

## KNOTS ARE DETERMINED BY THEIR COMPLEMENTS

C. McA. GORDON AND J. LUECKE

### INTRODUCTION

Two (smooth or PL) knots  $K$  and  $K'$  in  $S^3$  are equivalent if there exists a homeomorphism  $h: S^3 \rightarrow S^3$  such that  $h(K) = K'$ . This implies that their complements  $S^3 - K$  and  $S^3 - K'$  are homeomorphic. Here we prove the converse implication.

**Theorem 1.** *If two knots have homeomorphic complements then they are equivalent.*

This answers a question apparently first raised by Tietze [T, p. 83].

It was previously known that there were at most two knots with a given complement [CGLS, Corollary 3].

The notion of equivalence of knots can be strengthened by saying that  $K$  and  $K'$  are *isotopic* if the above homeomorphism  $h$  is isotopic to the identity, or, equivalently, orientation-preserving. The analog of Theorem 1 holds in this setting too: *if two knots have complements that are homeomorphic by an orientation-preserving homeomorphism, then they are isotopic.*

Theorem 1 and its orientation-preserving version are easy consequences of the following theorem concerning Dehn surgery.

**Theorem 2.** *Nontrivial Dehn surgery on a nontrivial knot never yields  $S^3$ .*

**Corollary 2.1.** *If two prime knots have isomorphic groups then they are equivalent.*

Theorem 2 has the following corollary; see [W, p. 26; J; FW].

The arguments used to prove Theorem 2 also lead to restrictions on when Dehn surgery on a knot yields a reducible manifold. (It is conjectured that this happens only with torus knots and cable knots.)

**Theorem 3.** *If a 3-manifold obtained by Dehn surgery on a nontrivial knot is reducible then it has a lens space as a connected summand.*

---

Received by the editors August 28, 1988.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 57M25.

The first author was partially supported by NSF grant NSF-DMS-8701366.

The second author was partially supported by an NSF post-doctoral Fellowship and NSF-DMS-8504683.

**Corollary 3.1.** *Any homology 3-sphere obtained by Dehn surgery on a knot is irreducible.*

Theorem 3 also gives a new proof of the following result of Gabai [Ga], which includes the Property R Conjecture.

**Corollary 3.2** (Gabai). *Any 3-manifold obtained by 0-framed surgery on a non-trivial knot is irreducible.*

Let  $K$  be a nontrivial knot in  $S^3$ , with tubular neighborhood  $N(K)$ , and let  $X = S^3 - \text{int } N(K)$  be the exterior of  $K$ . Let  $\rho$  be a slope on  $\partial X$ , that is, the unoriented isotopy class of an essential simple loop on  $\partial X$ . Let  $K(\rho)$  denote the closed 3-manifold obtained by  $\rho$ -Dehn surgery on  $K$ , in other words, the result of attaching a solid torus  $V$  to  $X$  so that  $\rho$  bounds a disk in  $V$ . Let  $\gamma$  be the slope of a meridian of  $K$ . Then the trivial Dehn surgery yields  $K(\gamma) \cong S^3$ . Let  $\pi$  be another slope on  $\partial X$ , having minimal geometric intersection number  $n \geq 1$  with  $\gamma$ .

§1 is devoted to the proof of the following proposition, which has been independently proved by David Gabai.

**Proposition 1.** *If  $K(\pi)$  is homeomorphic to  $S^3$ , then there exist planar surfaces  $P$  and  $Q$  properly embedded in  $X$  such that*

- (i)  $\partial P$  (resp.  $\partial Q$ ) consists of parallel copies of  $\pi$  (resp.  $\gamma$ );
- (ii)  $P$  and  $Q$  intersect transversely, and each component of  $\partial P$  intersects each component of  $\partial Q$  in  $n$  points;
- (iii) no arc of  $P \cap Q$  is boundary-parallel in either  $P$  or  $Q$ .

The construction of  $P$  and  $Q$  is based on [Ga, §4(A)], using the idea of a thin presentation of a knot.

In §§2 and 3 we prove the following proposition (in which we do not assume that  $K(\pi)$  is  $S^3$ ).

**Proposition 2.** *Suppose that  $X$  contains properly embedded planar surfaces  $P$  and  $Q$  satisfying conditions (i), (ii), and (iii) of Proposition 1, where  $Q$  is the intersection with  $X$  of a level sphere in a thin presentation of  $K$ . Then  $K(\pi)$  has a lens space as a connected summand.*

The proof of Proposition 2 is based on a combinatorial analysis of the intersection of the planar surfaces  $P$  and  $Q$ . Capping off the boundary components of  $P$  and  $Q$  with disks, we regard these disks as forming the “fat” vertices of graphs  $G_P$  and  $G_Q$  in  $S^2$ , the edges of  $G_P$  (resp.  $G_Q$ ) corresponding to the arcs of  $P \cap Q$  in  $P$  (resp.  $Q$ ). The (disk) faces of  $G_P$  correspond to subdisks of  $P$ , which we may regard as lying in  $K(\gamma)$  with their boundaries contained in  $Q \cup \partial N(K)$ . Similarly, the faces of  $G_Q$  may be regarded as lying in  $K(\pi)$ . This allows us to infer topological properties of  $K(\gamma)$  (resp.  $K(\pi)$ ) from graph-theoretic properties of  $G_P$  (resp.  $G_Q$ ). For  $n \geq 2$  this program

is already carried out in [CGLS, §§2.5 and 2.6]; in particular, Proposition 2 in this case follows from [CGLS, Proposition 2.5.6].

In §2 we therefore assume that  $n = 1$ , and develop the graph-theoretic techniques necessary to handle this case. The main result of the chapter is Proposition 2.0.1, which asserts that either the graph  $G_Q$  contains a special kind of face (a *Scharlemann cycle*) which implies that  $K(\pi)$  contains a punctured lens space, or else there is a certain desirable set of faces in the graph  $G_P$ . In §3 we show that under the hypotheses of Proposition 2, such a set of faces in  $G_P$  cannot exist, thus proving the proposition.

The idea of analyzing certain types of labeled graphs to study intersections of planar surfaces (in the context of Dehn surgery) is originally due to Litherland [L].

We conclude this introduction with the proofs of Theorems 2 and 3.

*Proof of Theorem 2.* This follows immediately from Propositions 1 and 2 and the fact that the surface  $Q$  in Proposition 1 does indeed arise as the intersection with  $X$  of a level sphere in a thin presentation of  $K$ .  $\square$

*Proof of Theorem 3.* Let  $\pi$  be a slope on  $\partial X$  such that  $K(\pi)$  is reducible. Then there exists a properly embedded, incompressible (and nonboundary-parallel) planar surface  $P$  in  $X$  whose boundary components have slope  $\pi$ . By [Ga, §4(A)] there is a planar surface  $Q$  in  $X$ , coming from a level sphere in a thin presentation of  $K$ , such that  $P$  and  $Q$  satisfy conditions (i), (ii), and (iii) of Proposition 1. Proposition 2 then gives the desired result.  $\square$

The first author would like to thank the members of the topology group at the University of California at Santa Barbara for providing the opportunity to present this work in their seminar during July 1988 and for their helpful comments.

The second author would like to thank the Courant Institute for its hospitality while this work was being carried out.

### 1. FINDING THE PLANAR SURFACES

This section is devoted to the proof of Proposition 1 stated in the Introduction.

The construction of the planar surfaces  $P$  and  $Q$  is based on [Ga, §4(A)], using the idea of a *thin presentation* of a knot in  $S^3$ . Given  $P$ , the argument in [Ga, §4(A)] produces  $Q$  such that (i) and (ii) hold and no arc of  $P \cap Q$  is boundary-parallel in  $P$ . Similarly, given  $Q$ , one can find  $P$  satisfying (i) and (ii) so that no arc on  $P \cap Q$  is boundary-parallel in  $Q$ . The additional content of Proposition 1 is that we can find  $P$  and  $Q$  so that these conditions hold simultaneously.

For the convenience of the reader we recall the definition of a thin presentation. Let  $\pm\infty$  be the north and south poles of  $S^3$ . Then  $S^3 - \{\pm\infty\}$  is naturally homeomorphic to  $S^2 \times R^1$ , and we have an associated height function  $h: S^3 - \{\pm\infty\} \rightarrow R^1$ . The *level 2-spheres* are the spheres  $S^2 \times \{t\}$ ,  $t \in R^1$ .

Let  $K$  be a nontrivial knot in  $S^3$ . By an isotopy of  $K$  we may assume that  $K \subset S^3 - \{\pm\infty\}$  and that  $h|_K$  is a Morse function (that is,  $h|_K$  has only finitely many critical points, all nondegenerate, and with all critical values distinct).

Given such a Morse presentation of  $K$ , let  $S_1, \dots, S_m$  be level 2-spheres, one between each consecutive pair of critical levels of  $h|_K$ . One then calls the number  $\sum_{i=1}^m |S_i \cap K|$  the *complexity* of the Morse presentation. A *thin presentation* of  $K$  is a Morse presentation of minimal complexity.

Let  $K$  be in Morse presentation, and let  $S$  be a level 2-sphere. If  $S$  meets the tubular neighborhood  $N(K)$  of  $K$  in meridian disks then the planar surface  $Q = S - \text{int } N(K) \subset X = S^3 - \text{int } N(K)$  is called a *level surface*.

More generally, let  $F$  be the intersection with  $X$  of any 2-sphere in  $S^3$  that meets  $N(K)$  in meridian disks.  $F$  separates  $X$  into two components, one of which we shall regard as lying *above*  $F$ , the other, *below*  $F$ .

Let  $F'$  be a surface properly embedded in  $X$  such that  $F$  intersects  $F'$  transversely and each component of  $\partial F$  intersects each component of  $\partial F'$  algebraically and geometrically the same number of times.

Suppose that  $F \cap F'$  contains an arc  $\alpha$  which, together with an arc  $\beta \subset \partial F'$ , bounds a disk  $\Delta$  in  $F'$  such that  $\Delta - \alpha$  contains no arc components of  $F \cap F'$  (but may contain circle components). If  $\beta$  lies above (below)  $F$  then we say that  $\Delta$  is a *high (low) disk* for  $F$ , and that  $F$  is *low (high) with respect to*  $F'$ .

**Lemma 1.1.** *Let  $F$  be isotopic in  $X$  to a level surface in a thin presentation of  $K$ . Then  $F$  cannot be both high and low with respect to the same or disjoint surfaces.*

*Proof.* We may assume that  $F$  is a level surface in a thin presentation of  $K$ .

Assume for contradiction that  $F$  is both high and low with respect to the same or disjoint surfaces. Then there exist high and low disks  $\Delta^+$  and  $\Delta^-$  for  $F$  such that  $\partial\Delta^\pm = \alpha^\pm \cup \beta^\pm$ ,  $\alpha^\pm = \partial\Delta^\pm \cap F$ ,  $\beta^\pm = \Delta^\pm \cap \partial X$ , and such that either  $\alpha^+ = \alpha^-$  or  $\alpha^+ \cap \alpha^- = \emptyset$ .

Extending radially inwards across  $N(K)$ ,  $F$  becomes a level 2-sphere  $\widehat{F}$  and  $\beta^+$  and  $\beta^-$  become arcs in  $K$ . See Figure 1 for the possible configurations.

We can now use  $\Delta^+$  and  $\Delta^-$  to isotop  $\Delta^+ \cap K = \beta^+$ ,  $\Delta^- \cap K = \beta^-$  to  $\alpha^+$ ,  $\alpha^-$  respectively. If  $\partial\beta^+ = \partial\beta^-$ , it follows that  $K$  is unknotted. Otherwise, perturbing  $\beta^+$  and  $\beta^-$  slightly to make them transverse to the level spheres then reduces the complexity of  $K$ , contradicting our hypothesis.  $\square$

Let  $F$  be a compact surface with boundary, and let  $f: F \times I \rightarrow S^3 - \{\pm\infty\}$  be a (smooth) embedding. Write  $F(\lambda) = f(F \times \{\lambda\})$ , etc.,  $\lambda \in I$ . We say the family  $\{F(\lambda)\}$  is *Cerf with respect to the height function  $h$*  if the following conditions hold.

- (i)  $f$  extends to an embedding of  $F^+ \times I$ , where  $F^+$  is  $F$  enlarged by the addition of an external open collar  $C^+$ , and there exists an internal collar  $C^-$  of  $\partial F$  in  $F$  such that  $h|(C^+ \cup C^-)(\lambda)$  has no critical points, for all  $\lambda \in I$ ;

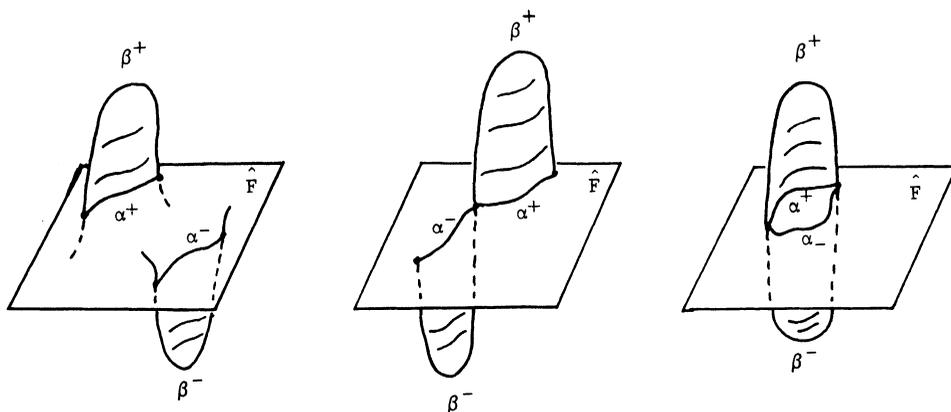


FIGURE 1

- (ii)  $h|F^+(\lambda)$  is a Morse function for all but finitely many values of  $\lambda$ ;
- (iii) if  $h|F^+(\lambda)$  is not a Morse function then exactly one of the following occurs:
  - (a)  $h|F^+(\lambda)$  has distinct critical values and a single birth-death degenerate critical point (all other critical points are nondegenerate).
  - (b)  $h|F^+(\lambda)$  has only nondegenerate critical points and distinct critical values except for a single value which corresponds to exactly two critical points.

Recall that the *graphic* is the image in  $I \times R^1$  of the singular points of the function  $h|f(F \times I)$  under the map  $f(v, \lambda) \mapsto (\lambda, hf(v, \lambda))$ . The graphic is locally modeled on the pictures in Figure 2.

Then the final condition we require in the definition of a Cerf family is

- (iv) the tangent directions on any branch of the graphic are never vertical, and never horizontal at a singularity of the graphic.

We require one more definition before proceeding with the proof of Proposition 1. Let  $K$  be in thin presentation, and let  $a < b$  be two adjacent critical values of  $h|K$ , such that  $a$  corresponds to a local minimum and  $b$  to a local maximum. Then a family  $Q \times I$  of level surfaces in  $X$  such that  $h(Q \times I) \subset (a, b)$  is called a *middle slab*.

*Proof of Proposition 1.* As in the introduction, let  $K \subset S^3$  be a nontrivial knot and  $X$  its exterior. Let the slope  $\gamma$  on  $\partial X$  represent the meridian of  $K$ . Let  $\pi$  be another slope on  $\partial X$ , with minimal geometric intersection number  $n \geq 1$  with  $\gamma$ , such that  $K(\pi) \cong S^3$ . Then the cores of the attached solid tori  $V_\gamma$  and  $V_\pi$  of  $K(\gamma)$  and  $K(\pi)$  become nontrivial knots  $K$  and  $K_\pi$  in copies  $S_\gamma^3$  and  $S_\pi^3$  of  $S^3$ . Let  $h_\gamma$  and  $h_\pi$  denote the height functions on  $S_\gamma^3 - \{\pm\infty\}$  and  $S_\pi^3 - \{\pm\infty\}$ .

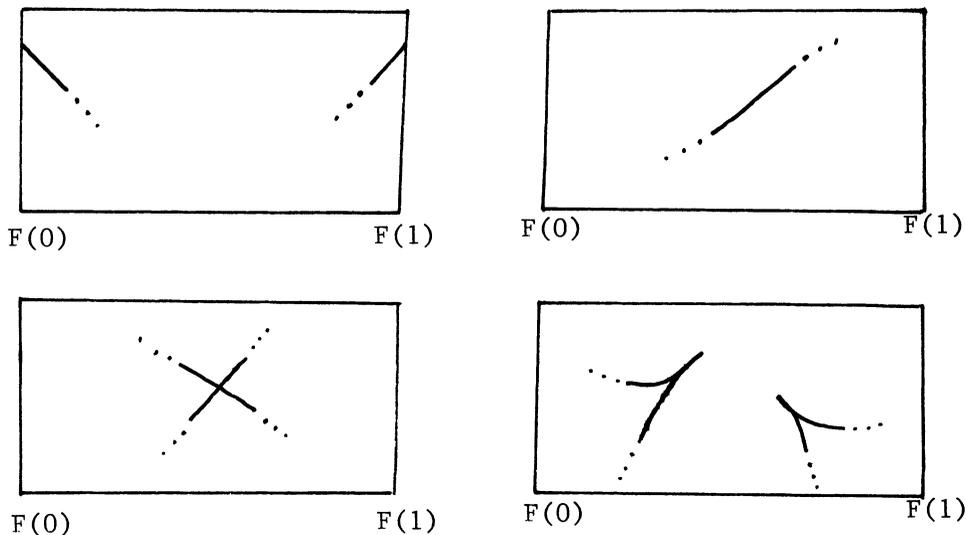


FIGURE 2

**Lemma 1.2.** *There are families of planar surfaces  $P \times I$  and  $Q \times I$  in  $X$  such that*

- (1)  $Q \times I$  is a middle slab in a thin presentation of  $K$ ;
- (2)  $P \times I$  is isotopic to a middle slab in a thin presentation of  $K_\pi$ ;
- (3) each component of  $\partial P(\lambda)$  intersects each component of  $\partial Q(\mu)$  transversely in  $n$  points;
- (4)  $P(0)$  is high and  $P(1)$  is low with respect to  $Q(\mu_0)$  for some  $\mu_0 \in (0, 1)$ ;
- (5)  $Q(0)$  is high and  $Q(1)$  is low with respect to  $P(\lambda_0)$  for some  $\lambda_0 \in (0, 1)$ ;
- (6)  $\{P(\lambda)\}$  is Cerf with respect to  $h_\gamma$ .

*Proof.* Choose thin presentations for  $K$  and  $K_\pi$ , and isotop  $V_\gamma$  and  $V_\pi$  so that they are locally modeled on Figure 3.

Let  $P \times I$  and  $Q \times I$  be middle slabs in these thin presentations of  $K$  and  $K_\pi$ , respectively.

Let  $C$  be a collar neighborhood of  $\partial P$  in  $P$ . First isotop  $P \times I$  so that in  $S^3_\gamma$   $C \times I$  looks like several parallel copies of product neighborhoods of the sheets shown in Figure 4.

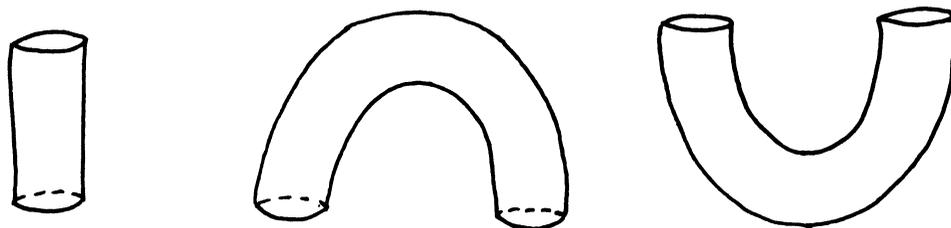


FIGURE 3

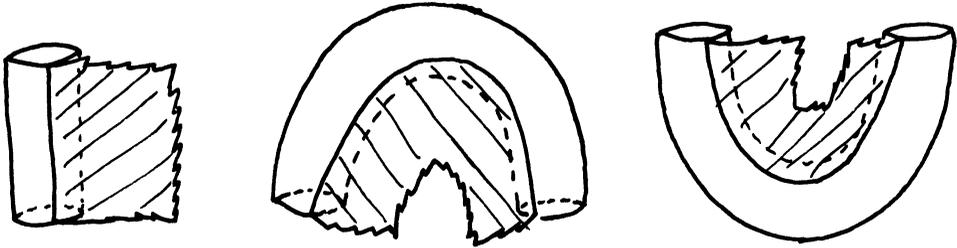


FIGURE 4



FIGURE 5

Also, ensure that immediately above (resp. below)  $Q \times I$ , near the first local maximum (resp. minimum) of  $K$  just above (resp. below)  $Q \times I$ ,  $C \times I$  is in the same sense locally modeled on Figure 5.

Now pick  $\lambda_0 \in (0, 1)$  and isotop  $P \times I$  by a small isotopy, fixed on  $C \times I$ , so that  $P(\lambda_0)$  is transverse to  $Q(0) \cup Q(1)$ . Then  $Q(0)$  is high and  $Q(1)$  is low with respect to  $P(\lambda_0)$ .

Pick  $\mu_0 \in (0, 1)$ . Mentally reversing the isotopy that has been applied to  $P \times I$  so far, consider the image of  $Q(\mu_0)$  in  $S^3_\pi$ . This may be isotoped by an isotopy fixed on  $C_0 \times I \cup P(\lambda_0)$ , where  $C_0$  is a subcollar of  $C$ , so that near the local maximum and minimum of  $K_\pi$  adjacent to  $P \times I$  it looks like Figure 5. Reversing *this* isotopy gives an isotopy of  $P \times I$ , making  $P(1)$  low and  $P(0)$  high with respect to  $Q(\mu_0)$ .

Conditions (1)–(5) of Lemma 1.2 are now satisfied.

Regarding (6), note that condition (i) in the definition of Cerf is already satisfied. Conditions (ii), (iii), and (iv) can now be achieved by an arbitrarily small isotopy of  $P \times I$ , by the transversality arguments in [Cerf, Chapter 2], together with the fact that for compact manifolds, the space of smooth embeddings is open in the space of smooth maps. Since (3), (4), and (5) are also open conditions, this can be done so that they continue to hold.  $\square$

By replacing  $Q \times I$  by a slightly thinner slab if necessary (and reparametrizing  $I$ ), we may also assume that for those values of  $\lambda$  at which the graphic intersects  $I \times \{0, 1\}$ , it does so transversely at a single point, and  $h_\gamma|P(\lambda)$  is a Morse function (see Figure 6).

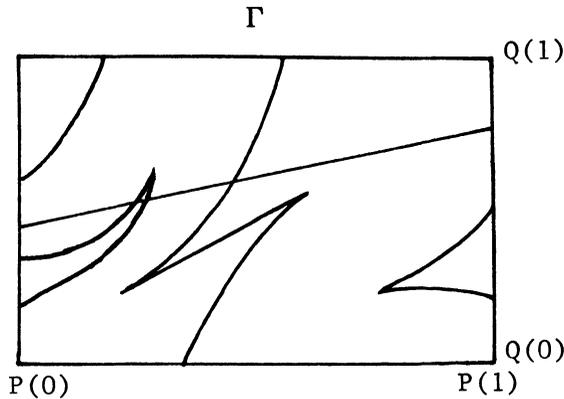


FIGURE 6

Let  $\Gamma$  be the intersection of the graphic with  $I \times I$  (Figure 6). Thus the singularities of  $\Gamma$  occur at a finite set of values  $s_i$  of  $\lambda$ ,  $0 < s_1 < \dots < s_n < 1$ , each corresponding to one of the following:

- (1) interchange of two critical values;
- (2) birth-death;
- (3) a point of intersection of  $\Gamma$  with  $I \times \{0, 1\}$ .

Note that for  $(\lambda, \mu) \in I^2 - \Gamma$ ,  $P(\lambda)$  intersects  $Q(\mu)$  transversely.

Define functions  $p, q: I^2 - \Gamma \rightarrow \{H, L, N\}$  as follows.

$p(\lambda, \mu) = H, L$ , or  $N$  according as  $P(\lambda)$  is high, low, or neither with respect to  $Q(\mu)$ ;

$q(\lambda, \mu) = H, L$ , or  $N$  according as  $Q(\mu)$  is high, low, or neither with respect to  $P(\lambda)$ .

We note the following properties of  $p$  and  $q$ .

(P.1) By Lemma 1.1, we cannot have  $p(\lambda, \mu_1) = H$  and  $p(\lambda, \mu_2) = L$ , nor  $q(\lambda_1, \mu) = H$  and  $q(\lambda_2, \mu) = L$ . In particular,  $p$  and  $q$  are well defined.

Also, we say that  $P(\lambda)$  is  $H$  (resp.  $L$ ) if  $p(\lambda, \mu) = H$  (resp.  $L$ ) for some  $\mu$ , and similarly for  $Q(\mu)$ .

(P.2) As  $(\lambda, \mu)$  varies continuously in  $I^2 - \Gamma$ ,  $P(\lambda) \cap Q(\mu)$  does not change up to isotopy (in  $P(\lambda)$  and  $Q(\mu)$ ). Hence  $p$  and  $q$  are locally constant.

(P.3) Let  $c_\lambda$  be a critical value of  $P(\lambda)$  of index 0 or 2. As  $\mu$  passes from one side of  $c_\lambda$  to the other,  $P(\lambda) \cap Q(\mu)$  either acquires or loses a single circle and is otherwise unchanged up to isotopy. Hence, for all sufficiently small  $\varepsilon > 0$ , we have  $p(\lambda, c_\lambda - \varepsilon) = p(\lambda, c_\lambda + \varepsilon)$ , and  $q(\lambda, c_\lambda - \varepsilon) = q(\lambda, c_\lambda + \varepsilon)$ .

(P.4) By Lemma 1.2(4),  $P(0)$  is  $H$  and  $P(1)$  is  $L$ .

(P.5) By Lemma 1.2(5),  $Q(0)$  is  $H$  and  $Q(1)$  is  $L$ .

For  $\lambda \neq s_i$ , let  $0 = c_0(\lambda) < c_1(\lambda) < \dots < c_m(\lambda) = 1$  be the critical values of  $P(\lambda)$  together with 0 and 1. Let  $I_j(\lambda)$  denote the interval  $(c_{j-1}(\lambda), c_j(\lambda))$ ,  $1 \leq j \leq m$ .

We say  $I_j(\lambda)$  is  $H$ ,  $L$ , or  $N$  according as for some (and hence by (P.2), all)  $\mu \in I_j(\lambda)$  we have  $q(\lambda, \mu) = H$ ,  $L$ , or  $N$ .

**Lemma 1.3.** *For  $\lambda \neq s_i$  there exist  $k, l$  with  $1 \leq k < l \leq m$  such that  $I_j(\lambda)$  is  $N$  for  $k \leq j \leq l$ ,  $I_{k-1}(\lambda)$  (if it exists) is  $H$ , and  $I_{l+1}(\lambda)$  (if it exists) is  $L$ .*

*Proof.* By (P.5),  $I_1(\lambda)$  is not  $L$ , and  $I_m(\lambda)$  is not  $H$ .

By the argument given in [Ga, §4], we cannot have one of the intervals  $I_j(\lambda)$ ,  $I_{j+1}(\lambda)$ ,  $H$  and the other  $L$ .

The lemma follows.  $\square$

Our goal is to find  $(\lambda, \mu) \in I^2 - \Gamma$  such that  $p(\lambda, \mu) = N = q(\lambda, \mu)$ . The surfaces  $P(\lambda)$  and  $Q(\mu)$  then satisfy the conclusion of Proposition 1. From now on, we will assume that no such  $(\lambda, \mu)$  exists, and ultimately derive a contradiction.

Consider a singular point  $(s, t)$  in the graphic  $\Gamma$  corresponding to the interchange of the values of two critical points. Then there is a neighborhood  $U$  of  $(s, t)$  in  $I^2$  such that  $\Gamma \cap U$  is as shown in Figure 7.

Recall that for small enough  $U$  neither branch of  $\Gamma \cap U$  has a horizontal or vertical tangent in  $U$ .

Let  $R_1, R_2, R_3,$  and  $R_4$  be the indicated open regions of  $U$ .

**Lemma 1.4.** *There exists  $(s, t) \in \Gamma$  as above, where both critical points have index 1, such that*

$$\begin{aligned} p(\lambda, \mu) &= H && \text{in } R_1, \\ p(\lambda, \mu) &= L && \text{in } R_3, \\ q(\lambda, \mu) &= H && \text{in } R_2, \\ q(\lambda, \mu) &= L && \text{in } R_4. \end{aligned}$$

*Proof.* For  $\lambda \neq s_i$ , let  $J_\lambda$  denote the union of all the intervals  $I_j(\lambda)$  that are  $N$ . By Lemma 1.3,  $J_\lambda$  is nonempty.

For  $\mu \in J_\lambda$ , since  $q(\lambda, \mu) = N$  we must have  $p(\lambda, \mu) = H$  or  $L$ . Thus  $P(\lambda)$  is correspondingly either  $H$  or  $L$ .

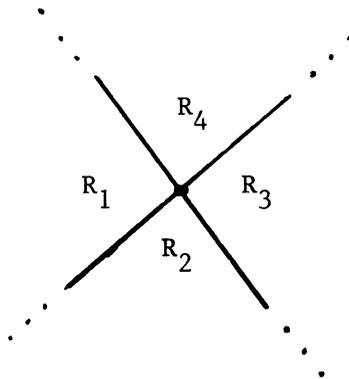


FIGURE 7

By (P.4),  $P(0)$  is  $H$  and  $P(1)$  is  $L$ . Therefore (by (P.2)) there exists  $s = s_i$  for some  $i$  and  $\delta > 0$  such that

$$P(\lambda) = \begin{cases} H, & \text{for } \lambda \in (s - \delta, s), \\ L, & \text{for } \lambda \in (s, s + \delta). \end{cases}$$

It follows that if  $|J_\lambda|$  denotes the sum of the lengths of the intervals in  $J_\lambda$ , then  $|J_\lambda| \rightarrow 0$ , as  $\lambda \rightarrow s_-$  and as  $\lambda \rightarrow s_+$ .

If  $s$  corresponds to a singularity of type (2) or (3), then  $|J_\lambda| \rightarrow 0$  as  $\lambda \rightarrow s$  from the side away from that part of  $\Gamma$  containing the singularity. Hence  $s$  corresponds to the interchange of two critical values  $c_\lambda$  and  $c'_\lambda$ . See Figure 8. Since  $|J_\lambda| \rightarrow 0$  as  $\lambda \rightarrow s_\pm$ , we must have for some  $\delta > 0$ :

$$J_\lambda = \begin{cases} (c_\lambda, c'_\lambda), & \text{for } \lambda \in (s - \delta, s), \\ (c'_\lambda, c_\lambda), & \text{for } \lambda \in (s, s + \delta). \end{cases}$$

Therefore, by (P.3), each of the critical points corresponding to  $c_\lambda$  and  $c'_\lambda$  has index 1.

The assertions about the values of  $p$  and  $q$  in the regions  $R_1, R_2, R_3$ , and  $R_4$  follow from the definition of  $s$ , Lemma 1.3, and (P.2).  $\square$

With respect to the lines  $\lambda = s, \mu = t$  there are a priori three possibilities for  $\Gamma$  in  $U$ , illustrated in Figure 9.

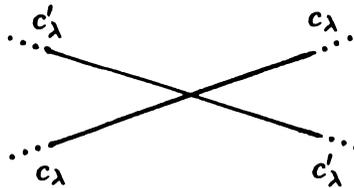


FIGURE 8

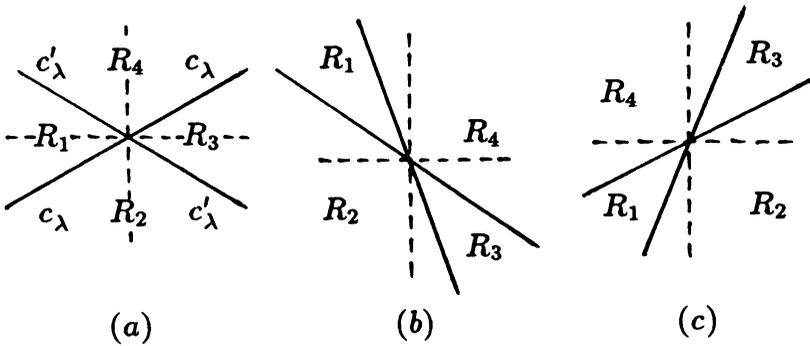


FIGURE 9

**Lemma 1.5.**  $\Gamma \cap U$  is as shown in case (a) of Figure 9.

*Proof.* In (b) and (c) there are horizontal lines  $\mu = \mu_0$  meeting both  $R_2$  and  $R_4$ . Then by Lemma 1.4 we would have  $q(\lambda_1, \mu_0) = H$  and  $q(\lambda_2, \mu_0) = L$ , contradicting (P.1).  $\square$

We now proceed to show that the situation described in Lemmas 1.4 and 1.5 is impossible.

The intersection  $P(s) \cap Q(t)$ , considered as lying in  $Q(t)$ , is a properly embedded 1-manifold with two singularities  $x$  and  $x'$  corresponding to the two saddle points of  $P(s)$  with critical value  $t$  ( $x$  and  $x'$  corresponding to the branches  $c_\lambda$  and  $c'_\lambda$  of the graphic). Figure 10 illustrates  $P(\lambda) \cap Q(\mu)$ , in  $Q(\mu)$ , in a neighborhood of these two saddle points, for  $(\lambda, \mu)$  belonging to the different regions  $R_1, R_2, R_3$ , and  $R_4$  and their boundaries.

For  $(\lambda, \mu) \in R_1$  (resp.  $R_3$ ) we have a continuous family of disks  $A_{\lambda, \mu}^-$  (resp.  $A_{\lambda, \mu}^+$ ) in  $Q(\mu)$  which are low (resp. high) disks for  $P(\lambda)$ .

Similarly, for  $(\lambda, \mu) \in R_2$  (resp.  $R_4$ ) we have disks  $B_{\lambda, \mu}^-$  (resp.  $B_{\lambda, \mu}^+$ ) in  $P(\lambda)$  which are low (resp. high) disks for  $Q(\mu)$ .

$A_{\lambda, \mu}^\pm \cap P(\lambda)$  contains a single arc component  $\alpha_{\lambda, \mu}^\pm$ , and  
 $B_{\lambda, \mu}^\pm \cap Q(\mu)$  contains a single arc component  $\beta_{\lambda, \mu}^\pm$ .

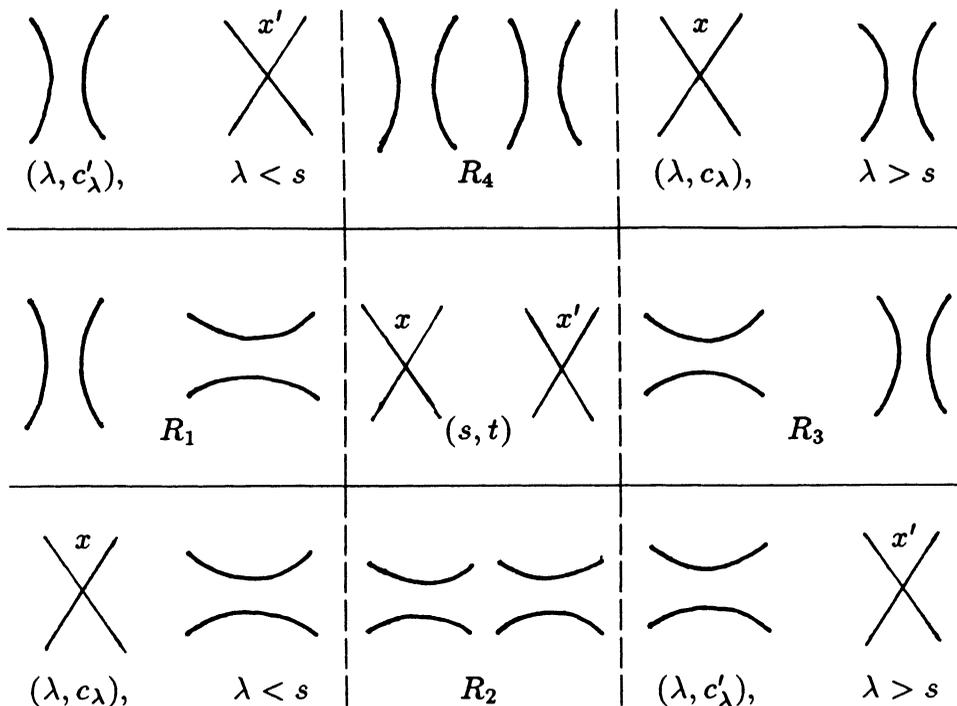


FIGURE 10

Consider these arcs  $\alpha_{\lambda,\mu}^\pm$  and  $\beta_{\lambda,\mu}^\pm$  as they lie in  $Q(\mu)$ . Note that each of  $\partial\alpha_{\lambda,\mu}^+$  and  $\partial\alpha_{\lambda,\mu}^-$  lies in a single boundary component of  $Q(\mu)$ , while  $\beta_{\lambda,\mu}^+$  and  $\beta_{\lambda,\mu}^-$  each join distinct boundary components of  $Q(\mu)$ .

Let  $\alpha^\pm$  and  $\beta^\pm$  denote the possibly singular arcs in  $Q(t)$  that are the limits of  $\alpha_{\lambda,\mu}^\pm$  and  $\beta_{\lambda,\mu}^\pm$  as  $(\lambda, \mu) \rightarrow (s, t)$  in the appropriate region  $R_1, R_2, R_3$ , or  $R_4$ .

**Lemma 1.6.**  $\beta^+ \cup \beta^-$  meets at least three distinct boundary components of  $Q(t)$ .

*Proof.* If not, then there are boundary components  $\partial_1$  and  $\partial_2$  of  $Q$  such that  $\beta_{\lambda,\mu}^+$  joins  $\partial_1 \times \{\mu\}$  and  $\partial_2 \times \{\mu\}$ , and  $\beta_{\lambda,\mu}^-$  joins  $\partial_1 \times \{\mu\}$  and  $\partial_2 \times \{\mu\}$ . Since  $B_{\lambda,\mu}^+$  is a high disk for  $Q(\mu)$  and  $B_{\lambda,\mu}^-$  is a low disk for  $Q(\mu)$ , this implies that  $\partial_1$  and  $\partial_2$  are the only boundary components of  $Q$ , and hence that  $K$  is trivial.  $\square$

**Lemma 1.7.** Each of  $\alpha^+, \alpha^-, \beta^+$ , and  $\beta^-$  contains  $x$  and  $x'$ .

*Proof.* Suppose, for example, that  $x \notin \alpha^+$ . Then, taking  $(\lambda, \mu) \in R_3$  and increasing  $\mu$  past  $c_\lambda$  until  $(\lambda, \mu) \in R_4$ ,  $A_{\lambda,\mu}^+$  remains a nonsingular disk in  $Q(\mu)$  and is therefore a high disk for  $P(\lambda)$ . Recalling Figure 9(a), this implies that for some  $\delta > 0$ ,  $P(\lambda)$  is  $L$  for  $\lambda \in (s - \delta, s)$ . But this contradicts the fact that  $p(\lambda, \mu) = H$  for  $(\lambda, \mu) \in R_1$ .

The other cases are completely analogous.  $\square$

For  $i = 1, 2, 3$ , or  $4$ , let  $G_i$  (resp.  $G'_i$ ) denote the smallest union of components of  $P(\lambda) \cap Q(\mu)$ , for  $(\lambda, \mu) \in R_i$ , that contains both branches of  $P(\lambda) \cap Q(\mu)$  in a small neighborhood of  $x$  (resp.  $x'$ ). (See Figure 10, the squares labeled  $R_1, R_2, R_3$ , and  $R_4$ .)

**Lemma 1.8.** Each of  $G_1, G'_1, G_3$  and  $G'_3$  meets at least two boundary components of  $Q(\mu)$ .

*Proof.* We prove this for  $G_1$ . The other cases are completely analogous.

Since  $x \in \beta^-$  by Lemma 1.7,  $\beta_{\lambda,\mu}^- \subset G_2$ . Since  $\beta_{\lambda,\mu}^-$  joins distinct boundary components of  $Q(\mu)$ , and since  $G_1$  is obtained from  $G_2$  by the transformation  $\asymp \mapsto ()$  in a neighborhood of  $x$ , the result follows.  $\square$

**Lemma 1.9.** Each of  $\alpha^+$  and  $\alpha^-$  involves only a single corner at each of  $x$  and  $x'$ .

*Proof.* Suppose, for example, that  $\alpha^+$  involves two corners at  $x$ . See Figure 10 (middle square). Then we would have  $\alpha_{\lambda,\mu}^+ = G_3$ . But  $\alpha_{\lambda,\mu}^+$  meets only one component of  $\partial Q(\mu)$ , contradicting Lemma 1.8.

The other cases are completely analogous.  $\square$

Let  $A^+, A^- \subset Q(t)$  denote the limits of the disks  $A_{\lambda,\mu}^+$  and  $A_{\lambda,\mu}^-$ .

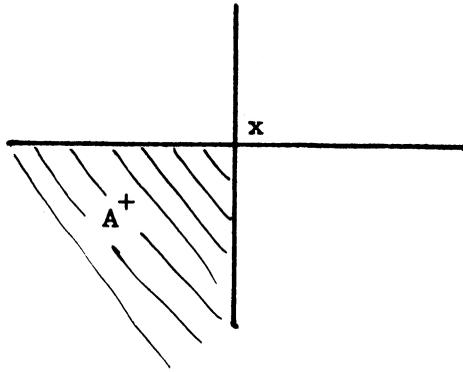


FIGURE 11

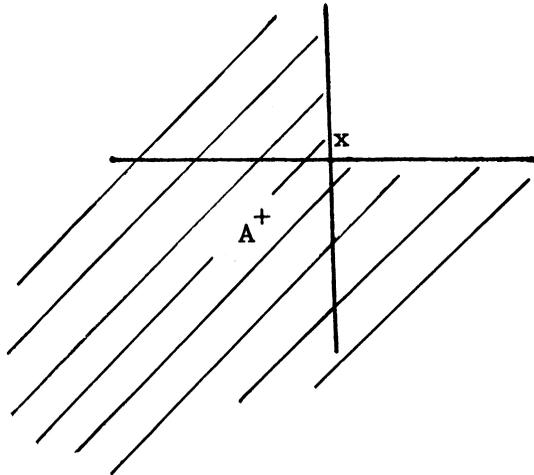


FIGURE 12

**Lemma 1.10.** *In a neighborhood of  $x$ ,  $(A^+, x)$  is as shown in Figure 11. Similarly for  $(A^+, x')$ ,  $(A^-, x)$ , and  $(A^-, x')$ .*

*Proof.* The alternative is as shown in Figure 12. But then  $G_3$  would lie in  $A_{\lambda, \mu}^+$ , contradicting Lemma 1.8.  $\square$

It follows from Lemmas 1.9 and 1.10 that  $A^+$  is as shown in Figure 13 with  $\alpha^+ = \gamma_1 \cup \gamma_2 \cup \gamma_3$  as shown. Similarly,  $A^-$  has the same form.

**Lemma 1.11.**  $\alpha^+ \cap \alpha^- = \gamma_2$ .

*Proof.* If not, then  $\alpha^+ \cap \alpha^-$  would be  $\gamma_1 \cup \gamma_3$  ( $A^+$  and  $A^-$  are locally on opposite sides of  $P(s)$  at  $x$  and  $x'$ ), and we would have the situation shown in Figure 14.

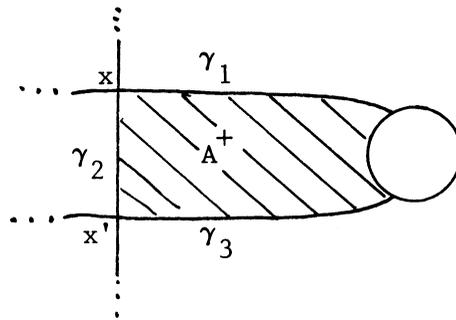


FIGURE 13

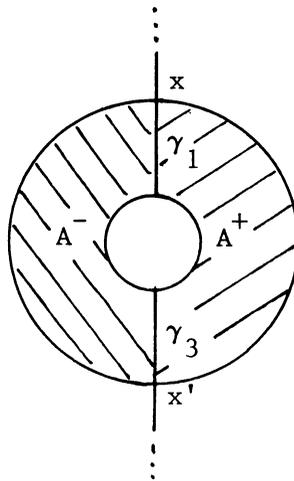


FIGURE 14

But this implies that  $P$  has only two boundary components, and hence that  $K_\pi$  is trivial.  $\square$

By Lemma 1.11,  $A^+$  and  $A^-$  must be as shown in Figure 15 (where the two boundary components  $\partial_1$  and  $\partial_2$  of  $Q(t)$  possibly coincide).

By Lemma 1.6 we have another boundary component  $\partial_3$  of  $Q(t)$  which is (without loss) as indicated in Figure 15.

Then  $\beta_{\lambda,\mu}^+$  and  $\beta_{\lambda,\mu}^-$  are contained in the submanifolds of  $P(\lambda) \cap Q(\mu)$  indicated in Figure 16.

But since  $\beta^+ \cup \beta^-$  meets  $\partial_3$  and each of  $\beta^+$  and  $\beta^-$  contains  $x$  and  $x'$  by Lemma 1.7, this is clearly impossible.

This contradiction completes the proof of Proposition 1.

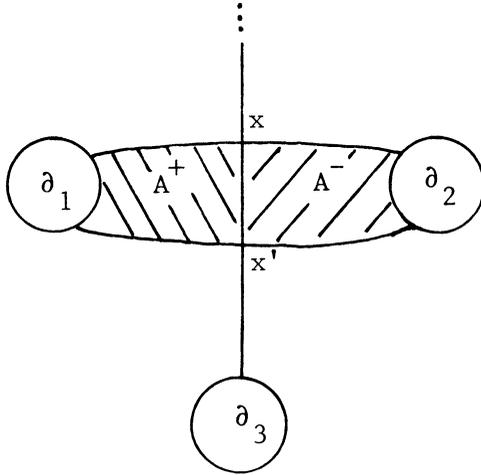


FIGURE 15

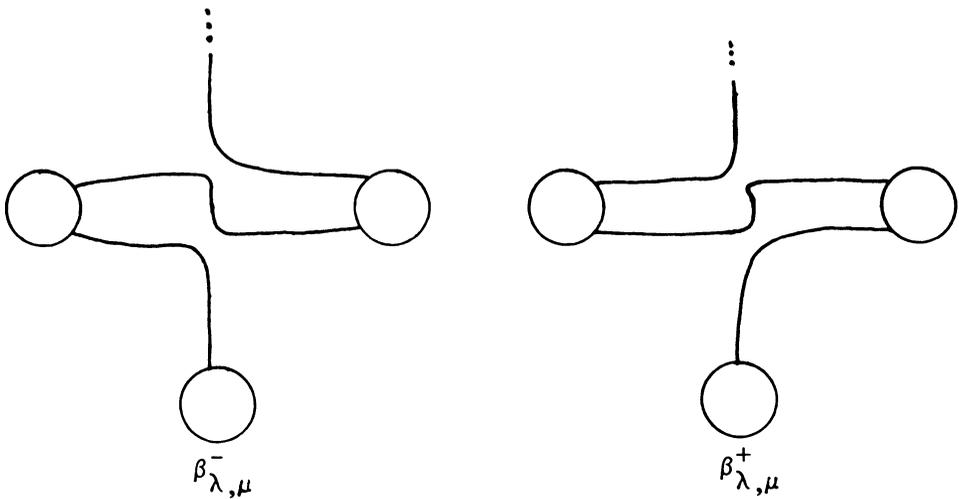


FIGURE 16

2. THE COMBINATORICS

2.0. **Basic definitions.** For the purposes of §2,  $X$  will be an orientable 3-manifold containing a toral boundary component  $T$ .  $T$  contains slopes  $\pi$  and  $\gamma$  with intersection number 1, and  $P$  and  $Q$  will be compact connected planar surfaces properly embedded in  $X$  with  $\partial P, \partial Q \subset T$ . Furthermore, each component of  $\partial P$  and  $\partial Q$  represents  $\pi$  and  $\gamma$  (resp.),  $P$  and  $Q$  intersect transversely, and each component of  $\partial P$  intersects each component of  $\partial Q$  exactly once. Finally, no arc of  $P \cap Q$  will be boundary-parallel in either  $P$  or  $Q$ .



FIGURE 17

Number the components of  $\partial P$ ,  $\{1, \dots, p\}$ , and the components of  $\partial Q$ ,  $\{1, \dots, q\}$ , in the order in which they appear on  $T$ . This allows us to label the end points of arcs of  $P \cap Q$  in  $P$  ( $Q$ ) with the corresponding boundary components of  $Q$  ( $P$ , resp.). Thus around each component of  $\partial P$  we see the labels  $\{1, \dots, q\}$  appearing sequentially (either clockwise or anticlockwise). See Figure 17.

Assigning (arbitrary) orientations to  $P$  and  $Q$  allows us to refer to  $+$  and  $-$  boundary components of  $P$  and  $Q$ , according to the direction of the induced orientation of a boundary component as it lies on  $T$ . For any arc,  $\alpha$ , of  $P \cap Q$  the orientability of  $P$ ,  $Q$ , and  $X$  give us the parity rule: the boundary components of  $P$  joined by  $\alpha$  on  $P$  have the same sign if and only if the boundary components of  $Q$  joined by  $\alpha$  on  $Q$  have opposite sign.

Capping off the boundary components of  $P$  ( $Q$ ) with disks, we regard these disks as forming the “fat” vertices of a graph  $G_P$  ( $G_Q$ ) in  $S^2$ , the edges of  $G_P$  ( $G_Q$ ) corresponding to the arcs of  $P \cap Q$  in  $P$  ( $Q$ ). We thus obtain two labeled graphs in  $S^2$ , whose edges are in one-one correspondence, such that the labeling satisfies the *parity rule* noted above, and such that neither graph contains a trivial loop, i.e., a 1-sided face (no arc is boundary-parallel). It is these graphs with which we will be working in this section. In such a graph,  $G$ , we then have the following concepts:

- (1) *fat vertex*  $v$  and *sign*  $v = \pm$ ;
- (2) *label*,  $x \in \{1, \dots, p\}$  or  $\{1, \dots, q\}$ , and *parity*  $x = \pm$  (= sign of the fat vertex of the other graph corresponding to  $x$ );
- (3)  $(x, v)$ : the occurrence of a label,  $x$ , at vertex  $v$ . *Character* of  $(x, v) = \text{char}(x, v) = (\text{parity } x) \cdot (\text{sign } v)$  (if  $v$  is understood we refer to *char*  $x$ );
- (4) *parity rule*: edges join pairs  $(x, v)$  of opposite character.

If two vertices have the same sign we say they are *parallel*, otherwise *antiparallel*.

Let  $G$  be such a graph. Let  $N(G)$  be a regular neighborhood of  $G$  in  $S^2$ . A *face*  $F$  of  $G$  is a component of  $S^2 - N(G)$ . Each component of  $\partial F$  gives rise (via the projection  $N(G) \rightarrow G$ ) to an alternating sequence of *corners* (the subarcs of  $\partial F$  that project to the vertices of  $G$ ) and edges of  $G$ . There are

no repetitions among the corners, but an edge may occur twice (once with each orientation).

We orient  $\partial F$  by giving a clockwise orientation to  $F$  (so each corner is oriented anticlockwise around its vertex).

For each vertex  $v$  of  $G$ , let  $L(v)$  be a set of labels at  $v$ .  $G(\{L(v)\})$ , the graph generated by  $\{L(v): v \text{ a vertex of } G\}$ , is defined to be the subgraph of  $G$  consisting of all edges  $e$  such that some end point of  $e$  is  $(l, v)$  where  $l \in L(v)$  and  $v$  is a vertex of  $G$ , together with all the vertices of  $G$ .

Let  $L$  be a set of labels. Define  $G(L)$  to be  $G(\{L(v)\})$  where  $L(v)$  is  $L$  for every vertex  $v$  of  $G$ . A label  $x$  is an *exceptional label* of  $G(L)$  if  $x \notin L$  but some end point of some edge of  $G(L)$  has label  $x$ . An  *$L$ -interval* is an interval on the boundary of an abstract fat vertex between adjacent labels of  $L$ .

An  *$m$ -type* is an ordered  $m$ -tuple  $(\epsilon_1, \epsilon_2, \dots, \epsilon_m) \in \{+, -\}^m$ . A *trivial  $m$ -type* is one for which  $\epsilon_1 = \epsilon_2 = \epsilon_3 = \dots = \epsilon_m$ .

Let  $l$  be the number of  $L$ -intervals. An  *$L$ -type* is an  $l$ -type in which each coordinate is associated to a distinct  $L$ -interval. If  $\mathbf{L}_0$  is a subset of  $L$ -intervals and  $\tau$  is an  $L$ -type we define  $\tau|_{\mathbf{L}_0}$  to be the  $|\mathbf{L}_0|$ -type obtained by restricting  $\tau$  to the coordinates associated to the elements of  $\mathbf{L}_0$ . A *trivial  $L$ -type* is one corresponding to a trivial  $l$ -type. Note that every corner of a face of  $G(L)$  belongs to a unique  $L$ -interval.

Let  $E$  be a disk face of  $G(L)$  and  $\tau$  be an  $L$ -type. Let  $\mathbf{L}_E$  be the set of  $L$ -intervals  $I$  with the property that some corner of  $E$  is contained in  $I$ . We say that  $E$  *represents*  $\tau$  if

- (1) for each  $L$ -interval  $I$  in  $\mathbf{L}_E$ , the vertices  $v$  of  $G$  such that a corner of  $E$  at  $v$  is contained in  $I$  all have the same sign, say  $\epsilon(I)$ ;
- (2)  $(\epsilon(I): I \in \mathbf{L}_E) = \pm\tau|_{\mathbf{L}_E}$ .

$G(L)$  *represents*  $\tau$  if and only if some disk face of  $G(L)$  represents  $\tau$ .

An  *$x$ -cycle* in  $G$ ,  $\Sigma$ , is a cycle of edges in  $G$  such that

- (1) if we think of the vertices of  $G$  as points,  $\Sigma$  is homeomorphic to a circle;
- (2) we can orient this circle so that the tail of each edge (under the induced orientation) is labeled  $x$ ;
- (3) the vertices of  $G$  in  $\Sigma$  are all parallel.

A *Scharlemann cycle* in  $G$  is an  $x$ -cycle  $\Sigma$  in  $G$  for some label  $x$  such that  $\Sigma$  bounds a disk  $D$  with  $G \cap \text{int} D = \emptyset$ . Since  $G$  contains no trivial loops, a Scharlemann cycle has at least two edges. Furthermore, notice that  $D$  represents all, say,  $\{1, \dots, q\}$ -types where  $\{1, \dots, q\}$  is the set of labels of  $G$ .

The goal of the present section is to prove the following proposition.

**Proposition 2.0.1.** *Let  $G_p$  and  $G_q$  be a pair of graphs as described above. Either  $G_q$  contains a Scharlemann cycle or  $G_p$  represents all  $\{1, \dots, q\}$ -types.*

Note that  $G_p$  has no exceptional labels so the generality of the above definition of representing an  $L$ -type is unnecessary in the statement of Proposition 2.0.1. However the proof of the proposition will require this generality.

A *great  $x$ -cycle* is an  $x$ -cycle such that all the vertices on one side of the  $x$ -cycle in  $S^2$  are the same sign as those in the  $x$ -cycle.

**Lemma 2.0.2.** *If  $G$  contains a great  $x$ -cycle for some label  $x$ , then it contains a Scharlemann cycle.*

*Proof.* See [CGLS, Lemma 2.6.2].

**2.1. Stars.** We consider two model fat vertices  $V_{\pm}$ , with  $\text{sign } V_{\pm} = \pm$  (see Figure 17).

The set of labels around  $V_{\pm}$  is  $\{1, \dots, q\}$ . Thus  $V_- = -V_+$ , where  $-$  denotes reflection.

A *star  $X$*  is an ordered triple  $(V(X), L(X), \omega(X))$ , where  $V(X) = V_{\pm}$ ,  $L(X)$  is a subset of the labels around  $V(X)$ , and  $\omega(X)$  is a *dual orientation* (indicated by an arrow denoting either a *sink* or *source*) on each  $L(X)$ -interval (see Figure 18).

The *sign* of  $X$  is the sign of  $V(X)$ . Let  $\tau = (\varepsilon_1, \dots, \varepsilon_{|L(X)|})$  be an  $L(X)$ -type. If there is a way of equating the two dual orientations with the signs  $+$  and  $-$  (i.e.  $\{\text{sink}, \text{source}\} = \{+, -\}$ ) such that  $\omega(X) = \tau$  (under the correspondence between the  $L(X)$ -intervals and the coordinates of  $\tau$ ) then we say that  $X$  *represents*  $\tau$  and write  $[X] = \tau$ .

We shall use the term *right (left)* to refer to the direction defined by proceeding around  $\partial V(X)$  in an *anticlockwise (clockwise)* direction.

A label  $x \in L(X)$  is a *clockwise switch* if  $\omega(X)$  is a sink on the  $L(X)$ -interval immediately to the left of  $x$  and a source on the  $L(X)$ -interval immediately to the right of  $x$ . Similarly for *anticlockwise switches* (see Figure 19).

$C(X)$  and  $A(X)$  denote the subsets of  $L(X)$  consisting of the clockwise and anticlockwise switches of  $X$ , respectively.  $S(X) = C(X) \cup A(X)$  is the set of switches of  $X$ . Note that the clockwise and anticlockwise switches alternate as we go around  $\partial V(X)$ .

$\bar{X} = X$  with  $\omega(X)$  reversed. Thus  $C(\bar{X}) = A(X)$ ,  $A(\bar{X}) = C(X)$ .

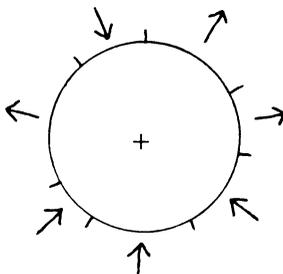


FIGURE 18

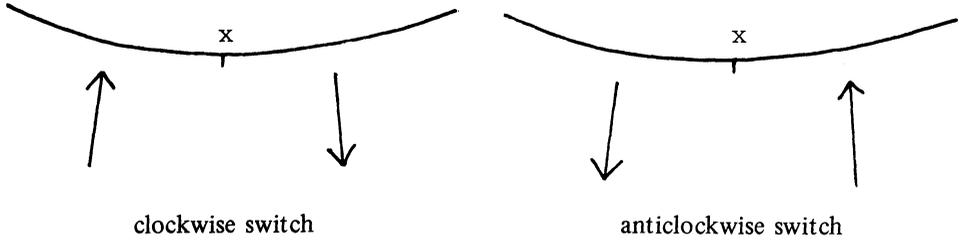


FIGURE 19

$-X$  denotes the reflection of  $X$  with  $\omega(X)$  reversed. Thus (as subsets of  $\{1, \dots, q\}$ ),  $C(-X) = C(X)$ ,  $A(-X) = A(X)$ .

We define the *positive* and *negative (clockwise) derivatives* of  $X$  by

$$d^\pm(X) = (V(X), C(X), \omega(d^\pm X))$$

where  $\omega(d^\pm X)$  on a  $C(X)$ -interval is determined by the character of the element  $a \in A(X)$  in that interval. Specifically (writing  $\text{char } a = \text{char}(a, V(X))$ ),

$$\begin{aligned} \omega(d^+ X) \text{ is a source if } \text{char } a = +, & \quad \omega(d^+ X) \text{ is a sink if } \text{char } a = -; \\ \omega(d^- X) \text{ is a source if } \text{char } a = -, & \quad \omega(d^- X) \text{ is a sink if } \text{char } a = +. \end{aligned}$$

We will use  $d$  to denote either  $d^+$  or  $d^-$ . Note that  $d^- X = \overline{(d^+ X)}$  and  $d(-X) = -dX$ . There is also a relative notion. Let  $L_0$  be a subset of the labels  $\{1, \dots, q\}$  around  $V(X)$ . Then the  $(\pm)$ -*derivative of  $X$  relative to  $L_0$*  is

$$d_{L_0} X = d_0 X = (V(X), C(X) \cup L_0, \omega(d_0 X))$$

where  $\omega(d_0 X)$  on a  $(C(X) \cup L_0)$ -interval  $I$  is defined as follows:

- (i) if there is an element  $a$  of  $A(X)$  in  $\text{int } I$ , let  $\omega(d_0 X)$  be determined by  $\text{char } a$ , exactly as for  $\omega(dX)$ ;
- (ii) if there is no element of  $A(X)$  in  $\text{int } I$ , then  $\omega(X)$  induces the same orientation on all  $L(X) \cup L_0$ -intervals contained in  $I$ ; let  $\omega(d_0 X)$  on  $I$  be that orientation.

In particular,  $d = d_\emptyset$ . Note also that if  $C(X) = C(Y)$  and  $A(X) = A(Y)$  (and  $V(X) = V(Y)$ ) then  $d_0 X = d_0 Y$ .

The following observation will be useful in §2.8.

**Lemma 2.1.1.** *If  $L_1$  and  $L_2$  differ only by elements of  $C(X)$  then  $d_{L_1} X = d_{L_2} X$ .*

*Proof.* It is enough to prove the lemma for the case  $L_2 = L_1 \cup \{c\}$ ,  $c \in C(X)$ . But this is clear from the definition.  $\square$

Again let  $X$  be a star and  $L_0 \subset \{1, \dots, q\}$ . Let  $\tilde{A}(X)$  denote  $A(X) - L_0$ . The main result of this subsection is the following proposition.

**Proposition 2.1.2.** *Let  $D$  be any composition of  $d^+$ 's and  $d^-$ 's, and  $D_0$  the corresponding composition of  $d_0^+$ 's and  $d_0^-$ 's. Then*

$$C(DX) \subset C(D_0X), \text{ and } \tilde{A}(DX) \supset \tilde{A}(D_0X).$$

The proof of Proposition 2.1.2 will involve a series of lemmas.

**Lemma 2.1.3.**  $C(d_0X) \subset C(X)$ .

*Proof.* Suppose  $x \in C(d_0X)$ . If  $x \notin C(X)$ , then  $x \in L_0$ .

Consider  $\omega(X)$  immediately to the right of  $x$ , and suppose it is a sink (see Figure 20). Let  $y$  be the first element of  $S(X) \cup L_0$  we encounter to the right of  $x$ . Then either  $y \in L_0$ , in which case  $\omega(d_0X)$  is a sink immediately to the right of  $x$ , or  $y \in C(X)$ , in which case  $\omega(d_0X)$  is also a sink immediately to the right of  $x$ . Similarly, immediately on the left of  $x$ , a source in  $\omega(X)$  will give rise to a source in  $\omega(d_0X)$ . Thus, if  $x \in C(d_0X)$ , we must have  $x \in C(X)$  (see Figure 21).  $\square$

We now want to compare  $dX$  with  $d_0X$ .

**Lemma 2.1.4.** *Immediately to the right (left) of an element of  $C(X)$ ,  $\omega(d_0X)$  either agrees with  $\omega(dX)$  or is a source (resp. sink).*

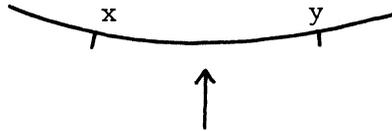


FIGURE 20

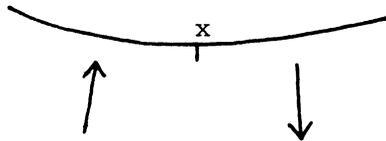


FIGURE 21

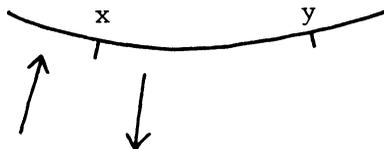


FIGURE 22

*Proof.* Suppose  $x \in C(X)$  and let  $y$  be the first element of  $S(X) \cup L_0$  we encounter to the right of  $x$  (see Figure 22).

Then either  $y \in \tilde{A}(X)$ , in which case  $\omega(d_0X)$  and  $\omega(dX)$  agree immediately to the right of  $x$ , or  $y \in L_0$ , in which case  $\omega(d_0X)$  is a source immediately to the right of  $x$ . We argue similarly to the left of  $x$ .  $\square$

We now want to compare  $d_0X$  with  $d_0Y$  for certain related pairs  $X, Y$ .  
(In the following lemmas, we assume that  $V(X) = V(Y)$ .)

**Lemma 2.1.5.** *Suppose  $C(X) \subset C(Y)$  and  $\tilde{A}(X) \supset \tilde{A}(Y)$ . Then immediately to the right (left) of an element of  $C(X)$ ,  $\omega(d_0Y)$  either agrees with  $\omega(d_0X)$  or is a source (resp. sink).*

*Proof.* Suppose  $x \in C(X) \subset C(Y)$ .

Let  $a$  be the first element of  $S(X) \cup S(Y) \cup L_0$  we encounter to the right of  $x$  (see Figure 23). There are three possibilities.

- (i)  $a \in L_0$ . Then  $\omega(d_0X)$  and  $\omega(d_0Y)$  are both sources immediately to the right of  $x$ .
- (ii)  $a \in \tilde{A}(Y) (\subset \tilde{A}(X))$ . Then  $\omega(d_0X)$  and  $\omega(d_0Y)$  immediately to the right of  $x$  are both determined by  $\text{char } a$ , and hence agree.
- (iii)  $a \in \tilde{A}(X)$ ,  $a \notin \tilde{A}(Y)$ . Let  $b$  be the first element of  $S(Y) \cup L_0$  we encounter to the right of  $a$ . There are two possibilities.
  - (1)  $b \in L_0$ . Then  $\omega(d_0Y)$  is a source immediately to the right of  $x$ .
  - (2)  $b \in \tilde{A}(Y)$ . Then  $b \in \tilde{A}(X)$  (since  $\tilde{A}(Y) \subset \tilde{A}(X)$ ) and thus there is a  $c \in C(X)$  between  $a$  and  $b$ . But  $c \in C(Y)$  (since  $C(X) \subset C(Y)$ ), contradicting the choice of  $b$ .

The argument to the left of  $x$  is similar.  $\square$

**Lemma 2.1.6.** *Suppose  $C(X) \subset C(Y)$  and  $\tilde{A}(X) \supset \tilde{A}(Y)$ . Then  $\tilde{A}(d_0Y) \subset C(X)$ .*

*Proof.* Suppose  $x \in \tilde{A}(d_0Y)$ . Thus  $x \in C(Y)$ , and the dual orientations on both sides of  $x$  get reversed in going from  $Y$  to  $d_0Y$ . Thus the first element in  $S(Y) \cup L_0$  we meet to the right of  $x$  is  $b$ , say, in  $\tilde{A}(Y)$ , and the first element of  $S(Y) \cup L_0$  we meet to the left of  $x$  is  $a$ , say, in  $\tilde{A}(Y)$ . Thus  $a, b \in \tilde{A}(X)$  and hence there is a  $c \in C(X)$  between  $a$  and  $b$ . It follows that  $c \in C(Y)$ , so  $c = x$ . Thus  $x \in C(X)$ .  $\square$

We can now prove the two main lemmas we need.

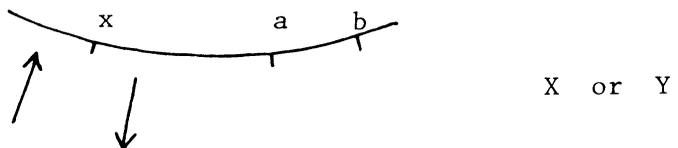


FIGURE 23

**Lemma 2.1.7.**  $C(dX) \subset C(d_0X)$ , and  $\tilde{A}(dX) \supset \tilde{A}(d_0X)$ .

*Proof.* Both statements follow from Lemma 2.1.4.  $\square$

**Lemma 2.1.8.** Suppose  $C(X) \subset C(Y)$  and  $\tilde{A}(X) \supset \tilde{A}(Y)$ . Then  $C(d_0X) \subset C(d_0Y)$ , and  $\tilde{A}(d_0X) \supset \tilde{A}(d_0Y)$ .

*Proof.* By Lemma 2.1.6,  $\tilde{A}(d_0Y) \subset C(X)$ , and by Lemma 2.1.3,  $C(d_0X) \subset C(X)$ . Both statements now follow from Lemma 2.1.5.  $\square$

*Proof of Proposition 2.1.2.* This follows from Lemmas 2.1.7 and 2.1.8 by induction on the number of  $d^\pm$ 's in  $D$ .  $\square$

**2.2. Graphs with dual orientation.** A *graph with dual orientation* (g.d.o.) is a pair  $\Gamma = (G(\Gamma), \omega(\Gamma))$ , where  $G(\Gamma)$  is a subgraph of a graph  $G$  of the type described in §2.0 and  $\omega(\Gamma)$  is an orientation (sink or source) on each corner of each face of  $G(\Gamma)$ .

We often suppress reference to  $G(\Gamma)$  and refer to edges, faces, etc. of  $\Gamma$ .

Given a g.d.o.  $\Gamma$ , the *dual graph*  $\Gamma^*$  is defined as follows. For each disk face  $F$  of  $\Gamma$ , choose a *dual vertex*  $v \in \text{int } F$ . Then  $\{\text{vertices of } \Gamma^*\} = \{\text{vertices of } \Gamma\} \cup \{\text{dual vertices}\}$ . For each corner of  $F$  at a (fat) vertex  $V$ , put in an edge of  $\Gamma^*$  joining  $V$  to  $v$ . Finally, we orient the edges of  $\Gamma^*$  according to the dual orientation  $\omega(\Gamma)$ . Thus  $\Gamma^*$  is a directed graph.

*Notes.* (1)  $\Gamma^*$  is uniquely determined (up to isotopy) by  $\Gamma$ .

(2) We could define  $\Gamma^*$  (nonuniquely) by putting in a dual vertex for every face of  $\Gamma$ . However, we choose not to do this.

(3) Every face of  $\Gamma$  is a disk if and only if  $\Gamma$  is connected.

Let  $\Gamma$  be a g.d.o. Then  $\Gamma|v$  is a star for all vertices  $v$ . Conversely, given a star  $X(v)$  at each vertex  $v$  of  $G$ , we have the *g.d.o. generated by*  $\{X(v)\}$ ,

$$\Gamma(\{X(v)\}) = (G(\{L(X(v))\}), \{\omega(X(v))\}),$$

where if  $c$  is a corner of a face of  $G(\{L(X(v))\})$  at a vertex  $v$  such that  $c$  is properly contained in an  $L(X(v))$ -interval, then assign to  $c$  the orientation defined on that  $L(X(v))$ -interval by  $\omega(X(v))$ .

In particular, if  $X$  is a star, we define

$$\Gamma(X) = \Gamma(\{X(v)\}) \quad \text{where } X(v) = \begin{cases} X, & \text{if } \text{sign } v = \text{sign } X, \\ -X, & \text{if } \text{sign } v = -\text{sign } X. \end{cases}$$

Thus  $G(\Gamma(X)) = G(L(X))$ .

Let  $L$  be a set of labels with  $|L| \geq 2$  and  $\tau$  be an  $L$ -type. Let  $T$  be a star with  $[T] = \tau$ . (Note that there are exactly four such stars,  $T$ ,  $\bar{T}$ ,  $-T$ , and  $-\bar{T}$ .)

The following lemma follows immediately from the definitions.

**Lemma 2.2.1.**  $G(L)$  represents  $\tau$  if and only if  $\Gamma(T)^*$  has a sink or source at a dual vertex.

The notion of *derivative* for stars extends naturally to g.d.o.'s, i.e. we define  $\delta\Gamma = \Gamma(\{d(\Gamma|v)\})$ , and if  $G_0$  is a subgraph of  $G(\Gamma)$ ,

$$\delta_{G_0}\Gamma = \delta_0\Gamma = \Gamma(\{d_{G_0|v}(\Gamma|v)\}),$$

(where  $G_0|v$  denotes the set of labels of endpoints of edges of  $G_0$  at  $v$ ). The following lemma will be needed in §2.8.

**Lemma 2.2.2.** *If the exceptional labels of  $G(L_0)$  are contained in  $C(X)$ , then  $\delta_0\Gamma(X) = \Gamma(d_0X)$ . (Here  $d_0 = d_{L_0}, \delta_0 = \delta_{G(L_0)}$ .)*

*Proof.* For all vertices,  $v$ ,  $C(\Gamma(X)|v) = C(X)$ , and  $A(\Gamma(X)|v) = A(X)$ , hence  $\delta_0\Gamma(X) = \Gamma(\{d_{G_0|v}X\})$  ( $G_0 = G(L_0)$ ). But by Lemma 2.1.1,  $d_{G_0|v}X = d_{L_0}X = d_0X$ . Therefore  $\delta_0\Gamma(X) = \Gamma(d_0X)$ .  $\square$

**2.3. Index.** Let  $\Delta$  be a directed graph in  $S^2$ . A *switch* at a vertex  $v$  of the graph is a pair of adjacent edges incident to  $v$  whose orientations are opposite at  $v$  (Figure 24).

A *switch* around a face  $F$  of the graph is a pair of adjacent edges of  $\partial F$  incident to the vertex  $v$ , say, whose orientations agree at  $v$  (Figure 25).

The *index* of a vertex  $I(v) = 1 - s(v)/2$  where  $s(v)$  is the number of switches at  $v$ . The *index* of a face  $I(F) = \chi(F) - s(F)/2$  where  $s(F)$  is the number of switches around  $F$ . The following lemma and its proof are taken from [G1].

**Lemma 2.3.1.**  $\sum_{vertices} I(v) + \sum_{faces} I(F) = 2$ .

*Proof.* Let  $V$  and  $E$  be the number of vertices and edges, respectively, of the graph  $\Delta$ .

$$\text{LHS} = V + \sum_F \chi(F) - \frac{1}{2} \left( \sum_v s(v) + \sum_F s(F) \right).$$



FIGURE 24

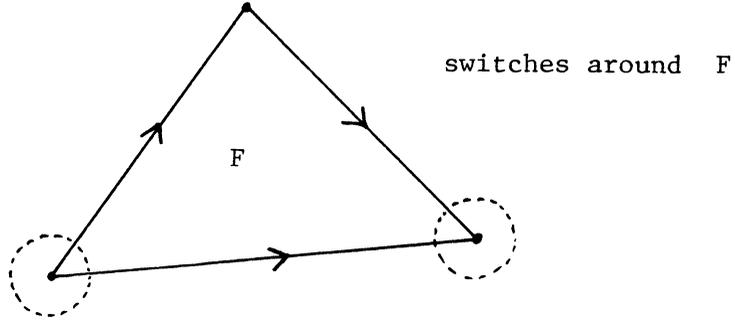


FIGURE 25

Note that each corner contributes 1 to exactly one of  $\sum_v s(v)$  or  $\sum_F s(F)$  (Figure 25). That is,

$$\sum_v s(v) + \sum_F s(F) = \text{the number of corners} = 2E.$$

Thus LHS above =  $V - E + \sum_F \chi(F) = \chi(S^2) = 2$ .  $\square$

Let  $L$  be a set of labels with  $|L| \geq 2$  and  $\tau$  a nontrivial  $L$ -type. Let  $T$  be a star with  $[T] = \tau$  and  $\Gamma = \Gamma(T)$ . Let  $p$  = the number of fat vertices of  $G$ .

A *switch-edge* is an edge of  $\Gamma$  whose endpoints are both in  $C(T)$  or both in  $A(T)$ . Let  $i = |S(T)|/2 - 1$  ( $\geq 0$ ),  $s$  be the number of switch-edges of  $\Gamma$ , and  $r$  be the number of disk faces of  $G(L)$  representing  $\tau$ .

**Lemma 2.3.2.**  $r + s \geq ip + 2$ .

*Proof.* Let  $\Lambda$  be a minimal connected graph containing  $\Gamma$ . The dual orientation of  $\Gamma$  induces a dual orientation on  $\Lambda$ . All faces of  $\Lambda$  are disks; let  $\Lambda^*$  be the (directed) dual graph of  $\Lambda$ .

$$\{\text{vertices of } \Lambda^*\} = V_1 \sqcup V_2 \sqcup V_3, \text{ where}$$

$$V_1 = \{\text{dual vertices of } \Lambda^* \text{ corresponding to disk faces of } \Gamma\},$$

$$V_2 = \{\text{dual vertices of } \Lambda^* \text{ corresponding to nondisk faces of } \Gamma\}, \text{ and}$$

$$V_3 = \{\text{fat vertices}\}.$$

$$\{\text{faces of } \Lambda^*\} \leftrightarrow \{\text{edges of } \Lambda\} = E_1 \cup E_2, \text{ where}$$

$$E_1 = \{\text{edges of } \Gamma\}, E_2 = \{\text{edges of } \Lambda - \Gamma\}.$$

For an edge  $e$  of  $\Lambda$ , write  $I(e)$  for the index of the corresponding face of  $\Lambda^*$ . Applying Lemma 2.3.1 to  $\Lambda^*$  gives

$$\sum_{V_1} I(v) + \sum_{V_2} I(v) + \sum_{V_3} I(v) + \sum_{E_1} I(e) + \sum_{E_2} I(e) = 2.$$

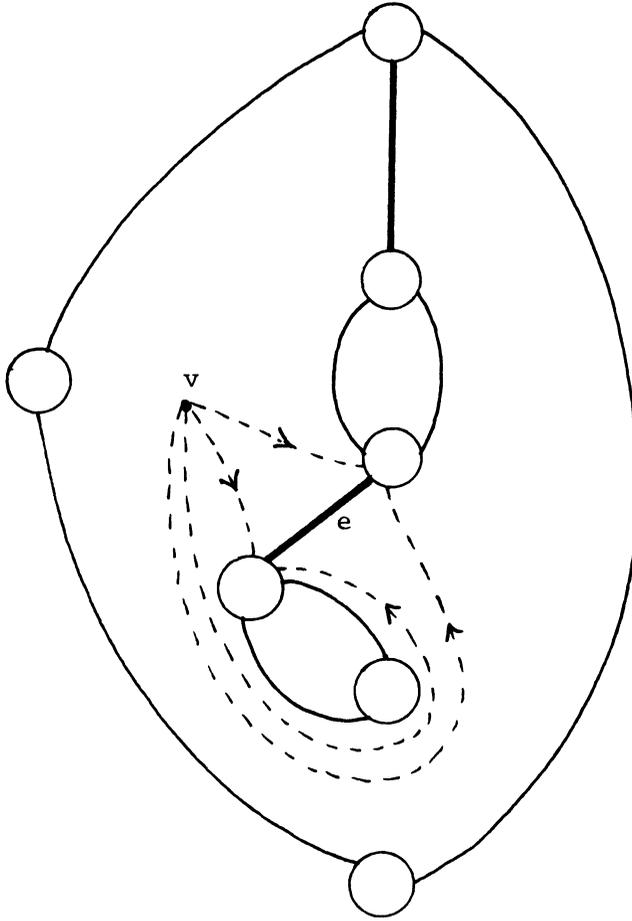


FIGURE 26

Note that

$$\sum_{V_1} I(v) = -ip;$$

$$\{v \in V_1 : I(v) = 1\} \leftrightarrow \{\text{disk faces of } G(L) \text{ representing } \tau\};$$

$$\{e \in E_1 : I(e) = 1\} = \{\text{switch-edges of } \Gamma\}.$$

Hence the lemma will follow if we show

$$(*) \quad \sum_{V_2} I(v) + \sum_{E_2} I(e) \leq 0.$$

To do this, first note that  $e \in E_2$  implies that  $I(e) \leq 0$  since at each end of  $e$  the dual orientations on either side of  $e$  agree (by definition of the induced dual orientation on  $\Lambda$ ). Thus if  $(*)$  fails there is a nondisk face  $F$  of  $\Gamma$  such that the corresponding dual vertex  $v$  of  $\Lambda^*$  has  $I(v) = 1$ . But then every  $e \subset F$  such that  $e \in E_2$  has index  $-1$  (see Figure 26), and there is at least one such edge in  $F$ . Therefore  $(*)$  holds here also.  $\square$

We have talked about the corners of a face of a graph. We now want to refer to the corners of a face of a subgraph of a g.d.o. We therefore make the following formal definition. Recall the model fat vertices  $V_+$  and  $V_-$  in the definition of a star. A *corner*  $X$  is  $(V(X), I(X), L(X), \omega(X))$  where  $V(X)$  is  $V_\pm$ ;  $I(X)$  is an interval at  $V(X)$ ;  $L(X)$  a subset of the labels in  $I(X)$ ;  $\partial I(X) \subset L(X)$ ; and  $\omega(X)$  a dual orientation on the  $L(X)$ -intervals in  $I(X)$ . As for stars, we have  $\bar{X}$  and  $-X$ ;  $C(X)$  and  $A(X)$  (switches in  $\text{int } I(X)$ ).  $Y$  is a *subcorner* of  $X$  if  $V(Y) = V(X)$ ,  $I(Y) \subset I(X)$ ,  $L(Y) = L(X) \cap I(Y)$ , and  $\omega(X) = \omega(Y)|I(Y)$ .

Let  $\Gamma$  be a g.d.o., and  $F$  a disk face of a subgraph of  $\Gamma$ . Thus  $\partial F$  can be expressed as a sequence of *edges*  $e$  of  $\Gamma$  and *corners*  $X$ . For a corner  $X$ , define  $\text{ind } X = \text{ind}_\Gamma X = 1 - s(X)$  where  $s(X)$  is the number of switches of  $\omega(\Gamma)|X$ . For an edge  $e$  of  $\partial F$ , define  $\text{ind } e = \text{ind}_\Gamma e$  to be  $-1$  or  $0$  according as the restrictions of  $\omega(\Gamma)$  immediately inside  $F$  at the two ends of  $e$  agree or disagree (see Figure 27).

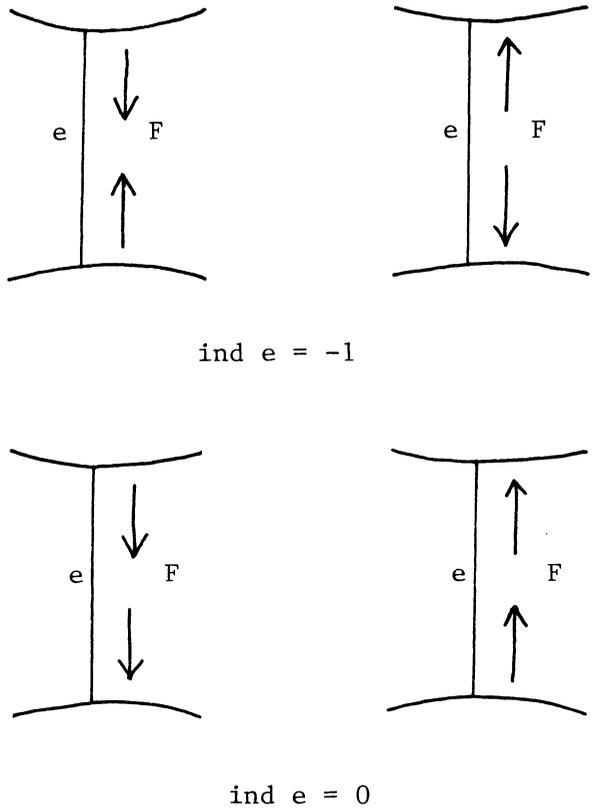


FIGURE 27

Finally, define

$$\text{index } \partial F = \text{ind}_\Gamma \partial F = \sum_{X \text{ a corner of } F} \text{ind } X + \sum_{e \subset \partial F} \text{ind } e.$$

Let  $\Gamma$  be a graph with dual orientation.

**Lemma 2.3.3.** *Let  $F$  be a disk face of a subgraph of  $\Gamma$  containing no vertices of  $\Gamma$  in its interior and such that  $\text{ind}_\Gamma \partial F \leq 0$ . Then  $\Gamma^*$  has a face of positive index or a dual vertex sink or source in  $F$ .*

*Proof.* Let  $2F$  denote the double of  $F$ ,  $= F \cup_{\partial F} -F$ , and let  $2\Gamma^*$  denote the double of  $\Gamma^* \subset 2F$ . Applying Lemma 2.3.1 to  $2\Gamma^* \subset 2F$  gives

$$2 \sum_{v \text{ dual vertex of } \Gamma^* \text{ in } F} I(v) + 2 \sum_{f \text{ face of } \Gamma^* \text{ in int } F} I(f) + \sum_{X \text{ corner of } F} \text{ind}_\Gamma X + \sum_{e \text{ edge of } \Gamma \subset \partial F} \text{ind}_\Gamma e = 2.$$

In other words,

$$2 \sum I(v) + 2 \sum I(f) + \text{ind}_\Gamma \partial F = 2.$$

Thus  $\text{ind}_\Gamma \partial F \leq 0$  implies  $\sum I(v) + \sum I(f) > 0$ , hence the result.  $\square$

Note that it follows from the above proof that  $\text{index } \partial F$  is always even.

**2.4. Pushing back sinks/sources.** Let  $G$  be a graph as in §2.0 and  $G_0$  a subgraph. Let  $\Gamma$  be a graph with dual orientation with  $G_0 \subset G(\Gamma) \subset G$ . Set  $\delta_0 = \delta_{G_0}$ .

**Lemma 2.4.1.** *If  $(\delta_0 \Gamma)^*$  has a sink or source at a dual vertex corresponding to a face  $E$  of  $\delta_0 \Gamma$ , then  $\Gamma^*$  has a sink or source at a dual vertex corresponding to a face of  $\Gamma$  contained in  $E$ .*

*Proof.*  $E$  is a disk face of  $\delta_0 \Gamma$  corresponding to a sink or source of  $(\delta_0 \Gamma)^*$ .

*Claim 1.*  $\Gamma^* \cap E$  contains no cycle that is a face of  $\Gamma^*$ .

*Proof.* Such a clockwise cycle would correspond to an edge of  $\delta_0 \Gamma$  in  $\text{int } E$ , giving a contradiction.

An anticlockwise cycle which is a face of  $\Gamma^*$  would correspond to an edge of  $\Gamma$  joining two anticlockwise switches of  $\Gamma$ ,  $a_1$  and  $a_2$  (Figure 28).

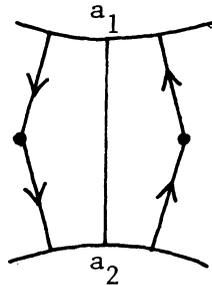


FIGURE 28

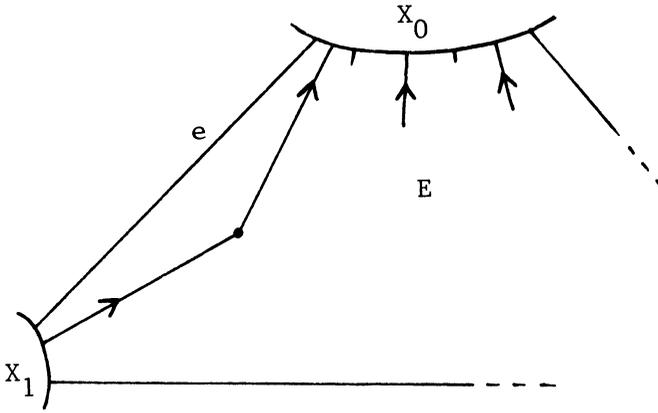


FIGURE 29

Thus  $a_1$  and  $a_2$  have distinct characters, and the corners of  $E$  containing  $a_1$  and  $a_2$  would acquire distinct orientations in  $\delta_0\Gamma$ , contradicting the fact that  $E$  corresponds to a sink or source of  $(\delta_0\Gamma)^*$ .  $\square$  (Claim 1)

*Claim 2.*  $\text{ind}_\Gamma \partial E \leq 0$ .

*Proof.* Suppose  $E$  corresponds to a source of  $(\delta_0\Gamma)^*$ . We show

$$\text{number of sink/source corners of } E \leq \text{number of edges of index } -1 \text{ in } \partial E$$

(all with respect to  $\Gamma$ ).

(a) Let  $X_0$  be a sink corner of  $E$  with respect to  $\Gamma$ . Let  $e$  be the edge of  $\partial E$  immediately anticlockwise around  $\partial E$  from  $X_0$ . Suppose  $\text{ind}_\Gamma e = 0$  (as shown in Figure 29). Then no endpoint of  $e$  is a clockwise switch. Hence  $e \subset G_0$ .

Now consider the corner  $X_1$  of  $E$  at the other end of  $e$ . If  $\Gamma$  had a switch in  $\text{int } X_1$ , it would have a clockwise switch, contradicting the fact that  $E$  is a face of  $\delta_0\Gamma$ . Thus  $X_1$  is a source corner for  $\Gamma$ . Since  $e \subset G_0$ ,  $X_1$  becomes a source corner for  $\delta_0\Gamma$ . This contradicts the fact that  $E$  corresponds to a source of  $(\delta_0\Gamma)^*$ . Therefore  $\text{ind } e = -1$ .

(b) Let  $X_0$  be a source corner of  $E$  with respect to  $\Gamma$ . Let  $e$  be the edge of  $\partial E$  immediately clockwise from  $X_0$  around  $\partial E$ . Suppose  $\text{ind } e = 0$  (Figure 30). Then  $e \subset G_0$ .

Thus as in (a) for  $X_1$ ,  $X_0$  would become a source corner for  $\delta_0\Gamma$ . Thus  $\text{ind } e = -1$ .

So the edge immediately clockwise from a source corner has index  $-1$  and the edge immediately anticlockwise from a sink corner has index  $-1$ . This gives a function  $\{\text{sink/source corners}\} \mapsto \{\text{index } -1 \text{ edges}\}$ , and this function is 1-1 (e.g., Figure 31). Thus  $\text{ind}_\Gamma \partial E \leq 0$ .

If  $E$  corresponds to a sink of  $(\delta_0\Gamma)^*$  the argument is similar.  $\square$  (Claim 2)

Claims 1 and 2 along with Lemma 2.3.3 now prove Lemma 2.4.1.  $\square$

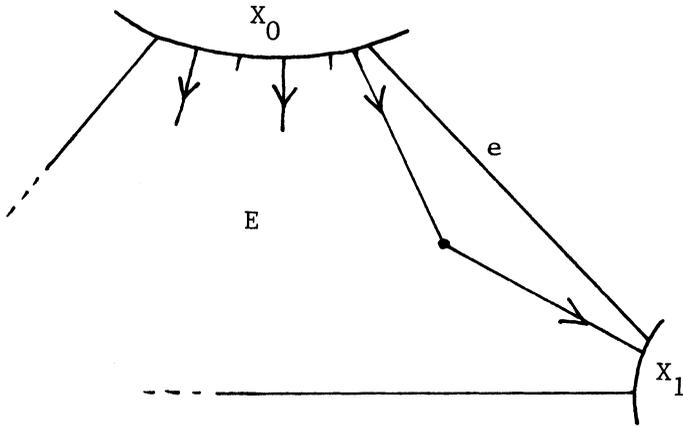


FIGURE 30

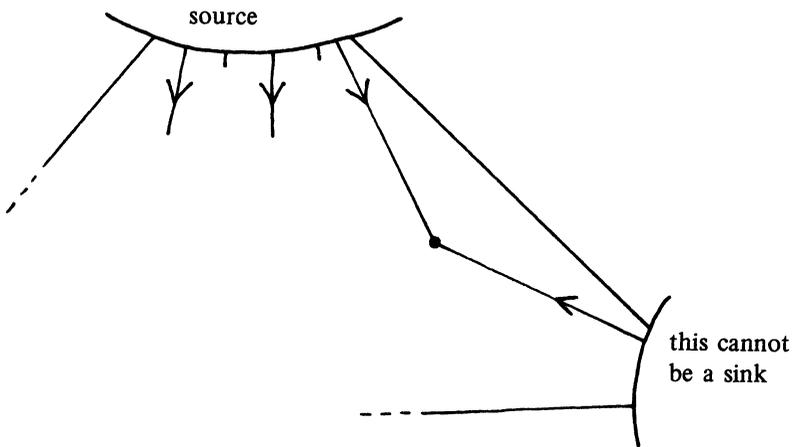


FIGURE 31

**Corollary 2.4.2.** *Let  $\tau$  be an  $L$ -type and  $T$  a star with  $[T] = \tau$ . If  $G(C(T))$  represents  $[dT]$ , then  $G(L)$  represents  $\tau$ .*

*Proof.* This follows by applying Lemma 2.4.1 with  $\Gamma = \Gamma(T)$ , and  $G_0 = \emptyset$ .  $\square$

**2.5.  $x$ -cycle existence.** Recall that  $G_P$  and  $G_Q$  are the graphs of intersection of planar surfaces  $P$  and  $Q$  in  $X$ . Let  $L$  be a set of labels with  $|L| \geq 2$  and  $\tau$  an  $L$ -type (nontrivial). Let  $T$  be a star with  $L(T) = L$  and  $[T] = \tau$ . Set  $G = G_P$ .

**Lemma 2.5.1.** *Suppose that*

- (i) *all elements of  $C(T)$  have the same parity;*
- (ii) *all elements of  $A(T)$  have the same parity; and*
- (iii)  *$G(L)$  does not represent  $\tau$ .*

Then there exists an  $x$ -cycle  $\Sigma$  in  $G_Q$  such that the set of vertices of  $\Sigma$  is a subset of either  $C(T)$  or  $A(T)$ .

*Proof.* Let  $i = |S(T)|/2 - 1$ ;  $s =$  number of switch-edges of  $\Gamma(T)$ . Then by (iii) and Lemma 2.3.2,  $s \geq ip + 2$  ( $p = |\{\text{vertices of } G\}|$ ). Hence there are greater than  $ip/2$  edges of  $G(L)$  all of whose endpoints are in (say)  $C(T)$ . Since  $|C(T)| = i + 1$ , there exists a vertex  $x$  of  $G$  such that every label  $c \in C(T)$  at  $x$  is the endpoint of a switch-edge  $e(c)$  whose other endpoint is  $\alpha(c) \in C(T)$ , where  $\alpha$  is some function  $C(T) \rightarrow C(T)$ . Let  $c_0, \alpha(c_0), \dots, \alpha^{m-1}(c_0)$ ,  $m \geq 1$ , be an orbit of  $\alpha$  (thus  $\alpha^m(c_0) = c_0, \alpha^i(c_0) \neq c_0, 0 < i < m$ ). Recall that all the elements of  $C(T)$  have the same parity. Then the edges  $e(c_0), e(\alpha(c_0)), \dots, e(\alpha^{m-1}(c_0))$  in  $G$  correspond to edges in  $G_Q$  which form an  $x$ -cycle with vertices  $c_0, \alpha(c_0), \dots, \alpha^{m-1}(c_0)$ .  $\square$

**2.6. Good corners.** Recall the abstract definition of corner in §2.3.

We partition the set of all corners  $\mathbf{C}$  into three mutually disjoint subsets:  $\mathbf{C} = \mathbf{G} \sqcup \mathbf{B} \sqcup \mathbf{U}$ , the good, bad, and ugly, as follows.

First, a corner  $X$  is ugly if and only if it fails to satisfy the condition

(\*)  $\text{all elements of } A(X) \text{ have the same parity.}$

Note that  $X$  satisfies (\*) if and only if  $-X$  satisfies (\*).

If  $X$  is not ugly we will write  $\text{char } A(X)$  for  $\text{char}(a, V(X))$  for any  $a \in A(X)$ .

We define good and bad relative to the following choices. We choose a clockwise character  $\eta_c (= \pm)$ , and an anticlockwise character  $\eta_a (= \pm)$ .

The graph  $G = G_p$  is the graph of  $P \cap Q$  in  $P$  where  $P$  and  $Q$  are two-sided surfaces. In particular, a small neighborhood of  $Q$  is divided into, say, black and white sides. Thus each edge  $e$  of  $G$  divides a small neighborhood of  $e$  in  $S^2$  into black and white sides. In particular, each pair  $(l, v)$  at vertex  $v$  locally separates white and black regions at  $v$ : as we go from left to right along  $v$  at  $(l, v)$  we either go from black to white or from white to black according to  $\text{char}(l, v)$ . We choose the B/W shading of the sides so that a pair  $(l, v)$  of character  $\eta_c$  is WB (going from left to right, see Figure 32) and will refer to pairs  $(l, v)$  as BW or WB.

We will call a corner,  $X$ , BW if the leftmost label of  $X$  ( $\in \partial I(X)$ ) is WB (i.e., has character  $\eta_c$ ) and the rightmost label is WB (e.g. Figure 33(c), (d)). Similarly, we define WB, BB, and WW corners (Figure 33(a) is BB, (b) is WW).

An atom is a corner with no switches. First we define good and bad for atoms.

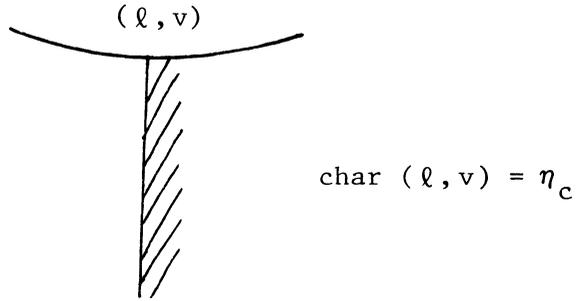


FIGURE 32

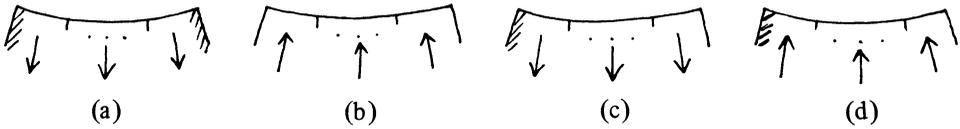


FIGURE 33

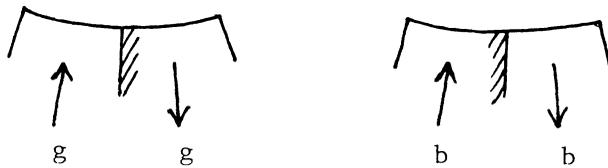


FIGURE 34

(a) The atoms of Figure 33 are good. The others are bad. Note that an atom  $X$  is good if and only if  $-X$  is bad.

We pause to make the following observation.

(\*\*) *Atoms on either side of a clockwise switch  $(l, v)$  are either both good or both bad (Figure 34). In particular, both sides are good if and only if  $\text{char}(l, v) = \eta_c$ .*

Now the general definition.

(b) If  $\text{char } A(X) = \eta_a$ , then  $X$  is good if and only if all maximal atoms in  $X$  are good.

(c) If  $\text{char } A(X) = -\eta_a$ , then  $X$  is good if and only if some maximal atom in  $X$  is good.

Finally, we say that  $X$  is bad if it is neither good nor ugly.

Note that (by (\*\*)) if a corner  $X$  has  $A(X) = \emptyset$  then all maximal atoms in  $X$  are good if and only if some maximal atom in  $X$  is good, so (b) and (c) are consistent. Also, since  $\text{char } A(-X) = -\text{char } A(X)$ , we clearly have that  $X$  is good if and only if  $-X$  is bad.

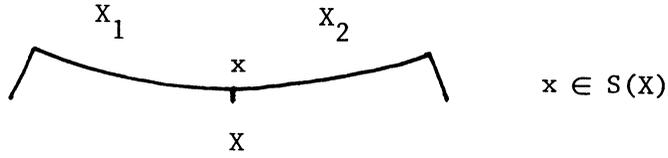


FIGURE 35

A clockwise (anticlockwise) switch is *double-sided* if it has character  $\eta_c$  ( $\eta_a$ ) and *single-sided* if it has character  $-\eta_c$  ( $-\eta_a$ ). Splitting corners at switches satisfies the following.

**Lemma 2.6.1.** *Let  $x \in S(X)$ , and  $X$ ,  $X_1$ , and  $X_2$  be as in Figure 35. Assume  $X$  is not ugly.*

- (i) *If  $x$  is double-sided then  $X$  is good if and only if  $X_1$  and  $X_2$  are good.*
- (ii) *If  $x$  is single-sided then  $X$  is good if and only if  $X_1$  or  $X_2$  is good.*

*Proof.* (i) There are two cases.

(a)  $\text{char } A(X) = \eta_a$ . Then  $X$  is good if and only if all maximal atoms in  $X$  are good if and only if  $X_1$  and  $X_2$  are good.

(b)  $\text{char } A(X) = -\eta_a$ . Then  $x \in C(X)$ . Hence the atoms on both sides of  $x$  are good (by (\*\*)). Thus  $X$ ,  $X_1$ , and  $X_2$  are all good.

(ii) This follows by applying (i) to  $-X$ .  $\square$

The following simple observation will be useful later.

**Lemma 2.6.2.** *If  $\text{char } A(X) = -\eta_a$  and there exists  $c \in C(X)$  with  $\text{char } c = \eta_c$  then  $X$  is good.*

*Proof.* This is just case (i)(b) above.  $\square$

**2.7. Good cycles.** Let  $G$  be as always and  $\Gamma$  a g.d.o. ( $G(\Gamma)$  a subgraph of  $G$ ).

**Lemma 2.7.1.** *Let  $F$  be a disk face of a subgraph of  $\Gamma$  such that each corner of  $F$  with respect to  $\Gamma$  is good. Then  $\text{ind}_\Gamma \partial F \leq 0$ .*

*Remark.* Since the index is unchanged under reflection and reversal of the dual orientations, clearly the conclusion of the lemma also holds if each corner of the face is bad.

*Proof.*  $\partial F$  gives rise to a sequence of corners  $X_1, \dots, X_k$  ( $k \geq 1$ ) such that  $\text{char}(\text{right endpoint of } X_i) = -\text{char}(\text{left endpoint of } X_{i+1})$  ( $i \bmod k$ ). We can therefore consider such an abstract cycle of corners  $\gamma$  (we formally insert edges  $e_i$  between  $X_i$  and  $X_{i+1}$ ). We then define

$$\text{ind } \gamma = \sum_{i=1}^k \text{ind } X_i + \sum_{i=1}^k \text{ind } e_i$$

as before. We now state the lemma in this more abstract setting (which will allow us a proof by induction).

**Claim 2.7.1.** A cycle of good corners has index  $\leq 0$ .

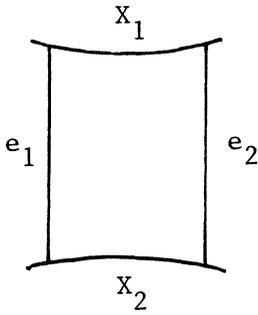


FIGURE 36

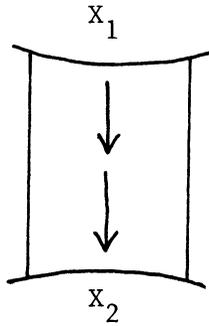


FIGURE 37

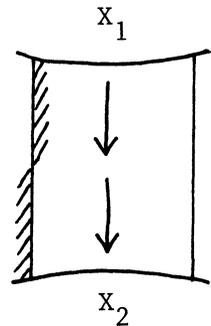


FIGURE 38

*Proof of Claim.* We prove the claim by induction on the number of good corners. Let  $\alpha$  be a cycle of good corners. If the length of  $\alpha$  is 1, and the single corner is a sink or source, then the edge of  $\alpha$  must have index  $-1$ . That is,  $\text{ind } \alpha \leq 0$ .

Assume length  $\alpha$  is 2 (see Figure 36).

If  $\text{ind } \alpha > 0$  then  $\text{ind } \alpha = 2$  (it is always even). Thus  $\text{ind } X_1 = \text{ind } X_2 = 1$ , and  $\text{ind } e_1 = \text{ind } e_2 = 0$ . Thus  $\alpha$  must be as in Figure 37. But then the corners cannot be good (Figure 38).

So we assume length  $\alpha \geq 3$ .

- (1) We shall construct a subcycle  $\beta$  of  $\alpha$  such that
  - (a) no BB corner is adjacent to a source corner; no WW corner is adjacent to a sink corner.
  - (b)  $\text{ind } \alpha \leq \text{ind } \beta$ .
- (2) We shall show any  $\beta$  satisfying (a) above has  $\text{ind } \beta \leq 0$ .

(1) We repeatedly perform the move described in Figure 39 to reduce the number of BB corners that are adjacent to source corners. Then  $\text{ind } X + \text{ind } e_1 + \text{ind } e_2 \rightarrow \text{ind } e$ , where  $(\text{ind } X + \text{ind } e_1 + \text{ind } e_2) \leq 1$  and  $\text{ind } e \geq -1$  and  $\text{ind } \alpha, \text{ind } \beta$  are even. Thus  $\text{ind } \beta$  can be  $< \text{ind } \alpha$  only if  $\text{ind } X + \text{ind } e_1 + \text{ind } e_2 = 1$

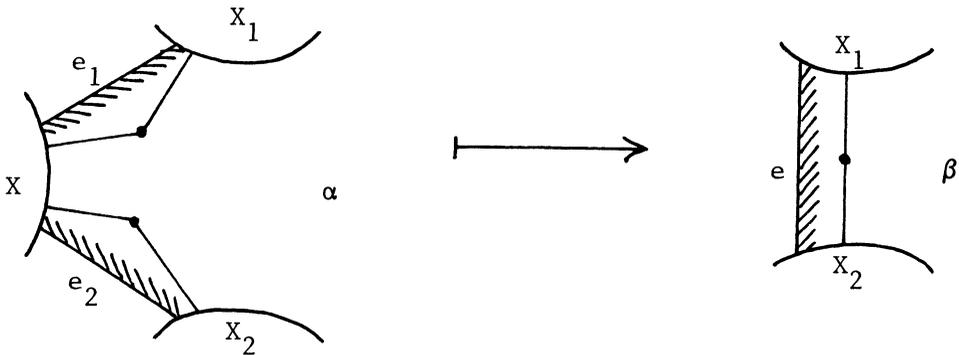


FIGURE 39

and  $\text{ind } e = -1$ . But  $\text{ind } X = 1$  means that  $X$  is a source (since  $X$  is good). Then, if either  $X_1$  or  $X_2$  is a source,  $\text{ind } e_1$  or  $\text{ind } e_2 = -1$ .

A similar move reduces the number of WW corners adjacent to sink corners.

(2) Let  $X$  be a source corner of  $\beta$ . So  $X$  is either BB or BW. By (a), the corner  $Y$  to the left of  $X$  is WB (Figure 40).

**Sublemma.** *Any good WB corner contains a clockwise switch.*

*Proof.* By inspection of those WB corners containing 0 switches or 1 anticlockwise and no clockwise switches.  $\square$

Thus either  $\text{ind } e = -1$ , where  $e$  is the edge between  $X$  and  $Y$ , or  $\text{ind } Y \leq -1$ . Thus anticlockwise from a source we have an object of index  $\leq -1$ .

Similarly, clockwise from a sink we have an object of index  $\leq -1$ .

Can these objects coincide? If so, they cannot be the edges. Therefore, we are as in Figure 41.

But  $Y$  contains a clockwise switch by the sublemma. Thus  $\text{ind } Y \leq -2$ . That is, if the objects associated to a sink and source coincide, they coincide in a corner of index  $\leq -2$ . Thus the  $+$  contribution to  $\text{ind } \beta$  from sink, source corners is cancelled by negative index contributions. Hence  $\text{ind } \beta \leq 0$ .  $\square$

**2.8. Main argument.** In this subsection we prove the following proposition, in which  $G$  denotes  $G_p$ . Proposition 2.0.1 is precisely the first case of this proposition (recall that a Scharlemann cycle in  $G$  represents all  $\{1, \dots, q\}$ -types).

**Proposition 2.8.1.** *Let  $D$  be a disk in  $S^2$  that is either the complement of a small open disk disjoint from  $G_Q$ , or a disk bounded by an  $x$ -cycle in  $G_Q$ . Let*

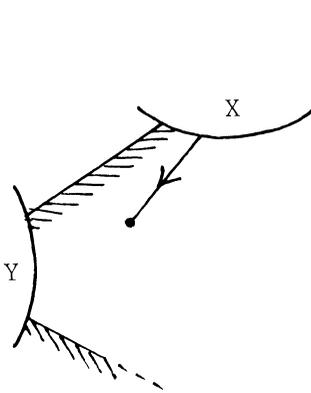


FIGURE 40

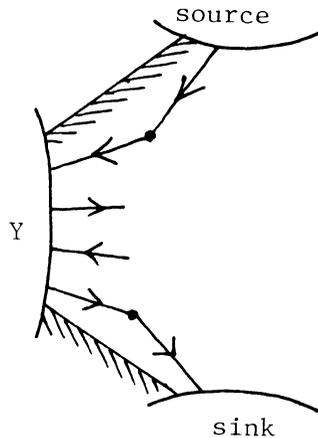


FIGURE 41

$L$  be the set of vertices of  $G_Q$  in  $\text{int } D$  and suppose  $|L| \geq 2$ . Suppose also that  $G_Q$  contains no Scharlemann cycle. Then either  $G(L)$  represents all  $L$ -types or  $G$  contains a Scharlemann cycle.

Given an  $L$ -type  $\tau$ , we show that either  $G(L)$  represents  $\tau$  or there exists an  $x_0$ -cycle in  $\text{int } D$ . The following lemma takes care of the special case when this  $x_0$ -cycle has fewer than two vertices in its interior.

**Lemma 2.8.2.** *Suppose that the hypotheses of Proposition 2.8.1 hold, and that  $G_Q$  contains an  $x_0$ -cycle in  $\text{int } D$  bounding a disk  $D_0$  in  $\text{int } D$ . Let  $L_0$  be the set of vertices of  $G_Q$  in  $\text{int } D_0$ . If  $|L_0| \leq 1$  then ( $|L_0| = 1$  and)  $G$  contains a Scharlemann cycle.*

*Proof of Lemma 2.8.2.* Let the  $x_0$ -cycle be  $\Sigma_0$ . If  $\Sigma_0$  is a great  $x_0$ -cycle then  $G_Q$  contains a Scharlemann cycle (Lemma 2.0.2), contrary to hypothesis. Hence we may suppose that  $L_0 = \{y_0\}$ , where  $y_0$  has opposite sign to the vertices of  $\Sigma_0$ . Therefore, for every vertex  $x$  of  $G$ , the edge of  $G$  with label  $y_0$  at  $x$  joins  $x$  to a vertex of the same sign (there are no trivial loops in  $G_Q$ ). Hence  $G$  contains a great  $y_0$ -cycle, and thus a Scharlemann cycle.  $\square$

*Proof of Proposition 2.8.1.* We proceed by induction on the number of  $x$ -cycles in  $\text{int } D$ . We thus need to establish

- (A) the case where there are no  $x$ -cycles in  $\text{int } D$ ,
- (B) the inductive step.

Let  $\tau$  be an  $L$ -type. We shall show that  $G(L)$  represents  $\tau$ . We distinguish two cases.

- (1)  $\tau$  trivial,
- (2)  $\tau$  nontrivial.

Let  $\Sigma$  be the  $x$ -cycle in the hypothesis of the proposition, or, in the first case of the hypothesis, let  $\Sigma = \emptyset$ .

(1)  $\tau$  trivial. Let  $J \subset L$  be the set of vertices of opposite sign to those in  $\Sigma$  (or if  $\Sigma = \emptyset$ , let  $J$  be the set of (say)  $+$  vertices of  $G_Q$ ).

Since  $\Sigma$  cannot be a great  $x$ -cycle of  $G_Q$  by Lemma 2.0.2,  $J \neq \emptyset$ .

(a) Suppose that for all vertices  $x$  of  $G$ , there is a  $y(x) \in J$  such that the edge of  $G$  with label  $y(x)$  at  $x$  joins  $x$  to a parallel vertex of  $G$ . Note that the label at the other end of this edge  $\notin J$ , by the parity rule. The subgraph of  $G$  consisting of the vertices of  $G$  along with these edges will contain a cycle  $\sigma$  all of whose vertices are parallel, that bounds a disk  $E$  where all the vertices in the interior of  $E$  are parallel to the vertices of  $\sigma$ . Then  $E$  contains a disk face of  $G(L)$  with all vertices parallel, showing that  $G(L)$  represents  $\tau$ .

(b) If the supposition in (a) fails, then there exists a vertex  $x_0$  of  $G$  such that for all  $y \in J$ , the edge of  $G$  with label  $y$  at  $x_0$  joins  $x_0$  to an antiparallel vertex. The corresponding edge of  $G_Q$  therefore joins  $y$  to a parallel vertex, i.e., to a vertex  $\in J$ . These edges give us an  $x_0$ -cycle  $\Sigma_0$  in  $G_Q$  with vertices  $\in J$ . Thus  $\Sigma_0 \subset \text{int } D$ . This immediately gives (A).

To prove (B), let  $L_0$  be the set of vertices in the interior of the disk in  $D$  bounded by  $\Sigma_0$ .

If  $|L_0| \leq 1$  then we are done by Lemma 2.8.2. If  $|L_0| \geq 2$ , then by the inductive hypothesis  $G(L_0)$  represents all  $L_0$ -types, in particular, the trivial  $L_0$ -type. Let  $E_0$  be a disk face of  $G(L_0)$  representing this type. Then any sub- $G(L)$ -face  $E$  of  $E_0$  represents  $\tau$ .

(2)  $\tau$  nontrivial. We construct a sequence of stars  $T_1, \dots, T_n, n \geq 1$  (with sign  $V(T_i) = +$ ) such that

- (i)  $[T_1] = \tau, [T_i]$  is nontrivial,  $i = 1, \dots, n$ ;
- (ii)  $T_i = d_i T_{i-1}$ , where  $d_i = d^\pm, 2 \leq i \leq n$ ;
- (iii) all elements of  $C(T_n)$  have the same parity;
- (iv) all elements of  $A(T_n)$  have the same parity.

To do this, first let  $T_1$  be any one of the two stars with sign  $V(T_1) = +$  and  $[T_1] = \tau$ .

If all elements of  $C(T_1)$  have the same parity and all elements of  $A(T_1)$  have the same parity, take  $n = 1$ .

If not, then by replacing, if necessary,  $T_1$  by  $\overline{T_1}$  we may assume that not all elements of  $A(T_1)$  have the same parity. Let  $m$  be the least integer  $\geq 1$  such that all elements of  $A((d^+)^m T_1)$  have the same parity.

If all elements of  $C((d^+)^m T_1)$  also have the same parity, take  $n = m + 1$  and let  $T_i = d^+ T_{i-1}, 2 \leq i \leq n$ . If not, take  $n = m + 2$ , with

$$\begin{aligned} T_i &= d^+ T_{i-1}, & 2 \leq i \leq m, \\ T_{m+1} &= d^- T_m & (= \overline{(d^+)^m T_1}), \\ T_{m+2} &= d^+ T_{m+1}. \end{aligned}$$

Since  $C(T_{m+2}), A(T_{m+2}) \subset C(T_{m+1}) = A((d^+)^m T_1)$ ,  $T_{m+2}$  satisfies (iii) and (iv). Let  $L_i = L(T_i)$ . Thus  $L_1 = L$ , and  $L_i = C(T_{i-1}), 2 \leq i \leq n$ . By construction  $|L_i| \geq 2, 1 \leq i \leq n$ .

If  $G(L_n)$  represents  $[T_n]$ , then by Corollary 2.4.2  $G(L)$  represents  $[T_1] = \tau$ , as desired.

So assume that  $G(L_n)$  does not represent  $[T_n]$ . Then by Lemma 2.5.1,  $G_Q$  contains an  $x_0$ -cycle  $\Sigma_0$  whose vertices are contained in either  $C(T_n)$  or  $A(T_n)$ . This immediately proves the proposition in case (A). For (B), proceed as follows. We may assume that the vertices of  $\Sigma_0$  are contained in  $C(T_n)$  by replacing (if necessary)  $T_n$  by  $\overline{T_n}$ . If  $n > 1$ , this can be achieved by replacing  $d_n (= d^+)$  by  $d^-$ . Thus, in addition to (i)-(iv), we may assume

- (v) the vertices of  $\Sigma_0$  are contained in  $C(T_n)$ .

Let  $L_0$  be the set of vertices of  $G_Q$  in the interior of the disk in  $D$  bounded by  $\Sigma_0$ .

If  $|L_0| \leq 1$  then we are done by Lemma 2.8.2. So assume  $|L_0| \geq 2$ . By the inductive hypothesis  $G(L_0)$  represents all  $L_0$ -types. We shall define a certain

$L_0$ -type  $\tau_0$  and use the fact that  $G(L_0)$  represents  $\tau_0$  to eventually conclude that  $G(L)$  represents  $\tau$ .

First, let  $R_1, \dots, R_n$  be the sequence of stars corresponding to  $T_1, \dots, T_n$  obtained by taking derivatives relative to  $L_0$ . That is, we define  $R_1 = T_1$ ,  $R_i = d_0^\pm R_{i-1}$ , according as  $d_i = d^\pm$ ,  $2 \leq i \leq n$  (and where, as usual,  $d_0 = d_{L_0}$ ). By Proposition 2.1.2,  $C(T_n) \subset C(R_n)$ , and  $\tilde{A}(T_n) \supset \tilde{A}(R_n)$ . Let  $I$  be an  $L_0$ -interval (at  $V_+$ ).

*Claim 1.* The corner  $R_n|I$  is not ugly.

*Proof of Claim 1.*  $\tilde{A}(R_n) \subset \tilde{A}(T_n) \subset A(T_n)$ . Hence by (iv) above all anticlockwise switches of  $R_n$  in the interior of  $I$  have the same parity.  $\square$

Hence the set of corners  $\{R_n|I : I \text{ an } L_0\text{-interval at } V_+\}$  can be partitioned into good and bad, once we choose characters  $\eta_c$  and  $\eta_a$ . We do this by setting

$$\eta_c = \text{char}(C(T_n), V(T_n)), \quad \eta_a = -\text{char}(\tilde{A}(R_n), V(R_n)).$$

(If  $\tilde{A}(R_n) = \emptyset$ ,  $\eta_a$  can be chosen arbitrarily.) Recall that  $V(T_n) = V(R_n) = V_+$ .

Now for each  $L_0$ -interval  $I$ , define  $\varepsilon(I) = \pm$  by the requirement that  $\varepsilon(I)(R_n|I)$  be good. Finally, define the  $L_0$ -type  $\tau_0$  by  $\tau_0 = (\varepsilon(I) : I \text{ an } L_0\text{-interval})$ . By our inductive hypothesis, there exists a disk face  $E$  of  $G(L_0)$  representing  $\tau_0$ . This means there exists  $\eta = \pm$  such that if a corner of  $E$  at a vertex  $v$  is contained in the  $L_0$ -interval  $I$  then  $\text{sign } v = \eta\varepsilon(I)$ . Let  $J$  be the subinterval of  $I$  corresponding to the corner.

*Claim 2.*  $\eta(\text{sign } v)R_n|J$  is good.

*Proof of Claim 2.* Note that  $\eta(\text{sign } v) = \varepsilon(I)$ . If  $J$  is a full  $L_0$ -interval, the claim follows from the definition of  $\tau_0$ . If not,  $J$  is a subinterval of an  $L$ -interval  $I$  with at least one endpoint of  $J$  in the set of exceptional labels of  $G(L_0)$ . Since this set is contained in  $C(T_n)$  by (v) above,  $\varepsilon(I) = +$  by Lemma 2.6.2. Now  $R_n|J$  is good by Lemma 2.6.1(i).  $\square$

Let  $\Gamma_i = \Gamma(R_i)$ ,  $1 \leq i \leq n$ , and let  $\delta_0 = \delta_{G(L_0)}$ . Since  $\{\text{exceptional labels of } G(L_0)\} \subset \{\text{vertices of } \Sigma_0\} \subset C(T_n) \subset C(T_i) \subset C(R_i)$ ,  $1 \leq i \leq n$ , Lemma 2.2.2 implies that  $\Gamma_i = \delta_0 \Gamma_{i-1}$ ,  $2 \leq i \leq n$ .

By Claim 2, if  $X$  is a corner of  $E$  with respect to  $\Gamma_n$ , then  $\eta X$  is good.

Now consider  $\Gamma_n^* \cap E$ . A face of  $\Gamma_n^* \cap E$  of positive index corresponds to a switch-edge  $e$  of  $\Gamma_n \cap E$ . Since the endpoints of  $e$  have opposite characters, one endpoint of  $e$  will be a double-sided switch, the other single-sided. Let  $E = E_1 \cup_c E_2$ . By Lemma 2.6.1, at least one of  $E_1, E_2$  has the property that for all its corners  $X$  with respect to  $\Gamma_n$ ,  $\eta X$  is good. Continuing in this way, we eventually get a disk  $F \subset E$ , bounded by edges of  $\Gamma_n$ , such that

- (a) for each  $\Gamma_n$ -corner  $X$  of  $F$ ,  $\eta X$  is good;
- (b)  $\Gamma_n^*$  has no faces of positive index in  $F$ .

By Lemma 2.7.1 and the remark following it, (a) implies that  $\text{ind}_{\Gamma_n} \partial F \leq 0$ . Lemma 2.3.3 and (b) now implies that  $\Gamma_n^*$  has a sink or source in  $F$ . Since  $\Gamma_i = \delta_0 \Gamma_{i-1}$ ,  $2 \leq i \leq n$ , Lemma 2.4.1 shows that  $\Gamma_1^* = \Gamma(T_1)^*$  has a sink or source at a dual vertex in  $F$ . Thus  $G(L)$  represents  $[T_1] = \tau$  as desired.  $\square$  (2.8.1)

### 3. TOPOLOGICAL CONSEQUENCES OF THE COMBINATORICS

In this section we examine the topological implications of the combinatorial Proposition 2.0.1 and show that it leads to a proof of Proposition 2 of the Introduction.

Let  $\tau$  be an  $n$ -type. We say that  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbf{Z}^n$  represents  $\tau$  if

- (i)  $\sum_{i=1}^n |\alpha_i| \geq 2$ , and
- (ii) if  $\tau = (\varepsilon_1, \dots, \varepsilon_n)$ , then there exists  $\eta = \pm 1$  such that  $\text{sign } \alpha_i = \eta \varepsilon_i$  for all  $i$  such that  $\alpha_i \neq 0$ .

A subset  $A$  of  $\mathbf{Z}^n$  represents all  $n$ -types if for every  $n$ -type  $\tau$  there exists  $\alpha \in A$  such that  $\alpha$  represents  $\tau$ .

We conjecture that if  $A$  represents all  $n$ -types, then it contains a subset  $A_0$  such that the abelian group on  $n$  generators presented by the matrix whose rows are the elements of  $A_0$  has nontrivial torsion. This conjecture along with Proposition 2.0.1 easily proves Proposition 2 (without the assumption that  $Q$  comes from a level sphere in a thin presentation of  $K$ ). In the absence of a proof of the conjecture<sup>1</sup>, we are lead to the argument of the present section.

The following observation will be useful in the sequel. Suppose  $\{1, \dots, n\} = N_0 \sqcup N_1 \sqcup N_2$ , and let  $\tau_0$  be some fixed  $|N_0|$ -type. Let  $A$  be a subset of  $\mathbf{Z}^n$ . For  $j = 1, 2$ , let  $A_j = \{\alpha \in A : \alpha_i = 0 \text{ for } i \notin N_j\}$ . Now suppose  $A$  has the property that if  $\alpha \in A$  represents any  $n$ -type  $\tau$  such that  $\tau|_{N_0} = (\tau \text{ restricted to the coordinates corresponding to } N_0) = \tau_0$ , then  $\alpha \in A_1$  or  $A_2$ .

**Lemma 3.1.** *Given the above data, suppose in addition that  $A$  represents all  $n$ -types. Then there exists  $j \in \{1, 2\}$  such that  $N_j \neq \emptyset$  and  $\{\alpha|_{N_j} : \alpha \in A_j\}$  represents all  $|N_j|$ -types.*

*Proof.* Suppose for a contradiction that for  $j = 1, 2$  we either have  $N_j \neq \emptyset$  and an  $|N_j|$ -type  $\tau_j$  such that  $\{\alpha|_{N_j} : \alpha \in A_j\}$  does not represent  $\tau_j$ , or  $N_j = \emptyset$  (in which case we let  $\tau_j$  be the empty type). Now let  $\tau$  be the  $n$ -type such that  $\tau|_{N_j} = \tau_j$  for  $j = 0, 1$ , and  $2$ . By hypothesis, there exists  $\alpha \in A$  such that  $\alpha$  represents  $\tau$ . Since  $\tau|_{N_0} = \tau_0$ ,  $\alpha \in A_1$  or  $A_2$ , say  $A_1$ . But this implies that  $N_1 \neq \emptyset$  and  $\alpha|_{N_1}$  represents  $\tau_1$ , a contradiction.  $\square$

Let  $B$  be a 3-ball in  $S^3$ , and write  $B' = \overline{S^3 - B}$ . Let  $\mathbf{H}$  be a disjoint union of  $n \geq 1$  1-handles  $H_1, \dots, H_n$  in  $B'$ , attached to  $B$ , such that  $\mathbf{H}$  is standard, i.e., there is an ambient isotopy of  $S^3$  taking  $(B, \mathbf{H})$  to the pair illustrated in Figure 42. Let  $C_i$  be the boundary of a cocore of  $H_i$ ,  $1 \leq i \leq n$ , with some orientation.

<sup>1</sup> Added in proof. This conjecture has now been proved by Walter Parry (private communication).

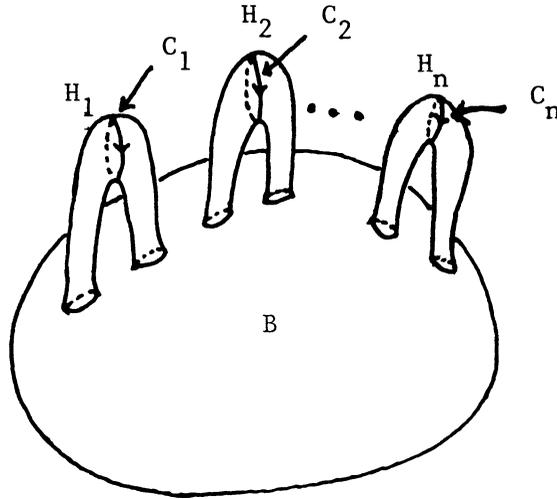


FIGURE 42

By an **H-type** we mean an  $n$ -type  $(\varepsilon_1, \dots, \varepsilon_n)$  where  $\varepsilon_i$  is formally associated with  $H_i$ ,  $1 \leq i \leq n$ . Let  $C$  be a simple closed curve in  $\partial(B \cup \mathbf{H})$ , and let  $\tau$  be an **H-type**. Then we say that  $C$  represents  $\tau$  if for some (and hence any) orientation of  $C$  we have

- (i) for each  $i$ ,  $1 \leq i \leq n$ ,  $C \cap H_i$  consists of  $|\alpha_i|$  arcs transverse to  $C_i$ , where  $\alpha_i$  is the algebraic intersection number of  $C$  with  $C_i$ ; and
- (ii)  $(\alpha_1, \dots, \alpha_n)$  represents  $\tau$ .

A disjoint union  $\mathbf{C}$  of simple closed curves in  $\partial(B \cup \mathbf{H})$  represents all **H-types** if for every **H-type**  $\tau$  there exists  $C \in \mathbf{C}$  such that  $C$  represents  $\tau$ . (Here, and in similar contexts, we abuse notation and write  $C \in \mathbf{C}$  to mean that  $C$  is a component of the disjoint union  $\mathbf{C}$ .) We shall always assume, without loss of generality, that every  $C \in \mathbf{C}$  satisfies condition (i) immediately above and the condition that  $|C \cap \mathbf{H}| \geq 2$ .

Let  $\mathbf{E}$  be a disjoint union of (embedded) disks in  $S^3$  such that

- (i)  $\partial \mathbf{E} \subset \partial(B \cup \mathbf{H})$ ;
- (ii) a collar of  $\partial \mathbf{E}$  is contained in  $\overline{B' - \mathbf{H}}$  and is transverse to  $\partial(B \cup \mathbf{H})$ ;
- (iii)  $(\text{int } \mathbf{E}) \cap \mathbf{H} = \emptyset$ .

**Proposition 3.2.** *For no collection of disks  $\mathbf{E}$  as above does  $\partial \mathbf{E}$  represent all **H-types**.*

*Proof.* We may assume that  $\text{int } \mathbf{E}$  intersects  $\partial B$  transversely.

We assume for a contradiction that there exists  $(B, \mathbf{H}, \mathbf{E})$  as above such that  $\partial \mathbf{E}$  does represent all **H-types**. We first show that if  $|(\text{int } \mathbf{E}) \cap \partial B| > 0$  then there exists another such triple,  $(\tilde{B}, \tilde{\mathbf{H}}, \tilde{\mathbf{E}})$ , with  $|(\text{int } \tilde{\mathbf{E}} \cap \partial \tilde{B})| < |(\text{int } \mathbf{E}) \cap \partial B|$ . This reduces us to the case  $(\text{int } \mathbf{E}) \cap \partial B = \emptyset$  where we proceed by induction on  $n = |\mathbf{H}|$ .

(A)  $|(\text{int } \mathbf{E}) \cap \partial B| > 0$ . Let  $D$  be a disk  $\subset \mathbf{E}$  that is innermost with respect to the disjoint simple closed curves  $(\text{int } \mathbf{E}) \cap \partial B$ . There are two subcases.

(a)  $D \subset B$ .  $D$  separates  $B$  into 3-balls  $B_1$  and  $B_2$ , say. Then  $\mathbf{H} = \mathbf{H}_0 \sqcup \mathbf{H}_1 \sqcup \mathbf{H}_2$ , where for  $j = 1$  or  $2$ ,  $\mathbf{H}_j$  consists of those  $H \in \mathbf{H}$  with both feet in  $\partial B_j$ , and  $\mathbf{H}_0$  consists of those  $H \in \mathbf{H}$  with one foot in each of  $\partial B_1$  and  $\partial B_2$ . We have a corresponding partition  $\{1, \dots, n\} = N_0 \sqcup N_1 \sqcup N_2$ , where  $H_i \in \mathbf{H}_j$  if and only if  $i \in N_j$  ( $1 \leq i \leq n$ ,  $0 \leq j \leq 2$ ).

For every  $i \in N_0$ , let  $\eta_i = \pm$  be defined by the condition that a transverse arc on  $H_i$  intersecting  $C_i$  with sign  $\eta_i$  is directed from the foot of  $H_i$  in  $\partial B_1$  to the foot in  $\partial B_2$ . Let  $\tau_0$  be the  $\mathbf{H}_0$ -type  $(\eta_i : i \in N_0)$ . Let  $\tau$  be any  $\mathbf{H}$ -type such that  $\tau|_{\mathbf{H}_0} = \tau_0$ . Then any disk  $E \in \mathbf{E}$  such that  $\partial E$  represents  $\tau$  must satisfy  $E \cap \mathbf{H} = E \cap \mathbf{H}_j$ ,  $j = 1$  or  $2$ . Hence by Lemma 3.1 there exists  $j \in \{1, 2\}$  such that  $\mathbf{H}_j \neq \emptyset$ , and  $\mathbf{E}_j \subset \mathbf{E}$  such that  $\partial \mathbf{E}_j$  meets only handles in  $\mathbf{H}_j$  and represents all  $\mathbf{H}_j$ -types. Finally, if  $\mathbf{E}_j$  contains the component of  $\mathbf{E}$  containing  $D$ , we move this component slightly off  $D$  away from  $B_j$ . Now  $(B_j, \mathbf{H}_j, \mathbf{E}_j)$  is a triple of the desired form with  $|(\text{int } \mathbf{E}_j) \cap \partial B_j| < |(\text{int } \mathbf{E}) \cap \partial B|$ .

(b)  $D \subset B'$ .  $D$  separates  $B'$  into 3-balls  $B'_1$  and  $B'_2$ , say. Let  $\mathbf{H}_j = \mathbf{H} \cap B'_j$ ,  $j = 1, 2$ . No disk  $E \in \mathbf{E}$  can meet both  $\mathbf{H}_1$  and  $\mathbf{H}_2$ . Therefore, by Lemma 3.1 (with  $N_0 = \emptyset$ ) there exists  $j \in \{1, 2\}$  such that  $\mathbf{H}_j \neq \emptyset$ , and  $\mathbf{E}_j \subset \mathbf{E}$  such that  $\partial \mathbf{E}_j$  meets only handles in  $\mathbf{H}_j$  and represents all  $\mathbf{H}_j$ -types. Let  $B_j = B \cup B'_{3-j}$ , and (if necessary) move  $\mathbf{E}_j$  slightly off  $D$  away from  $B'_j$ . Then as in subcase (a) above  $|(\text{int } \mathbf{E}_j) \cap \partial B_j| < |(\text{int } \mathbf{E}) \cap \partial B|$ .

(B)  $|(\text{int } \mathbf{E}) \cap \partial B| = 0$ . Note that in this case  $\mathbf{E} \subset \overline{B' - \mathbf{H}}$ . If  $n = 1$ , then for any  $E \in \mathbf{E}$  we have that  $N(B \cup \mathbf{H} \cup E)$  is a punctured lens space, since  $|\partial E \cap \mathbf{H}| \geq 2$ . Since  $N(B \cup \mathbf{H} \cup E)$  is embedded in  $S^3$ , this is absurd. So we assume that  $n > 1$  and show that there exists another triple  $(\tilde{B}, \tilde{\mathbf{H}}, \tilde{\mathbf{E}})$  of the stated type with  $|\tilde{\mathbf{H}}| < n$ . We say that  $E \in \mathbf{E}$  is a *bigon* if  $|\partial E \cap \mathbf{H}| = 2$ . Again we consider two subcases.

(a)  $\mathbf{E}$  contains a bigon. Let  $E \in \mathbf{E}$  be a bigon.

If  $\partial E$  meets only one  $H \in \mathbf{H}$ , then  $N(B \cup \mathbf{H} \cup E)$  is a punctured  $RP^3$  in  $S^3$ , which is absurd.

So suppose that  $\partial E$  meets distinct handles  $H_1$  and  $H_2 \in \mathbf{H}$ . We may assume without loss of generality that the orientations of the cocore boundaries  $C_1$  and  $C_2$  are chosen so that the intersections of  $\partial E$  with  $C_1$  and  $C_2$  both have the same sign. Then  $\partial E$  represents any  $\mathbf{H}$ -type of the form  $\pm (++\dots)$ . Any other  $\mathbf{H}$ -type is of the form  $\pm(- + \dots)$ ; let  $\mathbf{F} \subset \mathbf{E}$  be such that  $\partial \mathbf{F}$  represents all  $\mathbf{H}$ -types of this form. Orient each  $F \in \mathbf{F}$  so that if  $\alpha_i(F)$  is the algebraic intersection number of  $\partial F$  with  $C_i$ , then we have  $\alpha_1(F) \leq 0$ ,  $\alpha_2(F) \geq 0$ . Now successively attach parallel copies of  $E$  to  $\mathbf{F}$  along the arcs  $\mathbf{F} \cap H_1$ , and move the resulting disks slightly off  $H_1$ . This produces for each

$F \in \mathbf{F}$  an oriented disk  $F'$  such that

$$\begin{aligned} \partial F' \cap H_1 &= \emptyset; \\ |\partial F' \cap H_2| &= \alpha_2(F') = \alpha_2(F) - \alpha_1(F) = |\partial F \cap H_1| + |\partial F \cap H_2|; \\ \partial F' \cap H_i &= \partial F \cap H_i, \quad i \neq 1, 2. \end{aligned}$$

Therefore, if we let  $\mathbf{F}'$  be the union of the disks  $F'$  and  $\mathbf{H}' = \mathbf{H} - H_1$ , then  $\partial \mathbf{F}'$  represents all  $\mathbf{H}'$ -types. Then  $(B, \mathbf{H}', \mathbf{F}')$  has  $|\mathbf{H}'| = n - 1$ .

(b)  $\mathbf{E}$  contains no bigon. Let  $D' \subset \overline{B' - \mathbf{H}}$  be a disk that separates  $B'$  into  $B'_1$  and  $B'_2$  such that  $H_1 \subset B'_1$  and  $\mathbf{H} - H_1 \subset B'_2$ . Since  $\overline{B' - \mathbf{H}}$  is a handlebody, an isotopy of (say)  $\mathbf{E}$ , keeping  $\partial \mathbf{E}$  fixed, will ensure that  $\mathbf{E} \cap D'$  contains no simple closed curves. There are now two possibilities.

(i)  $\mathbf{E} \cap D' \neq \emptyset$ . Let  $\alpha$  be an arc of  $\mathbf{E} \cap D'$  that is outermost on  $D'$ , i.e., there exists a disk  $D_0 \subset D'$  such that  $D_0 \cap \mathbf{E} = \partial D_0 \cap \mathbf{E} = \alpha$  and  $\partial D_0 - \alpha \subset \partial D'$ . Let  $E$  be the component of  $\mathbf{E}$  containing  $\alpha$ . Then surgering  $E$  along  $D_0$  gives two new disks  $E'$  and  $E''$  such that

$$|\partial E' \cap \mathbf{H}| + |\partial E'' \cap \mathbf{H}| = |\partial E \cap \mathbf{H}| \geq 3.$$

Therefore  $\partial E'$  (say) has  $|\partial E' \cap \mathbf{H}| \geq 2$  and clearly represents any  $\mathbf{H}$ -type represented by  $\partial E$ . So if we let  $\mathbf{E}' = (\mathbf{E} - E) \cup E'$ , then  $\partial \mathbf{E}'$  represents all  $\mathbf{H}$ -types, and  $|\mathbf{E}' \cap D'| < |\mathbf{E} \cap D'|$ . If  $\mathbf{E}'$  contains a bigon, case (a) above applies. If not, and  $\mathbf{E}' \cap D' \neq \emptyset$ , then we may repeat the procedure just described. It thus only remains to consider case (ii) below.

(ii)  $\mathbf{E} \cap D' = \emptyset$ . If  $\mathbf{E} \cap H_1 \neq \emptyset$  then there exists  $E \in \mathbf{E}$  such that  $N(B \cup H_1 \cup E)$  is a punctured lens space, which, as usual, is impossible.

If  $\mathbf{E} \cap H_1 = \emptyset$ , let  $\mathbf{H}' = \mathbf{H} - H_1$ . Then  $\partial \mathbf{E}$  represents all  $\mathbf{H}'$ -types and  $(B, \mathbf{H}', \mathbf{E})$  has  $|\mathbf{H}'| = n - 1$ .  $\square$

For the proof of Proposition 2 we now specialize to the case where  $K \subset S^3$  is a nontrivial knot and  $X$  is its exterior. Let  $\gamma$  be the meridian of  $K$ .

**Proposition 2.** *Suppose that  $X$  contains properly embedded surfaces  $P$  and  $Q$  satisfying conditions (i), (ii), and (iii) of Proposition 1, where  $Q$  is the intersection with  $X$  of a level sphere in a thin presentation of  $K$ . Then  $K(\pi)$  has a lens space as a connected summand.*

Let  $G_p$  and  $G_q$  be the labeled graphs of  $P \cap Q$  in  $P$  and  $Q$ . If  $n \geq 2$  then it is proven in [CGLS, §§2.5 and 2.6; see Proposition 2.5.6] that either  $G_p$  or  $G_q$  contains a Scharlemann cycle. This implies that either  $K(\gamma)$  or  $K(\pi)$  contains a lens space summand (see the proof of Proposition 3.2). As  $K(\gamma) \cong S^3$ , we obtain the desired conclusion. Thus we hereafter assume that  $n = 1$  and we are in the context of §2. By Proposition 2.0.1, either  $G_q$  contains a Scharlemann cycle or  $G_p = G$  represents all  $L$ -types where  $L = \{1, \dots, q\}$ . Again, in the first case,  $K(\pi)$  has a lens space summand. So suppose  $G$  represents all

*L*-types. Because  $Q$  comes from a level sphere,  $Q$  separates  $X$  into a black region, say above  $Q$ , and a white region, below  $Q$ . Every *L*-interval of  $G$  is then either white or black. Let  $\mathbf{B}$  be the set of black *L*-intervals and  $\mathbf{W}$  the set of white *L*-intervals. Then  $\{L\text{-intervals}\} = \mathbf{B} \cup \mathbf{W}$ . A **B**-type (**W**-type) is a  $|\mathbf{B}|$ -type ( $|\mathbf{W}|$ -type) where each coordinate is formally associated to a different black (white) *L*-interval. Let  $f$  be a face of  $G$ . Then either every corner of  $f$  is a black *L*-interval or every corner is a white *L*-interval; in the former case  $f$  is called a *black face*, in the latter  $f$  is called a *white face*. This dichotomy of faces and *L*-intervals into black and white allows us to conclude that in fact there is a collection consisting either of black faces of  $G$  representing all **B**-types or of white faces of  $G$  representing all **W**-types (Lemma 3.1 with  $N_0 = \emptyset$ ). Without loss of generality we assume there is a collection,  $\tilde{\mathbf{E}}$ , of black faces in  $G$  representing all **B**-types.

Let  $a < b$  be adjacent critical levels of  $K$  in the thin presentation (from which  $Q$  comes), such that  $a$  is a relative minimum and  $b$  a relative maximum of  $K$ .  $(a, b)$  is called a *middle interval* of  $K$ . Let  $(a_1, b_1), (a_2, b_2), \dots, (a_k, b_k)$  be the middle intervals strictly above the level sphere  $\hat{Q}$  from which  $Q$  comes. We will argue by induction on  $k$ , using Proposition 3.2 to obtain a contradiction. (In the course of the inductive argument, we allow arcs of  $P \cap Q$  to be boundary-parallel on  $Q$ .)

Assume  $k = 0$ . Let  $B$  be the 3-ball bounded by  $\hat{Q}$  that is below  $\hat{Q}$  (containing the white region of  $X$ ) and  $B'$  be the 3-ball above  $\hat{Q}$ . Let  $\mathbf{H}$  be the union of the 1-handles in  $B'$  that are regular neighborhoods of the arcs  $K \cap B'$ . Applying Proposition 3.2 to  $(B, \mathbf{H}, \tilde{\mathbf{E}})$ , we arrive at a contradiction. Note that if  $E \in \tilde{\mathbf{E}}$  and  $(\alpha_1, \dots, \alpha_n)$  corresponds to  $\partial E$  then  $\sum_{i=1}^n |\alpha_i| \geq 2$  since there are no 1-sided faces in  $G$ .

So we assume  $k \geq 1$ . Let  $(a_1, b_1)$  be the first middle interval above  $Q$ . By [Ga, §4(A)], there is a level surface  $Q_1$ , coming from  $(a_1, b_1)$  such that, after an isotopy of  $P$  that is away from  $Q$ ,  $P \cap Q_1$  satisfies (i) and (ii) of Proposition 1 and has the property that no arc of  $P \cap Q_1$  is boundary-parallel on  $P$ .

Let  $G'$  be the graph defined by the arc components of  $P \cap (Q \cup Q_1)$  in  $P$ . Define the signs of the vertices, the parity of the labels, and the character function as usual for  $G'$ . Let  $L$  be the set of labels in  $G'$  corresponding to the components of  $\partial Q$  and  $L_1$  be the labels corresponding to the components of  $Q_1$ . Let  $L' = L \cup L_1$ . Note that  $G = G'(L)$ .

Denote the region of  $X$  above  $Q_1$ , the region between  $Q$  and  $Q_1$ , and the region below  $Q$  as the green, red, and white region, respectively. Thus the black region above  $Q$  is now divided into a green region and a red region. We can now write  $\{L'\text{-intervals}\} = \mathbf{R} \cup \mathbf{G} \cup \mathbf{W}$  where  $\mathbf{R}$  is the set of red  $L'$ -intervals,  $\mathbf{G}$  the set of green  $L'$ -intervals, and  $\mathbf{W}$  the set of white  $L'$ -intervals. As before, the *L*-intervals are divided into white and black intervals. The white *L*-intervals are exactly the white  $L'$ -intervals, and any black *L*-interval is a union of red

and green  $L'$ -intervals. A face of  $G'$  is called red, green, or white if all corners of the face are red, green, or white, respectively. Every face of  $G'$  is either red, green, or white. An  $\mathbf{R}$ -type is an  $|\mathbf{R}|$ -type where the coordinates are formally associated with the red  $L'$ -intervals. A red face of  $G'$  represents an  $\mathbf{R}$ -type in the same way a face of  $G'$  represents an  $L'$ -type. Similarly define a  $\mathbf{G}$ -type.

**Lemma 3.3.**  $G'$  represents all  $\mathbf{R}$ -types or all  $\mathbf{G}$ -types.

*Proof.* We find a collection of disk faces  $\mathbf{E}'$  of  $G'$  that represents every  $L'$ -type. Furthermore,  $\mathbf{E}'$  will consist entirely of red and green faces. Lemma 3.1 then implies that we can find a subcollection  $\mathbf{E}$  consisting entirely of red faces and representing all  $\mathbf{R}$ -types or entirely of green faces and representing all  $\mathbf{G}$ -types.

Let  $\tau$  be an  $L'$ -type.

*Claim.* There is a sequence of stars  $X_1, X_2, \dots, X_n$ , such that

- (1)  $[X_1] = \tau$ ,
- (2)  $X_i = d_L^+ X_{i-1}$  or  $\overline{d_L^+ X_{i-1}}$ ,  $2 \leq i \leq n$ ,
- (3)  $L(X_n) = L$ .

*Proof of Claim.* Let  $X_1$  be a star with  $[X_1] = \tau$ . Let  $m$  be the least integer  $\geq 1$  such that  $A((d_L^+)^m X_1) \subset L$ . Take  $n = m + 2$  with

$$\begin{aligned}
 X_{i+1} &= (d_L^+)^i X_1, & 0 \leq i < m, \\
 X_{m+1} &= \overline{d_L^+ X_m}, \\
 X_{m+2} &= d_L^+ X_{m+1}. \quad \square
 \end{aligned}$$

Consider the  $\mathbf{B}$ -type  $[X_n]|\mathbf{B}$ , obtained by restricting the  $L$ -type  $[X_n]$  to the set of black  $L$ -intervals. As  $G$  represents all  $\mathbf{B}$ -types, there is a black face  $F$  of  $G(L)$  representing  $[X_n]|\mathbf{B}$ . That is,  $F$  corresponds to a source or sink of  $(\Gamma(X_n))^*$ . Since  $G$  has no exceptional labels, by Lemma 2.2.2 we have  $\delta_G^+(\Gamma(X_{i-1})) = \Gamma(d_L^+ X_{i-1}) = \Gamma(X_i)$  or  $\Gamma(\overline{X_i})$ . Then successively applying Lemma 2.4.1 (noting that the sinks and sources of  $\Gamma(X_i)^*$  are exactly the sources and sinks of  $\Gamma(\overline{X_i})^*$ ) gives a subface  $F_1$  of  $\Gamma(X_1) \cap F$  corresponding to a sink or source of  $\Gamma(X_1)^*$ . Since  $F_1 \subset F$ ,  $F_1$  will be a red or green face of  $G(\Gamma(X_1)) = G'$ . Note that  $F_1$  is not a 1-sided face by the construction of  $Q_1$ .

Thus we have proven that every  $L'$ -type  $\tau$  is represented by a red or green face of  $G'$ . By Lemma 3.1,  $G'$  represents all  $\mathbf{R}$ -types or all  $\mathbf{G}$ -types.  $\square$  (3.3)

Assume  $G'$  represents all  $\mathbf{G}$ -types.  $Q_1$  divides  $X$  into two regions: green and red-white. By assumption  $G'(L_1) \subset G'(L')$  represents all  $\mathbf{G}$ -types. But  $G'(L_1)$  is the graph of intersection between the surfaces  $P$  and  $Q_1$  on  $P$ . Furthermore  $G'(L_1)$  contains no 1-sided faces. Since  $Q_1$  has fewer middle slabs above it (the green region) we are done by induction.

So we assume  $G'$  represents all  $\mathbf{R}$ -types and therefore have a collection  $\mathbf{E}$  of disks representing all  $\mathbf{R}$ -types. Let  $\widehat{Q}_1$  and  $\widehat{Q}$  be the level spheres obtained by

capping off the components of  $\partial Q_1$  and  $\partial Q$  by meridional disks in a tubular neighborhood of  $K$ . We will also call the region of  $K(\gamma) \cong S^3$  between  $\widehat{Q}_1$  and  $\widehat{Q}$  the red region. Then  $N(K) \cap (\text{red region})$  is a disjoint union  $\mathbf{H}$  of 1-handles  $\{H_1, \dots, H_n\}$ . Fix an orientation of the cocores of these 1-handles. For every  $E \in \mathbf{E}$ ,  $\text{int } E \cap \mathbf{H} = \emptyset$  and  $\partial E \subset Q_1 \cup \partial \mathbf{H} \cup Q$ .

Let  $\mathbf{H} = \mathbf{H}_0 \sqcup \mathbf{H}_1 \sqcup \mathbf{H}_2$  where

$$\begin{aligned} \mathbf{H}_1 &= \{H \in \mathbf{H} \mid H \text{ has both feet in } \widehat{Q}_1\}, \\ \mathbf{H}_2 &= \{H \in \mathbf{H} \mid H \text{ has both feet in } \widehat{Q}\}, \\ \mathbf{H}_0 &= \{H \in \mathbf{H} \mid H \text{ has one foot in } \widehat{Q} \text{ and one foot in } \widehat{Q}_1\}. \end{aligned}$$

Correspondingly  $\{1, \dots, n\} = N_0 \sqcup N_1 \sqcup N_2$  where  $H_i \in \mathbf{H}_j$  if and only if  $i \in N_j$  ( $j = 0, 1, 2$ ).

There is an obvious 1-1 correspondence between  $\mathbf{R}$  and  $\mathbf{H}$ . For each  $i \in N_0$ , let  $\eta_i = \pm$  be such that a transverse arc on  $H_i$  intersecting the cocore of  $H_i$  with sign  $\eta_i$  takes us from the foot of  $H_i$  in  $\widehat{Q}$  to the foot of  $H_i$  in  $\widehat{Q}_1$ . Let  $\tau$  be any  $\mathbf{R}$ -type  $(\varepsilon_1, \dots, \varepsilon_n)$  such that  $\varepsilon_i = \eta_i$  when  $i \in N_0$ . Then any disk  $E \in \mathbf{E}$  such that  $\partial E$  represents  $\tau$  must meet  $\mathbf{H}$  only in  $\mathbf{H}_1$  or only in  $\mathbf{H}_2$ . Note that if  $E \in \mathbf{E}$  and  $(\alpha_1, \dots, \alpha_n)$  corresponds to  $\partial E$  then  $\sum_{i=1}^n |\alpha_i| \geq 2$  since there are no 1-sided faces in  $G$ . Applying Lemma 3.1, there is a  $j = 1, 2$  such that  $\mathbf{H}_j \neq \emptyset$  and a subset  $\mathbf{E}_j$  of  $\mathbf{E}$  such that

- (1) for all  $E \in \mathbf{E}_j$ ,  $\partial E$  meets only the handles  $\mathbf{H}_j$ ,
- (2)  $\mathbf{E}_j$  represents all  $\mathbf{H}_j$ -types.

Without loss of generality assume  $j = 2$ . Let  $B$  be the ball bounded by  $\widehat{Q}$  that does not contain the red region. Setting  $\mathbf{E} = \mathbf{E}_2$  and  $\mathbf{H} = \mathbf{H}_2$  we are in the setting of Proposition 3.2 (the handles of  $\mathbf{H}_2$  are standard). But this proposition contradicts the fact that  $\mathbf{E}_2$  represents all  $\mathbf{H}_2$ -types.  $\square$

### REFERENCES

[CGLS] M. Culler, C. McA. Gordon, J. Luecke, and P. B. Shalen, *Dehn surgery on knots*, Ann. of Math. (2) **125** (1987), 237–300.

[Cerf] J. Cerf, *Sur les difféomorphismes de la sphère de dimension trois* ( $\Gamma_4 = 0$ ), Lecture Notes in Math., no. 53, Springer-Verlag, Berlin and New York, 1968.

[FW] C. D. Feustel and W. Whitten, *Groups and complements of knots*, Canad. J. Math. **30** (1978), 1284–1295.

[Ga] D. Gabai, *Foliations and the topology of 3-manifolds*. III, J. Differential 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.

[J] K. Johannson, *Équivalences d'homotopie des variétés de dimension 3*, C. R. Acad. Sci. Paris Sér. A **281** (1975), 1009–1010.

- [L] R. A. Litherland, *Surgery on knots in solid tori*, II, J. London Math. Soc. (2) **22** (1980), 559–569.
- [T] H. Tietze, *Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten*, Monatsh. Math. Phys. **19** (1908), 1–118.
- [W] F. Waldhausen, *Recent results on sufficiently large 3-manifolds*, Proc. Sympos. Pure Math., Vol. 32, Part 2, Amer. Math. Soc., Providence, RI, 1978, pp. 21–38.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TEXAS, AUSTIN, TEXAS 78712