COMBINATORIAL METHODS IN DEHN SURGERY

C. McA. Gordon

Abstract. This is an expository paper, in which we give a summary of some of the joint work of John Luecke and the author on Dehn surgery. We consider the situation where we have two Dehn fillings $M(\alpha)$ and $M(\beta)$ on a given 3-manifold M, each containing a surface that is either essential or a Heegaard surface. We show how a combinatorial analysis of the graphs of intersection of the two corresponding punctured surfaces in M enables one to find faces of these graphs which give useful topological information about $M(\alpha)$ and $M(\beta)$, and hence, in certain cases, good upper bounds on the intersection number $\Delta(\alpha, \beta)$ of the two filling slopes.

1. Introduction

This expository paper is an expanded version of part of the author's lectures at the Workshop. The remaining part of the lectures covered material that is discussed in Sections 2 and 3 of the survey article [Go].

Here, we give an introductory account of the combinatorial methods developed by John Luecke and the author (mainly in the proof of the Knot Complement Conjecture [GL2]) for studying various questions about Dehn surgery. For more details, see [GL2], [GL3] and [GL4].

The setting is that we have two Dehn fillings $M(\alpha)$, $M(\beta)$ on a 3-manifold M, which contain surfaces $\widehat{P}, \widehat{Q}$ respectively, and the main goal is to obtain sharp upper bounds on the intersection number of α and β, under certain hypotheses on $\widehat{P}$ and $\widehat{Q}$.

In Section 2 we describe how $\widehat{P}$ and $\widehat{Q}$ give rise to labelled intersection graphs $G_P \subset \widehat{P}$, $G_Q \subset \widehat{Q}$. The idea of using such graphs in the context of Dehn surgery is due to Litherland [L].

In Section 3 we introduce the algebraic notion of a *type*, which plays a key role in the theory.

In Section 4 we show how certain faces or collections of faces of G_P give topological information about $M(\beta)$ (and/or $\widehat{Q}$), and of course similarly for G_Q.

Partially supported by NSF Grant DMS-9626550.

Research at MSRI supported in part by NSF grant #DMS 9022140.

To appear in the Proceedings of the Workshop of *Knots '96*, the fifth MSJ International Research Institute in Knot Theory, Waseda University, Tokyo, Japan, 1996.

Section 5 contains the main combinatorial assertion, Theorem 5.1, which says that, if G_P and G_Q are graphs as in Section 2, then either G_P represents all types or G_Q contains a configuration we call a $(p - \chi(\widehat{P}))$-*web*.

Finally, Section 6 contains some applications, mainly to Dehn surgery on knots in S^3, although we also say something about the case where $M(\alpha)$ and $M(\beta)$ are reducible.

We have tried to keep the discussion as general as possible, allowing $\widehat{P}$ and $\widehat{Q}$ to have arbitrary genus until the last part of Section 5. It must be admitted that this generality is somewhat bogus, inasmuch as in all the applications in Section 6, $\widehat{Q}$ is a 2-sphere and $\widehat{P}$ is either a 2-sphere or a torus. (For statements which apply to surfaces of arbitrary genus see [T] and [R].) The main point is that if $\widehat{Q}$ is a 2-sphere then, if G_Q contains a $(p - \chi(\widehat{P}))$-web, it contains one that is innermost in the obvious sense, and in such a *great* web one can find useful faces, namely Scharlemann cycles. Nevertheless, we hope that presenting the material in this way clarifies the logical structure of the argument.

2. Graphs of surface intersections

Let M be a compact, orientable, irreducible 3-manifold with torus boundary. Let α be a *slope*, that is, the isotopy class of an essential embedded circle, on ∂M. Then the result of α-*Dehn filling* on M is the manifold $M(\alpha) = M \cup V_\alpha$, where V_α is a solid torus, glued to M via a homeomorphism $\partial M \to \partial V_\alpha$ taking α to the boundary of a meridian disk of V_α.

If α and β are two slopes on ∂M, then $\Delta(\alpha, \beta)$ will denote the minimal geometric intersection number of α and β.

One finds in practice that statements of the following form tend to hold, and this is the main type of result that we shall be seeking:

> *if $M(\alpha)$ and $M(\beta)$ have certain special properties, then $\Delta(\alpha, \beta)$ is bounded above by some Δ_0.*

The bound Δ_0 will depend on the properties in question. The situation in which we are particularly interested, and to which our methods apply, is when $M(\alpha)$ and $M(\beta)$ contain certain kinds of surfaces.

So suppose that $\widehat{P}, \widehat{Q}$ are (closed, orientable) surfaces in $M(\alpha)$, $M(\beta)$ respectively. We may assume that $\widehat{P}$ meets V_α in a finite collection of meridian disks, so that $P = \widehat{P} \cap M$ is a surface in M each of whose boundary components has slope α. Similarly, $\widehat{Q}$ gives rise to a surface Q in M whose boundary components have slope β.

For our machinery to be applicable, it is important that ∂P and ∂Q be non-empty. So, we may state our basic assumption as follows:

(A1) *there are properly embedded surfaces $(P, \partial P)$, $(Q, \partial Q) \subset (M, \partial M)$ such that ∂P and ∂Q are non-empty, and each component of ∂P (resp. ∂Q) has slope α (resp. β).*

By an isotopy of (say) P we may assume in addition that

(A2) *P and Q intersect transversely, and each component of ∂P meets each component of ∂Q in exactly $\Delta(\alpha,\beta)$ points.*

It is clear, however, that for any α and β we can always find surfaces P and Q, of arbitrary genus, satisfying (A1) and (A2), and therefore we must impose some additional conditions if we are to extract any non-trivial information from this situation. The technical assumption which enables us to do this is:

(A3) *each arc component of $P \cap Q$ is essential in P and in Q.*

By (A2), $P \cap Q$ consists of a finite disjoint union of circles and properly embedded arcs, the endpoints of the arcs being the points of intersection of ∂P with ∂Q. Condition (A3) says that for no arc component γ of $P \cap Q$ is there an arc $\delta \subset \partial P$, with $\partial\gamma = \partial\delta$, and a disk $D \subset P$ such that $\partial D = \gamma \cup \delta$ (and similarly for Q).

We now describe some conditions under which (A3) holds.

Recall that a closed orientable surface S in a 3-manifold N is *essential* if either S has positive genus, is incompressible in N, and is not parallel to a component of ∂N, or S is a 2-sphere which does not bound a 3-ball in N.

First suppose

(a) *$\widehat{P}$ is essential in $M(\alpha)$ and M contains no essential surface homeomorphic to $\widehat{P}$.*

Let $K_\alpha \subset M(\alpha)$ denote the core of the solid torus V_α. We choose $\widehat{P}$ (among all essential surfaces in $M(\alpha)$ homeomorphic to $\widehat{P}$) so that $p = |\partial P| = |\widehat{P} \cap K_\alpha|$ is minimal. Then $p > 0$ by hypothesis. Also, standard arguments show that P is incompressible and boundary incompressible in M. In particular, we may assume that no circle component of $P \cap Q$ bounds a disk in Q.

If $\widehat{Q}$ also satisfies (a) (with $M(\alpha)$ replaced by $M(\beta)$), and is chosen to minimize $q = |\partial Q| = |\widehat{Q} \cap K_\beta|$, then, again by standard arguments, we may assume that condition (A3) holds.

Another case of interest is

(b) *$\widehat{P}$ is a Heegaard surface for $M(\alpha)$ and K_α cannot be isotoped to lie on $\widehat{P}$.*

If $\widehat{P}$ satisfies (a) and $\widehat{Q}$ satisfies (b) (for $M(\beta)$), then Gabai [Ga2] shows that (A3) can be achieved if K_β is put in *thin position* with respect to $\widehat{Q}$. (Gabai explicitly treats the case when $M(\beta) \cong S^3$ and $\widehat{Q} \cong S^2$, but his argument carries over verbatim to any Heegaard surface $\widehat{Q}$.)

Finally, if $\widehat{P}$ and $\widehat{Q}$ both satisfy (b) (for $M(\alpha)$ and $M(\beta)$ respectively), then Rieck has shown [R], using thin position, that again (A3) can be assumed to hold. (The case $M(\alpha) \cong M(\beta) \cong S^3$, $\widehat{P}, \widehat{Q} \cong S^2$ is done in [GL2], and was also proved independently by Gabai (unpublished).)

Given surfaces P, Q in M satisfying (A1), (A2) and (A3), we focus on the arc components of $P \cap Q$, as they lie in P and in Q. We regard these arcs as defining graphs G_P and G_Q in $\widehat{P}$ and $\widehat{Q}$ respectively, in the obvious way. Thus the (fat) vertices of G_P are the disks $\widehat{P} - \text{int } P$, the edges of G_P are the arc components of $P \cap Q$ as they lie in P, and similarly for G_Q.

We now encode some additional structure, as follows. Number the components of ∂P $1, 2, \ldots, p$ in the order in which they appear on ∂M, and similarly number the components of ∂Q $1, 2, \ldots, q$. This gives a corresponding numbering of the vertices of G_P and G_Q. Consider an arc component of $P \cap Q$. Its endpoints are points of $\partial P \cap \partial Q$, the intersections of (say) the ith component of ∂P with the jth component of ∂Q, and the i'th component of ∂P with the j'th component of ∂Q. We then label the endpoints of the corresponding edge of G_P with j and j' at vertices i and i' respectively, and similarly for G_Q. See Figure 1.

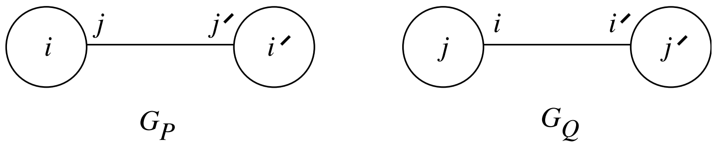


Figure 1

We next give a *sign*, $+$ or $-$, to each vertex of G_P, according to the direction on ∂M of the orientation of the corresponding component of ∂P, induced by some chosen orientation of P. Equivalently, this is the sign of the corresponding intersection point of K_α with $\widehat{P}$ (with respect to some chosen orientations of $M, \widehat{P}$ and K_α). In particular, if $\widehat{P}$ separates $M(\alpha)$ then the signs of the vertices $1, 2, \ldots, p$ of G_P alternate. Similarly, we give a sign to each vertex of G_Q. Note that around each vertex of G_P the edge-endpoint labels occur in order $1, 2, \ldots, q, 1, 2, \ldots, q, \ldots$, repeated $\Delta = \Delta(\alpha, \beta)$ times, the ordering being (say) anticlockwise at a positive vertex and clockwise at a negative vertex. See Figure 2. Similarly, around each vertex of G_Q we see the labels $1, 2, \ldots, p, 1, 2, \ldots, p, \ldots$, repeated Δ times.

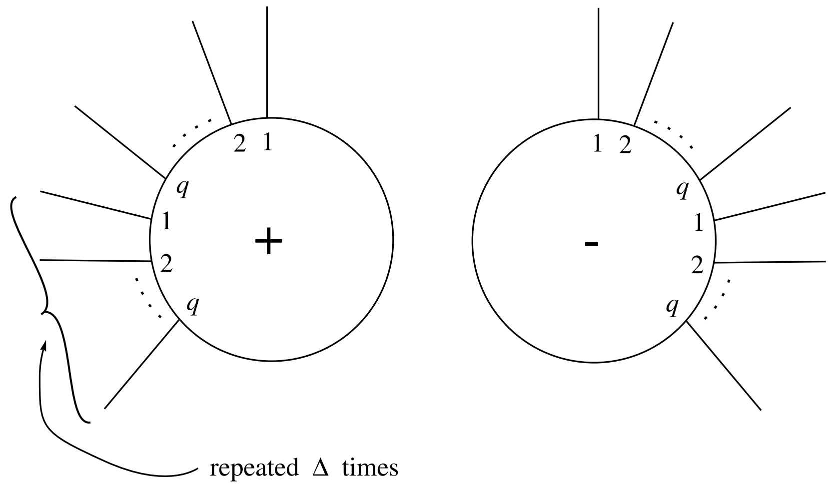


Figure 2

Since M, P and Q are orientable, an arc component of $P \cap Q$ joins points of intersection of ∂P with ∂Q of opposite sign. Hence we have the *parity rule*: if an edge of G_P joins vertices i and i' and the corresponding edge of G_Q joins vertices j and j', then i and i' have the same sign if and only if j and j' have opposite signs. (See Figure 1.)

Condition (A3) becomes, in graph-theoretic language, the condition that G_P and G_Q contain no trivial loops.

We will denote by $\mathbf{q}$ the set of (edge-endpoint) labels $\{1, 2, \ldots, q\}$ of G_P. We also have the associated $\mathbf{q}$-*intervals* $(1, 2), (2, 3), \ldots, (q-1, q), (q, 1)$.

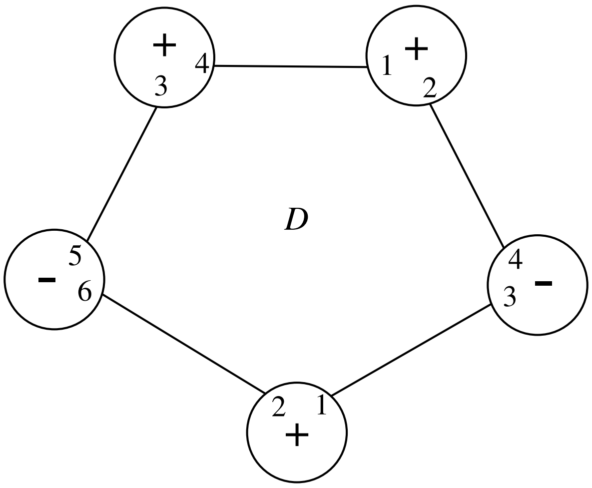


Figure 3

Note that if D is a disk face of G_P, then ∂D consists of an alternating sequence of *edges* and *corners*, where the edges are edges of G_P (i.e., arc components of $P \cap Q$), and the corners are $\mathbf{q}$-intervals $(i, i+1)$ (i.e., arcs in ∂P between consecutively labelled components $i, i+1$ of ∂Q. See Figure 3.

Similar remarks apply to G_Q.

We may describe the philosophy behind our approach as consisting of two parts:

Combinatorial: use the one-one correspondence between the edges of the labeled graphs G_P and G_Q to show that either G_P contains certain configurations or G_Q contains certain (perhaps other) configurations.

Topological: use the existence of certain collections of faces of G_P (resp. G_Q) to get topological information about the pair $(M(\beta), \widehat{Q})$ (resp. $(M(\alpha), \widehat{P})$).

These will be elaborated in Sections 5 and 4 respectively.

Remark. For simplicity we have assumed that ∂M is a torus, but everything carries over without much change if we allow M to have additional boundary components (with α and β being slopes on some fixed torus component of ∂M). In this more general setting, we may also allow $\widehat{P}$ and/or $\widehat{Q}$ to have non-empty boundary.

3. Types

We now introduce a purely algebraic concept, that of a *type*, which plays an important role in the theory under discussion. The reader may choose to skip to Section 4 and return to the present section when necessary.

Let q be a positive integer. A *q-type* is a q-tuple $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q)$, where $\varepsilon_i = \pm$, $1 \le i \le q$.

We say that $a = (a_1, a_2, \dots, a_q) \in \mathbb{Z}^q$ *represents* the q-type $\tau = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q)$ if and only if

(1) for some $\eta = \pm$, sign $a_i = \eta\varepsilon_i$ for each i such that $a_i \neq 0$; and
(2) $\sum\limits_{i=1}^{q} |a_i| \ge 2$.

Condition (2) merely excludes the zero vector, together with vectors all of whose co-ordinates are zero except one, which is ± 1.

Note that a represents τ if and only if it represents $-\tau$. Also, there are 2^{q-1} q-types up to sign, including the *trivial* type, which by definition has $\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_q$.

Example. There are $2^2 = 4$ 3-types (up to sign): $+++$, $-++$, $+-+$, and $++-$. $(1, 0, -2)$ represents $-++$ and $++-$.

A set $A \subset \mathbb{Z}^q$ *represents all q-types* if and only if, for each q-type τ, there exists $a \in A$ such that a represents τ.

Examples.

(1) If $|k| \ge 2$, then $\{(0, \dots, 0, k, 0, \dots, 0)\}$ represents all q-types.
(2) $\{(1, 0, 1), (-1, 1, 0), (1, 1, -1)\}$ represents all 3-types.
(3) If $1 \le r \le q$, then $A \subset \mathbb{Z}^q$ represents all q-types if and only if $\{(a_1, \dots, a_r, 0, a_{r+1}, \dots, a_q) : (a_1, \dots, a_q) \in A\}$ represents all $(q+1)$-types.

For $A \subset \mathbb{Z}^q$, let $c(A)$ be the number of co-ordinates $1 \le i \le q$ such that $a_i \neq 0$ for some $a \in A$. Then we may regard $A \subset \mathbb{Z}^{c(A)} \subset \mathbb{Z}^q$ in the obvious way. By (3) above, A represents all $c(A)$-types if and only if it represents all q-types. Thus we may use the phrase *A represents all types* without ambiguity.

The key property of types is the following "all types implies torsion" theorem due to Parry [P].

Theorem 3.1 (Parry). *If $A \subset \mathbb{Z}^q$ represents all types then there exists $A_0 \subset A$ such that $\mathbb{Z}^q$ modulo the subgroup generated by A_0 has non-trivial torsion.*

We remark that it is not true that A_0 can be chosen to represent all types; equivalently, it is not true that if $A \subset \mathbb{Z}^q$ is a minimal set of representatives of all types then $\mathbb{Z}^q/(A)$ has non-trivial torsion. (Here, *minimal* means that no proper subset of A represents all types.) For example, $A = \{(1, 2, 1), (-1, 1, 1), (1, -1, 2), (1, 1, -1)\}$ is clearly a minimal set of representatives of all types, but $\mathbb{Z}^3/(A) = 0$.

A useful fact, proved in [GL4, Lemma 4.4], is that if $A \subset \mathbb{Z}^q$ represents all types then either A contains a basis for $\mathbb{R}^q$ or there exists $A_1 \subset A$ such that $c(A_1) < q$ and A_1 represents all types. As an immediate consequence we have

Lemma 3.2. *If A is a minimal set of representatives of all types then A contains a basis for $\mathbb{R}^{c(A)}$.*

4. The topology

Consider the solid torus $V_\beta \subset M(\beta)$. The surface $\widehat{Q}$ meets V_β in q meridian disks, cutting V_β into q 3-balls. For each $\mathbf{q}$-interval $\lambda = (i, i+1)$, let H_λ denote the 3-ball (1-handle) consisting of that part of V_β between the meridian disks whose boundaries are components i and $i+1$ of ∂Q.

Let D be a disk face of G_P. Then $D \subset M \subset M(\beta)$, where the edges of ∂D are arcs in Q, and the corners $(i, i+1)$ of ∂D are arcs in ∂P running along the corresponding 1-handle $H_{(i,i+1)}$. See Figure 4, which shows ∂D for the face D illustrated in Figure 3.

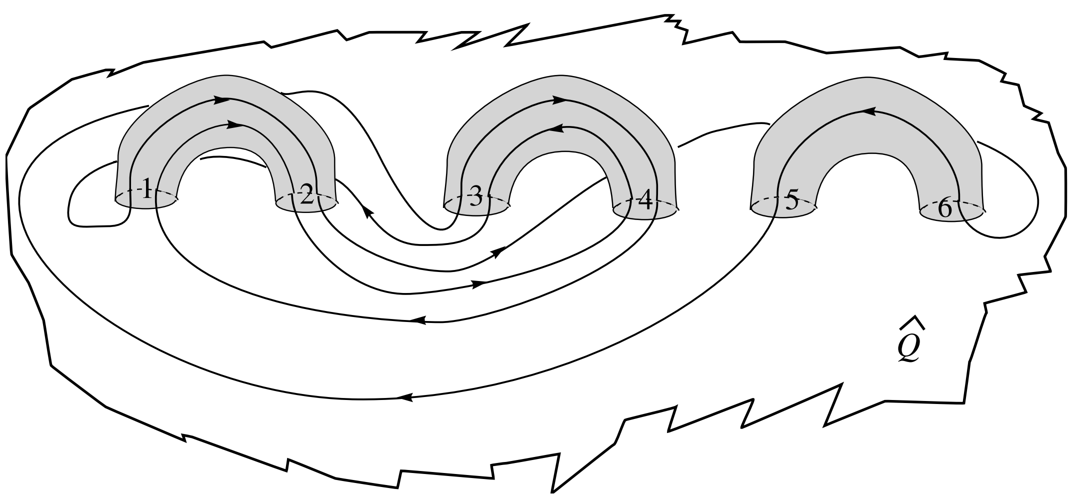


Figure 4

Let $\mathcal{D}$ be a set of disk faces of G_P. Let $c(\mathcal{D})$ be the set of $\mathbf{q}$-intervals $\lambda = (i, i+1)$ that occur as corners of faces in $\mathcal{D}$. Let $e(\mathcal{D})$ be the set of edges of G_Q corresponding to edges belonging to faces in $\mathcal{D}$.

Suppose that the edges $e(\mathcal{D})$ are contained in a disk $E \subset \widehat{Q}$. (Note that this will necessarily be the case if $\widehat{Q}$ is a 2-sphere.) We take E to also contain the fat vertices of G_Q at the endpoints of the edges in $e(\mathcal{D})$.

Recall that if $\widehat{Q}$ is an essential surface, then we may assume that no circle component of $P \cap Q$ bounds a disk in Q. Hence $(\text{int } \mathcal{D}) \cap \widehat{Q} = \emptyset$. If $\widehat{Q}$ is a Heegaard surface, this need not hold. Assume for the moment, however, that $(\text{int } \mathcal{D}) \cap E = \emptyset$; we shall see later (at the end of the present section) that in the cases of interest this can be achieved after some disk exchanges on int $\mathcal{D}$.

Define a 3-manifold $N(E, \mathcal{D}) \subset M(\beta)$ by

$$N(E, \mathcal{D}) = \text{nhd}\left(E \cup \bigcup_{\lambda \in c(\mathcal{D})} H_\lambda \cup \mathcal{D} \right).$$

Note that $N(E, \mathcal{D})$ has a natural handle decomposition, with a single 0-handle (corresponding to E), the 1-handles H_λ, and 2-handles whose cores are the elements of $\mathcal{D}$. Also, if W denotes the handlebody $\text{nhd}(E \cup \bigcup_{\lambda \in c(\mathcal{D})} H_\lambda)$, then we can read off the element of $\pi_1(W)$ (and hence of $H_1(W)$) represented by the boundary of a disk $D \in \mathcal{D}$ by reading the sequence of corners of D (with sign, given by the sign of the corresponding vertex of G_P) as we go around ∂D in some direction.

The idea is to find $\mathcal{D}$ such that $N(E,\mathcal{D})$, or perhaps the surface $\partial N(E,\mathcal{D})$, gives topological information about $M(\beta)$, or the surface $\widehat{Q}$, possibly a contradiction.

The simplest example of this philosophy is when $\mathcal{D}$ consists of a single face of G_P which is a *Scharlemann cycle*. This is a disk face D of G_P such that all the vertices of D have the same sign, and all the corners of D are the same $\mathbf{q}$-interval $\lambda = (i, i+1)$. See Figure 5. Scharlemann cycles first appeared (in a slightly different context) in [S1]. The following theorem, which is contained in [S2, Proof of Proposition 5.6], is immediate; see Figure 6. (Note that vertices i and $i+1$ of G_Q are of opposite sign, by the parity rule.)

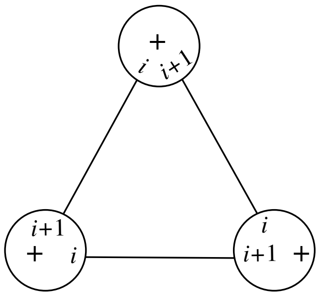


Figure 5

Theorem 4.1. *Let D be a Scharlemann cycle in G_P, such that the edges $e(D)$ lie in a disk $E \subset \widehat{Q}$. Then $N(E,D)$ is a punctured lens space.*

Note that $\pi_1(N(E,D)) \cong \mathbb{Z}_k$, where k is the *length* of D, i.e., the number of edges in ∂D.

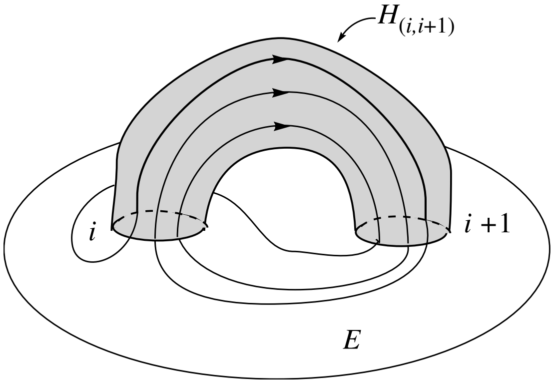


Figure 6

Let us examine the implications for $(M(\beta), \widehat{Q})$ of the existence of a Scharlemann cycle in G_P as in Theorem 4.1.

(Sa) If $\widehat{Q}$ is a Heegaard surface, we conclude tht $M(\beta)$ has a lens space summand.

Now suppose that $\widehat{Q}$ is an essential surface. Note that (after a small isotopy of $N(E, D)$) we may assume that $\partial N(E, D) \cap \widehat{Q} = E$. Let E' be the disk $\overline{\partial N(E, D) - E}$, and let $\widehat{Q}' = (\widehat{Q} - E) \cup E'$. Then $\widehat{Q}' \cong \widehat{Q}$. Also, $|\widehat{Q}' \cap K_\beta| = q - 2$, since the two points of intersection of K_β with $\widehat{Q}$ corresponding to the labels i and $i+1$ that appear in the corners of the Scharlemann cycle D, have been eliminated.

We therefore have the following.

(Sb) If $\widehat{Q}$ is an incompressible surface of positive genus, then $\widehat{Q}'$ is also incompressible. But this is a contradiction, because of the minimality of q.

(Sc) If $\widehat{Q}$ is an essential 2-sphere, then $M(\beta)$ has a lens space summand. In fact, by the minimality of q, $\widehat{Q}$ decomposes $M(\beta)$ as $M' \#$ (lens space).

(Sd) For a variant of (Sc), let us say that a 2-sphere S in a 3-manifold N is $\mathbb{Q}$-*essential* if S does not bound a $\mathbb{Q}$-homology ball in N. (In particular, this is the case if S is non-separating.) It is clear that if $\widehat{Q}$ is $\mathbb{Q}$-essential in $M(\beta)$, then so is $\widehat{Q}'$. Therefore, if we choose $\widehat{Q}$ so that q is minimal over all $\mathbb{Q}$-essential 2-spheres in $M(\beta)$, then the existence of a Scharlemann cycle in G_P is a contradiction.

Although Scharlemann cycles are very useful, their existence is not guaranteed in all situations of interest. A more global concept, which gives similar topological conclusions, is that a set of faces representing all types.

A $\mathbf{q}$-*type* is a q-type $\tau = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_q)$ (see Section 3), where the ith co-ordinate of τ is formally associated with the ith $\mathbf{q}$-interval $(i, i+1)$.

Let D be a disk face of G_P. By taking the algebraic sum of the corners of D we get an element $[D] \in \mathbb{Z}^q$, the free abelian group on the set $\mathbf{q}$-intervals.

We say that D *represents* the $\mathbf{q}$-type τ if and only if

(1) $[D]$ represents τ (in the sense of Section 3); and

(2) (*coherence condition*) no $\mathbf{q}$-interval $(i, i+1)$ appears as a corner of D at two vertices of opposite sign.

As an example, the disk D shown in Figure 3 fails to satisfy the coherence condition.

A set of (disk) faces $\mathcal{D}$ of G_P *represents all types* if and only if for each $\mathbf{q}$-type τ there exists $D \in \mathcal{D}$ such that D represents τ. If such a set of faces exists we say that G_P *represents all types*.

Remark. The role of the coherence condition (2) is twofold. First, it ensures that condition (2) in the algebraic definition of representing a type (see Section 3) is automatically satisfied (since G_P has no trivial loops). Second, it makes it possible to search for representatives of types in a more organized fashion, as we shall see in Section 5 (cf. Proposition 5.2).

If D is a face of G_P such that small neighborhoods in D of the edges $e(D)$ all lie on the same side of $\widehat{Q}$, we say that D *lies locally* on that side of $\widehat{Q}$. To get the topological conclusions that we want from a set of representatives of all types, we first need to observe the following (cf. [GL2, Proof of Proposition 3.2, part (A)]).

Lemma 4.2. *Let $\mathcal{D}$ be a set of faces of G_P representing all types. Then there exists $\mathcal{D}' \subset \mathcal{D}$ such that $\mathcal{D}'$ represents all types and the elements of $\mathcal{D}'$ all lie locally on the same side of $\widehat{Q}$.*

Proof. Label the two local sides of $\widehat{Q}$ in $M(\beta)$, B and W (black and white).

Let $\lambda = (i, i+1)$ be a $\mathbf{q}$-interval, and recall the definition of the 1-handle H_λ at the beginning of this section. Orient (the core of) H_λ from the end corresponding to component i of ∂Q to the end corresponding to component $i+1$ of ∂Q. Then we have the partition

$$\{\mathbf{q}\text{-intervals}\} = \mathcal{B} \amalg \mathcal{W} \amalg \mathcal{BW} \amalg \mathcal{WB}\ ,$$

where

$$\lambda \in \left\{ \begin{matrix} \mathcal{B} \\ \mathcal{W} \\ \mathcal{BW} \\ \mathcal{WB} \end{matrix} \right\} \text{ if } H_\lambda \text{ runs from } \left\{ \begin{matrix} B \text{ to } B \\ W \text{ to } W \\ B \text{ to } W \\ W \text{ to } B \end{matrix} \right\} .$$

Suppose the conclusion of the lemma is false. Then, if $\mathcal{B} \neq \emptyset$, there exists a $\mathcal{B}$-type τ_B such that no member of $\mathcal{D}$ represents τ_B. Similarly, if $\mathcal{W} \neq \emptyset$, there exists a $\mathcal{W}$-type τ_W such that no member of $\mathcal{D}$ represents τ_W. Define a $\mathbf{q}$-type τ by

$$\begin{aligned} &\tau|\mathcal{B} = \tau_B\ , \quad \text{if } \mathcal{B} \neq \emptyset \quad ; \quad \tau|\mathcal{W} = \tau_W\ , \quad \text{if } \mathcal{W} \neq \emptyset\ ; \\ &\tau|\lambda = +\ ,\ \forall\ \lambda \in \mathcal{BW} \quad ; \quad \tau|\lambda = -\ ,\ \forall\ \lambda \in \mathcal{WB}\ . \end{aligned}$$

By hypothesis, there exists $D \in \mathcal{D}$ such that D represents τ. Note that, for some orientation of ∂D, every corner of D in $\mathcal{BW} \amalg \mathcal{WB}$ runs from B to W, and hence D can contain no such corner. Therefore either all corners of D belong to $\mathcal{B}$, or all corners belong to $\mathcal{W}$. It follows that either $\mathcal{B} \neq \emptyset$ and D represents τ_B, or $\mathcal{W} \neq \emptyset$ and D represents τ_W, contradicting the definitions of τ_B and τ_W. $\square$

Theorem 4.3. *Let $\mathcal{D}$ be a set of faces of G_P representing all types, such that the edges $e(\mathcal{D})$ lie in a disk $E \subset \widehat{Q}$. Then there exists $\mathcal{D}' \subset \mathcal{D}$ such that the elements of $\mathcal{D}'$ all lie locally on the same side of $\widehat{Q}$, $H_1(N(E, \mathcal{D}'))$ is finite and non-zero, and $\partial N(E, \mathcal{D}') \cong S^2$.*

Proof. (cf. [GL4, proof of Proposition 4.5]). We may assume that $\mathcal{D}$ is minimal. Hence, by Lemma 4.2, the elements of $\mathcal{D}$ all lie locally on the same side of $\widehat{Q}$. Also, by Lemma 3.2, there exists $\mathcal{D}' \subset \mathcal{D}$ such that $\{[D] : D \in \mathcal{D}'\}$ is a basis for $\mathbb{R}^{c(\mathcal{D})}$. Let $N' = N(E, \mathcal{D}')$. Then it follows that $H_1(N')$ is finite and $\partial N' \cong S^2$.

It remains to show that $H_1(N') \neq 0$.

Let $N = N(E, \mathcal{D}) = \mathrm{nhd}(N' \cup (\mathcal{D} - \mathcal{D}'))$. Since $\partial N' \cong S^2$, attaching to N' the 2-handles whose cores are the elements of $\mathcal{D} - \mathcal{D}'$ is equivalent to removing from N' an equal number of open 3-balls. By Theorem 3.1 there exists $\mathcal{D}_0 \subset \mathcal{D}$ such that $H_1(N_0)$ has non-trivial torsion, where $N_0 = N(E, \mathcal{D}_0)$. Note that $N_0 \subset N \subset N'$.

Suppose $H_1(N') = 0$. Then N' is a homology 3-ball, and the Mayer-Vietoris exact sequence gives $H_1(\partial N_0) \cong H_1(N_0) \oplus H_1(N' - N_0)$, contradicting the fact that $H_1(N_0)$ has non-trivial torsion. Hence $H_1(N') \neq 0$, as desired. □

Remark. Since $H_1(N) \cong H_1(N')$, the above proof shows that if $\mathcal{D}$ is a minimal set of representatives of all types then $H_1(N(E, \mathcal{D}))$ is finite and non-zero. As was noted in Section 3, the analogous statement is not true in the purely algebraic setting.

It follows from Theorem 4.3 that if G_P represents all types, and the edges of the relevant faces are contained in a disk in $\widehat{Q}$, then we get conclusions almost identical to (Sa), (Sb), (Sc) and (Sd) above for Scharlemann cycles.

- (Ta) If $\widehat{Q}$ is a Heegaard surface, then $M(\beta)$ has a summand $\hat{N}'$ with $H_1(\hat{N}')$ finite and non-zero.
- (Tb) If $\widehat{Q}$ is an incompressible surface of positive genus, this is a contradiction. (We get a new surface $\widehat{Q}'$, obtained from $\widehat{Q}$ by a disk exchange, with $|\widehat{Q}' \cap K_\beta| = q - |c(\mathcal{D})|$.)
- (Tc) If $\widehat{Q}$ is an essential 2-sphere, then $\widehat{Q}$ decomposes $M(\beta)$ as $M' \# \widehat{N}'$, where $H_1(\widehat{N}')$ is finite and non-zero.
- (Td) If $\widehat{Q}$ is a $\mathbb{Q}$-essential 2-sphere, this is a contradiction.

Finally, we return to the assumption, when $\widehat{Q}$, is a Heegaard surface, that $(\text{int}\ \mathcal{D}) \cap E = \emptyset$.

Let $\mathcal{D}$ be a collection of faces of G_P representing all types (this includes the case where $\mathcal{D}$ consists of a single Scharlemann cycle). We may assume that $(\text{int}\ \mathcal{D}) \cap E$ consists of a finite number of disjoint circles and arcs, properly embedded in E. Let γ be such a circle or arc, and let E_1 and E_2 be the components of $E - \gamma$. There is a corresponding partition

$$c(\mathcal{D}) = \mathcal{C}_1 \amalg \mathcal{C}_2 \amalg \mathcal{C}_{12}\ ,$$

where $\lambda \in \mathcal{C}_i$ if the 1-handle H_λ has both ends in E_i, $i = 1, 2$, and $\lambda \in \mathcal{C}_{12}$ if H_λ has one end in E_1 and one in E_2.

Then, exactly as in the proof of Lemma 4.2 (the $c(\mathcal{D})$-type τ here will be defined by orienting all $\lambda \in \mathcal{C}_{12}$ so as to run from (say) E_1 to E_2), there exists $\mathcal{D}' \subset \mathcal{D}$ such that $\mathcal{D}'$ represents all types and $c(\mathcal{D}') \subset \mathcal{C}_1$ (say). Since we may assume that $\mathcal{D}$ is a minimal set of representatives of all types, we conclude that the vertices of G_Q corresponding to the labels that appear in the corners $c(\mathcal{D})$ all lie in E_1.

If γ is an arc, or if γ is a circle and $\bar{E}_1$ is a disk, we replace E by $\bar{E}_1$ (moved slightly so as to eliminate the intersection γ with int $\mathcal{D}$). If γ is a circle and $\bar{E}_2$ is a disk, we use a standard innermost circle argument to cut-and-paste $\mathcal{D}$, using subdisks of $\bar{E}_2$, so as to reduce $(\text{int}\ \mathcal{D}) \cap E$. Thus we eventually get a collection of disks $\widetilde{\mathcal{D}}$, whose boundaries are the same as those in $\mathcal{D}$, such that $(\text{int}\ \widetilde{\mathcal{D}}) \cap E = \emptyset$.

Remark. There are situations when it is useful to consider submanifolds of $M(\beta)$ of the form $\text{nhd}(E \cup \bigcup H_\lambda \cup \mathcal{D})$ where E is a subsurface of $\widehat{Q}$ other than a disk. For example, the case where $\mathcal{D}$ consists of a Scharlemann cycle and E is an annulus arises in [GL3, Section 3].

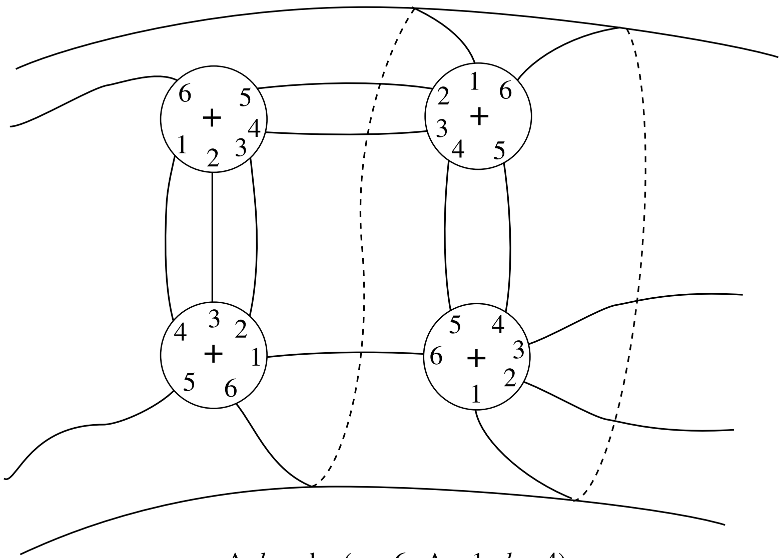
5. The combinatorics

To state the main result of this section, Theorem 5.1 below, we need the following definition.

A subgraph Λ of G_Q is a *k-web* (k a non-negative integer) if and only if

(1) all the vertices of Λ have the same sign, and
(2) at most k edge-endpoints of G_Q at vertices of Λ are not endpoints of edges of Λ.

See Figure 7 for an example.



A k-web ($p=6,\ \Delta=1,\ k=4$)

Figure 7

In the following theorem, G_P and G_Q are graphs as in Section 2.

Theorem 5.1. *Assume that $\Delta > 1 - \chi(\widehat{P})/p$. Then either G_P represents all types or G_Q contains a $(p - \chi(\widehat{P}))$-web.*

Most of the remainder of this section will be devoted to a sketch of the proof of Theorem 5.1.

First we deal with the case of the trivial type, which is different from the general case.

Define Λ to be the subgraph of G_Q consisting of all edges of G_Q joining vertices of the same sign. If some component of Λ is not a $(p-\chi(\widehat{P}))$-web, then G_Q contains more than $(p-\chi(\widehat{P}))$ edges joining vertices of opposite sign. By the parity rule, the corresponding edges of G_P join vertices of the same sign. Let Σ be the subgraph of G_P consisting of all vertices of G_P, and all edges of G_P joining vertices of the same sign. Then

$$\begin{aligned}\chi(\widehat{P}) &= V - E + \textstyle\sum \chi(f) \quad \text{(summed over all faces } f \text{ of } \Sigma\text{)}\\ &< p - (p - \chi(\widehat{P})) + \textstyle\sum \chi(f)\ .\end{aligned}$$

Hence $\sum \chi(f) > 0$, implying that Σ has a disk face. This face is then a face of G_P whose vertices all have the same sign, and which therefore represents the trivial $\mathbf{q}$-type.

For non-trivial $\mathbf{q}$-types we proceed as follows.

Define a graph $\Gamma \subset \widehat{P}$ by:

• the *vertices* of Γ consist of the *fat* vertices (i.e., the vertices of G_P), together with *dual* vertices $v(D)$, one in the interior of each face D of G_P;

• the *edges* of Γ join each dual vertex $v(D)$ to the fat vertices in the boundary of the corresponding face D.

See Figure 8.

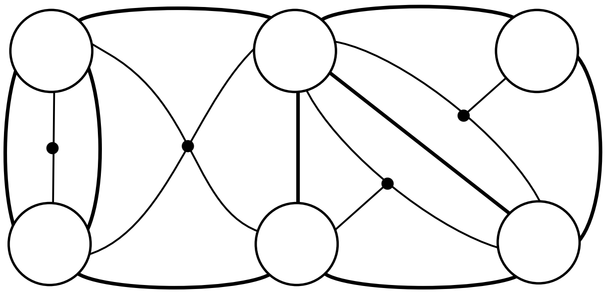


Figure 8

Now let τ be a $\mathbf{q}$-type. We define the directed graph $\Gamma(\tau) \subset \widehat{P}$ to be the graph Γ, with edges oriented according to the following rule. Let e be an edge of Γ, with one endpoint at the fat vertex v, lying in a $\mathbf{q}$-interval $(i, i+1)$ at that vertex. Then we orient e

inwards at v if $(\tau|(i, i+1)) \cdot \operatorname{sign} v = +$, and
outwards at v if $(\tau|(i, i+1)) \cdot \operatorname{sign} v = -$.

As an example, if $q = 6$ and $\tau = + - + - - -$, then around the vertices of G_P the edges of $\Gamma(\tau)$ are oriented as shown in Figure 9.

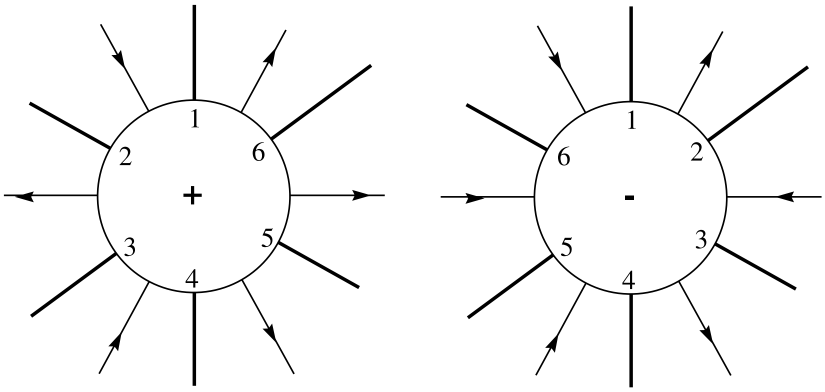


Figure 9

The reason for defining the orientation on the edges of $\Gamma(\tau)$ in this way is the following, which is immediate from the definition.

Proposition 5.2. *A disk face D of G_P represents τ if and only if the corresponding dual vertex $v(D)$ is a sink or source of $\Gamma(\tau)$.*

Our search for representatives of types is thus translated into the study of the directed graphs $\Gamma(\tau) \subset \widehat{P}$. Here we will use the following very elegant combinatorial formulation of the Poincaré-Hopf Index Theorem, due to Glass [Gl].

Consider any directed graph Ω in a closed surface S. For each vertex v of Ω, let $s(v)$ be the number of *switches* (i.e., changes in orientation of successive edges) around v, and for each face f of Ω, let $s(f)$ be the number of switches around ∂f. See Figure 10.

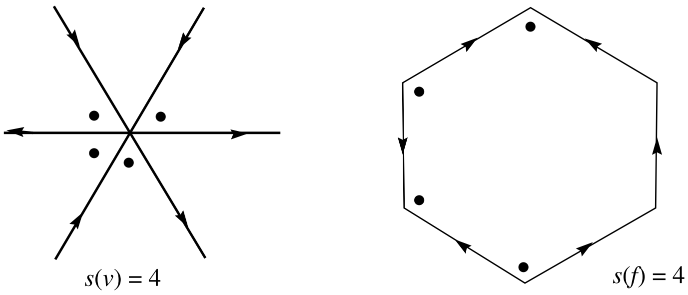


Figure 10

Define the *index* of a vertex or face by

$$I(v) = 1 - \frac{s(v)}{2} \quad , \quad I(f) = \chi(f) - \frac{s(f)}{2} \ .$$

In particular, a vertex of positive index is a sink or source ($I(v) = 1$), and a face of positive index is a disk whose boundary is a cycle ($I(f) = 1$).

Lemma 5.3. (Glass). *Let Ω be a directed graph in the closed surface S. Then $\sum I(v) + \sum I(f) = \chi(S)$.*

Proof. (See [Gl].) Each corner between adjacent edges at a vertex contributes exactly 1 to $\sum s(v) + \sum s(f)$. Hence

$$\begin{aligned}
\chi(S) &= V - E + \textstyle\sum \chi(f) \\
&= V - \frac{\#\text{ corners}}{2} + \textstyle\sum \chi(f) \\
&= V - \frac{(\sum s(v) + \sum s(f))}{2} + \textstyle\sum \chi(f) \\
&= \textstyle\sum I(v) + \sum I(f) \ .
\end{aligned}$$

□

A label i which at a fat vertex lies immediately between oppositely oriented edges of $\Gamma(\tau)$ is a *switch label* of τ. (Equivalently, i is a switch label if and only if $\tau|(i-1, i)$ and $\tau|(i, i+1)$ have opposite signs.) If the edges of $\Gamma(\tau)$ adjacent to i are oriented in a clockwise (resp. anticlockwise) direction around i, then i is a *clockwise* (resp. *anticlockwise*) switch label. Note that this is well-defined, i.e., independent of the sign of the fat vertex under consideration; see Figure 9.

Let $C(\tau), A(\tau) \subset \mathbf{q}$ be the set of clockwise and anticlockwise switch labels of τ, respectively.

As an example, for the type τ illustrated in Figure 9, $C(\tau) = \{2,4\}$, $A(\tau) = \{1,3\}$.

An edge e of G_P is a *clockwise* (resp. *anticlockwise*) *switch edge* if the labels at the endpoints of e both belong to $C(\tau)$ (resp. $A(\tau)$). A *switch edge* is an edge of G_P that is either a clockwise or an anticlockwise switch edge.

Note that the faces of $\Gamma(\tau)$ are in one-one correspondence with the edges of G_P, and that under this correspondence, faces of $\Gamma(\tau)$ of index 1 correspond to switch edges of G_P. See Figure 11.

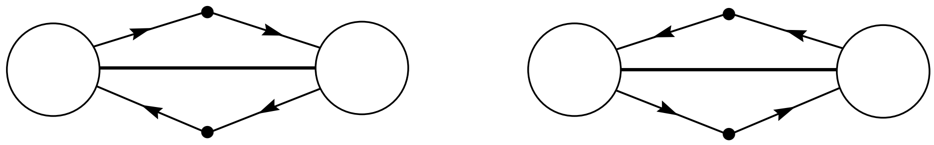


Figure 11

Now let τ be a non-trival $\mathbf{q}$-type, and write $c(\tau) = |C(\tau)|$, $a(\tau) = |A(\tau)|$. Thus $c(\tau) = a(\tau) > 0$.

If v is a fat vertex of $\Gamma(\tau)$, then

$$s(v) = \Delta(c(\tau) + a(\tau)) = 2\Delta c(\tau) \ .$$

Therefore

$$I(v) = 1 - \frac{s(v)}{2} = 1 - \Delta c(\tau) \ .$$

Assume that G_P does not represent τ. Then by Proposition 5.2, $\sum_{v \text{ dual}} I(v) \leq 0$. Therefore, by Lemma 5.3,

$$\textstyle\sum I(f) \geq \chi(\widehat{P}) - \sum_{v \text{ fat}} I(v) = \chi(\widehat{P}) - p(1 - \Delta c(\tau)) \ .$$

Note that since $c(\tau) \geq 1$ and $\Delta > 1 - \chi(\widehat{P})/p$ by hypothesis, the last quantity above is > 0.

By the remark above following the definition of a switch edge, the number of switch edges of $G_P \geq \sum I(f)$. Hence we may assume without loss of generality that there are at least $\frac{\sum I(f)}{2}$ clockwise switch edges, and so the number of endpoints of clockwise switch edges $\geq \sum I(f) \geq p\Delta c(\tau) - (p - \chi(\widehat{P}))$.

Now there are $\Delta c(\tau)$ clockwise switch labels at each vertex of G_P, giving a total of $p\Delta c(\tau)$ such labels in G_P, and so there are at most $(p - \chi(\widehat{P}))$ of these labels that are not endpoints of clockwise switch edges of G_P. Let Λ be the subgraph of G_Q consisting of those edges that correspond to clockwise switch edges of G_P. (In particular, the vertices of Λ correspond to a subset of $C(\tau)$.) As noted above, Λ is non-empty. Also, there are at most $(p - \chi(\widehat{P}))$ occurrences of labels at vertices of Λ that are not endpoints of edges of Λ. Hence if

$$\text{\textit{all elements of} } C(\tau) \text{ \textit{have the same sign,}} \qquad (*c)$$
$$\text{and \textit{all elements of} } A(\tau) \text{ \textit{have the same sign,}} \qquad (*a)$$

then all the vertices of Λ have the same sign, and so (a connected component of) Λ is a $(p-\chi(\widehat{P}))$-web, as desired. (Note that we need to make both assumptions $(*c)$ and $(*a)$, as we have no control over whether there are at least $\frac{\sum I(f)}{2}$ clockwise or anticlockwise switch edges.)

If at least one of $(*c)$ and $(*a)$ does not hold, then we take *derived types* until both do hold, and then work backwards. We now briefly discuss this inductive procedure.

For this, we have to work in the following more general setting. Consider a non-empty set of labels $L \subset \mathbf{q}$. Associated with L is the set of L-*intervals*, which are the intervals (ℓ_1, ℓ_2) between successive elements of L. An L-*type* is defined in the obvious way, to be an $|L|$-type where the ith co-ordinate is formally associated with the ith L-interval.

Define $G(L)$ to be the subgraph of G_P consisting of all edges of G_P with at least one endpoint label belonging to L. Thus $G(L)$ and G_P have the same vertices, and $G(\mathbf{q}) = G_P$.

Note that although the corners of the faces of $G(L)$ are not necessarily L-intervals, each is contained in a unique L-interval. Hence, for each disk face D of $G(L)$, by taking the algebraic sum of the L-intervals containing the corners of D we get an element $[D] \in \mathbb{Z}^{|L|}$. Then, very much as before, we say that D *represents* the L-type τ if and only if

(1) $[D]$ represents τ (in the sense of Section 3); and
(2) (*coherence condition*) no L-interval occurs as the L-interval containing a corner of D at two vertices of D of opposite sign.

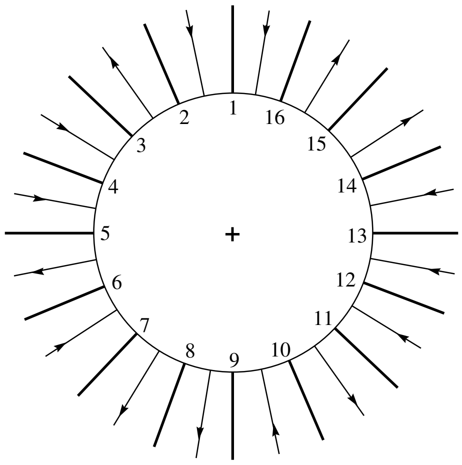
We define the directed graph $\Gamma(\tau)$ as before (starting with $G(L)$), orienting each edge according to the restriction of τ to the L-interval containing the corresponding corner (and the sign of the fat vertex in question). Proposition 5.2 continues to hold with G_P replaced by $G(L)$.

Let τ be a non-trivial L-type. We have $C(\tau), A(\tau) \subset L$, the set of clockwise (resp. anticlockwise) switches of τ, as before. The argument we have just given above carries over verbatim to our present more general setting, so we have the following lemma.

Lemma 5.4. *Let τ be a non-trivial L-type, satisfying $(*c)$ and $(*a)$, such that $G(L)$ does not represent τ. Then G_Q contains a $(p - \chi(\widehat{P}))$-web.*

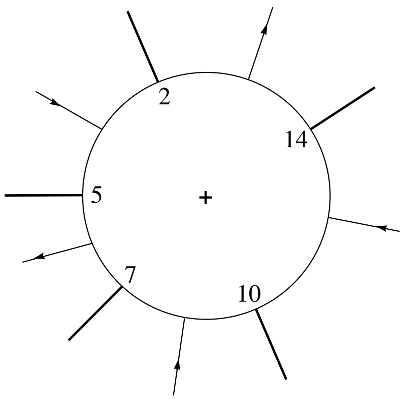
The *derivative* of a non-trivial L-type τ is the $C(\tau)$-type $d\tau$ defined as follows. Let (c, c') be a $C(\tau)$-interval. Then (c, c') contains a unique element $a \in A(\tau)$. We define $d\tau$ by

$$d\tau|(c, c') = \text{sign } a \ .$$

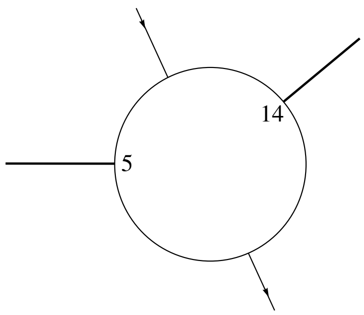


$C(\tau) = \{2,5,7,10,14\}$

$A(\tau) = \{3,6,9,11,16\}$



$C(d\tau) = \{5,14\}$, $A(d\tau) = \{2,7\}$



$C(d^2\tau) = \{5\}$, $A(d^2\tau) = \{14\}$

Figure 12

Example. Taking $q = 16$ and starting with the **q**-type $\tau = + - + + - + - - + - + + + - - +$, Figure 12 below illustrates how one obtains $d\tau$ and $d^2\tau$.

The key property of the definition of the derivative of a type is the following proposition, which says that to find a representative of τ it is enough to find a representative of $d\tau$.

Proposition 5.5. *Let τ be a non-trivial L-type. Then any face of $G(C(\tau))$ representing $d\tau$ contains a face of $G(L)$ representing τ.*

Sketch Proof. We shall give enough of the proof to at least explain the reason for the definition of $d\tau$.

Let D be a face of $G(C(\tau))$ representing $d\tau$. Let $\Gamma_D(\tau) = \Gamma(\tau) \cap D$.

Claim. *$\Gamma_D(\tau)$ has either a face of index 1 or a dual vertex of index 1.*

Proof of Claim. This is proved by doubling $(D, \Gamma_D(\tau))$ along the boundary, and applying Lemma 5.3 to the double $2\Gamma_D(\tau) \subset 2D \cong S^2$. Note that the faces of $2\Gamma_D(\tau)$ are the faces of $\Gamma_D(\tau)$ (each appearing twice), together with a face f_e for each edge $e \subset \partial D$ (with $e \subset f_e$). Hence the conclusion will follow if we show that

$$|\{\text{fat vertices of } 2\Gamma_D(\tau) \text{ of index } 1\}| \leq |\{\text{faces } f_e \text{ of index } -1\}| \ .$$

But this in turn follows from the fact that, by definition of $G(C(\tau))$, each $e \subset \partial D$ has at least one endpoint label $\in C(\tau)$ (see [GL2, p.398] for details).

A dual vertex of $\Gamma_D(\tau)$ of index 1 will correspond to a face of $G(L)$, contained in D, representing τ. So, by the claim above, we are done once we show that no face of $\Gamma_D(\tau)$ has index 1. Such a face corresponds to a switch edge e of $G(L)$. Now e cannot be a clockwise switch edge, for such edges are edges of $G(C(\tau))$, and hence cannot lie in the interior of the face D of $G(C(\tau))$. So assume that e is an anticlockwise switch edge, with endpoint labels $a_1, a_2 \in A(\tau)$ at vertices v_1, v_2 respectively; see Figure 13. Then, by the parity rule, (sign v_1)(sign a_1) $\neq$ (sign v_2)(sign a_2). But, recalling the definition of $d\tau$, this contradicts the fact that D represents $d\tau$. $\square$

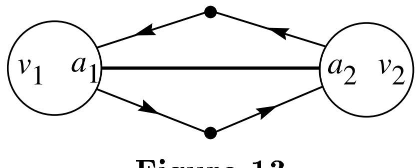


Figure 13

Returning to the proof of Theorem 5.1, suppose that one of the conditions $(*c), (*a)$ fails for the $\mathbf{q}$-type τ. Note that $C(\tau) = A(-\tau)$. Hence, if we define

$$\tau_0 = \begin{cases} \tau \ , & \text{if } \tau \text{ does not satisfy } (*a) \\ -\tau \ , & \text{if } \tau \text{ satisfies } (*a) \text{ but not } (*c), \end{cases}$$

then $(*a)$ fails for τ_0. Hence $d\tau_0$ is a non-trivial type. If $d\tau_0$ satisfies $(*c)$ and $(*a)$, stop. If not, define

$$\tau_1 = \begin{cases} d\tau_0 \ , & \text{if } d\tau_0 \text{ does not satisfy } (*a) \\ -d\tau_0 \ , & \text{if } d\tau_0 \text{ satisfies } (*a) \text{ but not } (*c) \end{cases}$$

Continuing in this way, we obtain a sequence $\tau_0, \tau_1, \dots, \tau_n$, $n \geq 1$, where $\tau_0 = \pm\tau$, $\tau_i = \pm d\tau_{i-1}$ is a non-trivial $C(\tau_{i-1})$-type, $1 \leq i \leq n$, and τ_n satisfies $(*c)$ and $(*a)$.

If $G(C(\tau_{n-1}))$ does not represent τ_n, then, by Lemma 5.4, G_Q contains a $(p - \chi(\widehat{P}))$-web.

If $G(C(\tau_{n-1}))$ does represent τ_n, then successive applications of Proposition 5.5 show that G_P represents τ.

This completes our sketch of the proof of Theorem 5.1.

We have seen, in Section 4, that if G_P represents all types then we get useful topological information about $(M(\beta), \widehat{Q})$. What if G_Q contains a $(p - \chi(\widehat{P}))$-web? One idea is that a web might give rise to Scharlemann cycles in G_Q (which then in turn give useful topological information about $(M(\alpha), \widehat{P})$). In order to make this work, however, we must impose a further condition. Namely, we define a *great* k-web in G_Q to be a k-web Λ which is contained in the interior of a disk $D_\Lambda \subset \widehat{Q}$ with the property that any vertex of G_Q lying in D_Λ is a vertex of Λ. See Figure 14.

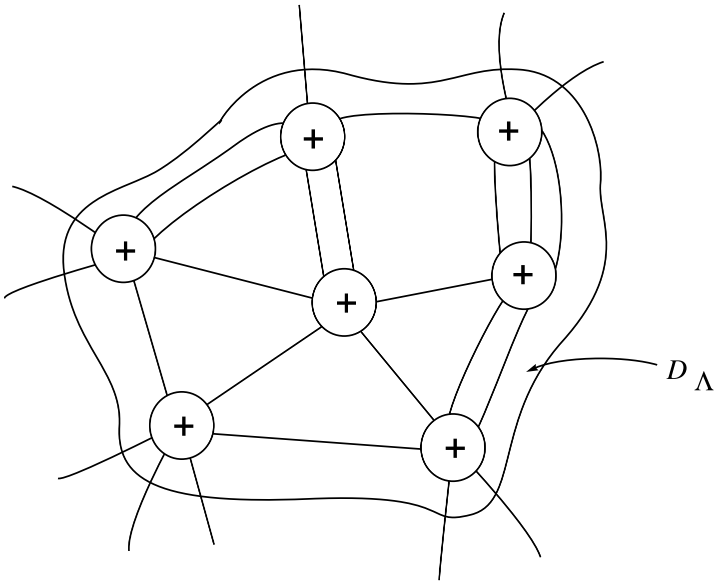


Figure 14

To guarantee that such objects exist, we need to specialize to the case where $\widehat{Q}$ is a 2-sphere. It turns out that one can then prove the following theorem.

Theorem 5.6. *Assume that $\Delta > 1 - \chi(\widehat{P})/p$, and that $\widehat{Q} \cong S^2$. Then either G_P represents all types or G_Q contains a great $(p - \chi(\widehat{P}))$-web.*

The proof proceeds as follows. By Theorem 5.1, if G_P does not represent all types then G_Q contains a $(p-\chi(\widehat{P}))$-web Λ. Let U be a component of $\widehat{Q} - \mathrm{nhd}(\Lambda)$, and let D be the disk $\widehat{Q} - U$, containing Λ. If Λ is not a great web, then there are vertices of G_Q in D that are not vertices of Λ. These vertices correspond to a non-empty subset $L \subset \mathbf{q}$. One then proves by induction on $|L|$ that either G_Q contains a great $(p - \chi(\widehat{P}))$-web or $G(L)$ represents all L-types. (The last step of the induction, which is the desired statement, is the degenerate case where $\Lambda = \emptyset$, $L = \mathbf{q}$, and D is a disk in $\widehat{Q}$ containing G_Q.) This inductive argument is quite subtle, and in particular involves relativizing the notion of the derivative of a type. We refer to [GL3, proof of Theorem 2.5] (which does the case where $\widehat{P}$ is a torus), and thence to [GL2, Section 2], for more details.

6. Applications

In this section we sketch the proofs of some results which use the methods we have discussed. The first three are about Dehn surgery on knots in S^3, and the proofs of these will all be based on the following statement, which is an immediate consequence of Theorem 5.6 and (Ta) of Section 4.

Theorem 6.1. *Suppose that $M(\beta) \cong S^3$ and that $\widehat{Q}$ is a Heegaard 2-sphere. If $\Delta > 1 - \chi(\widehat{P})/p$ then G_Q contains a great $(p - \chi(\widehat{P}))$-web.*

When $M(\beta) \cong S^3$, we shall make a slight change of notation, writing K for K_β, a non-trivial knot in S^3, M_K for M, the exterior of K, and μ_K for β, the meridian of K.

The following theorem is proved in [GL2]. It implies that knots in S^3 are determined by their complements.

Theorem 6.2. *Let K be a non-trivial knot in S^3. If $M_K(\alpha) \cong S^3$ then $\alpha = \mu_K$.*

Proof. Suppose that $M_K(\alpha) \cong S^3$, with $\Delta(\alpha, \mu_K) \geq 1$. Let $\widehat{P}$ be a Heegaard 2-sphere in $M_K(\alpha)$.

By Theorem 6.1, G_Q contains a great $(p-2)$-web, Λ. Hence there exists a label (in fact at least two labels) $i \in \mathbf{p}$ such that every occurrence of i at a vertex of Λ is the endpoint of an edge of Λ. Moreover, since all the vertices of Λ have the same sign, no edge of Λ can have the same label at both endpoints, by the parity rule. Hence, by always leaving a vertex of Λ along an edge of Λ with label i at that vertex, we see that Λ contains an *i-cycle*, that is, a cycle of edges which can be consistently oriented so that the tail of each edge has label i (and all the vertices in the cycle have the same sign). See Figure 15.

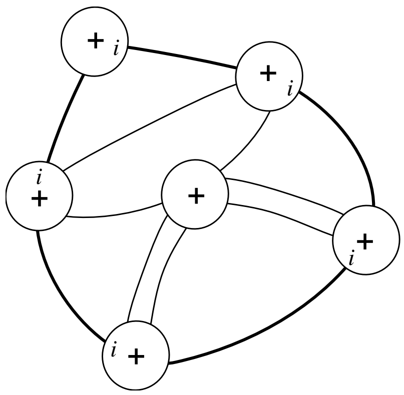


Figure 15

Now it is easy to see that a cycle of this kind which is innermost in the disk D_Λ (allowing the label i to be an arbitrary element of $\mathbf{p}$) is a Scharlemann cycle in G_Q. Since $\widehat{P}$ is a 2-sphere the edges of this Scharlemann cycle necessarily lie in a disk in $\widehat{P}$. But since $\widehat{P}$ is a Heegaard 2-sphere, this is a contradiction, by (Sa) in Section 4 (here, the roles of P and Q are interchanged). □

Remark. We observed in the above proof that a great $(p-2)$-web in G_Q always contains a Scharlemann cycle. So we may state

Theorem 6.3. *If $\widehat{P}$ and $\widehat{Q}$ are 2-spheres, and $\Delta \geq 1$, then either G_P represents all types or G_Q contains a Scharlemann cycle.*

If $\Delta \geq 2$ then the following stronger conclusion holds.

Theorem 6.4. *If $\widehat{P}$ and $\widehat{Q}$ are 2-spheres, and $\Delta \geq 2$, then either G_P or G_Q contains a Scharlemann cycle.*

This is an immediate consequence of [CGLS, Proposition 2.5.6] (which deals with the case where $\widehat{P}$ and $\widehat{Q}$ are disks).

Theorem 6.4 already shows that if $M(\alpha) \cong M(\beta) \cong S^3$ then $\Delta(\alpha,\beta) \leq 1$, implying that there are at most two inequivalent knots with homeomorphic complements (a fact first proved, by a different argument, in [CGLS]). However, there are examples of graphs G_P, G_Q with $\Delta = 1$ for which the conclusion of Theorem 6.4 fails, illustrating the necessity of introducing the more general concept of representing all types.

The next application addresses the question of when a reducible manifold can be obtained by Dehn surgery on a knot in S^3.

Theorem 6.5. *Let K be a non-trivial knot in S^3. If $M_K(\alpha)$ is reducible then*

(1) *$M_K(\alpha)$ has a lens space summand; and*
(2) *$\Delta(\alpha, \mu_K) = 1$.*

Part (1) was proved in [GL2]; part (2) was proved in [GL1] (by an argument different from that outlined below).

Part (1) implies the Property R Conjecture, proved by Gabai [Ga2].

Corollary 6.6 (Gabai)**.** *If K is a non-trivial knot in S^3 then $M_K(0)$ is irreducible.*

Part (1) also implies

Corollary 6.7. *Any homology 3-sphere obtained by Dehn surgery on a knot in S^3 is prime.*

We remark that Auckly, using results from gauge theory, has shown that there are prime homology 3-spheres which cannot be obtained by Dehn surgery on any knot in S^3 [A].

The situation described in Theorem 6.5 can occur: if K is a torus knot or cable knot then there is a slope α such that $M_K(\alpha)$ is reducible. However, the Cabling

Conjecture [GS] asserts that these are the only knots with this property. This is still open.

Sketch Proof of Theorem 6.5. Here we take $\widehat{P}$ to be an essential 2-sphere in $M_K(\alpha)$. (Note that M_K is irreducible.)

By Theorem 6.1, G_Q contains a great $(p-2)$-web Λ. As in the proof of Theorem 6.2 above, this implies that G_Q contains a Scharlemann cycle, and hence that $M_K(\alpha)$ has a lens space summand (see (Sc) in Section 4). This proves (1).

To prove (2), assume that $\Delta(\alpha, \mu_K) > 1$. Then, for homological reasons, $\widehat{P}$ separates $M_K(\alpha)$, and hence p is even.

If $p = 2$ then P is an annulus, and hence K is a cable of a knot K' (which may be trivial). Now it is not hard to show (see the first four lines of the proof of Theorem 3 in [GL1]) that if $M_K(\alpha)$ is reducible for some α with $\Delta(\alpha, \mu_K) > 1$ then $M_{K'}(\alpha')$ is reducible for some α' with $\Delta(\alpha', \mu_{K'}) > 1$. Hence, by induction it suffices to prove the result when K is not cabled.

So assume $p \geq 4$. As in the proof of Theorem 6.2 above, there is a label in $\mathbf{p}$ such that each of the Δ occurrences of that label at each vertex of Λ is the endpoint of an edge of Λ. One can then show, using the fact that $\Delta \geq 2$, that Λ contains Scharlemann cycles of G_Q on distinct $\mathbf{p}$-intervals $(i, i+1)$, $(j, j+1)$ (see [GL4, Proof of Theorem 2.3]). This allows one to construct a new essential 2-sphere $\widehat{P}' \subset M_K(\alpha)$ such that $|\widehat{P}' \cap K_\alpha| < p$ (see [GL4, Theorem 2.4]) contradicting the minimality of p. $\square$

For our last application to Dehn surgery on knots in S^3 we consider the situation where $M_K(\alpha)$ contains an incompressible torus. Here, we must assume that M_K contains no essential torus, or equivalently, that K is not a satellite knot.

Theorem 6.8 [GL3]**.** *Let K be a knot in S^3 that is not a satellite knot. If $M_K(\alpha)$ contains an incompressible torus then $\Delta(\alpha, \mu_K) \leq 2$.*

Examples with $\Delta(\alpha, \mu_K) = 2$ have been constructed by Eudave-Muñoz [E2]. There are many examples with $\Delta(\alpha, \mu_K) = 1$.

Sketch Proof of Theorem 6.8. Let $\widehat{P}$ be an incompressible torus in $M_K(\alpha)$, and assume (for a contradiction) that $\Delta(\alpha, \mu_K) \geq 3$. In particular, this implies that $\widehat{P}$ separates $M_K(\alpha)$, so that p is even.

If $p = 2$ then one can show that the knot K is strongly invertible (see [GL3, Section 8]) in which case the theorem is proved by Eudave-Muñoz [E1].

So assume that $p \geq 4$. By Theorem 6.1, G_Q contains a great p-web Λ. An easy euler characteristic argument (see [GL3, Section 4]) shows that there are at least four labels $i \in \mathbf{p}$ such that Λ contains an *i-bigon*; see Figure 16. Now such an i-bigon is either a Scharlemann cycle, or contains within it an *extended Scharlemann cycle*, that is, a Scharlemann cycle flanked by bigons; see Figure 17. However, a topological argument shows that an extended Scharlemann cycle in G_Q would give rise to a new incompressible torus $\widehat{P}'$ in $M_K(\alpha)$ with $|\widehat{P}' \cap K_\alpha| < p$, contradicting the minimality of p (see [GL3, Theorem 3.2]). We conclude that Λ contains two

Scharlemann cycles D, D' of length 2 on disjoint $\mathbf{p}$-intervals $(i, i+1)$ and $(i', i'+1)$. By (Sb) in Section 4, the edges of neither D nor D' lie in a disk in $\widehat{P}$, hence each pair of edges forms an essential loop on $\widehat{P}$; see Figure 18. Shrinking the 1-handle $H_{(i,i+1)}$ to its core has the effect of gluing together the two $(i, i+1)$ corners of D (see Figure 19), giving a Möbius band B. Similarly, D' gives rise to a Möbius band B'. Since ∂B and $\partial B'$ are parallel on $\widehat{P}$ we may join B and B' by an annulus in $\widehat{P}$ to get a Klein bottle F in $M_K(\alpha)$. We now replace the torus $\widehat{P}$ by the boundary $\widehat{P}_0$ of a regular neighborhood of F. (Note that $\widehat{P}_0$ is incompressible, since $M_K(\alpha)$ contains an incompressible torus by hypothesis, and is irreducible by Theorem 6.5(2).) But similar combinatorial arguments, applied to $P_0 \cap Q$, in this case lead to a contradiction (see [GL3, Section 6]). $\square$

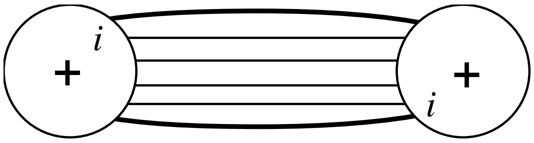


Figure 16

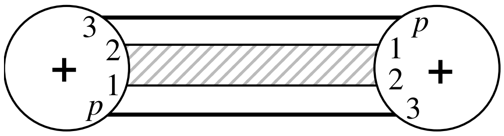


Figure 17

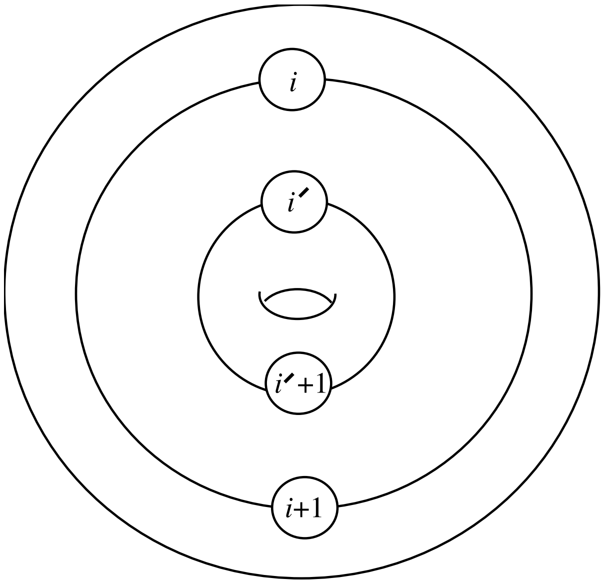


Figure 18

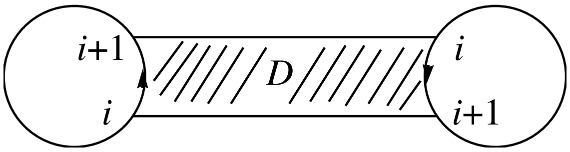


Figure 19

Finally, we show how the methods we have discussed give some information about the question of when $M(\alpha)$ and $M(\beta)$ are reducible. Let us say that a 3-manifold is $\mathbb{Q}$-*reducible* if it contains a $\mathbb{Q}$-essential 2-sphere.

Theorem 6.9. (i) *If $M(\alpha)$ and $M(\beta)$ are $\mathbb{Q}$-reducible then $\alpha = \beta$.*

(ii) *If $M(\alpha)$ is reducible and $M(\beta)$ is $\mathbb{Q}$-reducible, with $\alpha \neq \beta$, then $M(\alpha)$ has a lens space summand.*

(iii) *If $M(\alpha)$ and $M(\beta)$ are reducible, with $\alpha \neq \beta$, then either both $M(\alpha)$ and $M(\beta)$ have summands M_α, M_β with $H_1(M_\alpha)$ and $H_1(M_\beta)$ finite and non-zero, or one of $M(\alpha), M(\beta)$ has a lens space summand.*

Remarks. (1) Part (i) is a special case of a result of Scharlemann [S3] proved using the theory of sutured manifolds. The case where $M(\alpha)$ and $M(\beta)$ are connected sums of copies of $S^1 \times S^2$ was proved by Gabai [Ga1].

(2) Scharlemann [S3] proves that under the hypotheses of (ii), $M(\beta) \cong S^1 \times S^2 \# W$, and $M(\alpha) \cong W' \# W''$, where W, W' and W'' are $\mathbb{Q}$-homology spheres, and that either W' is a lens space or $H_1(W) \neq 0$.

(3) It is shown in [GL4], by an elaboration of the methods discussed here, that if $M(\alpha)$ and $M(\beta)$ are reducible then $\Delta(\alpha, \beta) \leq 1$. A different proof has been given by Boyer and Zhang [BZ].

Proof of Theorem 6.9. (i) Let $\widehat{P}, \widehat{Q}$ be $\mathbb{Q}$-essential 2-spheres in $M(\alpha)$, $M(\beta)$ respectively with p and q minimal. By Theorem 6.3, either G_P represents all types or G_Q contains a Scharlemann cycle. But the first contradicts the minimality of q (see (Td) in Section 4), while the second contradicts the minimality of p (see (Sd) in Section 4).

(ii) Let $\widehat{Q}$ be as in (i), and let $\widehat{P}$ be an essential 2-sphere in $M(\alpha)$. By Theorem 6.3 (and (Td)) G_Q contains a Scharlemann cycle. Hence $M(\alpha)$ has a lens space summand.

(iii) Let $\widehat{P}, \widehat{Q}$ be essential 2-spheres in $M(\alpha)$, $M(\beta)$ respectively. The result then follows from Theorem 6.3, applied as stated and also with P and Q interchanged, (together with (Sc) and (Tc) in Section 4). $\square$

References

[A] D. Auckly, *Surgery numbers of 3-manifolds: a hyperbolic example*, Geometric Topology (W.H. Kazez, ed.), Proceedings of the 1993 Georgia International Topology Conference, AMS/IP Studies in Advanced Mathematics Vol.2, 1997 (Part 1), pp. 21–34.

[BZ] S. Boyer and X. Zhang, *The semi-norm and Dehn filling*, preprint.

[CGLS] M. Culler, C.McA. Gordon, J. Luecke and P.B. Shalen, *Dehn surgery on knots*, Ann. of Math. **125** (1987), 237–300.

[E1] M. Eudave-Muñoz, *Essential tori obtained by surgery on a knot*, Pacific J. Math. **167** (1995), 81–117.

[E2] ———, *Non-hyperbolic manifolds obtained by Dehn surgery on hyperbolic knots*, Geometric Topology (W.H. Kazez, ed.), Proceedings of the 1993 Georgia International Topology Conference, AMS/IP Studies in Advanced Mathematics Vol.2, 1997 (Part 1), pp. 35–61.

[Ga1] D. Gabai, *Foliations and the topology of 3-manifolds. II*, J. Diff. Geom. **26** (1987), 461–478.

[Ga2] ———, *Foliations and the topology of 3-manifolds. III*, J. Diff. Geom. **26** (1987), 479–536.

[Gl] L. Glass, *A combinatorial analog of the Poincaré Index Theorem*, J. Combin. Theory Ser. B **15** (1973), 264–268.

[GS] F. González-Acuña and H. Short, *Knot surgery and primeness*, Math. Proc. Camb. Phil. Soc. **99** (1986), 89–102.

[Go] C.McA. Gordon, *Dehn filling: a survey*, Proceedings of the Mini Semester in Knot Theory, Stefan Banach International Mathematical Center, Warsaw, Poland, 1995 (to appear).

[GL1] C.McA. Gordon and J. Luecke, *Only integral Dehn surgeries can yield reducible manifolds*, Math. Proc. Camb. Phil. Soc. **102** (1987), 97–101.

[GL2] ———, *Knots are determined by their complements*, J. Amer. Math. Soc. **2** (1989), 371–415.

[GL3] ———, *Dehn surgeries on knots creating essential tori, I*, Communications in Analysis and Geometry **3** (1995), 597–644.

[GL4] ———, *Reducible manifolds and Dehn surgery*, Topology **35** (1996), 385–409.

[L] R.A. Litherland, *Surgery on knots in solid tori, II*, J. London Math. Soc. (2) **22** (1980), 559–569.

[P] W. Parry, *All types implies torsion*, Proc. Amer. Math. Soc. **110** (1990), 871–875.

[R] Y. Rieck, *Heegaard structures of manifolds in the Dehn filling space*, preprint.

[S1] M. Scharlemann, *Smooth spheres in* $\mathbb{R}^4$ *with four critical points are standard*, Invent. Math. **79** (1985), 125–141.

[S2] ———, *Unknotting number one knots are prime*, Invent. Math. **82** (1985), 37–55.

[S3] ———, *Producing reducible 3-manifolds by surgery on a knot*, Topology **29** (1990), 481–500.

[T] I. Torisu, *Boundary slopes for knots*, Osaka J. Math. **33** (1996), 47–55.

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712-1082

Current address: MSRI, 1000 Centennial Dr., Berkeley, CA 94720

E-mail address: `gordon@math.utexas.edu`