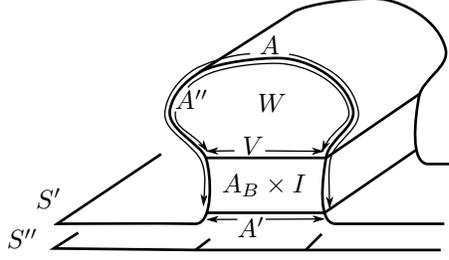


FIGURE 24. Schematic picture for when S' admits an essential annular simplification.

Because the annular simplification is essential, $A \cup A'$ bounds a 3-manifold P that forms a product region. This is a solid torus. The thickened annulus $A_B \times I$ forms a regular neighbourhood of A' in that solid torus. The remainder $P \setminus (A_B \times I)$ is a solid torus that we denote by W .

When performing annular simplifications, we will be isotoping annuli across solid tori. The following sequence of results leads to a bound on the number of edge swaps required under these moves.

Theorem 9.4 (Winding number in triangulated solid torus). *Let \mathcal{T} be a triangulation of a solid torus M . Suppose that some simple closed longitudinal curve λ in ∂M is simplicial. Let C be a simplicial simple closed curve in ∂M that intersects λ once transversely. Then the winding number of C in M is at most $10\Delta(\mathcal{T}) + 3$.*

Proof. We can assume that C runs over at most one edge in each 2-simplex in ∂M , as otherwise we may shorten C . Its length is therefore at most half the number of triangles in ∂M ; hence it is at most $(3/2)\Delta(\mathcal{T})$. Similarly, λ has length at most $(3/2)\Delta(\mathcal{T})$. Let w be the winding number of C . If w is odd, we replace C by a simplicial, but possibly non-embedded curve, with winding number $w + 1$, by Dehn twisting about λ . The resulting curve has length at most $3\Delta(\mathcal{T})$. Attach a 2-cell to M along this curve. We can triangulate this 2-cell by using at most $3\Delta(\mathcal{T})$ triangles. Then attach a 3-cell. The number of triangles in the boundary of this 3-cell is at most $9\Delta(\mathcal{T})$. So we can form a triangulation of this 3-manifold using at most $10\Delta(\mathcal{T})$ tetrahedra. This manifold is the lens space $L(p, 1)$, where $p = w$ or $w + 1$. It is a theorem of Jaco, Rubinstein and Tillmann [10] that, when p is even, $\Delta(L(p, 1)) = p - 3$. Hence, $10\Delta(\mathcal{T}) \geq p - 3 \geq w - 3$. \square

Lemma 9.5 (Triangulation from handle structure, solid torus). *Let the setup be as in Figure 24, as above. That is:*

- \mathcal{H} is a pre-tetrahedral handle structure on a compact orientable 3-manifold M containing a closed, transversely oriented surface S that is normal or almost normal.
- \mathcal{B} is a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points.
- S' is a transversely oriented surface that is obtained from S by a sequence of clean annular simplifications. It is disjoint from the interior of \mathcal{B} , with an essential annular component $A_B \times I$ of \mathcal{B} meeting S' in $A_B \times \partial I$.
- V is the vertical boundary component of $A_B \times I$ such that ∂V bounds an annulus A'' in S' , and $V \cup A''$ bounds a solid torus W with interior disjoint from $A_B \times I$.
- S' has a cell structure, where each 2-cell is either a component of intersection between $S' \cap S$ and \mathcal{H} , or a 2-cell in $\partial_v \mathcal{B}$.

If S is normal, define Δ to be the sum of the complexities of the 0-handles of $M \setminus S$ that intersect W . If S is almost normal, define Δ to be the sum of the complexities of the 0-handles of M that intersect W .

Suppose that the only components of \mathcal{B} in W are I -bundles over discs. Then the solid torus W admits a triangulation with at most $1320\Delta = 3t\Delta$ tetrahedra, where $t = 440$. Each 2-cell of $S' \cap W$

is a union of simplices in this triangulation, as is each 2-cell of V . There are at most $440\Delta = t\Delta$ 1-cells in the cell structure on $\partial W \setminus \partial_h \mathcal{B}$.

Proof. Suppose first that S is normal. Then $M \setminus S$ inherits a handle structure \mathcal{H}' from \mathcal{H} . Consider any 0-handle H of \mathcal{H}' that intersects W and is not a parallelity handle. Its boundary has a cell structure, obtained by viewing each of the associated handles on ∂H as 2-cells; the possible cell structures are shown in Figure 6.

Triangulate each 2-cell by adding a vertex to its interior and coning off. Then triangulate the 0-handle by adding a vertex to its interior and coning off. Colour the resulting tetrahedra red. In the case of a tetrahedral 0-handle, we use 72 tetrahedra. In the case of a semi-tetrahedral 0-handle, we use 60 tetrahedra. Note that a parallelity handle of length three or four does not occur, as it is a parallelity handle. Thus it only remains to consider the case of a product 0-handle (of length three), which uses 36 tetrahedra. Considering the associated complexity of these types of handles, as in Definition 4.13, we see that the number of red tetrahedra is at most 288 times the complexity of the 0-handle.

When two 0-handles in W are incident to the same 1-handle, glue them together along the relevant triangles in their boundary. This process does not require any more tetrahedra in our triangulation. Each component of \mathcal{B} in W is, by assumption, an I -bundle over a disc. Its vertical boundary has already been triangulated by triangles coloured red. We triangulate the horizontal boundary by adding a vertex to its interior and then coning. We then triangulate the whole I -bundle by adding a vertex to its interior and coning. For those tetrahedra meeting a red triangle, adjacent to the vertical boundary, colour the tetrahedron orange. Colour the rest of the tetrahedra, adjacent to the horizontal boundary, yellow. Note each red tetrahedron is adjacent to at most one orange, and each orange tetrahedron is adjacent to at most one yellow.

For each 2-handle of \mathcal{H}' that is not a parallelity handle, first triangulate its horizontal boundary by adding a vertex to its interior. Then triangulate the whole 2-handle by adding a vertex to its interior and coning off. Again colour tetrahedra adjacent to the vertical boundary orange, and the others yellow.

The boundary of each 3-handle has then been triangulated, and so we triangulate the 3-handle by adding a vertex to its interior and coning. For any resulting tetrahedra that are adjacent to a red tetrahedron, colour the tetrahedron orange. The remaining tetrahedra will be adjacent to a yellow tetrahedron on the boundary of the 3-handle, but also adjacent to an orange tetrahedron in the interior, so colour these yellow.

Now, there are at most 288Δ red tetrahedra. Each orange tetrahedron is adjacent to a red tetrahedron, but each red tetrahedron is adjacent to at most one orange. Hence there are at most 288Δ orange tetrahedra. Similarly, each yellow tetrahedron is adjacent to an orange tetrahedron, and each orange is adjacent to at most one yellow, so there are at most 288Δ yellow tetrahedra. Thus in total there are $3 * 288\Delta = 864\Delta$ tetrahedra in the triangulation of the solid torus.

When S is almost normal, this construction needs to be adjusted a little, because S contains an almost normal piece. Let H_0 denote the handle of \mathcal{H} containing this almost normal piece. Note S still cuts handles of \mathcal{H} into a pre-tetrahedral handle structure, except for two components of $H_0 \setminus S$ adjacent to the almost normal piece. Then away from H_0 , the complexity of M agrees with the complexity of $M \setminus S$, which is well-defined. The complexity of $H_0 \setminus S$ is not defined for exactly two components. However, note that if these two components are glued along the almost normal piece of S , the result is a pre-tetrahedral handle H with well-defined complexity. Note also that because S is separating, at most one of these two pieces meets W . Hence for our bound, we may use the bounds above in the normal case away from H , and we will bound the number of tetrahedra in one of the two components of $H_0 \setminus S$ in terms of the complexity of H to complete the bound.

When the almost normal piece is an octagon, then we triangulate the octagon by adding a vertex to interior and coning off. We then triangulate the 3-ball lying to one side of it, as described above, with red tetrahedra. This uses 84 tetrahedra: 16 adjacent to the octagon, 7×4 adjacent to 1-handles, 6×2 adjacent to 2-handles, and $(6 \times 2) + (4 \times 4)$ adjacent to remaining 0-handles. The octagon lies

in a tetrahedral 0-handle. The union H of the two components of $H_0 \setminus S$ adjacent to the octagon must be tetrahedral, and H has complexity at least $1/2$. Thus the number of red tetrahedra is at most 168 times the complexity of H .

When the almost normal piece is a tube, there are several cases. The handle H , which is the union of the two components of $H_0 \setminus S$ adjacent to the tube, could be tetrahedral, semi-tetrahedral, or even a parallelity or product handle, since S may run past the tube in triangles or squares on either side. The solid torus W might lie on the 3-ball side of the tube or the solid torus side of the tube. Finally, the tube might run between two squares, between a triangle and a square, or between two triangles. All these cases need to be considered.

In the case W lies on the 3-ball side of the tube, we triangulate the 3-ball by adding an essential arc to the tube, which cuts it into a disc with either 18 sides (in the case the tube joins two squares), 16 sides (for a triangle and a square), or 14 sides (when the tube joins two triangles). We then cone off the disc and cone off all other 2-cells, then cone the 3-ball, and colour these tetrahedra red. The different cases outlined above give different numbers of tetrahedra. The highest bound on tetrahedra per complexity occurs in the case that the tube connects a square to a square and H is semi-tetrahedral, with complexity at least $1/4$. In this case there are 110 tetrahedra produced from the coning procedure, so the number of tetrahedra is bounded by at most 440 times the complexity of H .

In the case W lies on the solid torus side of the tube, we triangulate the solid torus by attaching an edge compression disc, and then triangulating the resulting 3-ball. The edge compression disc cuts the tube into a disc with 18, 16, or 14 sides again, depending on whether the tube runs from a square to a square, a square to a triangle, or a triangle to a triangle, respectively. The edge compression disc itself appears twice on the boundary of the resulting 3-ball; each time it is coned into two triangles, hence coned to a total of four tetrahedra. Again considering the other 2-cells in all the cases outlined above, the highest bound on tetrahedra per complexity occurs either in the case that the tube runs between a triangle and a square, or between two triangles, and H is semitetrahedral. In that case, there are 66 red tetrahedra, so the number of tetrahedra is bounded by at most 264 times the complexity of H .

As in the normal case, obtaining a triangulation of W can be achieved by triangulating components of \mathcal{B} that are I -bundles over discs, remaining 2-handles, and remaining 3-handles, and this adds at most three times the number of tetrahedra. So in all cases, the result is a triangulation \mathcal{T} of the solid torus W with at most $1320\Delta = 3t\Delta$ tetrahedra, where $t = 440$ is the maximum number of triangles per complexity on the boundary of a 0-handle.

Finally, the number of 1-cells in $\partial W \setminus \partial_h \mathcal{B}$ is at most the number of triangles on the boundary of the 0-handles in $W \setminus \mathcal{B}$, which is at most $t\Delta$ for $t = 440$, as above. \square

Proposition 9.6 (Edge swap bound, essential annular simplification). *Let \mathcal{H} , M , S , \mathcal{B} be as in Lemma 9.5, and let S' , $A_B \times I$, V , A'' and the solid torus W be as in that lemma; see Figure 24.*

Suppose that the only components of \mathcal{B} in W are I -bundles over discs. Let S'' be the result of performing an annular simplification to S' , by removing $A = A'' \cup (A_B \times \partial I)$ from S' and replacing it by $A' = (\partial A_B \times I) - V$. Let w be the width of $A_B \times I$. Let ℓ be the length of $(\partial A_B \times \partial I) - \partial V$.

In the case where S is normal, let Δ be the sum of the complexities of the 0-handles of $M \setminus S$ that intersect W . In the case where S is almost normal, let Δ be the sum of the complexities of the 0-handles of M that intersect W . Then the following hold.

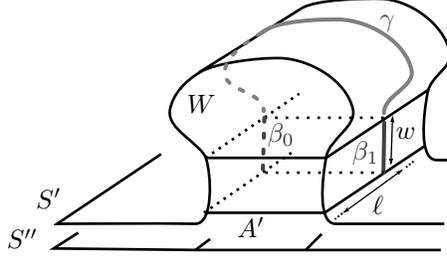
- (1) *Any cellular spine in S' may be converted to a cellular spine in S'' using at most*

$$6g(S) + 34t\Delta + 8w + \ell(6g(S) + 50) + 5 \text{ edge swaps,}$$

- (2) *Any cellular spine in S'' may be converted to a cellular spine in S' using at most*

$$6g(S) + 2\ell + 30t\Delta + 9 \text{ edge swaps,}$$

where in both cases, $t = 440$ is the constant of Lemma 9.5.


 FIGURE 25. Curves γ , β_0 , β_1 in the proof of Proposition 9.6.

Proof. Use Lemma 9.5 to obtain a triangulation for W with at most $3t\Delta$ tetrahedra, where t is the constant of that lemma, and such that the corresponding cell structure on $\partial W \setminus \partial_h \mathcal{B}$ contains at most $t\Delta$ 1-cells.

We now wish to use Lemma 9.3. This deals with situation where a cell structure on a surface is modified by changing it within a subsurface that is homeomorphic to an annulus. In our situation, one surface is S' , containing the annulus $A = A'' \cup (A_B \times I)$. We remove the interior of A , and replace it by the cell structure on $A' = (\partial A_B \times I) - V$. This gives a cell structure \mathcal{C}'' on the surface S'' . This is not quite the canonical cell structure on S'' , rather it is a refinement of it.

In order to apply Lemma 9.3, we need to specify essential arcs α and α' that are cellular in the canonical cell structure on S' and \mathcal{C}'' , respectively. We do this as follows. The annular region $A_B \times I$ has width w . So there is some essential arc β properly embedded in A_B such that $\beta \times I$ is vertical in the parallelity handles in $A_B \times I$ and where each component of $\beta \times \partial I$ has length w . Denote the components of $\beta \times \partial I$ by β_0 and β_1 ; see Figure 25.

Join the two points $(\beta \times \partial I) \cap A''$ by an embedded arc γ in A'' . We may ensure that γ is cellular in the cell structure on S' and simplicial in the triangulation of W . We may also ensure that it avoids the interior of $\partial_h \mathcal{B}$ and so its length in the cell structure is at most the number of 1-cells in $\partial W \setminus \partial_h \mathcal{B}$, which is at most $t\Delta$ by Lemma 9.5. Let α be $\beta_0 \cup \beta_1 \cup \gamma$. This is cellular in S' and has length at most $t\Delta + 2w$. Let α' be the component of $\partial\beta \times I \subset A'$ with the same endpoints as α . Note since α' is a fibre in the vertical boundary, it has length 1.

We need to find an upper bound on the intersection number between α' and the image of α under the isotopy from S' to S'' . Let C be the simple closed curve in ∂W obtained from γ by attaching the component of $\partial\beta \times I \subset V$ to which it is incident. This has winding number at most $10 \times 3t\Delta + 3 = 30t\Delta + 3$ in W , by Theorem 9.4 and Lemma 9.5. So, after the isotopy taking S' to S'' , the image of α and α' have intersection number at most $30t\Delta + 3$.

By Lemma 9.3, a cellular spine in S' may be converted to a cellular spine in \mathcal{C}'' using at most

$$6g(S) + 2\ell + 4 \cdot (t\Delta + 2w) + (30t\Delta + 3) + 2$$

edge swaps.

It may then be converted to a cellular spine in the canonical handle structure on S'' using at most a further $6g(S)\ell + 48\ell$ edge swaps, as follows. Recall from Lemma 6.13 that to get from \mathcal{C}'' to the canonical handle structure on S'' , we remove the edges that lie in the boundary of the annulus $A' = (\partial A_B \times I) - V$. This has the effect of combining each 0-handle in A' with the pair of 0-handles adjacent to it, and doing the same thing to the 1-handles. If this process creates any regions that are not discs, then they are replaced by discs. Therefore, let D_1, \dots, D_n be the 0-handles and 1-handles of S'' that intersect the interior of A' . The number of these is at most the length ℓ of the annulus. The length of the boundary of each D_i is at most 24 by Lemma 5.11. Hence, the number of edges swaps needed to move the spine off the interior of these discs is at most $6g(S)\ell + 48\ell$, by Corollary 8.8.

Thus for part (1), we have used at most

$$6g(S) + 34t\Delta + 8w + \ell(6g(S) + 50) + 5$$

edge swaps.

For part (2), a cellular spine in the canonical handle structure S'' is already cellular in \mathcal{C}'' . By Lemma 9.3, it may be converted to a cellular spine in the canonical handle structure on S' using at most

$$6g(S) + 2\ell + 4 + 30t\Delta + 5$$

edge swaps. Note the bounds are different here, because when we apply Lemma 9.3 in this situation, what is relevant is the length of α' , which is 1, rather than the length of α , which is at most $t\Delta(W) + 2w$. \square

Note that there is an asymmetry between the statements (1) and (2) in the above theorem. In particular, when a spine in S'' is transferred to one in S' , the width of the annular bundle is not relevant. This will be important and useful later.

At this point, we can also deal with trivial annular simplifications.

Lemma 9.7 (Edge swap bound, trivial annular simplification). *Let \mathcal{H} , M , S , \mathcal{B} and S' be as in Lemma 9.5. Let S'' be obtained from S' by performing a trivial annular simplification. This has the effect of removing an annulus A' in S' and replacing it by an annulus A'' that is a vertical boundary component of \mathcal{B} . Since the annular simplification is trivial, both boundary components of A' bound discs in S' , which are nested. Let D be the larger of these discs and let ℓ be the length of its boundary. Give S' and S'' the cell structures associated with their canonical handle structures. Then the following hold.*

- (1) Any cellular spine in S' may be converted to a cellular spine in S'' using at most

$$6g(S)(\ell + 1) + 50\ell \text{ edge swaps.}$$

- (2) Any cellular spine in S'' may be converted to a cellular spine in S' using at most $6g(S) + 2\ell$ edge swaps.

Proof. The canonical handle structure of S'' is obtained from that of S' as follows:

- (1) Remove the handles in A' and replace them with the cell structure on A'' . This gives a cell structure \mathcal{C}'' on S'' .
- (2) Remove the 1-cells in $\partial A''$. Any disc regions that are created become 2-cells. However, if any regions are not discs, then these are combined with the disc subsets of S'' that they surround to form 0-handles.

Let Γ' be a cellular spine for S' . Using Lemma 8.7, we can move Γ' off D using at most $6g(S) + 2\ell$ edge swaps. It is then cellular in \mathcal{C}'' . To make it cellular with respect to the canonical handle structure of S'' , we need to move it off the 1-cells in $\partial A''$. Let D_1, \dots, D_n be the 0-handles and 1-handles of S'' that intersect the interior of A'' . The number of these is at most the length ℓ of ∂D . The length of the boundary of each D_i is at most 24 by Lemma 5.11. So, the number of edge swaps needed to move the spine off the interior of these discs is at most $6g(S)\ell + 48\ell$ by Corollary 8.8. It is then cellular in S'' .

Now let Γ'' be a cellular spine for S'' . We refine the cell structure of S'' by making ∂D cellular. This keeps Γ'' cellular. We then move Γ'' off D using at most $6g(S) + 2\ell$ edge swaps. It is then cellular in S' . \square

9.3. The width of annular bundles. To apply Proposition 9.6, we will need to bound the width w in that lemma. This subsection contains results bounding width.

Lemma 9.8 (Existence of core curve). *For $1 \leq j \leq k$, let A_j be an annulus and let A'_j be the result of removing some disjoint discs from A_j . Let \mathcal{C} be a cell structure for $\bigcup_j A'_j$. Suppose certain 2-cells in \mathcal{C} are marked with an X . Define numbers as follows.*

- Let ℓ_j be the length of $\partial A'_j - \partial A_j$.
- Let w_j denote the length of the shortest cellular curve in A'_j joining the components of ∂A_j , or ℓ_j if there is no such curve.

- Let k be the maximal number of edges in the boundary of any 2-cell, counted with multiplicity.
- Let n_j be the number of cells in \mathcal{C} marked with an X .

Then, provided $\sum_j (n_j(k-2) + \ell_j) < \sum_j w_j$, there is a core curve of some A_j lying in A'_j that lies in just the 1-cells and 2-cells of \mathcal{C} and that misses all the 2-cells marked with an X and all the 1-cells in $\partial A'_j$.

Proof. Since $\sum_j (n_j(k-2) + \ell_j) < \sum_j w_j$, there is some j such that $n_j(k-2) + \ell_j < w_j$. Let $A = A_j$ and let $A' = A'_j$. Let $n = n_j$, $\ell = \ell_j$ and $w = w_j$.

First, subdivide \mathcal{C} to a cell structure \mathcal{C}' , without introducing any new vertices, so that each 2-cell has at most 3 edges. We define path metric on A' by declaring that each 1-cell has length 1, each 2-cell with three edges is an equilateral Euclidean triangle and each 2-cell with one or two edges is a spherical hemisphere. Define $f: A' \rightarrow [0, \infty)$ by letting $f(x)$ be the distance from x to a specific boundary component of A . Let L_i be $f^{-1}(i + (1/3))$ for integers i satisfying $0 \leq i < w - 1$.

Since we have chosen the inverse image of a real number that is neither an integer nor a half-integer, L_i is a union of disjoint properly embedded 1-manifolds. It separates one component of ∂A from the other. Each edge intersects at most one L_i . Hence, as each 2-cell of \mathcal{C}' has at most 3 sides, each 2-cell of \mathcal{C}' intersects at most one L_i . So, each 2-cell of \mathcal{C} intersects at most $(k-2)$ of the curves L_i . Therefore, the number of the L_i going through a 2-cell with an X is at most $n(k-2)$. The number intersecting $\partial A' - \partial A$ is at most ℓ . Since we are assuming that $n(k-2) + \ell < w$, there is some L_i that is disjoint from the 2-cells marked with an X and from the 1-cells in $\partial A'$. It is a closed 1-manifold separating the two components of ∂A . It must therefore contain the required core curve. \square

Lemma 9.9 (Width of incoherent essential annular components). *Let \mathcal{H} be a pre-tetrahedral handle structure of $M = S \times [0, 1]$, where S is not a torus. Let $\mathcal{B}(\mathcal{H})$ be a maximal generalised parallelity bundle for \mathcal{H} .*

- If \mathcal{H} contains a normal fibre that is not normally parallel to a boundary component, then let S be one such fibre with least weight.
- If every normal fibre in \mathcal{H} is normally parallel to a boundary component, then let S be an almost normal fibre with least weight.

Suppose that S is disjoint from the incoherent components of $\mathcal{B}(\mathcal{H})$ that are not I -bundles over discs. Let \mathcal{B} be a maximal generalised parallelity bundle for the handle structure obtained by cutting along S . Then the total width of the incoherent essential annular components of \mathcal{B} that are incident to S is at most $7640\Delta(\mathcal{H})$.

Proof. Let $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$ and \mathcal{B}_A be as in Lemma 7.7. By that lemma, \mathcal{B}_A is disjoint from $\mathcal{B}_I^{\text{coh}}(\mathcal{H})$. By assumption, S is disjoint from the incoherent components of $\mathcal{B}(\mathcal{H})$ that are not I -bundles over discs. For each component of \mathcal{B}_A , pick a horizontal boundary component. Let $A_1 \cup \dots \cup A_k$ be the union of these annuli. We will show that their total width is at most $7640\Delta(\mathcal{H})$.

We now form \mathcal{B}_A^- from \mathcal{B}_A by removing its intersection with $\mathcal{B}_D(\mathcal{H})$ and by removing any non-parallelity handles of \mathcal{B} . The result is disjoint from $\mathcal{B}(\mathcal{H})$, and consists only of parallelity handles of \mathcal{B} , so is an I -bundle. For $1 \leq j \leq k$, let $A'_j = A_j \cap \mathcal{B}_A^-$. Then A'_j is obtained from A_j by removing some discs. As in Lemma 9.8, let ℓ_j be the length of $\partial A'_j - \partial A_j$. Let w_j be the length of a shortest curve in A'_j joining the two components of ∂A_j , or if there is no such curve, let $w_j = \ell_j$. We will show that $\sum_j w_j \leq 7640\Delta(\mathcal{H})$. This will imply that the total width of $A_1 \cup \dots \cup A_k$ is at most $7640\Delta(\mathcal{H})$. For if there is no curve in some A'_j joining the two components of ∂A_j , then the width of A_j is at most ℓ_j .

Since $A'_1 \cup \dots \cup A'_k$ misses the parallelity handles of $\mathcal{B}(\mathcal{H})$, it lies entirely in the non-parallelity handles of \mathcal{H} .

We give S a cell structure, by declaring that each component of intersection between S and a handle of \mathcal{H} is a 2-cell. There is one exception to this, which is when S has a tubed piece. In this

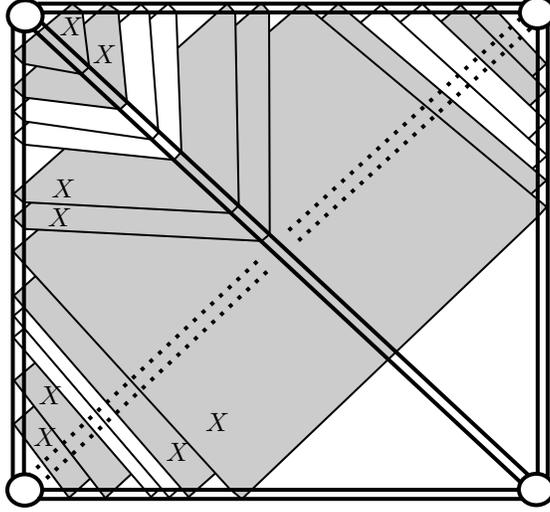


FIGURE 26. The two outermost triangles and squares for each type in a 0-handle are marked by an X .

case, we add in a further 1-cell, which cuts the annular tube to a disc. Note that the boundary of each 2-cell then has length at most 18, where the length is the number of 1-cells in the boundary.

We now place an X on various of these 2-cells, first in 0-handles, then 1-handles, then 2-handles. Within each non-parallelity 0-handle of \mathcal{H} , for each type of triangle and square of S in such a 0-handle, place an X on the four outermost ones in the collection; see Figure 26. Similarly, if S contains an almost normal piece, place an X on this. A tetrahedral 0-handle without an almost normal piece has at most $4 * 5 = 20$ cells marked with an X , a semi-tetrahedral 0-handle has at most 12 cells marked with an X , and a product handle has at most four. Multiplying by the minimal complexity of the cell, as in Definition 4.13, there are at most $48\Delta(\mathcal{H})$ such X 's that are not arising from almost normal pieces. (The extremal case is a semi-tetrahedral 0-handle containing a square and two triangle types.)

Within each non-parallelity 1-handle of \mathcal{H} , there are at most three types of disc of S , where all discs of the same type are normally parallel. Place an X on the outermost four discs of each type. In a 1-handle arising as the dual of a triangular face, there are at most 12 such X 's. In a 1-handle arising as the dual of a bigon, there are at most four such X 's. Note that a tetrahedral 0-handle meets four 1-handles coming from triangles, so it meets at most 48 such X 's, or at most 96 times the complexity. A semi-tetrahedral 0-handle meets two 1-handles from triangles, and two from bigons, so at most 32 such X 's, or at most 128 times the complexity. A non-parallelity product handle has three 1-handles from bigons, so meets at most a number of X 's equal to 96 times the complexity. Thus there are at most $128\Delta(\mathcal{H})$ such X 's.

Finally, there are the non-parallelity 2-handles. Each is incident to at least one non-parallelity 1-handle. So the number of non-parallelity 2-handles is at most 3 times the number of non-parallelity 1-handles, which is no more than $24\Delta(\mathcal{H})$. Within such a 2-handle, all components of intersection with S are normally parallel. We place an X on the outermost four. Thus there are at most $3 * 24\Delta(\mathcal{H}) * 4 = 288\Delta(\mathcal{H})$ of these.

In total, the number of X 's we have placed is at most

$$48\Delta(\mathcal{H}) + 1 + 128\Delta(\mathcal{H}) + 288\Delta(\mathcal{H}) = 464\Delta(\mathcal{H}) + 1 \leq 472\Delta(\mathcal{H}).$$

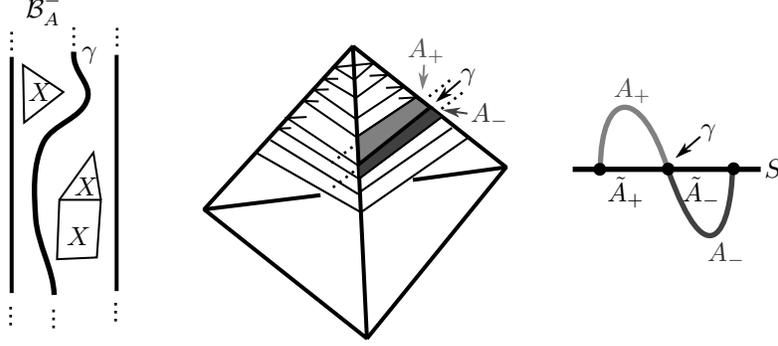


FIGURE 27. Left: a core curve γ misses all triangles and squares labeled X . Middle: Therefore it is adjacent to only triangles and squares that lie in the interior of the parallelity bundle; the I -bundles over γ form annuli A_- and A_+ . Right: A_- and A_+ are parallel to annuli \tilde{A}_- and \tilde{A}_+ on S .

The total length of $(\partial A'_1 - \partial A_1) \cup \dots \cup (\partial A'_k - \partial A_k)$ is at most $88\Delta(\mathcal{H})$, by Lemma 8.15. Lemma 9.8 implies that if $\sum_j w_j$ is more than

$$\sum_j (n_j(k-2) + \ell_j) = (472\Delta(\mathcal{H})) * (18-2) + 88\Delta(\mathcal{H}) = 7640\Delta(\mathcal{H}),$$

then there is a core curve γ of one of these punctured annuli that lies in just the 1-cells and 2-cells of the cell structure and that misses all the 2-cells marked with an X and the boundary of $\partial_h \mathcal{B}_A^-$. Thus γ lies deep inside the parallelity bundle, away from outermost triangles and squares in each 0-handle.

Suppose by way of contradiction that such a γ exists. Because γ does not meet outermost triangles and squares in each 0-handle, it is adjacent to two annuli A_- and A_+ , lying on the positive and negative sides of S , such that $A_- \cap S = \partial A_-$ and $A_+ \cap S = \partial A_+$ and that are vertical in the parallelity bundle for $M \setminus S$. These are parallel to annuli \tilde{A}_- and \tilde{A}_+ in S , by the incompressibility of S . Let γ_- and γ_+ be the curves $\partial A_- - \gamma$ and $\partial A_+ - \gamma$. See Figure 27.

Note that, near the two components of $\partial \tilde{A}_-$, \tilde{A}_- emanates from the same side of A_- , because otherwise $A_- \cup \tilde{A}_-$ would be a Klein bottle in $S \times [0, 1]$. Similarly, near the two components of $\partial \tilde{A}_+$, \tilde{A}_+ emanates from the same side of A_+ . Hence, \tilde{A}_- and \tilde{A}_+ emanate from opposite sides of γ in S .

We claim that γ_- and γ_+ are disjoint curves in S . Suppose that, on the contrary, they intersect. Since γ avoids the 2-cells labelled X , γ_+ also lies in a component of \mathcal{B}_A on the negative side of S . Let H be its horizontal boundary component that contains γ_+ . Since we are assuming that γ_- and γ_+ intersect, then γ_- also lies in H . Hence, H is an annular subsurface of S that contains the two curves γ_- and γ_+ but does not contain the curve γ between them. Therefore, $S \setminus \gamma$ is an annulus. One way to see this is to note that the boundary component of $S \setminus \gamma$ on the γ_- side is homotopic to γ_- , and this is homotopic to γ_+ since they both are core curves of the annulus H , and γ_+ is homotopic to the other boundary component of $S \setminus \gamma$. Hence, S is a torus, contrary to assumption.

As a result of this claim, the annuli \tilde{A}_- and \tilde{A}_+ intersect only along γ . Therefore, as in the proof of Lemma 7.7, we may remove $\tilde{A}_- \cup \tilde{A}_+$ from S and reglue its boundary components to form a new fibre S' . Then S is the normal sum of S' and the torus formed from \tilde{A}_- and \tilde{A}_+ by gluing its boundary components. In particular, S' is not normally parallel to a boundary component of $S \times [0, 1]$. If S was normal, then so is S' , and it has smaller weight than S , which is a contradiction. If S was almost normal, then S' is either almost normal or normal, and again it has smaller weight than S . Again, this is a contradiction. \square

We also need a version of this for surfaces that are normally cylindrical on exactly one side.

Lemma 9.10 (Width, S cylindrical one side). *Let \mathcal{H} be a pre-tetrahedral handle structure of $M = S \times [0, 1]$. Let S be a normal fibre that is normally cylindrical on exactly one side. Let $\mathcal{B}(\mathcal{H})$ be a maximal generalised parallelity bundle for \mathcal{H} . Suppose that S is disjoint from the incoherent components of $\mathcal{B}(\mathcal{H})$ that are not I -bundles over discs. Let \mathcal{B} be a maximal generalised parallelity bundle for the handle structure obtained by cutting along S . Then the total width of the incoherent essential annular components of \mathcal{B} that are incident to S is at most $7640\Delta(\mathcal{H})$.*

Proof. Suppose that width of the incoherent essential annular components of \mathcal{B} is more than $7640\Delta(\mathcal{H})$. The argument above gives that there are annuli A_- and A_+ on opposite sides of S with $A_- \cap S = \partial A_-$ and $A_+ \cap S = \partial A_+$, with essential boundary curves and that are vertical in the parallelity bundle for $M \setminus S$. This contradicts the assumption that S is normally cylindrical on exactly one side. \square

10. INTERPOLATING SPINES

Road map: At this stage, we know how to interpolate between almost normal and normal surfaces, using nearly normal surfaces. We know how to transfer spines on the surfaces, and we have bounds on the number of edge swaps required to transfer a spine for each generalised isotopy move. In this section, we put that information together to complete the proof of Theorem 1.4. The proof involves four results on transferring spines across $S \times [0, 1]$ that have somewhat subtly different hypotheses, namely Propositions 10.1 and 10.2, and Theorems 10.3 and 10.4. Before diving into the details, we summarise the argument, and how these fit together.

Recall that Theorem 1.4 gives upper and lower bounds on the minimal number of tetrahedra in a triangulation of $S \times [0, 1]$ with given 1-vertex triangulations of $S \times \{0\}$ and $S \times \{1\}$. It will be a fairly rapid consequence of Theorem 10.4, which provides an upper bound on the number of edge swaps required to take a cellular spine in $S \times \{0\}$ to one in $S \times \{1\}$ when $S \times [0, 1]$ has a coherently bundled pre-tetrahedral handle structure. Theorem 10.4 is proved by induction on the complexity of the handle structure. The inductive step implies that whenever the handle structure contains a normally acylindrical fibre that is not normally parallel to a boundary component, we may cut along it, giving two coherently bundled handle structures with smaller complexity. Thus Theorem 10.4 follows quickly from a similar result, Theorem 10.3, with the additional hypothesis that the only normally acylindrical fibres are the ones that are boundary parallel.

Theorem 10.3 is also proved by induction, but the argument is more delicate. The proof divides into several cases; in most we consider the existence of a normal fibre that is not normally parallel to the boundary. Necessarily, this is normally cylindrical. Depending on the case, we may pick such a fibre that is innermost (as in Definition 7.15) or of least weight. If there is no such fibre, then we pick an almost normal fibre with least weight, which exists by Theorem 4.24. Inductively, we get a collection of normal fibres. We know how to interpolate between these normal fibres, using almost normal and nearly normal surfaces. We also have upper bounds on the number of edge swaps required to transfer a spine in one fibre to the next. In Propositions 10.1 and 10.2, we use these to bound the total number of edge swaps in terms of the complexity of the handle structure and the width of certain annular bundles. In Proposition 10.1, these annular bundles are components of the generalised parallelity bundles \mathcal{B}_i given by Lemma 6.31. However, we really need the generalised parallelity bundles \mathcal{B} for $M \setminus S$, where S is a normal surface in M . In Proposition 10.2, we obtain an upper bound on the number of edge swaps in terms of the width of the annular components of \mathcal{B} . The bounds are then used in Theorem 10.3.

Proposition 10.1 (Edge swap bounds under generalised isotopy). *Let \mathcal{H} be a pre-tetrahedral handle structure for $M = S \times [0, 1]$. Let S be a normal or almost normal fibre with a transverse orientation. Let $S = S_0, \dots, S_n = S'$ be a sequence of surfaces as in Lemma 6.31. Let M_i be the manifold between S_i and S_{i+1} . Let \mathcal{B}_i be a maximal generalised parallelity bundle for the manifold between S_i and S_{i+1} . In the case where S is normal, let \mathcal{H}' be the resulting handle structure for the component of $M \setminus S$ into which S points. When S is almost normal, let \mathcal{H}' equal \mathcal{H} .*

- (1) Let Γ be a cellular spine for S . Let w denote the total width of the incoherent essential annular components of $\mathcal{B}_0 \cup \dots \cup \mathcal{B}_{n-1}$. Then there is a sequence of at most $k\Delta(\mathcal{H}') + 8w$ edge swaps taking the cellular spine Γ to a spine Γ' that is cellular in S' .
- (2) Let Γ' be a cellular spine for S' . Then there is a sequence of at most $k\Delta(\mathcal{H}')$ edge swaps taking Γ' to a spine Γ that is cellular in S .

In both cases, $k = 4224g(S) + 32472$, or $4224g(S) + 17512 + 34t$, where $t = 440$ is the constant of Proposition 9.6.

Proof. For each i , there is a sequence of generalised isotopy moves taking S_i to S_{i+1} , as in Lemma 6.31. Let $S_i = S_i^0, \dots, S_i^{m(i)} = S_{i+1}$ be these surfaces. Each of these surfaces S_i^j inherits its canonical handle structure. We will work with the associated cell structure for this, which we denote \mathcal{C}_i^j . For each surface S_i^j , we will pick a spine Γ_i^j that is cellular with respect to \mathcal{C}_i^j . For each j , the product region between S_i^j and S_i^{j+1} provides a homeomorphism $S_i^j \rightarrow S_i^{j+1}$ that is well-defined up to isotopy. We will ensure that the image of Γ_i^j under this homeomorphism will be related to Γ_i^{j+1} by a controlled number of edge contractions and expansions. Note that depending whether we are proving (1) or (2) of the proposition, we either create Γ_i^{j+1} from Γ_i^j or we create Γ_i^j from Γ_i^{j+1} .

For $i > 0$, let \mathcal{H}_i be the handle structure on M_i . When $i = 0$ and S is normal, let \mathcal{H}_0 be the handle structure on the manifold M_0 between S_0 and S_1 . When $i = 0$ and S is almost normal, let \mathcal{H}_0 denote the handle structure on the component of $M \setminus S_1$ containing S . Thus, $\sum_i \Delta(\mathcal{H}_i) \leq \Delta(\mathcal{H}')$ by Lemma 4.14. Many of the bounds below will be in terms of the length of the vertical boundary of \mathcal{B}_i . This length is at most $88\Delta(\mathcal{H}_i)$ by Lemma 8.14 or Lemma 8.15. So the sum of these lengths over all i is at most $88\Delta(\mathcal{H}')$.

We consider the various types of move in turn.

Case 1. S_i^{j+1} is obtained from S_i^j by a generalised isotopy move across an I -bundle over a disc.

Denote the I -bundle over the disc by $E \times I$, where the length of E in S_i^j is ℓ . By Lemma 8.12, if S_i^j is equipped with a cellular spine Γ_i^j , then we may build a cellular spine Γ_i^{j+1} on S_i^{j+1} by performing at most $6(2\ell + 1)g(S) + 92\ell - 64 \leq 18\ell g(S) + 92\ell$ edge swaps. By the same lemma, if S_i^{j+1} has a cellular spine Γ_i^{j+1} , then we may build a cellular spine Γ_i^j on S_i^j by performing at most $6g(S) + 8\ell + 40 \leq 6\ell g(S) + 48\ell$ edge swaps. Note that the sum of all such ℓ over all these moves is at most the length of the vertical boundary of $\mathcal{B}_0, \dots, \mathcal{B}_{n-1}$, which is at most $88\Delta(\mathcal{H}')$ by Lemma 8.14 or Lemma 8.15. Therefore, the total number of edge swaps, over all these generalised isotopy moves across I -bundles over discs, is at most $(1584g(S) + 8096)\Delta(\mathcal{H}')$ when going from S to S' , and at most $(528g(S) + 4224)\Delta(\mathcal{H}')$ when going in the other direction.

Case 2. S_i^{j+1} is obtained from S_i^j by a clean annular simplification across an essential annular bundle.

In this situation, there is an annular region $A \times I$ of \mathcal{B}_i such that $(A \times I) \cap S_i^j = A \times \partial I$. Let w' denote the width of $A \times I$. A component of $(M \setminus S_i^j) \setminus (A \times I)$ is a solid torus W . The only components of \mathcal{B}_i lying in W are I -bundles over discs. Let A' be the annulus $\partial W \setminus (\partial A \times I)$. The annular simplification removes $A' \cup (A \times \partial I)$ from S_i^j and replaces it by $V' = (\partial A \times I) - \partial W$, giving the surface S_i^{j+1} . Let ℓ be the length of the annulus V' .

By Proposition 9.6, if S_i^j is equipped with a cellular spine, then we may build a cellular spine on S_i^{j+1} by a sequence of at most $6g(S) + 34t\Delta + 8w' + \ell(6g(S) + 50) + 5$ edge swaps, where $t = 440$ is the constant from Proposition 9.6. This is at most $12g(S)\ell + 55\ell + 34t\Delta + 8w'$. If instead S_i^{j+1} is equipped with a cellular spine, then we may build a cellular spine on S_i^j by a sequence of at most $6g(S) + 2\ell + 30t\Delta + 9 \leq 6g(S)\ell + 11\ell + 30t\Delta$ edge swaps. Again, the sum of all such ℓ over all these moves is at most $88\Delta(\mathcal{H}')$ by Lemma 8.14 or Lemma 8.15. The sum of Δ over all such solid tori W , one for each annular simplification, is at most $\Delta(\mathcal{H}')$. Therefore, the total number of edge swaps,

over all these annular simplifications, is at most $(1056g(S) + (34t + 4840))\Delta(\mathcal{H}') + 8w$ when going from S to S' and at most $(528g(S) + (30t + 968))\Delta(\mathcal{H}')$ when going in the other direction.

Case 3. S_i^{j+1} is obtained from S_i^j by a trivial annular simplification.

This replaces an annulus in S_i^j with a vertical boundary component of \mathcal{B}_i . Let ℓ be the length of this vertical boundary component. By Lemma 9.7, we may convert a cellular spine in S_i^j to one in S_i^{j+1} using at most $6g(S)(\ell + 1) + 50\ell \leq 12g(S)\ell + 50\ell$ edge swaps. Alternatively, we may convert a cellular spine in S_i^{j+1} to one in S_i^j using at most $6g(S) + 2\ell \leq 6g(S)\ell + 2\ell$ edge swaps. Again, the sum of all such ℓ over all these moves is at most $88\Delta(\mathcal{H}')$ by Lemma 8.14 or Lemma 8.15. Thus, the total number of edge swaps is at most $(1056g(S) + 4400)\Delta(\mathcal{H}')$ when going from S to S' , and at most $(528g(S) + 176)\Delta(\mathcal{H}')$ when going from S' to S .

Case 4. S_i^{j+1} is obtained from S_i^j by a parallelisation isotopy.

Lemma 8.13 states that we may convert a cellular spine in one surface to a cellular spine in the other using at most $6g(S) + 2\ell$ edge swaps, which is at most $(6g(S) + 2)\ell$. So, in total, the number of edge swaps is at most $(528g(S) + 176)\Delta(\mathcal{H}')$

Case 5. S_i^{j+1} is obtained from S_i^j by a compression isotopy.

By Lemma 5.18, the compression isotopy takes the canonical handle structure for S_i^j to that for S_i^{j+1} . So, we may transfer a spine from one surface to the other without using any edge swaps.

So, adding the total number of edge swaps in each of Cases 1 to 5, we obtain the required upper bounds on the number of edge swaps. \square

The above proposition bounds edge swaps in terms of the width of the essential incoherent annular components of the parallelity bundles $\mathcal{B}_0, \dots, \mathcal{B}_{n-1}$. We now replace this with a bound that is solely in terms of the parallelity bundle for S .

Proposition 10.2 (Edge swap bounds, simplified version). *Let $M = S \times [0, 1]$, and let \mathcal{H} be a pre-tetrahedral handle structure for M . Let S be a normal fibre with a transverse orientation. Let $S = S_0, \dots, S_n = S'$ be a sequence of surfaces as in Lemma 6.31. Let \mathcal{B} be a maximal generalised parallelity bundle for the component of $M \setminus S$ into which S points. Let w be the sum of the widths of the essential incoherent annular components of \mathcal{B} that are incident to S . Let Γ be a cellular spine for S . Then Γ can be converted to a cellular spine for S' using at most $k\Delta(\mathcal{H}) + 122240\Delta(\mathcal{H}) + 8w$ edge swaps. Here, k is the constant from Proposition 10.1.*

Proof. For $0 \leq i \leq n-1$, let M_i be the manifold between S_i and S_{i+1} . Let \mathcal{H}_i be the handle structure on M_i . Now, when S is isotoped a little into $M \setminus S$ in the direction specified by its transverse orientation, the resulting surface is acylindrical in $M \setminus S$ on the side into which it does not point, since it is normally parallel to the boundary. Hence, by Proposition 7.11, each of the surfaces S_i is acylindrical in $M \setminus S$ on the side into which it does not point. In particular, for $i > 0$, S_i is acylindrical on that side in the manifold $M_{i-1} \cup M_i$. By the way S_1 is constructed, specifically conclusion (4) and (7) in Lemma 6.31, the essential incoherent annular components of \mathcal{B}_0 are a subset of the essential incoherent annular components of \mathcal{B} . Hence, their total width is at most w . By Lemma 9.10, applied to the surface S_i in $M_{i-1} \cup M_i$ for $i > 0$, the total width of the essential annular components of \mathcal{B}_i is at most $7640(\Delta(\mathcal{H}_{i-1}) + \Delta(\mathcal{H}_i))$. Summing this over all i , the total width of the incoherent essential annular components of $\mathcal{B}_0, \dots, \mathcal{B}_{n-1}$ is at most at most $15280\Delta(\mathcal{H}) + w$. Hence, by Proposition 10.1, the number of edge swaps that one must apply to the spine in S to make it cellular in S' is at most $k\Delta(\mathcal{H}) + 122240\Delta(\mathcal{H}) + 8w$. \square

Recall from Definition 7.2 that \mathcal{H} is coherently bundled if no vertical boundary component of its maximal generalised parallelity bundle \mathcal{B} is an incompressible annulus with both boundary components on $S \times \{0\}$ or $S \times \{1\}$. It follows that \mathcal{H} is coherently bundled if and only if the

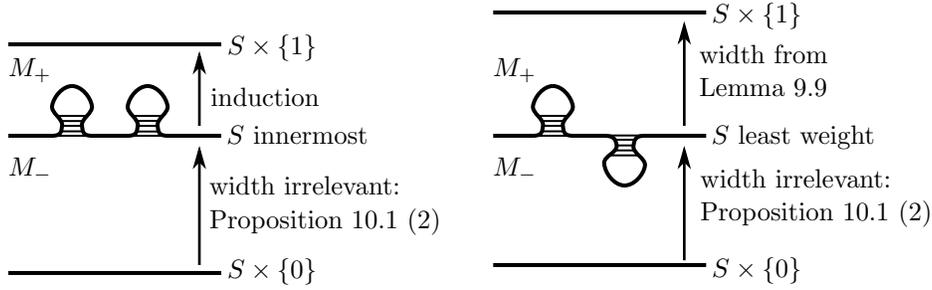


FIGURE 28. Left: Case 1. Right: Case 2A.

only components of \mathcal{B} with horizontal boundary disjoint from $S \times \{0\}$ or disjoint from $S \times \{1\}$ are I -bundles over discs or boundary-trivial; see Theorem 6.18.

Theorem 10.3 (Edge swaps across product, no acylindrical fibre). *Let S be an orientable, closed surface with genus at least two. Let \mathcal{H} be a coherently bundled pre-tetrahedral handle structure of $S \times [0, 1]$, such that every normally acylindrical fibre in \mathcal{H} is normally parallel to a boundary component. Let Γ be a cellular spine for $S \times \{0\}$. Then there is a sequence of at most $c\Delta(\mathcal{H})$ edge swaps taking Γ to a spine Γ' that is cellular with respect to $S \times \{1\}$. Here, $c = 9k$, where k is the constant from Proposition 10.1.*

Proof. We prove Theorem 10.3 by induction on $\Delta(\mathcal{H})$. The induction starts with $\Delta(\mathcal{H}) = 0$, in which case $S \times \{0\}$ and $S \times \{1\}$ are normally parallel. Then Γ becomes a spine for $S \times \{1\}$ using no swaps.

We now consider the inductive step. By assumption, every normal fibre in $M = S \times [0, 1]$ is normally cylindrical on at least one side, other than the two fibres that are normally parallel to the boundary.

Case 1. There exists a normal fibre that is normally cylindrical on the $S \times \{0\}$ side only. This case is illustrated schematically in Figure 28, left.

Then let S be one that is innermost, as in Definition 7.15. This divides M into two manifolds M_+ and M_- , where M_+ contains $S \times \{1\}$. These have handle structures \mathcal{H}_+ and \mathcal{H}_- such that $\Delta(\mathcal{H}_-) + \Delta(\mathcal{H}_+) = \Delta(\mathcal{H})$ by Lemma 4.14.

Observe first that \mathcal{H}_+ is coherently bundled by Lemma 7.6. Because S is innermost, no fibre in \mathcal{H}_+ is normally acylindrical other than those that are boundary parallel, by Lemma 7.16. Finally, $\Delta(\mathcal{H}_+) < \Delta(\mathcal{H})$. Hence, the inductive hypothesis applies to \mathcal{H}_+ . Thus, using at most $c\Delta(\mathcal{H}_+)$ edge swaps, we convert a cellular spine in S into a spine that is cellular in $S \times \{1\}$.

As for \mathcal{H}_- , apply generalised isotopy moves to S , all in the $S \times \{0\}$ direction, satisfying the conclusions of Lemma 6.31. By Corollary 7.12, the final surface in this sequence is normally parallel to $S \times \{0\}$.

We do not have a good bound on the width of the incoherent essential annular regions in \mathcal{H}_- . However, note that a bound on width is not required when going from $S \times \{0\}$ to S : by Proposition 10.1 (2), applied to \mathcal{H}_- , we obtain a cellular spine for S from Γ using at most $k\Delta(\mathcal{H}_-)$ edge swaps, which is less than $c\Delta(\mathcal{H}_-)$ edge swaps. So in total the number of edge swaps taking Γ to a cellular spine in $S \times \{1\}$ is at most $c\Delta(\mathcal{H}_+) + c\Delta(\mathcal{H}_-) = c\Delta(\mathcal{H})$.

Case 2. Every normal fibre is normally cylindrical on the $S \times \{1\}$ side at least, other than those that are boundary parallel, or there are no normal fibres that are not boundary parallel.

Case 2A. Every normal fibre is normally cylindrical on the $S \times \{0\}$ side also, other than those that are boundary parallel, or there are no normal fibres that are not boundary parallel. This case is illustrated schematically in Figure 28, right.

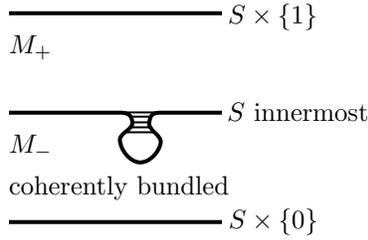


FIGURE 29. Case 2B.

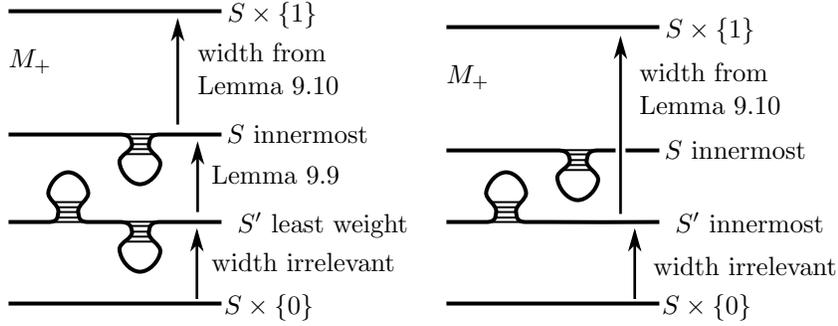


FIGURE 30. Left: Case 2B(i). Right: Case 2B(ii).

Pick a normal fibre S with least weight that is not boundary parallel, or if there is no normal fibre that is not boundary parallel, then pick an almost normal fibre with least weight, which exists by Theorem 4.24. Again $M \setminus S$ has components M_+ and M_- , with M_+ containing $S \times \{1\}$. Let \mathcal{B}_\pm be a maximal generalised parallelity bundle for M_\pm . By Proposition 7.14, we can interpolate from S to $S \times \{0\}$ and $S \times \{1\}$ using generalised isotopy moves satisfying the conclusions of Lemma 6.26 and 6.28 (in the case where S is almost normal) or 6.30 (in the case where S is normal). By Lemma 9.9, the total width of the incoherent essential annular components of \mathcal{B}_+ is at most $7640 \Delta(\mathcal{H})$. Hence, by Proposition 10.1 parts (1) and (2), there is a sequence of at most

$$k\Delta(\mathcal{H}) + k\Delta(\mathcal{H}) + 8 \times 7640 \Delta(\mathcal{H}) < c\Delta(\mathcal{H})$$

edge swaps taking Γ to a cellular spine in $S \times \{1\}$. We are using here the fact that $8 \times 7640 < 7k$.

Case 2B. There is some normal fibre that is normally cylindrical on the $S \times \{1\}$ side only. This case is illustrated in Figure 29.

Let S be one that is innermost in M . Let M_- be the component of $M \setminus S$ containing $S \times \{0\}$ and let \mathcal{H}_- be the handle structure that it inherits. It is coherently bundled by Lemma 7.4. Note that in M_- , every fibre is normally cylindrical in M_- on the $S \times \{0\}$ side at least, other than those that are boundary-parallel, by Lemma 7.16.

Case 2B(i). Every normal fibre in M_- is normally cylindrical in M_- on both sides, other than those that are boundary parallel. This case is illustrated in Figure 30, left.

Then pick a normal fibre S' in M_- that is not boundary parallel and that is of least weight. If there is no such fibre, then let S' be an almost normal fibre in M_- of least weight. By Proposition 7.14, we can interpolate from S' to $S \times \{0\}$ using generalised isotopy moves as in Lemma 6.26 and 6.28 (in the case where S' is almost normal) or 6.30 (in the case where S' is normal). By the same proposition, we can also interpolate from S' to S using generalised isotopy moves also as in those lemmas. Let \mathcal{B} be a maximal generalised parallelity bundle for $M_- \setminus S'$. The total width of the

essential incoherent annular components of \mathcal{B} is at most $7640 \Delta(\mathcal{H}_-)$ by Lemma 9.9. Hence, the total number of edge swaps taking Γ to a cellular spine in S is at most $2k\Delta(\mathcal{H}_-) + 8 \times 7640 \Delta(\mathcal{H}_-)$ by Proposition 10.1.

We can also perform generalised isotopy moves to S in the $S \times \{1\}$ direction, as in Lemma 6.31. The resulting normal surface admits no further generalised isotopy moves. So it is normally parallel to $S \times \{1\}$. (Recall that we are in Case 2.) If \mathcal{B}' is a maximal generalised parallelity bundle for $M \setminus S$, then the total width of the incoherent essential annular components of \mathcal{B}' is at most $7640 \Delta(\mathcal{H})$ by Lemma 9.10. Hence, by Proposition 10.2, the number of edge swaps that one must apply to the spine in S to make it cellular in $S \times \{1\}$ is at most $k\Delta(\mathcal{H}) + 122240 \Delta(\mathcal{H}) + 8 \times 7640 \Delta(\mathcal{H})$. So in total, the number of edge swaps is at most $9k\Delta(\mathcal{H}) = c\Delta(\mathcal{H})$. Here, we are using that $244480 \leq 6k$.

Case 2B(ii). There is a normal fibre in M_- that is normally cylindrical in M_- on the $S \times \{0\}$ side only. This case is illustrated in Figure 30, right.

Let S' be one that is innermost in M_- . We can perform generalised isotopy moves taking S' to $S \times \{0\}$. This is because, as mentioned above, every normal fibre in M_- is normally cylindrical on the $S \times \{0\}$ side at least, and so Proposition 7.13 applies. Also, by Proposition 7.13, we can get from S' to $S \times \{1\}$ in M using generalised isotopy moves, because every normal surface in M is cylindrical on this side, as we are in Case 2. We need to find an upper bound on the total width of the incoherent essential annular regions lying on the $S \times \{1\}$ side of S' . Any such region must contain an incoherent essential annular region for S , since S' is acylindrical in $M \setminus S$ on the S side. The total width of these annular regions for S is at least that of S' . By Lemma 9.10, the total width for S is at most $7640 \Delta(\mathcal{H})$, hence the same bound applies to S' . So, by Proposition 10.1 (2) and Proposition 10.2, the number of edge swaps that one must apply to take Γ from $S \times \{0\}$ to S' to $S \times \{1\}$ is at most $2k\Delta(\mathcal{H}) + 122240 \Delta(\mathcal{H}) + 8 \times 7640 \Delta(\mathcal{H}) < c\Delta(\mathcal{H})$. \square

The following is the main technical theorem in the paper.

Theorem 10.4 (Edge swaps across product). *Let S be a closed orientable surface with genus at least two. Let \mathcal{H} be a coherently bundled pre-tetrahedral handle structure of $S \times [0, 1]$. Let Γ be a cellular spine for $S \times \{0\}$. Then there is a sequence of at most $c\Delta(\mathcal{H})$ edge swaps taking Γ to a spine Γ' that is cellular with respect to $S \times \{1\}$. Here, c is the constant from Theorem 10.3.*

Proof. Let \mathcal{H} be a coherently bundled pre-tetrahedral handle structure of $S \times [0, 1]$. Let S_1, \dots, S_n be a maximal collection of disjoint normal fibres, none of which is normally parallel to a boundary component, no two of which are normally parallel and all of which are normally acylindrical. Note this collection may be empty. Suppose the surfaces are labelled so that they appear in the order $S \times \{0\} = S_0, S_1, \dots, S_n, S_{n+1} = S \times \{1\}$ in the product structure. Let \mathcal{H}_i be the handle structure inherited by the submanifold between S_i and S_{i+1} . Then by Lemma 7.10, \mathcal{H}_i contains no normal fibre that is normally acylindrical and not normally parallel to a boundary component. By Lemma 7.5, each \mathcal{H}_i is coherently bundled.

We build cellular spines $\Gamma_0, \dots, \Gamma_{n+1}$ for S_0, \dots, S_{n+1} as follows. We start with $\Gamma_0 = \Gamma$. Each \mathcal{H}_i is a coherently bundled pre-tetrahedral handle structure of $S \times [0, 1]$ such that every normally acylindrical fibre is normally parallel to a boundary component. Thus, inductively, we may apply Theorem 10.3 to \mathcal{H}_i and Γ_i to obtain Γ_{i+1} in S_{i+1} by at most $c\Delta(\mathcal{H}_i)$ edge swaps. We set $\Gamma' = \Gamma_{n+1}$. The total number of edge swaps is at most $c\Delta(\mathcal{H}_0) + \dots + c\Delta(\mathcal{H}_n)$, which is $c\Delta(\mathcal{H})$, by Lemma 4.14. \square

We are now able to prove Theorem 1.4.

Proof of Theorem 1.4. Suppose that \mathcal{T}_0 and \mathcal{T}_1 are 1-vertex triangulations of S . As in the proof of Proposition 2.8, we can triangulate $S \times [0, 1]$ so that $S \times \{0\}$ and $S \times \{1\}$ both have the triangulation \mathcal{T}_0 , using at most $18g(S)$ tetrahedra. A sequence of 2-2 Pachner moves relating \mathcal{T}_0 and \mathcal{T}_1 specifies a sequence of attachments of tetrahedra onto $S \times \{1\}$. We end with a triangulation of $S \times [0, 1]$ satisfying the required properties. It follows that the minimal number of 2-2 Pachner moves relating \mathcal{T}_0 to \mathcal{T}_1 gives an upper bound on the minimal number of tetrahedra.

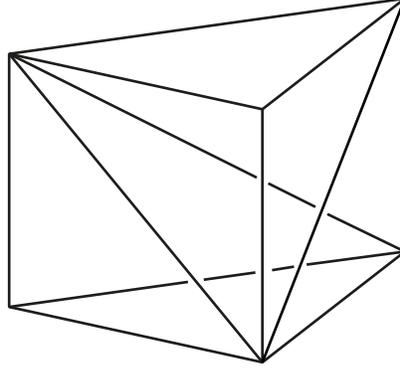


FIGURE 31. The triangulation of the product of a 2-simplex and an interval.

Conversely, suppose that we have a triangulation \mathcal{T} of $S \times [0, 1]$ so that \mathcal{T} restricted to $S \times \{0\}$ is \mathcal{T}_0 and \mathcal{T} restricted to $S \times \{1\}$ is \mathcal{T}_1 . We attach onto each of $S \times \{0\}$ and $S \times \{1\}$ the following triangulation of $S \times I$. Each triangle in $S \times \{0\}$, say, determines a prism in $S \times I$. This prism can be triangulated using three tetrahedra, as shown in Figure 31.

Gluing these prisms along their vertical faces gives a triangulation of $S \times I$. If we attach these onto \mathcal{T} , one on each boundary component, we end with a new triangulation \mathcal{T}' . This has the same triangulations on the boundary as \mathcal{T} did. However, \mathcal{T}' now has the property that for any tetrahedron, its intersection with $S \times \{0, 1\}$ is either a vertex, an edge or a face. So, when we dualise \mathcal{T}' to form a handle structure \mathcal{H} , this is pre-tetrahedral, for the following reason. We have one handle of \mathcal{H} for every simplex of \mathcal{T}' that does not lie wholly in $S \times \{0, 1\}$. So when a tetrahedron of \mathcal{T}' intersects $S \times \{0, 1\}$ in the empty set or in a vertex, the resulting 0-handle is tetrahedral. When a tetrahedron intersects $S \times \{0, 1\}$ in an edge, the resulting 0-handle is semi-tetrahedral. When a tetrahedron intersects $S \times \{0, 1\}$ in a face, the resulting 0-handle is a product handle with length 3. Note that $\Delta(\mathcal{H}) \leq \Delta(\mathcal{T}) + 12g(S)$, as follows. Suppose the top of the product in Figure 31 is glued to $S \times \{0\}$. Then the three vertices on the top of the product give 3-handles, whereas the three vertices on the bottom do not. So, we have the following complexities: The tetrahedron on the bottom becomes a product 0-handle with length 3 incident to a single 3-handle, with complexity $1/8$. The tetrahedron in the middle becomes a semi-tetrahedral 0-handle incident to two 3-handles, with complexity $1/4 + 2/8$. The tetrahedron on the top becomes a tetrahedral 0-handle incident to three 0-handles, with complexity $1/2 + 3/8$. So we have added complexity $3/2$ for each triangle of S . There are less than $4g(S)$ triangles of S by Lemma 2.3. We need to attach these triangulated products onto both $S \times \{0\}$ and $S \times \{1\}$. So in total, we add complexity $8g(S) * 3/2 = 12g(S)$.

Now, \mathcal{H} has empty parallelity bundle. For if it has any parallelity handles, then it would have to have a parallelity 2-handle. This would correspond to an edge of \mathcal{T}' not in $S \times \{0, 1\}$ but with endpoints lying in $S \times \{0, 1\}$. There is no such edge in our triangulation. In particular, \mathcal{H} is coherently bundled.

Pick a cellular spine Γ in $S \times \{0\}$ that is isotopic to the dual of \mathcal{T}_0 . Applying Theorem 10.4, we obtain a cellular spine Γ' in $S \times \{1\}$ that is obtained from Γ by at most $c\Delta(\mathcal{H})$ edge swaps. By Lemma 8.3, Γ' is therefore obtained from Γ by at most $24cg(S)\Delta(\mathcal{H})$ edge contractions and expansions.

The spine Γ' intersects each edge of \mathcal{T}_1 in at most 2 points. Hence, the total number of intersections with the edges of \mathcal{T}_1 is at most $12g(S)$ points by Lemma 2.3. Hence, by Lemma 2.9, one can change Γ' into the dual of \mathcal{T}_1 using at most $96g(S)^2$ edge contractions and expansions. So, we have related the dual of \mathcal{T}_0 to the dual of \mathcal{T}_1 using edge contractions and expansions, where the number of these is linearly bounded above by $\Delta(\mathcal{T})$. It is then straightforward to convert this to a sequence of 2-2 Pachner moves joining \mathcal{T}_0 to \mathcal{T}_1 . \square

11. PROOF OF THE MAIN THEOREM

We now have all the ingredients to give the proof of Theorem 1.3.

Proof of Theorem 1.3. As discussed in Section 1, the quantities (2) and (3) of Theorem 1.3 are known to lie within a bounded ratio of each other by the Švarc-Milnor lemma. We showed in Theorem 3.5 that quantities (3) and (4) lie within a bounded ratio of each other. We showed in Proposition 2.8 that $\Delta((S \times [0, 1])/\phi)$ is at most a constant times $\ell_{\text{MCG}(S)}(\phi)$. Hence, it remains to show that $\ell_{\text{MCG}(S)}(\phi)$ is at most a constant times $\Delta((S \times [0, 1])/\phi)$. In fact, in Proposition 2.7, it was shown that the spine graph $\text{Sp}(S)$ and the mapping class group $\text{MCG}(S)$ are quasi-isometric. So, it suffices to show that $\ell_{\text{Sp}(S)}(\phi)$ is at most a constant times $\Delta((S \times [0, 1])/\phi)$.

Let $M_n = (S \times [0, 1])/\phi^n$. In particular, M_n is an n -fold cover of M_1 . Let \mathcal{T}_1 be a triangulation for M_1 with $\Delta(\mathcal{T}_1) = \Delta(M_1)$. Let \mathcal{H}_1 be the dual handle structure, so $\Delta(\mathcal{H}_1) = \Delta(\mathcal{T}_1)$. Let S_1 be a normal fibre in \mathcal{H}_1 that has least weight in its isotopy class. By Lemma 7.8, S_1 exists and is normally acylindrical. Choose a spine Γ that is cellular in S_1 .

Now consider lifts to the finite cyclic covers M_n . The triangulation \mathcal{T}_1 lifts to a triangulation \mathcal{T}_n , with dual handle structure \mathcal{H}_n . The surface S_1 lifts to a normal fibre S_n in \mathcal{H}_n . The spine Γ lifts to a cellular spine identical to Γ on S_n . By cutting along all lifts of S_1 in \mathcal{H}_n , we deduce that $\Delta(\mathcal{H}_n) = n\Delta(\mathcal{H}_1) = n\Delta(M_1)$.

We claim that S_n is normally acylindrical. Suppose it is not and that there is annulus A in \mathcal{H}_n with $\partial A = A \cap S_n$ being essential curves in S_n , and with A vertical in the parallelity bundle of $M_n \setminus S_n$, and near ∂A , emanating from the same side of S_n . We may orient the n lifts of S to M_n coherently. Hence, when M_n is cut along these lifts, each component of the resulting 3-manifold has one boundary component oriented inwards and one oriented outwards. These lifts intersect A in a collection of parallel core curves. The outermost two are inconsistently oriented. Hence, there are two adjacent core curves in A that are inconsistently oriented. Between them lies an annulus that projects homeomorphically to an annulus in $M_1 \setminus S_1$ that makes S_1 normally cylindrical, which is a contradiction. This proves the claim.

Let \mathcal{H}'_n denote the handle structure obtained from cutting \mathcal{H}_n along S_n . By Lemma 4.14, $\Delta(\mathcal{H}'_n) = \Delta(\mathcal{H}_n)$.

By Lemma 7.4, \mathcal{H}'_n is coherently bundled. Thus Theorem 10.4 implies there is a sequence of at most $c\Delta(\mathcal{H}'_n) = c\Delta(\mathcal{H}_n)$ edge swaps taking Γ in $S_n \times \{0\}$ to a cellular spine Γ' in $S_n \times \{1\}$. By Lemma 8.3, this gives a bound of at most $12cg(S)\Delta(\mathcal{H}_n)$ edge contractions and expansions to take Γ to Γ' .

Apply the gluing map ϕ^n to obtain a spine $\phi^n(\Gamma)$ that is cellular in $S_n \times \{1\}$. By Lemma 2.10, at most $48g(S)^2L$ edge contractions and expansions are required to take Γ' to $\phi^n(\Gamma)$, where L is the number of 1-cells in Γ . Note L is independent of n since Γ is the same for all n . Thus the total number of edge contractions and expansions we have used is at most

$$12cg(S)\Delta(\mathcal{H}_n) + 48g(S)^2L = 12cng(S)\Delta(M_1) + 48g(S)^2L.$$

At this point, we do not have a bound on L . However, we know it is independent of n . Hence there exists N such that for all $n \geq N$, $48g(S)^2L \leq 12cng(S)\Delta(M_1)$. For such n , the total number of edge expansions and contractions required to take Γ to Γ' to $\phi^n(\Gamma)$ is at most $24cng(S)\Delta(M_1)$.

By Theorem 3.5, there exists a constant $k > 0$ depending only on $g(S)$ such that the translation length of ϕ^n in the spine graph is at least kn times the translation length of ϕ . Thus

$$kn\ell_{\text{Sp}(S)}(\phi) \leq \ell_{\text{Sp}(S)}(\phi^n) \leq 24cng(S)\Delta(M_1),$$

or

$$\Delta(M_1) \geq \frac{k}{24cg(S)}\ell_{\text{Sp}(S)}(\phi). \quad \square$$

REFERENCES

1. Ian Agol, *Ideal triangulations of pseudo-Anosov mapping tori*, Topology and geometry in dimension three, Contemp. Math., vol. 560, Amer. Math. Soc., Providence, RI, 2011, pp. 1–17. [9, 11]
2. Mark Bell, *flipper*, <https://markbell.github.io/build/html/software.html>. [11]
3. M. Bestvina and M. Handel, *Train-tracks for surface homeomorphisms*, Topology **34** (1995), no. 1, 109–140. [11]
4. Martin R. Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999. [1, 2, 7, 12]
5. Jeffrey F. Brock, *The Weil-Petersson metric and volumes of 3-dimensional hyperbolic convex cores*, J. Amer. Math. Soc. **16** (2003), no. 3, 495–535. [3]
6. ———, *Weil-Petersson translation distance and volumes of mapping tori*, Comm. Anal. Geom. **11** (2003), no. 5, 987–999. [1, 2]
7. David Futer and Saul Schleimer, *Cusp geometry of fibered 3-manifolds*, Amer. J. Math. **136** (2014), no. 2, 309–356. [2]
8. Ursula Hamenstädt, *Geometry of the mapping class groups. I. Boundary amenability*, Invent. Math. **175** (2009), no. 3, 545–609. [9, 11]
9. William Jaco and Ulrich Oertel, *An algorithm to decide if a 3-manifold is a Haken manifold*, Topology **23** (1984), no. 2, 195–209. [42]
10. William Jaco, Hyam Rubinstein, and Stephan Tillmann, *Minimal triangulations for an infinite family of lens spaces*, J. Topol. **2** (2009), no. 1, 157–180. [1, 3, 54]
11. William Jaco, J. Hyam Rubinstein, Jonathan Spreer, and Stephan Tillmann, *\mathbb{Z}_2 -Thurston norm and complexity of 3-manifolds, ii*, arXiv:1711.10737, 2017. [1]
12. ———, *On minimal ideal triangulations of cusped hyperbolic 3-manifolds*, arXiv:1808.02836, 2018. [1]
13. William Jaco, J. Hyam Rubinstein, and Stephan Tillmann, *Coverings and minimal triangulations of 3-manifolds*, Algebr. Geom. Topol. **11** (2011), no. 3, 1257–1265. [1]
14. ———, *\mathbb{Z}_2 -Thurston norm and complexity of 3-manifolds*, Math. Ann. **356** (2013), no. 1, 1–22. [1]
15. Marc Lackenby, *Taut ideal triangulations of 3-manifolds*, Geom. Topol. **4** (2000), 369–395. [7, 9]
16. ———, *The crossing number of composite knots*, J. Topol. **2** (2009), no. 4, 747–768. [5, 27, 28, 34, 42]
17. Marc Lackenby and Jessica S. Purcell, *The triangulation complexity of lens spaces and sol manifolds*, In preparation, 2019. [3]
18. W. B. R. Lickorish, *A finite set of generators for the homeotopy group of a 2-manifold*, Proc. Cambridge Philos. Soc. **60** (1964), 769–778. [2]
19. Bruno Martelli and Carlo Petronio, *Three-manifolds having complexity at most 9*, Experiment. Math. **10** (2001), no. 2, 207–236. [1]
20. Howard Masur, Lee Mosher, and Saul Schleimer, *On train-track splitting sequences*, Duke Math. J. **161** (2012), no. 9, 1613–1656. [3, 9, 11]
21. S. V. Matveev, *Tabulation of three-dimensional manifolds*, Uspekhi Mat. Nauk **60** (2005), no. 4(364), 97–122. [1]
22. Sergei Matveev, *Algorithmic topology and classification of 3-manifolds*, second ed., Algorithms and Computation in Mathematics, vol. 9, Springer, Berlin, 2007. [42]
23. Sergei Matveev, Carlo Petronio, and Andrei Vesnin, *Two-sided asymptotic bounds for the complexity of some closed hyperbolic three-manifolds*, J. Aust. Math. Soc. **86** (2009), no. 2, 205–219. [1]
24. R. C. Penner and J. L. Harer, *Combinatorics of train tracks*, Annals of Mathematics Studies, vol. 125, Princeton University Press, Princeton, NJ, 1992. [11]
25. Carlo Petronio and Andrei Vesnin, *Two-sided bounds for the complexity of cyclic branched coverings of two-bridge links*, Osaka J. Math. **46** (2009), no. 4, 1077–1095. [1]
26. Michelle Stocking, *Almost normal surfaces in 3-manifolds*, Trans. Amer. Math. Soc. **352** (2000), no. 1, 171–207. [18, 19, 24]
27. Abigail Thompson, *Thin position and the recognition problem for S^3* , Math. Res. Lett. **1** (1994), no. 5, 613–630. [19]
28. William P. Thurston, *The geometry and topology of three-manifolds*, Princeton Univ. Math. Dept. Notes, 1979, Available at <http://www.msri.org/communications/books/gt3m>. [1]