ISSN 1464-8997 (on line) 1464-8989 (printed)

Geometry & Topology Monographs
Volume 7: Proceedings of the Casson Fest
Pages 213–233

Local surgery formulas for quantum invariants and the Arf invariant

Robion Kirby
Paul Melvin

Abstract A formula for the Arf invariant of a link is given in terms of the singularities of an immersed surface bounded by the link. This is applied to study the computational complexity of quantum invariants of 3–manifolds.

AMS Classification 57M27; 68Q15

Keywords Arf invariants, quantum invariants, mu invariants, surgery, immersed surfaces, P/NP, complexity

0 Introduction

The quantum 3–manifold invariant of Witten [38] and Reshetikhin–Turaev [28] with gauge group $SU(2)$ at the fourth root of unity is given by the formula [16]

$$\tau_4(M) = \sum_\theta \omega^{\mu(M_\theta)}$$

where ω is a primitive sixteenth root of unity, and the sum is over all spin structures θ on the closed oriented 3–manifold M. Here $\mu(M_\theta)$ is Rokhlin's invariant of M with its spin structure θ, that is, the signature modulo 16 of any compact spin 4–manifold with spin boundary M_θ. The set of spin structures on M is parametrized by $H^1(M;\mathbb{Z}_2)$, so at first sight the complexity of computing $\tau_4(M)$ grows *exponentially* with $b_1(M) = \operatorname{rk} H^1(M;\mathbb{Z}_2)$.

This note originated when Mike Freedman, motivated by the $\mathcal{P}$ versus $\mathcal{NP}$ problem in theoretical computer science, observed that the formulas in our paper [16] lead to a *polynomial time* algorithm for the computation of τ_3, and asked us what difficulties arise in trying to find such an algorithm to evaluate τ_4. As it turns out the computation of τ_4 is $\mathcal{NP}$–hard (and conjecturally not even in $\mathcal{NP}$) as we shall explain in section 2, although a polynomial time algorithm exists for the restricted class of 3–manifolds of "Milnor degree" greater than three.

Published 19 September 2004:

We thank Freedman for several discussions on this topic which led naturally to the "local" formulas given below. We are also grateful to László Lovász for sharing his computational complexity insights with us.

In the process of investigating this complexity question, we found a new formula for the Arf invariant of a classical link L in terms of data derived from an immersed surface F bounded by L whose singularities S are *internal*, ie away from ∂F. This formula, discussed in section 4 after some algebraic preliminaries in section 3, depends only on linking numbers of curves near $\partial F \cup S$, and on the Arf invariants (or Brown invariants if F is nonorientable) of quadratic forms defined on $H_1(F)$. For example, if F is a union of Seifert surfaces F_i for the individual components L_i of L, then the formula is expressed in terms of the Arf invariants of the L_i, the linking numbers between the L_i, $F_i \cap F_j$ and their push-offs, and the total number of triple points $\cup(F_i \cap F_j \cap F_k)$.

The authors thank the National Science Foundation and the Microsoft Research Group for support.

1 Local surgery formulas

It was observed by Casson (see [7]) and independently by Rokhlin [30] that $\mu(M_\theta)$ can be computed using *any* compact oriented 4–manifold W bounded by M by

$$\mu(M_\theta) = \sigma(W) - F \cdot F + 8\alpha(F) \pmod{16}.$$

Here $F \subset W$ is an oriented *characteristic surface* for θ — meaning θ extends over $W \setminus F$ but not across any component of F — with self intersection $F \cdot F$, and $\alpha(F) \in \mathbb{Z}_2$ is the Arf invariant of a suitable quadratic form on $H_1(F;\mathbb{Z}_2)$. (See the appendices of [31] or [15] for generalities on the Arf invariant.) If F is nonorientable, there is an analogous formula due to Guillou and Marin [9], replacing $8\alpha(F)$ by $2\beta(F)$ where $\beta(F) \in \mathbb{Z}_8$ is Ed Brown's generalization of the Arf invariant [2].

In particular, M can be described as the boundary of a 4–manifold W obtained from the 4–ball by adding 2–handles along a framed link L in S^3. Then the spin structures on M correspond to *characteristic sublinks* C of L, that is sublinks C satisfying $C \cdot L_i \equiv L_i \cdot L_i \pmod 2$ for all components L_i of L. Here $\cdot$ denotes linking or self-linking number, ie framing (see for example [16, Appendix C]). Note that linking numbers are only defined for *oriented* links, so we fix an orientation on L; the family of characteristic sublinks of L is independent of this choice, but we shall need this orientation for other purposes below.

For any given characteristic sublink C of L, an associated characteristic surface $F \subset W$ can be constructed by taking the union of an oriented Seifert surface for C, with its interior pushed into B^4, with the cores of the 2–handles attached to C. The choice of Seifert surface is immaterial since the invariants $F \cdot F$ and $\alpha(F)$ that appear in the formula for the μ-invariant are independent of this choice; indeed, $F \cdot F = C \cdot C$, the sum of all the entries in the linking matrix of C, and $\alpha(F) = \alpha(C)$, the Arf invariant of the oriented proper link C. (Recall that a link is *proper* if each component evenly links the union of the other components.) It follows that

$$\tau_4(M) = \omega^{\sigma(L)} \sum_C (-1)^{\alpha(C)} \omega^{-C \cdot C}$$

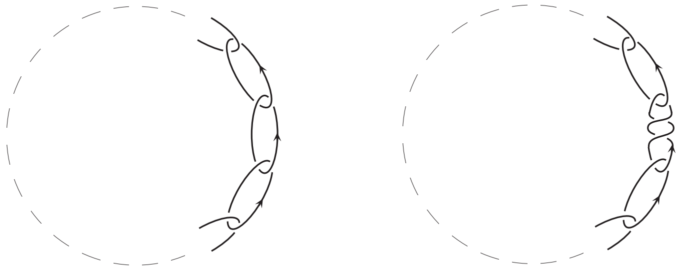
where $\sigma(L)$ is the signature of the linking matrix of L, and the sum is over all characteristic sublinks $C \subset L$. Since there are $2^{b_1(M)}$ such sublinks, this yields an exponential time evaluation of τ_4.

In fact the exponential nature of this formula is due solely to the Arf invariant factors, for without these, the formula could be evaluated in polynomial time. To see this note that the linking matrix of L can be stably diagonalized over $\mathbb{Z}$ (eg by [25]), which corresponds to adjoining a suitably framed unlink to L and then sliding handles [15]. Once L has been diagonalized, its characteristic sublinks are exactly those that include all the odd-framed components. It follows that if there are b_i components of L with framings congruent modulo 16 to $2i$, then

$$\sum_C \omega^{-C \cdot C} = \omega^{-s} \prod_{0 \le i \le 7} (1 + \omega^{-2i})^{b_i}$$

where s is the sum of all the odd framings in L.

Unfortunately the Arf invariant of a proper link C is global in the sense that its value depends simultaneously on *all* the components of C. For example the circular daisy chains in Figure 1a and 1b have different Arf invariants but identical families of sublinks (excluding the whole link). This casts doubt on the existence of a polynomial time algorithm for computing τ_4. However if C is *totally proper* — meaning all the pairwise linking numbers of C are even — then there exist *local formulas* for $\alpha(C)$, ie formulas that depend only on the sublinks of C with k or fewer components for some k (such a formula is called *k–local*). It is reasonable to attempt to exploit such local formulas in the search for an optimal algorithm for computing τ_4.



(a) (b)

Figure 1

One such formula arises from the work of Hoste [10] and Murakami [23]. They showed (independently) that the Arf invariant of a totally proper link C can be written as a sum

$$\alpha(C) = \sum_{S<C} c_1(S) \pmod 2$$

over all sublinks S of C.[†] Here $c_1(S)$ is the coefficient of z^{s+1} in the Conway polynomial of the s–component link S. It is known that $c_1(S) = 0$ if $s > 3$ (see for example [11]) and so this formula is in fact 3–local.

This formula can be expressed in geometric terms using familiar homological interpretations for the mod 2 reductions of the Conway coefficients $c_1(S)$. As noted above,

$$c_1(S) \equiv \alpha(S) \pmod 2$$

if S is a knot.

If S has two components, then $c_1(S)$ is determined by the linking number of the components and an unoriented version of the Sato–Levine invariant, as follows. The oriented Sato–Levine invariant [33] is defined for any oriented 2–component *diagonal* link (meaning pairwise linking numbers vanish) as the self-linking of the curves of intersection of any pair of Seifert surfaces for the components that meet transversely in their interiors; it was shown equal to c_1 by Sturm [34] and Cochran [4]. This invariant was extended to unoriented (totally) proper links S by Saito [32] by allowing nonorientable bounding surfaces for the components, meeting transversely in their interiors in a link C. One then defines

$$\lambda(S) = \mathrm{lk}(C, C^{\times}) \pmod 8 \ \in \ \mathbb{Z}_8$$

[†]This was first observed for knots by Kauffman [13], following Levine [18], and for 2–component links by Murasugi [24].

where $C^\times$ is the "quadruple push-off" of C (the union of the boundaries of tubular neighborhoods of C in the two surfaces, oriented compatibly with any chosen orientation on C). This is shown independent of the choice of bounding surfaces by a standard bordism argument (see [32] for details). It is clearly an even number, and a multiple of 4 for diagonal links; Saito's invariant is $\lambda/2 \in \mathbb{Z}_4$, and the oriented invariant is $\mathrm{lk}(C, C^\times)/4 \in \mathbb{Z}$. (In section 3 we study a closely related invariant δ, which can take on odd values as well.) Saito shows that in general $\lambda(S)$ is congruent mod 4 to the linking number $\mathrm{lk}(S)$ of the two components of S (with any chosen orientation), and that

$$c_1(S) \equiv \tfrac{1}{4}(\lambda(S) + \mathrm{lk}(S)) \pmod 2.$$

There is also a Sato–Levine invariant for 3–component diagonal links which counts the number of signed triple points of intersection of three oriented Seifert surfaces meeting only in their interiors. This invariant clearly depends on an orientation and ordering of the components. In fact it is equivalent (up to sign) to Milnor's triple linking number [22] [35], and its *square* is equal to the Conway coefficient c_1 for diagonal links [4]. To extend this invariant to totally proper links, one must reduce mod 2. Thus we allow nonorientable bounding surfaces and then count triple points mod 2. This gives a $\mathbb{Z}_2$–valued link concordance invariant τ by the usual bordism argument. In fact τ is a link homotopy invariant (eg by Cochran's argument in [4, Lemma 5.4]) which coincides with the mod 2 reduction of Milnor's triple linking number (cf [20]), and so we shall call it the *Milnor invariant*. Now it is not hard to show that

$$c_1(S) \equiv \tau(S) \pmod 2.$$

Indeed $c_1 \pmod 2$ is a link homotopy invariant for totally proper links of at least 3 components (by the proof of [4, Lemma 4.2]), and the congruence above is easily checked on Milnor's generators for 3–component link homotopy [21, page 23] (as in [4, page 539] in the diagonal case).

Putting these geometric evaluations of c_1 into the Hoste–Murakami formula yields the following:

Theorem 1.1 *If C is a totally proper link with components $C_1, \cdots, C_n$, then*

$$\alpha(C) = \sum_i \alpha(C_i) + \sum_{i<j} \tfrac{1}{4}(\lambda(C_i, C_j) + \mathrm{lk}(C_i, C_j)) + \sum_{i<j<k} \tau(C_i, C_j, C_k) \pmod 2$$

where α, λ and τ denote the Arf invariant, unoriented Sato–Levine invariant and Milnor triple point invariant, respectively.

In section 4 this theorem will be rederived as a corollary of a more general result expressing the *Brown invariant* of a link in terms of linking properties of the singularities of any immersed surface that it bounds.

2 Complexity

First recall, in rough terms, the complexity classes $\mathcal{P}$ = *polynomial time* and $\mathcal{NP}$ = *nondeterministic polynomial time* (see [26] for a more rigorous discussion). A computational problem is said to be in $\mathcal{P}$ if it can be solved by an algorithm whose run time on any given instance of the problem is bounded by a polynomial function of the size of the instance. If *answers* to the problem can be *checked* in polynomial time, then it is said to be in $\mathcal{NP}$. Of course any problem in $\mathcal{P}$ is in $\mathcal{NP}$, but the converse is not known; this is one of the central open problems in theoretical computer science.

There are a number of well-known $\mathcal{NP}$ problems, such as the travelling salesman and Boolean satisfiability (SAT) problems, whose polynomial time solution would yield polynomial time solutions for all $\mathcal{NP}$ problems, thus showing $\mathcal{P} = \mathcal{NP}$. These are called $\mathcal{NP}$*–complete* problems. Any problem (whether or not in $\mathcal{NP}$) whose polynomial time solution would yield polynomial time solutions for all $\mathcal{NP}$ problems is said to be $\mathcal{NP}$*–hard*.

With the formula in Theorem 1.1 (in fact the diagonal case is all that is needed) it is easy to show that the problem of calculating $\tau_4(M)$ for all 3–manifolds M is $\mathcal{NP}$–hard. The idea is to construct a class of 3–manifolds indexed by cubic forms over $\mathbb{Z}_2$ whose quantum invariants are given by counting the zeros of the associated forms, a well-known $\mathcal{NP}$–hard computational problem. In principle this construction goes back to Turaev's realization theorem for "Rokhlin functions" [36], but it can be accomplished more efficiently in the present setting as follows.

As a warmup, start with the 3–manifolds M_r obtained by zero-framed surgery on the links L_r (for $r = 1, 2, 3$) where L_1 is the trefoil, L_2 is the Whitehead link, and L_3 is the Borromean rings. Then M_r has 2^r spin structures given by the 2^r sublinks C of L_r. The μ-invariant is zero in all cases except when $C = L_r$, when it is 8 (coming from the Arf invariant if $r = 1$, the Sato–Levine invariant if $r = 2$, and the Milnor invariant if $r = 3$). Thus $\tau_4(M_r) = 2^r - 2$.

More generally, for any cubic form

$$c(x_1, \ldots, x_n) = \sum_i c_i\, x_i + \sum_{i,j} c_{ij}\, x_i x_j + \sum_{i,j,k} c_{ijk}\, x_i x_j x_k$$

in n variables $x_1, \ldots, x_n$ over $\mathbb{Z}_2$, let L_c be the framed link obtained from the zero-framed n–component unlink by tying a trefoil knot in each component L_i for which $c_i = 1$, a Whitehead link into any two components L_i, L_j for which $c_{ij} = 1$, and Borromean rings into any three components L_i, L_j, L_k for which $c_{ijk} = 1$. If M_c is the 3–manifold obtained by surgery on L_c, then the spin structure corresponding to any of the 2^n characteristic sublinks $C \subset L$ has μ–invariant $8c(x)$, where $x \in \mathbb{Z}_2^n$ is the n–tuple with 1's exactly in the coordinates corresponding to the components of C. Thus

$$\tau_4(M_c) = \sum_{x \in \mathbb{Z}_2^n} (-1)^{c(x)} = 2\#c - 2^n$$

where $\#c$ denotes the number of zeros of c (ie solutions to $c(x) = 0$).

Theorem 2.1 *For any cubic form c, the calculation of $\tau_4(M_c)$ is equivalent to the calculation of the number $\#c$ of zeros of c. The problem $\#\mathrm{C}$ of computing $\#c$ for all cubic forms is $\mathcal{NP}$–hard.*

Proof The first statement was proved above, and the last is presumably well known to complexity theorists. We thank L. Lovász for suggesting the following argument.

It is a fundamental result in complexity theory that the Boolean satisfiability decision problem SAT is $\mathcal{NP}$–complete (Cook's Theorem [6]) as is its "cubic" specialization 3–SAT. It follows that the associated counting problem #3–SAT is $\mathcal{NP}$–hard. This problem asks for the number $\#e$ of solutions to logical expressions in n variables $x_1, \ldots, x_n$ of the form

$$e = a_1 \wedge a_2 \wedge \cdots \wedge a_r$$

where each a_i is of the form $(x_j^{\pm} \vee x_k^{\pm} \vee x_\ell^{\pm})$. Here, each x_i can take the value T (true) or F (false); $x_i^+ = x_i$ and x_i^- is the negation of x_i; $\vee$ means *or* and $\wedge$ means *and*. Thus $\#e$ is the number of ways to assign T or F to each x_i so that the expression e is true.

To complete the proof of the theorem, it suffices to produce a polynomial time reduction of the problem #3–SAT to #C. To achieve this, consider any logical expression e as above, and rewrite it as a system of cubic equations over $\mathbb{Z}_2$ by setting $T = 0$ and $F = 1$ and replacing x_i^- by $1 - x_i$. Thus each a_i becomes an equation, eg $(x_j^- \wedge x_k^- \wedge x_\ell)$ becomes $(1 - x_j)(1 - x_k)x_\ell = 0$. The resulting system of r cubic equations in n variables has exactly $\#e$ solutions.

Now change this cubic system into a system of $k = 2r$ *quadratic* equations in $m \leq n + r$ variables

$$q = \begin{cases} q_1(x_1, \ldots x_m) = 0 \\ \vdots \\ q_k(x_1, \ldots x_m) = 0 \end{cases} \qquad (*)$$

also with exactly $\#e$ solutions. In particular, replace each cubic equation by two quadratics, the first assigning a new variable to the product of any two of the variables in the cubic, and the second obtained by substituting this into the cubic. For example $(1-x_j)(1-x_k)x_\ell = 0$ is replaced by $x_{jk} = x_j x_k$ and $(1-x_j-x_k+x_{jk})x_\ell = 0$.

Finally, convert q into a cubic equation by introducing k new variables z_i:

$$c = \sum_{i=1}^{k} z_i q_i(x_1, \dots x_n) = 0.$$

The number of solutions $\#c$ is equal to $2^k\#e + 2^{k-1}(2^m - \#e) = 2^{m+k-1} + 2^{k-1}\#e$, since any solution to $(*)$ allows any of 2^k choices for the z_i, and any non-solution to $(*)$ allows only 2^{k-1} choices for the z_i. Thus an algorithm to evaluate $\#c$ would yield one of the same complexity for $\#e$, and so #C is at least as hard as #3–SAT. $\square$

Corollary 2.2 *The calculation of $\tau_4(M)$ for all 3–manifolds is $\mathcal{NP}$–hard.*

In particular, Theorem 2.1 shows that this calculation for the special class of 3–manifolds M_c arising from cubic forms c is already $\mathcal{NP}$–hard (and presumably not in $\mathcal{NP}$, cf [26, section 18]).

Added in proof: In fact one need only consider the class of 3–manifolds M_q arising from *quadratic* forms $q(x_1,\dots,x_n) = \sum_i c_i\, x_i + \sum_{i,j} c_{ij}\, x_i x_j$, so the vanishing of the triple linking numbers does not reduce the complexity of the calculation if there are still pairwise Whitehead linkings. This follows by essentially the same proof, using the surprising result of Valiant [37] (brought to our attention by Sanjeev Khanna) that #2–SAT is also $\mathcal{NP}$–hard, although 2–SAT is in $\mathcal{P}$!

Remark There is a 3–manifold invariant that captures the complexity of the calculation of $\tau_4(M)$, namely the *Milnor degree* $d(M) \in \mathbb{N}$ introduced in [5, page 116]. This invariant can be defined by the condition $d(M) > n$ if M can be obtained by surgery on an integrally framed link whose $\overline{\mu}$–invariants of order $\le n$ vanish (where the *order* of a $\overline{\mu}$ invariant is one less than its length, eg the order–2 invariants are Milnor's triple linking numbers) [22]. It follows from the discussion in section 1 that there is a polynomial time algorithm for computing $\tau_4(M)$ for all 3–manifolds of Milnor degree > 3, and from the discussion above that the computation for 3–manifolds of Milnor degree ≤ 3 is $\mathcal{NP}$–hard.

3 The Brown invariant: algebra

The Brown invariant [3], which is a generalization of the Arf invariant, classifies $\mathbb{Z}_4$–enhanced inner product spaces over $\mathbb{Z}_2$. There are many excellent treatments of this subject in the literature (see eg [3, 27, 9, 19, 17]) but generally in the context of *nonsingular* spaces. For the reader's convenience, with apologies to the experts, we give an exposition which includes the case of singular forms (cf [14, 16, 8]).

The example to keep in mind is the space $H_1(F)$ with its intersection pairing, where F is a compact surface with boundary. (Throughout this paper, $\mathbb{Z}_2$–coefficients will be assumed.) The enhancements in this case arise from immersions of F in S^3, and these give rise to Brown invariants of the links on the boundary of F, as will be discussed in the next section.

3.1 Enhanced spaces

Let V be a finite dimensional $\mathbb{Z}_2$–vector space with a possibly singular inner product $(x, y) \mapsto x \cdot y$. Then V splits as an orthogonal direct sum

$$V = U \oplus V^{\perp}$$

where $\cdot$ is nonsingular on U and vanishes identically on $V^{\perp} = \{x \in V \mid x \cdot y = 0 \text{ for all } y \in V\}$. (For surfaces, the splitting of $H_1(F)$ arises from a decomposition $F = C\#D$ where C is closed and D is planar, and so $H_1(F)^{\perp}$ is the image of the map $H_1(\partial F) \to H_1(F)$ induced by inclusion.)

A standard diagonalization argument shows that U splits as a sum of indecomposables of one of two types: the 1–dimensional space P defined by $x \cdot x = 1$ on any basis x (corresponding to the real projective plane) and the 2–dimensional space T defined by $x \cdot x = y \cdot y = 0$ and $x \cdot y = 1$ on any basis x, y (the torus). Similarly $V^{\perp}$ is a sum of trivial 1–dimensional spaces A where $x \cdot x = 0$ (a boundary connected sum of annuli). Thus V is built from the indecomposables P, T and A, given by the matrices

$$P = (1) \qquad T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad A = (0).$$

The only relations among these spaces follow from the well-known isomorphism $P \oplus T \cong 3P$ $(= P \oplus P \oplus P)$. Hence V is uniquely expressible as a sum of copies of T and A if it is *even* (ie $x \cdot x = 0$ for all $x \in V$, which corresponds to orientability for surfaces), and of copies of P and A otherwise.

Now equip V with a $\mathbb{Z}_4$–valued *quadratic enhancement*, that is a function

$$e\colon V \to \mathbb{Z}_4$$

satisfying $e(x+y) = e(x) + e(y) + 2(x \cdot y)$ for all $x, y \in V$.

If e vanishes on $V^\perp$, then it is called a *proper* enhancement. The pair (V, e), also denoted V_e, is called an *enhanced space*. Observe that e is determined by its values on a basis for V, and these values can be arbitrary as long as they satisfy $e(x) \equiv x \cdot x \pmod 2$. Thus there are $2^{\dim V}$ distinct enhancements of V. However many of these may be isomorphic, and the Brown invariant

$$\beta\colon \{\text{enhanced spaces}\} \to \mathbb{Z}_8^*,$$

where $\mathbb{Z}_8^* = \mathbb{Z}_8 \cup \{\infty\}$, provides a complete isomorphism invariant.

3.2 The Brown invariant of an enhanced space

Let V_e be an enhanced space. Perhaps the simplest definition of the Brown invariant $\beta(V_e)$ is based on the relative values of e_0 and e_2, and of e_1 and e_3, where e_i denotes the number of $x \in V$ with $e(x) = i$, according to the scheme indicated in Figure 2.† For example $\beta = 7$ iff $e_0 > e_2$ and $e_1 < e_3$, and $\beta = \infty$ iff $e_0 = e_2$ and $e_1 = e_3$.

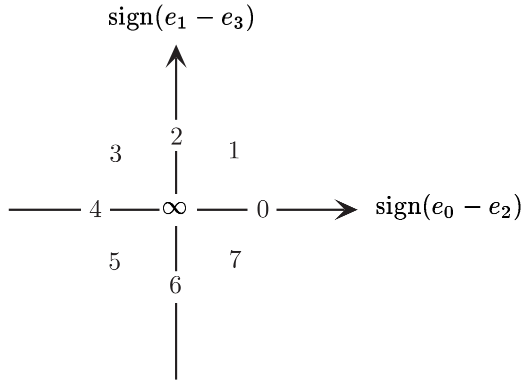


Figure 2: The Brown invariant

†This definition is in the spirit of the characterization of the Arf invariant for $\mathbb{Z}_2$–valued enhancements e (where $e(x+y) = e(x) + e(y) + x \cdot y$) in terms of the relative values of e_0 and e_1 (see eg [1]): $\alpha = 0$, 1 or ∞ according to whether $e_0 - e_1$ is positive, negative or zero. Observe that such an e can also be viewed as a $\mathbb{Z}_4$–valued enhancement by identifying $\mathbb{Z}_2$ with $\{0, 2\} \subset \mathbb{Z}_4$, and then $\beta = 4\alpha$.

Now T has four enhancements $T^{0,0}, T^{0,2}, T^{2,0}, T^{2,2}$ (where the superscripts give the values on a basis) which fall into two isomorphism classes, $T_0 = \{T^{0,0}, T^{0,2}, T^{2,0}\}$ and $T_4 = \{T^{2,2}\}$ (the subscripts specify the Brown invariants). The spaces P and A each have two nonisomorphic enhancements P_1, P_{-1} and A_0, A_∞ (where once again the subscripts are the Brown invariants). Thus V_e decomposes as a sum of copies of T_0, T_4, $P_\pm 1$, A_0 and A_∞, and it is proper if and only if there are no A_∞ summands.

The isomorphism $P \oplus T \cong 3P$ above induces isomorphisms

$$P_{\pm 1} \oplus T_0 \cong P_{\pm 1} \oplus P_1 \oplus P_{-1}$$

and

$$P_{\pm 1} \oplus T_4 \cong 3P_{\mp 1},$$

and the latter implies the first of the following basic relations (the others are left as exercises):

(a) $4\,P_1 \cong 4\,P_{-1}$ and $2\,T_0 \cong 2\,T_4$ (see [3] or [19] for details)

(b) $P_1 \oplus A_\infty \cong P_{-1} \oplus A_\infty$, $T_0 \oplus A_\infty \cong T_4 \oplus A_\infty$ and $A_0 \oplus A_\infty \cong A_\infty \oplus A_\infty$.

It follows from (b) that any two improper enhancements on V are isomorphic, and from (a) that for proper enhancements e,

- V even $\implies$ V_e is a sum of copies of T_0, T_4, A_0, with at most one T_4
- V odd $\implies$ V_e is a sum of copies of $P_{\pm 1}, A_0$, with at most three P_{-1}'s.

In fact these decompositions are unique since the Brown invariant adds under orthogonal direct sums. This additivity can be seen using the Gauss sum

$$\gamma(V_e) = \sum_{x \in V} i^{e(x)}$$

which clearly multiplies under $\oplus$. One readily computes $\gamma(P_{\pm 1}) = 1 \pm i$, $\gamma(T_0) = 2$, $\gamma(T_4) = -2$, $\gamma(A_0) = 2$ and $\gamma(A_\infty) = 0$. It follows that $\gamma(V_e) = 0$ if e is improper, and by the definition of β

$$\gamma(V_e) = \sqrt{2}^{\,m+n} \exp(\pi i \beta(V_e)/4)$$

if e is proper, where $m = \dim V$ and $n = \dim V^\perp$. The additivity of β now follows from the multiplicativity of γ.

4 The Brown invariant: topology

4.1 The Brown invariant of an immersion

Let

$$f\colon F \looparrowright S^3$$

be an immersion of a compact surface F. The immersion is assumed to be *regular*, meaning that the only singularities of f are interior transverse double curves with isolated triple points. Then there is an associated quadratic enhancement

$$f_*\colon H_1(F) \to \mathbb{Z}_4$$

(recall that $\mathbb{Z}_2$–coefficients are used throughout) which, in rough terms, counts the number of half-twists modulo 4 in the images of band neigborhoods of cycles on F. This is defined precisely below. The *Brown invariant* of f is defined to be the Brown invariant of this enhanced space,

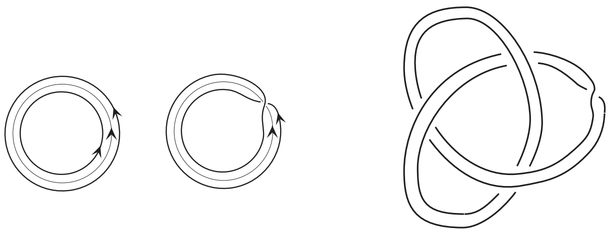
$$\beta_f = \beta(H_1(F)_{f_*}).$$

If f is an embedding then it can be identified with its image, and we write $\beta(F)$ for β_f.

To make this precise, we follow an approach suggested by Sullivan [3, Example 1.28] and later developed by Pinkall [27, section 2] and Siebenmann [9, Appendix]. (Also see Guillou and Marin [9] or Matsumoto [19] for the analogous theory for closed surfaces in simply-connected 4–manifolds.) Define a *band* to be a union of annuli and Möbius strips, and consider the function

$$\widehat{h}\colon \{\text{embedded bands in } S^3\} \to \mathbb{Z}$$

given by $\widehat{h}(B) = \mathrm{lk}(C, \partial B)$, where C is the core of the embedded band B (its zero-section when viewed as an I–bundle) and ∂B is its boundary. Here C and ∂B should be oriented compatibly on components, as shown in Figure 3a.



(a) compatible orientations (b) $\widehat{h} = 1 + 2 \cdot 3 = 7$

Figure 3

If B is connected, then $\widehat{h}(B)$ is just the number of right half-twists in the band relative to the corresponding zero-framed annular band, computed from a projection as "twist" (number of half-twists) plus twice the "writhe" (signed sum of the self-crossings of the core). An example is shown in Figure 3b.

Observe that the mod 4 reduction $h(B)$ of $\widehat{h}(B)$ is unaffected by linking among the components of B, and is in fact invariant under any regular homotopy of B, since a band pass changes $\widehat{h}$ by 4. It follows that there is a well-defined function

$$h\colon \{\text{immersed bands in } S^3\} \to \mathbb{Z}_4,$$

which will be called the *half-twist map*. This map is additive under unions (meaning $h(B \cup B') = h(B) + h(B')$, where B and B' may intersect) and is a complete regular homotopy invariant for connected bands.

Now define the enhancement f_* induced by f as follows (being careful, at least when f is not an embedding, to distinguish between subsets $S \subset F$ and their images $S' = f(S) \subset S^3$): For $x \in H_1(F)$, choose a regularly immersed cycle C in F representing x, and set

$$f_*(x) \;=\; h(B') + 2d(C) \quad (\text{mod } 4)$$

where B is an immersed band neighborhood of C (with image $B' \subset S^3$) and $d(C)$ is the number of double points of C in F.

To check that this definition is independent of the choice of C, first observe that small isotopies of C do not change the right hand side. Thus we may assume that C is transverse to the double curves of f, and that f embeds B onto an immersed band B'.

Now consider the special case in which C is embedded and null-homologous in F, and so in particular B' is an embedded band neighborhood of C'. We must show $h(B') = 0$. But C bounds a surface in F whose interior E has image E' transverse to C' at an *even* number of points (an easy exercise) and so $h(B') \equiv \mathrm{lk}(C', \partial B') \equiv 2C' \cdot E' \equiv 0 \pmod 4$.

In general, if C_1 and C_2 are two regular cycles representing x, then after a small isotopy into general position, $C = C_1 \cup C_2$ is a regular null-homologous cycle. Smoothing crossings $\times \rightsquigarrow \asymp$ converts C into an embedded cycle without changing $h + 2d$ (each smoothing changes both terms by 2) and so $h(B') + 2d(C) \equiv 0 \pmod 4$ by the special case above. Since h is additive, $h(B') = h(B_1') + h(B_2')$, while $d(C) = d(C_1) + d(C_2) + C_1 \cdot C_2$. Rearranging terms gives

$$\begin{aligned} h(B_1') + 2d(C_1) &\equiv (-h(B_2') - 2C_1 \cdot C_2) - 2d(C_2) \\ &\equiv h(B_2') + 2d(C_2) \pmod 4 \end{aligned}$$

since $h(B_2') \equiv x \cdot x \equiv C_1 \cdot C_2 \pmod 2$. Thus f_* is well-defined (compare Propositions 1 and 2 in [9] and Lemma 5.1 in [19]).

It is now immediate from the definitions that f_* is quadratic. Furthermore, it is readily seen that f_* is proper if and only if the link $L = f(\partial F)$ is proper, ie each component K of L links $L-K$ evenly. Indeed, if K^+ is a parallel copy of K (the image of a push off in F), then $f_*([K]) = 2\,\mathrm{lk}(K, K^+) = 2\,\mathrm{lk}(K, L-K)$ (since K^+ and $L-K$ are homologous in $S^3 - K$ across F') and so $f_*([K]) = 0$ if and only if $\mathrm{lk}(K, L-K)$ is even. In this case (when L is proper) we shall refer to f as a *proper immersion.*

4.2 The Brown invariant of a proper link

Observe that the Brown invariant of a proper *embedded* surface $F \subset S^3$ depends only on the *framed* link $L = \partial F$, where the framing is given by a vector field normal to L in F. (Note that each component gets an *even* framing since F is proper.) For if F' is any other surface in S^3 bounded by L with the *same* framing, then the closed surface $S \subset S^4$ obtained from $F \cup F'$ by pushing $\mathrm{int}(F)$ and $\mathrm{int}(F')$ to opposite sides of an equatorial S^3 has Brown invariant $\beta(S) = \beta(F) - \beta(F')$ and self-intersection $S \cdot S = 0$ (defined to be the twisted Euler class of the normal bundle of S in S^4 when S is nonorientable, cf [9].) But $\beta(S) = 0$ since $2\beta(S) \equiv \sigma(S^4) - S \cdot S \pmod{16}$ by a theorem of Guillou and Marin [9], where σ denotes the signature, and so $\beta(F) = \beta(F')$.

Thus one is led to define the Brown invariant for any *even framed proper link* L by $\beta(L) = \beta(F)$ where $F \subset S^3$ is any embedded surface bounded by L which induces the prescribed framing on L. Such a surface can be constructed from an arbitrary surface bounded by L by stabilizing (adding small half-twisted bands along the boundary to adjust the framings), and any two have the same Brown invariant by the discussion above.

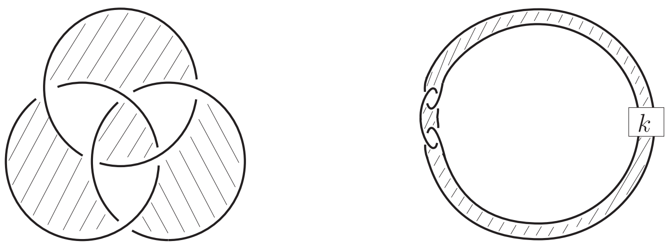
If no framing is specified on L, then the zero framing is presumed. In other words, the *Brown invariant of a proper link* (unframed) is defined to be the Brown invariant of the link with the zero framing on each component. It can be computed from *any* embedded surface F bounding L (possibly nonorientable) by the formula

$$\beta(L) = \beta(F) - \phi(F)$$

where $\phi(F)$ denotes half the sum of the framings on L induced by F. (Note that these framings are all even since L is proper.)

Examples (1) The Borromean rings have Brown invariant 4. This can be seen using the bounded checkerboard surface F in the minimal diagram for the link shown in Figure 2a. The enhanced homology is $P_1 \oplus 2A_0$, and the induced framings are all -2, so the Brown invariant is $1 - \frac{1}{2}(-6) = 4$.

(2) The k–twisted Bing double of any knot (with k full twists in the parallel strands) has Brown invariant $4k$. To see this use the obvious banded Seifert surface of genus 1 (shown for the double of the unknot in Figure 4b) which has enhancement $T_{4k} \oplus A_0$ and induces the 0–framing on both boundary components.



(a) Borromean rings (b) k-twisted Bing double

Figure 4

Remarks (1) If L is given an orientation, then it has an Arf invariant $\alpha(L) \in \mathbb{Z}_2$ which is related to the Brown invariant of any (oriented) Seifert surface F for L by the identity $\beta(F) = 4\alpha(L)$. Adjusting for the framings one obtains the formula

$$\beta(L) = 4\,\alpha(L) + \mathrm{lk}(L)$$

where $\mathrm{lk}(L)$ denotes the sum of all the pairwise linking numbers of L. Note that both terms on the right hand side depend on the orientation of L, while their sum does not. For example the $(2,4)$-torus link L has $\beta(L) = 6$, while $(\alpha(L), \mathrm{lk}(L)) = (1,2)$ or $(0,-2)$ according to whether the components are oriented compatibly or not.

(2) (see [17]) The Brown invariant of L can also be defined using a surface F in B^4 bounded by L for which there exists a Pin^- structure on $B^4 - F$ which does not extend over F. The Pin^- structure descends to a Pin^- structure on F which determines an element in 2–dimensional Pin^- bordism which, using the Brown invariant, is isomorphic to $\mathbb{Z}_8$.

For an immersion $f\colon F \looparrowright S^3$ bounded by L, a correction term coming from the singularities of f is needed to compute $\beta(L)$ in terms of β_f. This is most easily expressed using the *quarter-twist map*

$$q\colon \{\text{immersed doublebands in } S^3\} \to \mathbb{Z}_8,$$

defined analogously to the half-twist map h above: A *doubleband* is a union of $\times$–bundles over circles; for embedded doublebands B with core C, define $q(B) = \mathrm{lk}(C, \partial B) \pmod 8$, where ∂B is the compatibly oriented $::$–bundle ($S^0 \times S^0$–bundle) and then observe that this is a regular homotopy invariant. For example, any double curve C in $F' = f(F)$ has an immersed doubleband neighborhood B, and $q(B)$ (also denoted $q(C)$ by abuse of notation) is odd or even according to whether $f^{-1}(C)$ consists of one or two curves in F; we say C is *orientation-reversing* in the former case, and *orientation-preserving* in the latter.

Now consider the entire singular set of f. It consists of a collection of double curves which intersect in some number τ_f of triple points. A neighborhood of this singular set is an immersed doubleband B (generally not connected) and we define $\delta_f = q(B)$, and (as for the case of embedded surfaces) ϕ_f to be half the sum of the framings on L induced by F.

Theorem 4.1 *Let L be a proper link bounded by a regularly immersed surface $f\colon F \looparrowright S^3$. Then the Brown invariant*

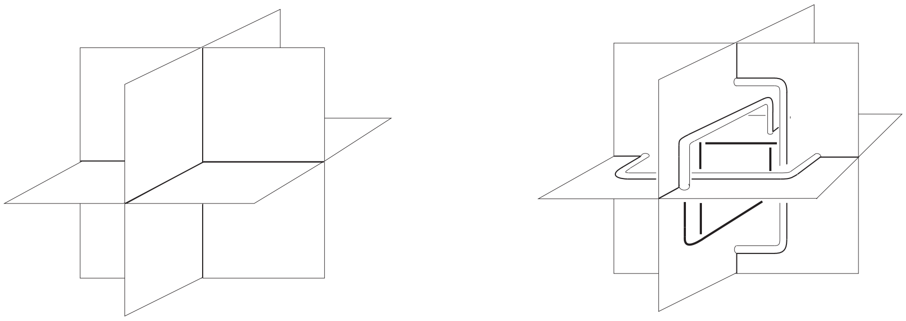
$$\beta(L) = \beta_f - \phi_f + 3\delta_f + 4\tau_f,$$

and so (by Remark 1 above) the Arf invariant

$$\alpha(L) = \tfrac{1}{4}(\beta_f - \phi_f - \mathrm{lk}(L) + 3\delta_f + 4\tau_f)$$

for any chosen orientation on L.

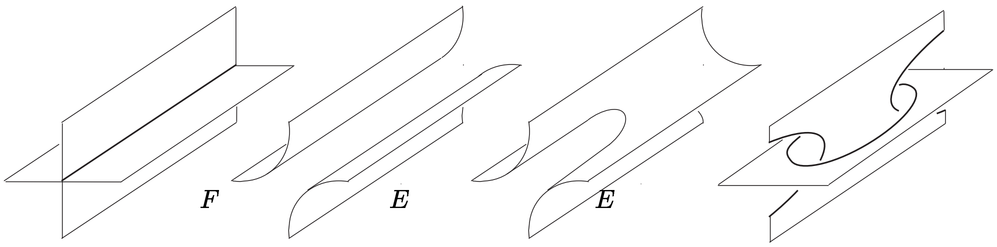
Proof The strategy is to reduce to the embedded case by local modifications of f. We first eliminate triple points by *Borromean cuts* as shown in Figure 5. This calls for the removal of three disks (bounded by the Borromean rings) and the addition of three tubes to maintain *interior* singularities (a condition for regularity of the immersion) in the three sheets near each triple point. The effect on the boundary is to add τ_f copies of the Borromean rings, which changes the left hand side of the formula by $4\tau_f$ (see Example 1 above). The terms β_f and ϕ_f on the right hand side are unchanged, since the effect of this modification is to add copies of $T_0 \oplus A_0$ to the enhanced homology, and to give the 0–framing to the Borromean rings on the boundary. Likewise δ_f is unchanged, since the double curves have simply undergone a regular homotopy, so the net change on the right hand side is also $4\tau_f$.



(a) triple point (b) Borromean cut

Figure 5

Now fix a double curve C, and set $n = q(C)$, the number of quarter-twists in the normal $\times$-bundle of C. Note that C is embedded since there are no triple points. Let $g\colon E \looparrowright S^3$ be the immersion obtained from f by "smoothing" along C. In other words, proceed along C, replacing each fiber in the normal $\times$-bundle of C (see Figure 6a) by two arcs (Figure 6b); if C is orientation reversing (ie n is odd) then one must insert a saddle at some point to allow the fibers to match up (Figure 6c). This can be done in two ways, starting with $\asymp$ or)(, and either will do. In any case C disappears from set of double curves, and so $\delta_g = \delta_f - n$. Since this construction adds no new boundary components or triple points, it suffices to show that $\beta_g = \beta_f + 3n$. There are several cases to consider depending on the value of n.



(a) double curve (b) even smoothing (c) odd smoothing (d) Bing cut

Figure 6

If $n = 4k \pm 1$, for some k, then E is obtained from F by removing a Möbius band, whose image wraps twice around C, and replacing it with a boundary-connected sum of two Möbius bands. The effect on the enhanced homology is to delete a P^{n+2}–summand (note that a regular homotopy of the doubly wrapped Möbius band introduces a kink which adds 2 to the number of half-twists) and to add two $P^{2k\pm1}$–summands. As above, the superscript is the value of the enhancement on a generator, and so the Brown invariant is obtained by reducing mod 4 to ±1. In other words, we are deleting a $P_{\mp1}$-summand and adding two $P_{(-1)^k}$–summands. Thus $\beta_g \equiv \beta_f \pm 1 \pm 2(-1)^k \equiv \beta_f + 3n \pmod 8$.

If $n = 4k+2$, then E is obtained by removing two Möbius bands from F and then identifying the resulting boundary components to form a circle C'. The effect on the enhanced homology is to delete two $P^{n/2} = P_{(-1)^k}$–summands, and to add either nothing (if the Möbius bands lie in distinct components of F) or one T_0 or K_0–summand† (a punctured torus or Klein bottle is seen as a neighborhood of the union of C' and a dual circle). Thus $\beta_g \equiv \beta_f - 2(-1)^k \equiv \beta_f - n \equiv \beta_f + 3n \pmod 8$.

Finally consider the case $n = 4k$. Let C_1, C_2 denote the two circles in $f^{-1}(C)$. An argument similar to the one above can be given by analyzing several cases depending on the homology classes of C_1, C_2 and $C_1 \cup C_2$, but there is a simpler argument using a different modification of f near C which we call a *Bing cut*. It is obtained by removing two discs from F, thereby introducing a 0–framed k–twisted Bing double of C on the boundary (see Example 2 above) as shown in Figure 6d. This adds $4k$ to $\beta(L)$ since C links L evenly; indeed $\mathrm{lk}(C, L) \equiv \mathrm{lk}(C, L \cup C^+ \cup C^-) \equiv 0 \pmod 2$, where $C^\pm$ are pushoffs of C in the image of a neighborhood of C_1, since $L \cup C^+ \cup C^-$ bounds in $S^3 - C$ across $f(F - C_2)$. The terms on the right hand side remain unchanged except for $3\delta_f$ which also changes by $12k \equiv 4k \pmod 8$. The proof is completed by induction on the number of double curves. □

A formula for the Brown invariant of a totally proper link C can be deduced by applying this theorem to a family of connected surfaces (bounding the components C_i of C) which meet only in their interiors. In this case, $\beta_f - \phi_f = \sum \beta(C_i)$, $\delta_f = \sum \lambda(C_i, C_j)$ and $\tau_f = \sum \tau(C_i, C_j, C_k)$, where λ and τ are the Sato–Levine and Milnor invariants defined in section 1. Noting that λ is even valued, we have

$$\beta(C) = \sum_i \beta(C_i) - \sum_{i<j} \lambda(C_i, C_j) + 4 \sum_{i<j<k} \tau(C_i, C_j, C_k) \pmod 8.$$

†By definition $K_0 = P_1 \oplus P_{-1}$. This corresponds to a Klein bottle $K \cong P \oplus P$. The four enhancements $K^{0,\pm1}, K^{2,\pm1}$ of K, with respect to the homology basis x, y with $x \cdot x = 0$ and $x \cdot y = y \cdot y = 1$, fall into three isomorphism classes $K_0 = \{K^{0,1}, K^{0,-1}\}$, and $K_{\pm2} = K^{2,\pm1}$.

The formula

$$\alpha(C) = \sum_i \alpha(C_i) + \sum_{i<j} \tfrac{1}{4}(\lambda(C_i, C_j) + \mathrm{lk}(C_i, C_j)) + \sum_{i<j<k} \tau(C_i, C_j, C_k) \pmod 2$$

for the Arf invariant of C with any chosen orientation (Theorem 1.1) follows, since $\beta(C) = 4\alpha(C) + \mathrm{lk}(C)$. The dependence on the orientation is captured by $\mathrm{lk}(C)$, the sum of the pairwise linking numbers of the components of C.

References

[1] **William Browder**, *Surgery on Simply-Connected Manifolds*, Ergebnisse series 65, Springer–Verlag, Berlin (1972)

[2] **Edgar H Brown Jr**, *The Kervaire invariant of a manifold*, from: "Algebraic Topology (Madison 1970)", Proc. Sympos. Pure Math. XXII, Amer. Math. Soc. Providence, RI (1971) 65–71

[3] **Edgar H Brown Jr**, *Generalizations of the Kervaire invariant*, Ann. Math. 95 (1972) 368–383

[4] **Tim D Cochran**, *Concordance invariance of coefficients of Conway's link polynomial*, Invent. Math. 82 (1985) 527–541

[5] **Tim D Cochran**, **Paul Melvin**, *Quantum cyclotomic orders of 3–manifolds*, Topology 40 (2001) 95–125

[6] **Stephen A Cook**, *The complexity of theorem proving procedures*, from: "Proc. 3rd Ann. ACM Symp. on Theory of Computing", Assoc. for Comput. Mach., New York (1971) 151–158

[7] **Michael Freedman**, **Robion Kirby**, *A geometric proof of Rochlin's theorem*, from: "Algebraic and Geometric Topology (Stanford (976)", Proc. Sympos. Pure Math. XXXII (Part 2), Amer. Math. Soc. Providence, RI (1978) 85–97

[8] **Patrick Gilmer**, *A method for computing the Arf invariants of links*, Quantum Topology, Ser. Knots Everything 3, World Sci. Publishing, River Edge, NJ (1993) 174–181

[9] **Lucien Guillou**, **Alexis Marin**, *Une extension d'un theoreme de Rohlin sur las signature*, from: "À la Recherche de la Topologie Perdue", (Lucien Guillou and Alexis Marin, editors), Progr. Math. 62, Birkhäuser, Boston, MA (1986) 97–118

[10] **Jim Hoste**, *The Arf invariant of a totally proper link*, Topology Appl. 18 (1984) 163–177

[11] **Jim Hoste**, *A formula for Casson's Invariant*, Trans. Amer. Math. Soc. 297 (1986) 547–562

[12] **Steve J Kaplan**, *Constructing framed 4–manifolds with given almost framed boundaries*, Trans. Amer. Math. Soc. 254 (1979) 237–263

[13] **Louis H Kauffman**, *The Conway polynomial*, Topology 20 (1981) 101–108

[14] **Viatcheslav Kharlamov**, **Oleg Viro**, *Extensions of the Gudkov–Rohlin congruence*, from: "Topology and Geometry – Rohlin Seminar", Lecture Notes in Math. 1346, Springer–Verlag, Berlin (1988) 357–406

[15] **Robion C Kirby**, *The Topology of 4–manifolds*, Lecture Notes in Math. 1374, Springer–Verlag, 1989.

[16] **Robion Kirby**, **Paul Melvin**, *The 3–manifold invariants of Witten and Reshetikhin–Turaev for sl(2,$\mathbb{C}$)*, Invent. Math. 105 (1991) 473–545

[17] **Robion C Kirby**, **Laurence R Taylor**, *Pin structures on low-dimensional manifolds*, from: "Geometry of Low-dimensional Manifolds, Vol. 2 (Durham, 1989)", (S K Donaldson and C B Thomas, editors), London Math Soc. Lecture Note Ser. 151, Cambridge Univ. Press (1990) 177–242

[18] **Jerome Levine**, *Knot cobordism groups in codimension two*, Comm. Math. Helv. 44 (1969) 229–244

[19] **Yukio Matsumoto**, *An elementary proof of Rochlin's signature theorem and its extension by Guillou and Marin*, from: "À la Recherche de la Topologie Perdue", (Lucien Guillou and Alexis Marin, editors), Progr. Math. 62, Birkhäuser, Boston, MA (1986) 119–139

[20] **Blake Mellor**, **Paul Melvin**, *Geometric interpretation of Milnor's triple invariants*, Algebr. Geom. Topol. 3 (2003) 557–568

[21] **John Milnor**, *Link groups*, Ann. of Math. 59 (1954) 177–195

[22] **John Milnor**, *Isotopy of links*, from: "Algebraic geometry and topology (Lefschetz symposium)", Princeton Math. Series 12 (1957) 280–306

[23] **Hitoshi Murakami**, *The Arf invariant and the Conway polynomial of a link*, Math. Sem. notes Kobe Univ. 11 (1983) 335–344

[24] **Kunio Murasugi**, *On the Arf invariant of links*, Math. Proc. Camb. Phil. Soc. 95 (1984) 61–69

[25] **Tomotada Ohtsuki**, *A polynomial invariant of rational homology spheres*, Invent. Math. 123 (1996) 241–257

[26] **Christos H Papadimitriou**, *Computational Complexity*, Addison–Wesley, Reading, Mass., (1994)

[27] **Ulrich Pinkall**, *Regular homotopy classes of immersed surfaces*, Topology 24 (1985) 421–434

[28] **Nikolai Yu Reshetikhin**, **Vladimir G Turaev**, *Invariants of 3–manifolds via link polynomials and quantum groups*, Invent. Math. 103 (1991) 547–597

[29] **Raymond A Robertello**, *An invariant of knot cobordism*, Comm. Pure Appl. Math. 18 (1965) 543–555

[30] **Vladimir A Rokhlin**, *Proof of a conjecture of Gudkov*, Functional Analysis and its Applications 6 (1972) 136–138; transl. from Funkcional. Anal. i Priložen. 6 (1972) 62–64

[31] **Colin P Rourke**, **Dennis P Sullivan**, *On the Kervaire obstruction*, Ann. of Math. 94 (1971) 397–413

[32] **Masahico Saito**, *On the unoriented Sato–Levine invariant*, J. Knot Theory Ramifications 2 (1993) 335–358

[33] **Nobuyuki Sato**, *Cobordism of semi-boundary links*, Topology Appl. 18 (1984) 225–234

[34] **Rachel Sturm**, *Link concordance invariants*, Thesis, New York University (1984)

[35] **Vladimir G Turaev**, *The Milnor invariants and Massey products*, J. Soviet Math. 12 (1979) 128–137; transl. from Zap. Naučn. Sem. LOMI 66 (1976) 189–203, 209–210

[36] **Vladimir G Turaev**, *Cohomology rings, linking coefficient forms and invariants of spin structures in three-dimensional manifolds*, Math. USSR Sbornik 48 (1984) 65–79; transl. from Mat. Sb. 120(162) (1983) 68–83, 143

[37] **Leslie G Valiant**, *The complexity of enumeration and reliability problems*, SIAM J. Comput. 8 (1979) 410–421

[38] **Edward Witten**, *Quantum field theory and the Jones polynomial*, Commun. Math. Phys. 121 (1989) 351–399

University of California, Berkeley, CA 94720, USA
Bryn Mawr College, Bryn Mawr, PA 19010, USA

Email: `kirby@math.berkeley.edu, pmelvin@brynmawr.edu`