TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 306, Number 2, April 1988

PRIMENESS AND SUMS OF TANGLES

MARIO EUDAVE-MUÑOZ

ABSTRACT. We consider knots and links obtained by summing a rational tangle and a prime tangle. For a given prime tangle, we show that there are at most three rational tangles that will induce a composite or splittable link. In fact, we show that there is at most one rational tangle that will give a splittable link. These results extend Scharlemann's work.

1. Introduction. A *tangle* (B, t) is a pair that consists of a 3-ball B and a pair of disjoint arcs t properly embedded in B. Two tangles are *equivalent* if there is a homeomorphism h between the pairs. Two tangles are *equal* if there is a homeomorphism $h\colon (B, t) \to (B, t')$ of pairs such that $h|\partial B = \mathrm{id}$.

A *trivial tangle* is a tangle equivalent to the standard pair $(D^2 \times I, \{u, v\} \times I)$, u, v in the interior of D^2, $u \neq v$. A *rational tangle* is an element of an equivalence class of trivial tangles under the equality relation; there is a one to one correspondence between the rational tangles and $\mathbf{Q} \cup \{1/0\}$ (see [**C**, $\mathbf{M}_1$]). Let $(B, p/q)$ denote the tangle determined by p/q, the standard pair is $(B, 1/0)$; we denote the homeomorphism between $(B, 1/0)$ and $(B, p/q)$ as trivial tangles by $h_{p/q}\colon (B, 1/0) \to (B, p/q)$.

Let J be a meridian of $(B, 1/0)$ as in Figure 1. Let $(B, p/q)$ and $(B, r/s)$ be two rational tangles. The distance between them denoted $d((B, p/q), (B, r/s))$ or more simply $d(p/q, r/s)$, is defined to be the minimum (over all the representatives) of $\frac{1}{2}\#(h_{p/q}(J) \cap h_{r/s}(J))$. It can be shown that $d(p/q, r/s) = |ps - qr|$.

A tangle (B, t) is *prime* if has the following properties: (a) It has no local knots, that is, any S^2 which meets t transversely in two points, bounds in B a ball meeting t in an unknotted spanning arc; (b) there is no disc properly embedded in B which separates the strings of (B, t). We refer to [**L**] for definitions and facts about tangles not found here.

Let k be a knot or link in S^3. k is *splittable* if there is a S^2 in $S^3 - k$ that separates the components of k. k is *composite* if there is a S^2 in S^3, which meets k transversely in two points, such that neither of the closures of the components of $S^3 - S^2$ meets k in a single unknotted spanning arc. k is *prime* if it is neither splittable, nor composite, nor trivial.

To sum a rational tangle (B', r) to a tangle (B, t) means the following: take an embedding of (B, t) into S^3 and also an embedding of $(B', 1/0)$ and join them as in Figure 2(a), now replace $(B', 1/0)$ by $h_r((B', 1/0)) = (B', r)$ as in Figure 2(b).

Received by the editors February 26, 1987. Presented at the XIX congress of the Mexican Mathematical Society, held at Guadalajara, Mexico, on November 16–21, 1986.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 57M25; Secondary 57N10.

Key words and phrases. Prime tangle, rational tangle, prime knot and link, composite knot and link, splittable link.

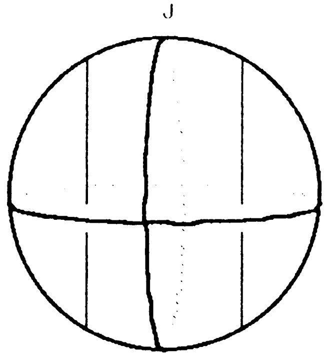


FIGURE 1

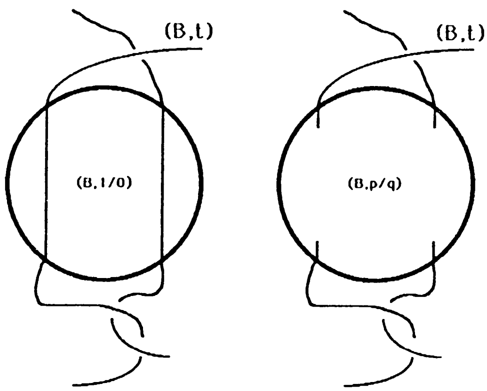


FIGURE 2

Let (B, t) be any tangle, fix an embedding of (B, t) in S^3. Let (B', r_i), $i = 1, 2$, be two rational tangles, and let k_i be the knot or link obtained by summing (B, t) and (B', r_i). Our results are the following:

THEOREM 1. *Let (B, t) be a prime tangle. If k_1 and k_2 are composite then* $d(r_1, r_2) \leq 1$.

THEOREM 2. *Let (B, t) be a prime tangle. If k_1 is composite and k_2 is splittable, then* $d(r_1, r_2) \leq 1$.

THEOREM 3. *Let (B, t) be any tangle. If k_1 and k_2 are splittable, then $r_1 = r_2$.*

In this same direction there are the following results.

THEOREM 4. *Let (B, t) be a prime tangle. If k_1 and k_2 are trivial knots, then* $r_1 = r_2$ [$\mathbf{BS}_1, \mathbf{BS}_2$].

THEOREM 5. *Let (B, t) be any tangle. If k_1 is a trivial knot and k_2 is splittable, then $d(r_1, r_2) \leq 1$, and (B, t) is a trivial tangle* [$\mathbf{S}_1$].

THEOREM 6. *Let (B, t) be a prime tangle. If k_1 is a trivial knot and k_2 is composite, then $d(r_1, r_2) \leq 1$* [**E**].

COROLLARY 1. *Given a prime tangle (B,t) there are at most three rational tangles (B',r_i), $i=1,2,3$, such that the knot or link k_i that results summing (B,t) and (B',r_i) is nonprime. Furthermore $d(r_i,r_j)\leq 1$.*

Theorems 1 and 6 and some results about branched double covers of S^3 branched over a knot or link (see [**M**$_2$, **KT**, **B**]) imply the following corollary.

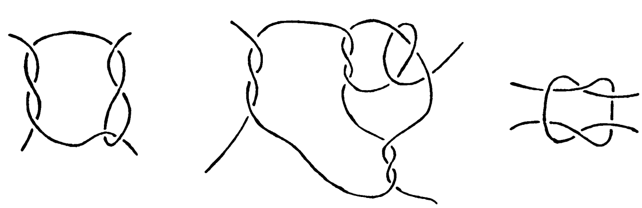
COROLLARY 2. *Let k be a strongly invertible knot in S^3, and $M(k,r)$ the manifold obtained by doing surgery with coefficient r on k. If $M(k,r_1)$ and $M(k,r_2)$ are reducible then r_1 and r_2 are integers and $|r_1-r_2|\leq 1$. If k is also amphicheiral, then $M(k,r)$ is irreducible for all coefficients r.*

W. B. R. Lickorish conjectured in [**L**] that given a prime tangle there is at most one rational tangle such that summing gives a nonprime knot; but S. A. Bleiler [**B**] found counterexamples, and he conjectured that given a prime tangle there is at most one rational tangle in each 'string attachment class' such that summing gives a nonprime knot or link. The truth of this conjecture is a consequence of Corollary 1. It can be observed that if for a given prime tangle there are three rational tangles such that summing yields three nonprime knots or links, then two of them must be knots and the third must be a link. It is unknown if there is a prime tangle that admits three distinct rational tangles such that summing yields three nonprime knots or links.

Theorem 6 is a generalization of Scharlemann's theorem "Unknotting number one knots are prime" [**S**$_2$]. In [**GL**] it is proved that for any knot k in S^3 the manifold $M(k,r)$ can be reducible only if r is an integer; this implies Theorem 6.

Theorems 1–6 are best possible, as is shown in the prime tangles of Figure 3. Abusing the terminology of Conway [**C**], the tangle 3(a) has 'numerator' the unknot and 'denominator' $3_1\#4_1$, the tangle 3(b) has numerator the square knot and denominator $3_1\#9_{37}$, and the tangle 3(c) has numerator the square knot and denominator the unlink.

In §§2, 3, 4, we prove Theorems 1, 2, 3 respectively. The techniques used here to prove the theorems are globally much the same as those of [**S**$_1$ and **S**$_2$]. The argument consists in converting the problem into a combinatorial problem on planar graphs, and contrasts conclusions based on the topology of the underlying problem with conclusions based on the combinatorics of the graph. We refer frequently to [**S**$_1$ and **S**$_2$], and their arguments are indispensable for the understanding of this paper.



(a) (b) (c)

FIGURE 3

2. Main topological and combinatorial arguments.

2.1 In this section we prove Theorem 1.

Let (B, t) be a prime tangle, and (B', r_1), (B', r_2) two rational tangles, such that $r = d(r_1, r_2) \geq 2$. In this section we use the indices a, b to denote 1 or 2, with the convention that when used together $\{a, b\} = \{1, 2\}$.

Let k_a be the knot or link obtained by summing (B, t) and (B', r_a). Suppose that k_a is composite, that is, there is a S^2 that meets k_a in two points, such that neither of the closures of the two components of $S^3 - S^2$ meets k_a in an unknotted spanning arc. Suppose that k_a is not a splittable link, we consider this case in §§3 and 4. We can suppose the following, after isotopies: (a) the strings of (B', r_a) are contained in $\partial B'$, this can be done because (B', r_a) is a trivial tangle; (b) S^2 meets k_a on the strings of (B, t); (c) the intersections of S^2 and ∂B are all essential circles in ∂B-{strings of (B', r_a)}, such that each of these circles is the boundary of a disc in S^2 whose interior does not meet ∂B. Let S_a be a sphere in S^3, with the above mentioned properties such that the number of intersection circles in $S_a \cap \partial B$ is minimized.

Now let $P_a = B \cap S_a$, hence P_a is a planar surface in B. ∂P_a is formed of n_a circles, parallel to $h_r(J)$ denoted by $a_1, a_2, \ldots, a_{n_a}$, labelled so that a_i and a_{i+1} cobound an essential annulus contained in ∂B-{strings of (B', r_a)} whose interior does not meet S_a, for $1 \leq i \leq n_a - 1$. The points of $P_a \cap k_a$ are denoted a_+ and a_-.

Let a_i and b_j be components of ∂P_a and ∂P_b, respectively. We can suppose that $\#(a_i \cap b_j)$ is minimum; that is, equal to $2r$. The circle a_i meets circles of ∂P_b as follows: first it meets b_1, then $b_2, \ldots, b_{n_b}$, then $b_{n_b}, \ldots, b_1$, then again $b_1, \ldots, b_{n_b}$, and so on successively until we return to the starting point. b_j meets ∂P_a similarly, see Figure 4. Label the points of intersection between a_i and b_j with j in a_i, and with i in b_j.

The intersection of k_a and k_b consists of two arcs, the strings of (B, t), together with $2r$ points, transversal crossings on ∂B. k_b meets P_a in the following points: a_+ and a_-, and $2r$ points in each one of the components of ∂P_a, the latter occur between the labels 1-1 and n_b-n_b.

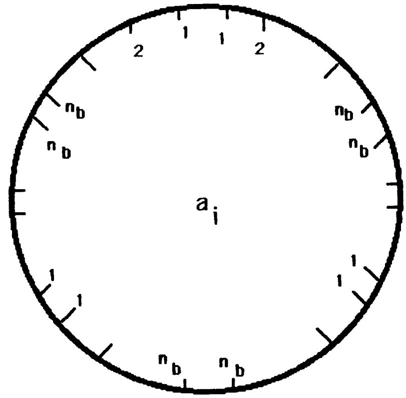
2.2 CLAIM. Both P_1 and P_2 are incompressible in B-{strings of (B, t)}.

PROOF. Suppose that P_a is compressible, and let D be a compression disc. If ∂D is essential in $S_a - \{a_+, a_-\}$ then do disc surgery on S_a with D, giving a sphere that meets k_a in one point, but this is impossible because S^3 is irreducible. If ∂D is not essential in $S_a - \{a_+, a_-\}$, because k_a is not splittable, an isotopy of S_a reduces $\#(S_a \cap \partial B)$, contradicting minimality.This completes the proof.

$P_a \cap P_b$ consists of arcs and circles, and by the incompressibility of P_a and P_b, it can be supposed that all the intersection circles are essential in both P_a and P_b.

CLAIM. There is no intersection circle between P_a and P_b, such that one of the discs determined by it in S_a has in its interior a_+ (a_-) but has neither a_- (a_+) nor any of the a_i's.

PROOF. Suppose there is one such curve, take an innermost, let this be c, and let D be the disc determined by c in S_a as above; look at c in S_b, if c is not essential in $S_b - \{b_+, b_-\}$ we can construct a sphere meeting k_b transversely in one point, but this is impossible. If c is essential in $S_b - \{b_+, b_-\}$, there are two possibilities:



labelling about a_i

FIGURE 4

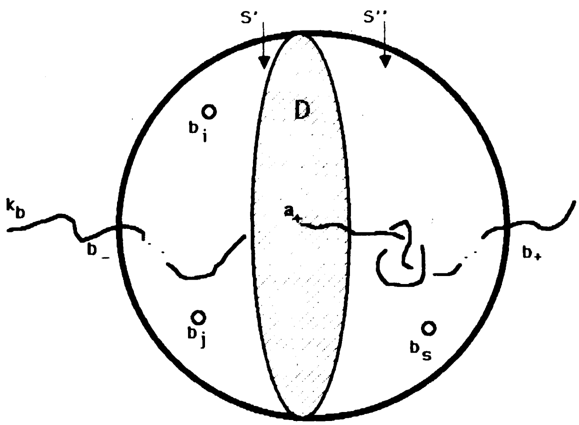


FIGURE 5

(a) c is the boundary of a disc in S_b that contains in its interior b_+ (or b_-) but it does not contain a b_j.

In this case, since (B,t) is a prime tangle, an isotopy removes the intersection.

(b) The two discs determined by c in S_b have in their interior some of the b_j's.

Doing surgery in S_b with D gives two spheres S' and S'', both spheres meet k_b in two points, but at least one of them separates k_b in two nontrivial parts, since otherwise S_b would separate k_b in two parts, one of them a trivial arc; but this is not possible because S' and S'' have less intersection circles with ∂B than S_b (see Figure 5). This completes the proof.

2.3. We construct a graph in S_a as follows: the vertex set is formed by the a_i's, and the edges are the intersection arcs between P_a and P_b. We denote the graph by G_a (similarly G_b). The ends of each edge are labelled by some number; if the two labels are different, orient the edge from the higher label to the lower. We do not consider a_+, a_- as vertices, because there is no intersection arc incident to them.

We define *circuit, cycle, semicycle, source, sink, loop, unicycle, level edge, interior vertex, chord, label sequence, interior label*, as in [$\mathbf{S}_1$, 2.4] In addition, we allow an edge of a circuit to be a loop.

The *interior* of a circuit in G_a is the component of its complement that does not contain a_-. A circuit of G_a is *bad* if it contains a_+ in its interior, otherwise it is *good.* A *double loop* is a circuit formed by two loops c_1, c_2 based at same vertex, and such that c_1 is in the interior of c_2.

2.4 LEMMA. (1) *No chord of an innermost cycle or semicycle is oriented.*

(2) *If an innermost cycle or semicycle has an interior vertex it must have an interior source or sink.*

(3) *Any loop which has interior vertices has in its interior either a sink or source or a cycle.*

(4) *A semicycle with exactly one level edge has in its interior either a source or sink, or a cycle, or a loop, or a semicycle with exactly one level edge and without interior vertices or chords.*

PROOF. It is similar to [$\mathbf{S}_1$, 2.5].

2.5 LEMMA. $n_a > 0$.

PROOF. If $n_a = 0$, S_a does not meet ∂B, hence it is contained in the interior of B; but (B, t) is a prime tangle, and so S_a is the boundary of a 3-ball meeting k_a in an unknotted spanning arc, contradicting the choice of S_a. This completes the proof.

2.6 LEMMA. *A good loop in G_a has interior vertices.*

PROOF. Suppose that in G_a there is a good loop without interior vertices. Take an innermost such loop, let this be γ based at a_i; its labels at a_i are adjacent and are $j, j+1$, or n_b, n_b, or 1, 1 (see Figure 4). In the first case the disc determined by the interior of γ together with the annulus in ∂B bounded by b_j and b_{j+1} can be used to obtain a compression disc for $P_b - \{b_+, b_-\}$, but this is not possible.

So suppose that the ends of γ are labeled 1 (the remaining case is identical). Let D be the disc in S_a determined by the interior of γ, and let α be the arc of a_i contained in the interior of γ, then $\partial D = \gamma \cup \alpha$. Consider the arc γ in G_b, a loop based at b_1 with ends labeled i. Let E be the disc in S_b determined by the interior of γ in G_b, and β be the arc of b_1 contained in the interior of γ. Then $\partial E = \gamma \cup \beta$, $\alpha \cap \beta = \partial\alpha = \partial\beta = \partial\gamma$. There is disc F properly embedded in B' with interior disjoint of S_b, and such that $\partial F = \alpha \cup \beta$, as in Figure 6. k_b meets $D \cup F$ only in one point, this intersection occurs over α. There are two subcases.

(a) γ in G_b is a good loop.

In this case k_b does not meet E, so $D \cup E \cup F$ is a sphere which intersects k_b in one point, but this is not possible.

(b) γ in G_b is a bad loop.

Let E' be the other disc in S_b determined by $\gamma \cup \beta$. k_b intersects each one of E and E' in one point. Doing surgery on S_b with $D \cup F$ gives two spheres, $S = D \cup F \cup E$ and $S' = D \cup F \cup E'$ (Figure 7); each one of them meets k_b in two points, and at least one of them, say S, must separate k_b into two nontrivial parts (that is, none of the parts is an unknotted spanning arc), since otherwise S_b

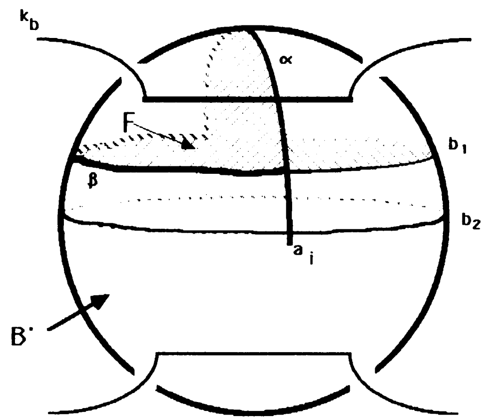


FIGURE 6

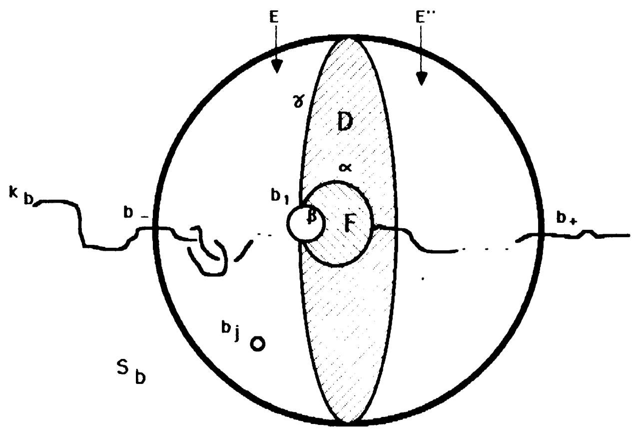


FIGURE 7

would separate k_b so that one of the parts would be a trivial arc. After an isotopy we can suppose that $S \cap \partial B$ consists of essential circles in ∂B-{strings of (B', r_b)}, and that $S \cap k_b$ lies in the strings of (B, t). There are two such possible isotopies, choose the one which eliminates b_1, so that $\#\partial P \leq n_b - 1$ $(P = S \cap B)$, but this contradicts the minimality of $\#\partial P_b$. This completes the proof.

2.7 LEMMA. *Let c be a bad level loop in G_a without vertices or edges in its interior. Then the corresponding level loop c in G_b is a good loop.*

PROOF. Suppose that the claim is false, that is, c is also a bad loop in G_b. The ends of c in G_a are labeled n_b or 1, w.l.o.g. suppose that they are labeled 1. c in G_b is a loop based at b_1. Let $D(E)$ be the disc in $S_a(S_b)$ determined by the interior of c in $G_a(G_b)$; $D \cap E = c$, because by 2.2 there is no intersection circle of S_a and S_b in D. As in the previous lemma, there is a disc F properly embedded in B', with interior disjoint of S_b, such that $F \cap (D \cup E) = \partial F = \partial(D \cup E)$. k_b meets the sphere $F \cup D \cup E$ in three points, one in the interior of $D(a_+)$, one in the interior of $E(b_+)$, and the other in $\partial F \cap \partial D$, but this is not possible, because S^3 is irreducible. This completes the proof.

2.8 LEMMA. *A semicycle in G_a with neither chords nor interior vertices cannot have exactly one level edge.*

PROOF. See [$\mathbf{S}_2$, 5.4].

2.9 LEMMA. *An innermost cycle in G_a with more than one edge has interior vertices.*

PROOF. Suppose this is false, let c be an innermost cycle without interior vertices. Let D be the disc determined by the interior of c; by 2.4 there is no oriented edge in D, and an application of 2.8 shows that there is no level edge in D; so there is no edge in D, and by 2.2 there is no intersection circle between S_a and S_b in D. Because c has at least two edges we can construct a punctured lens space as in [$\mathbf{S}_2$, 5.6], even if a_+ is in D. But this is not possible. Therefore the only cycles that may have no interior vertices are those cycles which have a bad unicycle in its interior.

2.10 LEMMA. *Let c be a cycle or a loop in G_a, then either*

(a) *c has in its interior a source or sink at which no loop is based*; *or*

(b) *c is or c has in its interior a bad level loop without chords or interior vertices*; *or*

(c) *c is or c has in its interior a bad unicycle without chords or interior vertices.*

PROOF. Take a cycle or loop σ contained in the interior of c, such that σ has no cycle or loop in its interior. If σ is a good loop or a cycle with more than two edges, then by 2.6 and 2.9 σ has vertices in its interior, and by 2.4 it has a source or sink in its interior, the election of σ implies this source or sink has no loops. So we have (a) unless σ is a bad loop. If σ is a bad loop but it has interior vertices, then again by 2.4 and the election of σ, there is a source or sink where no loop is based. If σ has no interior vertices, then σ has no chords, and σ is oriented or level, so we have (b) or (c). This completes the proof.

2.11 LEMMA. *If G_a has a bad level loop without chords or interior vertices, then G_b has a source or sink at which no loop is based.*

PROOF. Let c be this loop in G_a, by 2.7, c in G_b is a good loop. By 2.10 c in G_b must have an interior source or sink at which no loop is based, because as c is a good loop, (b) and (c) of 2.10 cannot happen. This completes the proof.

Two edges in G_a are parallel if they bound a disk in P_a, thus either they are loops based at the same vertex, or they join two distinct vertices, but in any case the circuit they form has no interior vertices.

2.12 LEMMA. *Let $e_1, e_2, \dots, e_p$ be parallel edges in G_a, then either*

(a) *each e_i is level*; *or*

(b) *each e_i is oriented.*

If (b) *then either*

(b′) *$p \leq n_b$*; *or*

(b″) *there are two edges e_i, e_j that form a cycle.*

PROOF. The edges $e_1, \dots, e_p$ join u and v (possibly $u = v$); if some edge is oriented and the other is level, then there are two consecutive edges e_i and e_{i+1}, e_i is oriented and e_{i+1} level, i.e. a semicycle with exactly one level edge and with

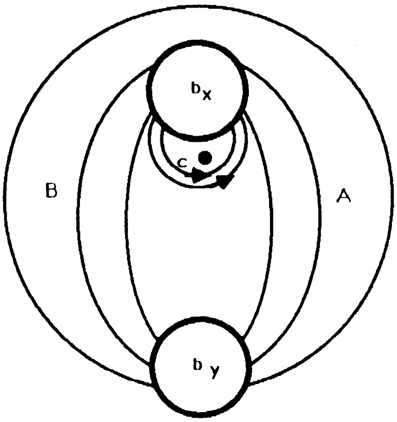


FIGURE 8

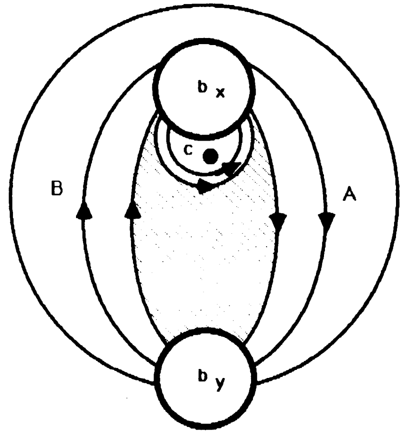


FIGURE 9

neither chords nor interior vertices, contradicting 2.8. Hence all the edges are level or all are oriented. Suppose that all the edges are oriented, that no two of them form a cycle, that $p > n_b$ and that the edges are oriented from u toward v. Take a subset of them that consists of exactly $n_b + 1$ consecutive edges, call them $e_1, \ldots, e_{n_b+1}$. Now the label of each e_i at $u(v)$ must be greater than 1 (less than n_b), since if the contrary occurs some e_i points toward u (points away from v). If the labels of e_1 are k and s at u and v respectively, the labels of e_{n_b+1} must be $n_b - k + 1$ and $n_b - s + 1$ at u and v, therefore $k > s$ and $n_b - k + 1 > n_b - s + 1$, that is $k > s$ and $k < s$, but this is not possible. This completes the proof.

2.13 LEMMA. *For each vertex v in G_a there is an i, $1 \leq i \leq n_b$, such that all the edges at v with label i are oriented. In particular $n_b > 1$.*

PROOF. Suppose this is false, then there is a vertex v in G_a such that for each i, $1 \leq i \leq n_b$, there is a level edge adjacent to v with label i. This implies that each vertex b_j in G_b is the base of a loop; if one of these loops is good, it is possible to find a good loop without interior vertices, contradicting 2.6; therefore suppose all the loops in G_b are bad. Then there is a vertex b_x in G_b such that the loops there have no interior vertices.

If $n_b = 1$, b_x is the only vertex in G_b, and all the edges are bad loops, by 2.12 all these loops are level or all are oriented. If all the loops are level, in G_a each vertex is the base of a loop and since in G_b there is a bad level loop without chords or interior vertices, in G_a there is a good loop by 2.7, and therefore it is possible to find a good loop in G_a without interior vertices, contradicting 2.6; if all these loops are oriented, by 2.12 we must have a cycle in G_b, but this cycle has no interior vertices, contradicting 2.9, therefore $n_b > 1$.

As $n_b > 1$ there is another vertex b_y in G_b such that the loops based there have only b_x as an interior vertex, hence all the edges at b_x are either loops or arcs joining b_x and b_y. Let c be an innermost loop based at b_x; there are two cases:

(1) c is level.

By 2.7 the corresponding loop c in G_a is good, so it is sufficient to prove that for each i, $1 \leq i \leq n_a$, there is incident to b_x one level edge with label i, because this

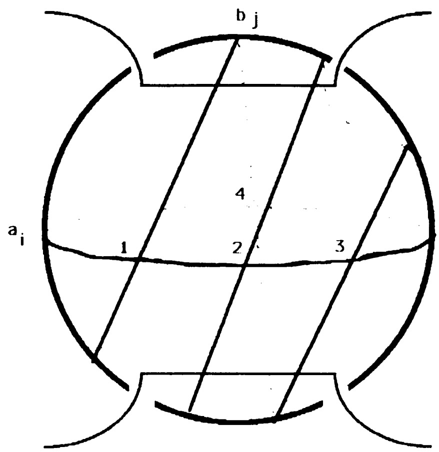


FIGURE 10

implies each vertex in G_a is the base of a loop, and as there is a good loop, there will be a good loop without interior vertices, contradicting 2.6.

The labels of c in b_x are $1, 1$ or n_a, n_a. If there are n_a or more loops in b_x, for each i there will be a level loop based at b_x with labels i. If there are less than n_a loops in b_x, there are at least $2n_a + 2$ edges connecting b_x and b_y; there are one or two sets of parallel edges, A and B, connecting b_x and b_y (see Figure 8); one of these sets, say A, has at least $n_a + 1$ edges, by 2.12 these edges are level, otherwise there is a good cycle without interior vertices; the labels of the edges of A and the labels of the loops in b_x are consecutive, so we have for each label i, $1 \leq i \leq n_a$, at least one level edge with label i in b_x.

(2) c is oriented.

By 2.12 there are at most n_a oriented loops in b_x. Let $r = d(r_1, r_2)$. There are at least $2(r-1)n_a$ edges connecting b_x and b_y, hence one or two sets of parallel edges, A and B, connecting b_x and b_y (see Figure 8). If one of these sets has more than n_a edges, these edges must be level by 2.12. Suppose $r \geq 3$, so $|A \cup B| \geq 4n_a$; if $|A| > n_a$ and $|B| > n_a$, these edges are level; if $|A| \geq 3n_a$ and $|B| \leq n_a$, the edges of A are level and the edges of B may or may not be level; anyway there are at least $3n_a$ level edges connecting b_x and b_y, and for each i, $1 \leq i \leq n_a$, there is at least one level edge with label i connecting these vertices. Then in G_a each vertex is the base of an oriented loop with labels x and y, so all the loops in G_a are bad. All the loops based on a vertex in G_a are parallel, so there are at most n_b loops at each vertex of G_a, so there are at most two loops based in a vertex with labels x, y. As there are at least $3n_a$ level edges connecting b_x and b_y which have consecutive lables in b_x, there is at least one label i, such that there are three level edges with label i connecting b_x and b_y. Then a_i in G_a is the base of three loops with labels x, y, but this is a contradiction.

Suppose now $r = 2$. We wish to prove that for each i, $1 \leq i \leq n_a$, there is one level edge with label i connecting b_x and b_y, and that for the labels $1, n_a$ there are two such level edges.

There are at most n_a loops at b_x. Suppose with no loss of generality the winding number of these loops with respect to a_+ is 1. There are at least $2n_a$ edges connecting b_x and b_y, all these edges can be oriented only if $|A| = |B| = n_a$, by 2.12, and in this case there are n_a loops at b_x. Suppose these edges are oriented. We have a situation like in Figure 9; at the right of c must be an edge e_a of A with label n_a in b_x, and at left of c must be an edge e_b of B with label 1 in b_x. Then $e_a(e_b)$ is oriented from b_x into b_y (b_y into b_x); a good cycle is formed with these edges and one loop in b_x, like in Figure 9, but this cycle does not have interior vertices, which is a contradiction.

So suppose the edges of A are level (the other case is similar). If the edges of B are level we are finished. So suppose the edges of B are oriented; these edges must be oriented from b_x into b_y, otherwise there would be a semicycle with exactly one level edge and without interior vertices or chords. There are four labels 1 in b_x, these labels cannot be ends of edges at B because these edges are oriented from b_x to b_y, and at most two of these labels are ends of the loops at b_x, so at least two of these labels are ends of edges of A. There are also two labels n_a in the end of edges of A, because of existence of the labels 1 in A and the orientation of the loops. So for each label i, there is one level edge with labels i connecting b_x and b_y, and there are two such level edges with label 1, n_a (these edges may not be parallel). Then each vertex in G_a is the base of a loop, and all these loops are bad.

Label the four points of intersection between a_i and b_j as j_1, j_2, j_3, j_4 (i_1, i_2, i_3, i_4) in a_i (b_j), so that a_i runs through them in the cyclic order j_1, j_2, j_3, j_4. The full set of labels in a_i is $1_1, 2_1, \ldots, n_{b_1}, n_{b_2}, \ldots, 1_2, 1_3, \ldots, 1_4$. Observe that b_j runs through the labels i_k in the cyclic order i_1, i_2, i_3, i_4, or its inverse, as is shown in Figure 10. If an edge α in G_a connects a_i and a_k with labels j_s and g_t respectively, then the corresponding edge α in G_b connects b_j and b_g with labels i_s and k_t respectively.

Consider only the vertices b_x, b_y, a_1, a_{n_a}. The labels of a_i when it meets b_x and b_y are ordered as follows: $x_1, x_2, y_2, y_3, x_3, x_4, y_4, y_1$, or $x_1, y_1, y_2, x_2, x_3, y_3, y_4, x_4$. The labels in b_x and b_y are ordered as $1_1, 1_2, n_{a_2}, n_{a_3}, 1_3, 1_4, n_{a_4}, n_{a_1}$, or $1_1, n_{a_1}$, $n_{a_2}, 1_2, 1_3, n_{a_3}, n_{a_4}, 1_4$, but equal or inverse in both b_x and b_y. There are two bad loops in a_1 with labels x, y, these loops have the same orientation (otherwise they form a good cycle without interior vertices), so we use exactly three subindices (e.g. 1, 2, 2, 3). The corresponding edges to these loops in G_b are two level edges connecting b_x and b_y with labels 1, furthermore we can suppose the labels 1 are adjacent in b_x. If also in b_y the ends are adjacent, we are using two or four subindices (e.g. 1, 2 in both b_x and b_y, or 1, 2 in one and 3, 4 in the other), but this is a contradiction. Now if the ends of these edges are not adjacent in b_y, then the ends of the two level edges with label n_a are adjacent in both b_x and b_y, so using n_a instead of 1 and repeating the argument, a contradiction is obtained; this is shown in Figure 11. This completes the proof.

2.14 LEMMA. *Let v be a vertex of G_a, suppose there is a family A of consecutive oriented edges that point into v, and a family B of consecutive oriented edges that point out from v; furthermore the last edge of A and the first of B (or vice versa) are adjacent at v. Then there is a set $\mathcal{L} = \{1, \ldots, n_b\}$ of n_b consecutive labels of v at which no edge of $A \cup B$ is incident, and the label 1 (n_b) is closer than n_b (1) to the labels of B (A) (that is, there is an arc of v, with interior disjoint from $\mathcal{L}$, A and B joining 1 (n_b) to a label of B (A)).*

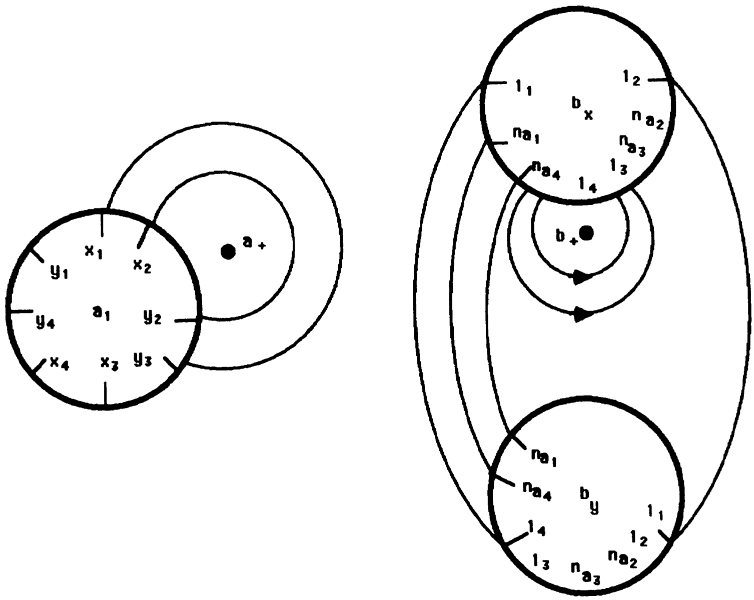


FIGURE 11

PROOF. Suppose with no loss of generality we have a situation as in Figure 12. Go through the labels of v in the counterclockwise direction and pick up the first label n_b (denote it by n_b^*) found after crossing the labels of A. Consider the set $\mathcal{K} = \{n_b^*, \dots, 1, 1, \dots, n_b, n_b, \dots, 1)$ of $3n_b$ consecutive labels of v, beginning with n_b^*, going in the clockwise direction. The ends of A (B) in v cannot be labeled with n_b (1), due to its orientation. So the labels of A in v are contained in the portion $n_{b-1}, \dots, 1, 1, \dots, n_{b-1}$ of $\mathcal{K}$, and the labels of B in the portion $n_{b-2}, \dots, 2$ or in $2, \dots, n_b, n_b, \dots, 2$; so the labels of $A \cup B$ are contained in $\mathcal{K}$. Take the set $\mathcal{L} = \{1, \dots, n_b\}$ of n_b consecutive labels of v, which is after $\mathcal{K}$ going in the clockwise direction. $\mathcal{L}$ does not overlap $\mathcal{K}$ because $r \geq 2$. Clearly $\mathcal{L}$ has the desired properties. This completes the proof.

2.15 LEMMA. *Let v be a vertex in G_a at which is based a bad unicycle without interior vertices. Then in G_a there is a source or sink where no loops are based.*

PROOF. All the loops based at v without interior vertices are oriented and parallel. Let c_1 be the bad unicycle without interior vertices or chords based at v. There are at most n_b bad unicycles based at v without interior vertices, let these be $c_1, c_2, \dots, c_m$. We can suppose there is no good loop in G_a, because if there is one by 2.10 we finish. Suppose that there is another bad loop, say c', based at u, so that there are no loops other than $c_1, \dots, c_m$ in its interior. Let D be the interior of c'. If $v \neq u$ an analogous argument to that of 2.14 shows that there are vertices in D other than v; if $v = u$ the choice of c' implies there are vertices in D. If there is no loop other than $c_1, \dots, c_m$ the proof is similar.

If there is a good cycle in D we are finished (by 2.10), so suppose all the cycles in D are bad. Let C be the set of all the bad cycles in D that have no interior vertices. Note that c_1 is in C, v is a vertex of each one of these cycles, and all

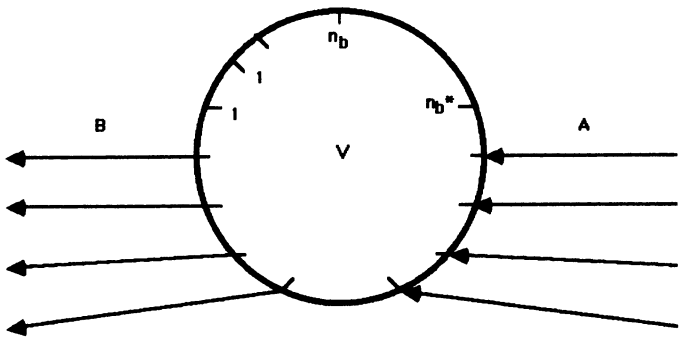


FIGURE 12

these cycles have the same winding number with respect to a_+ as c_1. We have a situation as in Figure 13. Let H be the subgraph of G_a defined as follows: {vertices of H}={vertices of G_a which are in D (u included)}, {edges of H}={edges of G_a which are in D except the edges of the cycles of C}. We have two cases.

(1) There is no source or sink in H (except possibly u).

Because there is at most one source or sink in H (but not both), there are cycles in H. Take one innermost, let this be σ. By the selection of C, σ has interior vertices, and by 2.4 σ has an interior source or sink, u cannot be in the interior of σ, so this is a contradiction.

(2) There is a source or sink in H (other than u).

If one vertex of H which is not a vertex of the cycles of C is a source or sink in H we are finished, so suppose none of these vertices is a source or sink. Suppose with no loss of generality that there is a vertex x in H, such that x is a source in H, and it is a vertex of the cycles of C. Let A (B) be the set of edges that belongs to the cycles of C which point into (out of) x. It is not difficult to see that the sets A, B satisfy the hypothesis of 2.14; so there is a set $\mathcal{L} = \{1, \ldots, n_b\}$ of consecutive labels of x at which edges of H are incident, and the label 1 (n_b) is closer than n_b (1) to the labels of B (A). Because x is a source in H, a level edge is incident to the label 1, let this be e'_x. By 2.13 there is a label i in $\mathcal{L}$ at which is incident an oriented edge, let this be e_x, this edge point out of x. We can suppose we have a situation as in Figure 14, so that the winding number of c_1 with respect to a_+ is 1, and e'_x is at the left of e_x.

Construct a path γ in H, starting with e_x, through oriented edges always consistent with its orientations. Finish the path when a vertex is repeated or when γ reaches u or a vertex of the cycles of C. Construct another path γ' in H, starting with e'_x, through oriented edges (except e'_x) always inconsistent with its orientations. Finish the path when a vertex is repeated, or when γ' reaches u, or a vertex of γ, or a vertex of the cycles of C. We have the following cases.

(a) The path γ repeats a vertex.

Then a cycle σ is formed, this cycle must be a bad one and contain all the vertices of the cycles of C in its interior. There is a path σ' which joins e_x with σ ($\sigma \cup \sigma' = \gamma$). Consider the path γ', if γ' finishes at a vertex of γ or at a vertex of the cycles of C, then with the path γ', a part of an outermost cycle of C (possibly empty), and a part of γ (possibly empty) a good semicycle in G_a with exactly one level edge is formed; this is ensured by the existence of σ' (see Figure 14). No vertex of the cycles of C is in the interior of this semicycle. So by 2.4 there is a good semicycle with exactly one level edge and without interior vertices or chords, but this contradicts 2.8. If the path γ' repeats a vertex, then a good cycle or a good semicycle with exactly one level edge is formed (this is ensured by the existence of σ'), the same argument as above yields a contradiction.

(b) γ finishes at u.

The same argument as in case (a) yields a contradiction.

(c) γ finishes at a vertex of the cycles of C.

γ together with a part of a cycle of C form a cycle in G_a, this cycle either is good or it is bad and contains e'_x in its interior, now we proceed as in case (a).

In the above argument it was important that e'_x be at the left of e_x, because if e'_x had been at the right of e_x, then no contradiction would be obtained. This completes the proof.

2.16 LEMMA. *G_a or G_b has a source or sink where no loop is based.*

PROOF. By 2.13 there are oriented edges in G_a; if G_a has no cycles or loops, then there is a source or sink with the desired properties. If there is a cycle or a loop in G_a, then by 2.10, 2.11, and 2.15 there is a source or sink in G_a or G_b where no loop is based. This completes the proof.

Let p be an integer, $1 \leq p \leq n_b$, define a *p-biflow* to be a circuit in G_a with the following properties:

(a) All edges are oriented, with heads (tails) labeled p.

(b) All interior labels are integers greater than (less) p.

(c) There is precisely one vertex of the circuit (called the *base*) for which both incident edges point out (in) and one (called the *apex*) for which both incident edges point in (out).

(d) There are interior labels at the apex, in fact at least two.

This definition is equal to that of [$\mathbf{S}_2$, 4.4], except by the property (d), this property is necessary for the proof of 2.17. Define a *p-loop* to be a loop with one end labeled p and either all interior labels greater than, or all less than p. Define a *p-double loop* to be a double loop that is a cycle and such that the two edges have heads (tails) labeled p and all interior labels greater than (less than) p.

2.17 LEMMA. *Suppose that b_p is a source or sink in G_b and c is either a good p-biflow, or a good p-loop, or a good p-double loop in G_a, then in the interior of c there is a p-loop or a p-biflow.*

PROOF. It is similar to that of [$\mathbf{S}_2$, 6.2, 6.3].

2.18 LEMMA. *If b_p is a source or sink in G_b, then in G_a there are neither good p-biflows nor good p-loops nor good p-double loops.*

PROOF. If there is one of these circuits, there is an innermost, but this contradicts 2.17. This completes the proof.

2.19 LEMMA. *Suppose that b_p is a source or sink in G_b at which no loop is based, then in G_a there is either a good p-loop or a good p-biflow or a good p-double loop.*

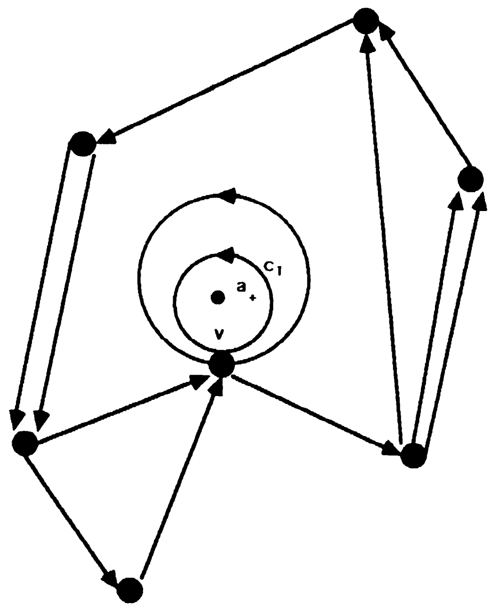


FIGURE 13

PROOF. It is essentially equal to [$\mathbf{S}_2$, 6.7].

The contradictions between the Lemmas 2.16, 2.18 and 2.19 complete the proof of Theorem 1.

3. Further applications of combinatorial techniques. I. Now we prove Theorem 2. Let (B,t) be a prime tangle, (B',r_1) and (B',r_2) two rational tangles such that $r = d(r_1,r_2) \geq 2$. Let k_1 be the knot or link obtained by summing (B,t) and (B',r_1), suppose that k_1 is composite. There is a S^2 that meets k_1 in two points, such that neither of the closures of the two components of $S^3 - S^2$ meets k_1 in an unknotted spanning arc. Suppose that k_2 is not a splittable link, we consider this case in §4. As in the previous section suppose that: (a) the strings of (B',r_1) are contained in $\partial B'$; (b) S^2 meets k_1 on the strings of (B,t); (c) the intersections of S^2 and ∂B are all essential circles in ∂B-{strings of (B',r_1)}, such that each of these circles is the boundary of a disk in S^2 whose interior does not meet ∂B. Let S_1 be a sphere in S^3, with the above-mentioned properties such that the number of intersection circles between it and ∂B is minimized.

Let k_2 be the link obtained by summing (B,t) and (B',r_2), suppose that k_2 is a splittable link, that is there is a S^2 disjoint of k_2 that separates the components of k_2. As before suppose that the strings of (B',r_2) are on ∂B and that the intersection circles between S^2 and ∂B are essential in ∂B-{strings of (B',r_2)}, and each of these circles is the boundary of a disk in S^2 whose interior does not meet ∂B. Let S_2 be a sphere as above which minimizes the number of intersection circles with ∂B.

Let $P_1 = S_1 \cap B$ and $P_2 = S_2 \cap B$, these are planar surfaces in B. ∂P_1 is formed by n circles denoted by $a_1, \ldots, a_n$, parallel to $h_{r_1}(J)$, labeled so that a_i and a_{i+1} cobound an essential annulus in ∂B-{strings of (B',r_1)} whose interior does not

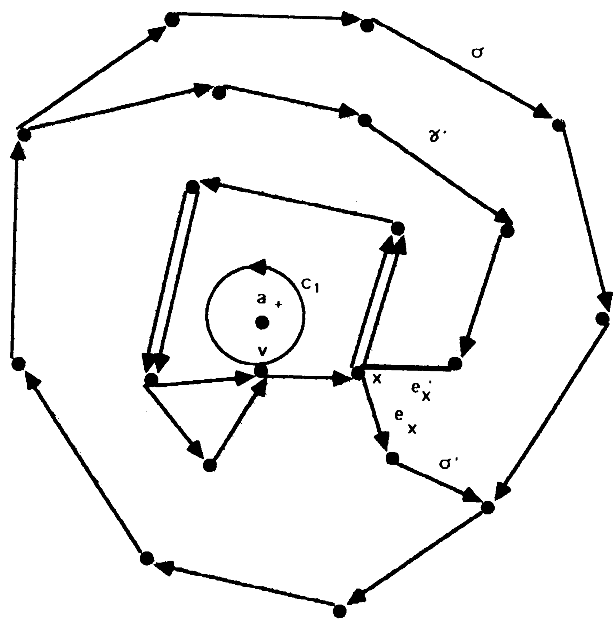


FIGURE 14

meet S_1, for $1 \leq i \leq n-1$. ∂P_2 is formed by m circles denoted by $b_1, \ldots, b_m$, parallel to $h_{r_2}(J)$, and labeled as in P_1. Furthermore m is odd. Denote the points of intersection between P_1 and k_1 by a_+ and a_-. The way that an a_i meets the b_j's is similar to that of §2.1 (see Figure 4).

P_1 and P_2 are incompressible in B-{strings of (B, t)}, hence we can suppose that all the intersection circles between P_1 and P_2 are essential in both $P_1 - \{a_+, a_-\}$ and P_2.

We construct graphs in S_1 and S_2 as before; the vertices are the a_i's and the b_j's respectively, and the edges are the intersection arcs between P_1 and P_2. Denote the graphs by G_1 and G_2. Label the ends of the edges and orient them as in the previous section.

The interior of a circuit in G_1 is the component of the complement of this circuit that does not contain a_-. A circuit in G_1 is good if it does not contain a_+ in its interior. Take a point $x \in P_2 - P_1$, define the interior of a circuit in G_2 to be the component of the complement of this circuit that does not contain x.

We have the following facts:

(1) $n > 0$.

(2) A loop (good loop) in G_2 (G_1) has interior vertices.

(3) A cycle (good cycle) in G_2 (G_1) has interior vertices.

The proofs of these facts are similar to those of §2.

(4) There are oriented edges in G_2. If all the edges of G_2 are level, then in G_1 all the edges are loops. All of them are bad loops, otherwise there is a good loop without interior vertices. Take any vertex in G_1, all the edges incident to it are bad loops; if they are level then in G_2 each vertex is the base of a loop, so there is a loop without interior vertices, a contradiction. If all the loops are oriented, then

because there are at least $2m$ loops, by 2.12 two of them form a good cycle with no interior vertices, a contradiction.

An easy application of these facts show that in G_2 there is a source or sink at which no loop is based. The Lemmas 2.17, 2.18, and 2.19 can be applied without difficulty in this case. In those Lemmas G_1 plays the role of G_a and G_2 the role of G_b.

This proves Theorem 2. The proof of this theorem is easier than the earlier one because in G_2 there are no bad circuits.

4. Further applications of combinatorial techniques. II. In this section we prove Theorem 3. Let (B,t) be any tangle, (B',r_1) and (B',r_2) two rational tangles. Suppose that summing (B,t) to (B',r_i), $i=1,2$, gives a link k_i, which is splittable. Suppose $r_1 \neq r_2$, so we have $d(r_1,r_2) \geq 2$ (any rational tangle to distance 1 of (B',r_1) will give a knot when summing to (B,t)). We use the indices a,b to denote 1 or 2, as in §2.

As k_a is splittable, there is a S^2 that does not meet k_a and that separates the components of k_a. As in the previous sections we can suppose the following: The strings of (B',r_a) are on ∂B; the intersections of S^2 and ∂B are all essential circles in ∂B-{strings of (B',r_a)}, such that each one of these circles is the boundary of a disk in S^2 whose interior does not meet ∂B. Let S_a be a sphere as above which minimizes the number of intersections circles with ∂B.

Let $P_a = S_a \cap B$, this is a planar surface. ∂P_a is formed by n_a circles denoted by $a_1,\dots,a_{n_a}$, parallel to $h_r(J)$, labeled so that a_i and a_{i+1} cobound an essential annulus in ∂B-{strings of (B',r_a)} whose interior does not meet S_a, for $1 \leq i \leq n_a - 1$. Both n_a and n_b are odd. The way that an a_i meets the b_j's is similar to that of §2.1, as in Figure 4.

P_a is incompressible in B-{strings of (B,t)}. We construct a graph G_a in P_a, as before. Take a point $x \in P_a - P_b$, define the interior of a circuit in G_a to be the component of the complement of this circuit that does not contain x.

We have the following facts: A loop in G_a has interior vertices; a cycle in G_a has interior vertices; there are oriented edges in G_a. The proofs of these facts are similar to the proofs of the previous sections. An easy application of those facts show that in G_a there is a source or sink at which no loop is based. Let v be a source (sink) in G_a at which no loop is based, all the edges incident to v with label 1 (n_b) are level, therefore in G_b, b_1 (b_n) is the base of several loops, all with one label i. An innermost such loop will be a i-loop. Lemma 2.18 can be applied in the present case, and hence we find a contradiction. This completes the proof of Theorem 3.

I would like to express my sincerest gratitude to J. C. Gomez Larrañaga for his supervision of this work which formed part of my Masters thesis at U.N.A.M. I would like to thank M. Scharlemann for his discovery of an error in one of the lemmas. I am grateful to H. Short and the referee for their suggestions.

REFERENCES

[B] S. A. Bleiler, *Prime tangles and composite knots*, Knot Theory and Manifolds, Lecture Notes in Math., vol. 1144, Springer-Verlag, Berlin and New York, 1985, pp. 1–13.

[BS$_1$] S. A. Bleiler and M. Scharlemann, *Tangles, property P, and a problem of J. Martin*, Math. Ann. **273** (1986), 215–225.

[$\mathbf{BS}_2$] ——, *A projective plane in $\mathbf{R}^4$ with three critical points is standard*, MSRI preprint.

[**C**] J. H. Conway, *An enumeration of knots and links, and some of their algebraic propertis*, Computational Problems in Abstract Algebra, Pergamon Press, Oxford and New York, 1969, pp. 329–358.

[**E**] M. Eudave-Muñoz, *Cirugía en nudos fuertemente invertibles*, An. Inst. Mat. Univ. Nac. Autónoma México **26** (1986), 41–57.

[**GL**] C. McA. Gordon and J. Luecke, *Only integral Dehn surgeries can yield reducible manifolds*, preprint.

[**KT**] P. K. Kim and J. L. Tollefson, *Splitting the P. L. involutions of nonprime 3-manifolds*, Michigan Math. J. **27** (1980), 259–274.

[**L**] W. B. R. Lickorish, *Prime knots and tangles*, Trans. Amer. Math. Soc. **267** (1981), 321–332.

[$\mathbf{M}_1$] J. M. Montesinos, *Variedades de Seifert que son recubridores cíclicos ramificados de dos hojas*, Bol. Soc. Mat. Mexicana (2) **18** (1973), 1–32.

[$\mathbf{M}_2$] ——, *Surgery on links and double branched covers of S^3*, Ann. of Math. Studies, no. 84, Princeton Univ. Press, Princeton, N. J., 1975, pp. 227–260.

[$\mathbf{S}_1$] M. Scharlemann, *Smooth spheres in $\mathbf{R}^4$ with four critical points are standard*, Invent. Math. **79** (1985), 125–141.

[$\mathbf{S}_2$] ——, *Unknotting number one knots are prime*, Invent. Math. **82** (1985), 37–55.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, SANTA BARBARA, CALIFORNIA 93106