Algebraic & Geometric Topology

Volume 17 (2017)

Issue 5 (pages 2565–3212)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY
msp.org/agt

See inside back cover or msp.org/agt for submission instructions.

The subscription price for 2017 is US $410/year for the electronic version, and $650/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) at Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published 6 times per year and continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

AGT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

msp
Algebraic & Geometric Topology 17 (2017) 2565–2602

Categorical models for equivariant classifying spaces

BERTRAND J GUILLOU
J PETER MAY
MONA MERLING

Starting categorically, we give simple and precise models for classifying spaces of equivariant principal bundles. We need these models for work in progress in equivariant infinite loop space theory and equivariant algebraic K–theory, but the models are of independent interest in equivariant bundle theory and especially equivariant covering space theory.

55R91, 55R35; 55P92, 55R91

Introduction

Let Π and G be topological groups and let G act on Π, so that we have a semidirect product $\Gamma = \Pi \rtimes G$ and a split extension

$$1 \to \Pi \xrightarrow{\subset} \Gamma \xrightarrow{q} G \to 1 \tag{0-1}$$

The underlying space of Γ is $\Pi \times G$, and the product is given by

$$(\sigma, g)(\tau, h) = (\sigma(g \cdot \tau), gh).$$

There is a general theory of (G, Π_G)–bundles — see Lashof [6], Lashof and May [7], May [13] and tom Dieck [3] — corresponding to such extensions. Here Π_G denotes Π together with its given action of G. We shall only be interested in principal (G, Π_G)–bundles $p\colon E \to B$.

Definition 0.2 Let $p\colon E \to B$ be a principal Π–bundle, where B is a G–space. Then p is a principal (G, Π_G)–bundle if the (free) action of Π on E extends to an action of Γ and p is a Γ–map, where Γ acts on B through the quotient map $\Gamma \to G$.

The more general theory of $(\Pi;\Gamma)$–bundles applicable to nonsplit extensions Γ is included in Lashof and May [7] and May [12; 13]. The theory is especially familiar when G acts trivially on Π, so that $\Gamma = G \times \Pi$. With $\Pi = O(n)$ or $U(n)$, the trivial action case gives classical equivariant bundle theory and equivariant topological K–theory.

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2565

Definition 0.3 A principal (G, Π_G)–bundle $p\colon E \to B$ is universal if for all G–spaces X of the homotopy types of G–CW complexes, pullback of p along G–maps $f\colon X \to B$ induces a natural bijection from the set of homotopy classes of G–maps $X \to B$ to the set of equivalence classes of (G, Π_G)–bundles over X.

For applications in equivariant infinite loop space theory and equivariant algebraic K–theory, we need to understand classifying G–spaces for (G, Π_G)–bundles as classifying spaces of categories. Nonequivariantly, it was already emphasized in Segal's classical paper [19, Section 3] that the universal principal Π–bundle of a topological group Π can be constructed on the level of topological categories, and the intuition is that we are giving the equivariant generalization of his classical construction.

One motivation is to give new constructions of E_∞ operads of G–categories and G–spaces. This much only requires trivial actions of G on Π. By definition, the j^{th} space of an E_∞ operad of G–spaces is a universal principal (G, Σ_j)–bundle. Having various category level models for such classifying spaces allows us to construct examples of E_∞ G–spaces from E_∞ categories, and these feed into equivariant infinite loop space machines to construct interesting G–spectra. This is discussed in Guillou and May [4] and in work in progress by Guillou, May, Merling and Osorno.

The examples relevant to the equivariant algebraic K–theory of G–rings, namely rings with G–action by automorphisms, require more general split extensions. If R is a G–ring, then G acts entrywise on $\mathrm{GL}(n, R)$. The classifying spaces of $(G, \mathrm{GL}(n, R)_G)$–bundles are central to the definition of the genuine equivariant algebraic K–theory spectrum $\mathbb{K}_G(R)$ of R; see Guillou and May [4] and Merling [14]. Our treatment of the fixed point spaces of the classifying spaces of equivariant bundles is crucial to determining the fixed point spectra of the $\mathbb{K}_G(R)$. The paradigmatic example is a finite Galois extension E/F with Galois group G. As explained in [4], it is an immediate application of examples in this paper, which demonstrate the relevance of Hilbert's Theorem 90, that the fixed point spectrum $\mathbb{K}_G(E)^H$ is the classical nonequivariant K–theory spectrum of the fixed field E^H. The use of genuine G–spectra in algebraic K–theory is new and is explored in [14].

The results we need are close to those of [6; 7; 12] and those stated by Murayama and Shimakawa [16],[1] but we require a more precise and rigorous categorical and topological understanding than the literature affords. This is intended as a service paper that displays the relevant constructions in their fullblown simplicity.

We start with the topologized equivariant version of the elementary theory of chaotic categories in Section 1. We analyze a general construction that specializes to give

[1] But see Scholium 3.12.

our classifying G–spaces in Section 2. We show how it gives universal equivariant bundles in Section 3. Our explicit description of the classifying spaces of (G, Π_G)–bundles as classifying spaces of categories allows us to compute their fixed point spaces categorically in Section 4. This gives precise information already on the category level, before passage to classifying spaces, and that is essential to our applications.

The main results of the paper are summarized in the following two theorems; the first gives a categorical model for equivariant universal bundles and their classifying spaces, and the second gives a description of the fixed points of the classifying spaces of equivariant bundles. Details of the first are in Theorems 3.10 and 3.11 and details of the second are in Theorems 4.18, 4.23, and 4.24. We need some preliminary definitions and notations to state these results.

Let G be discrete and let $\mathcal{E}G$ denote the unique contractible groupoid with object set G. It is a (right) G–category, meaning that G acts on both objects and morphisms, and it has a unique morphism between any two objects. We agree to identify the topological group Π with the topological groupoid with a single object and morphism space Π. Then the action of G on Π makes it a G–groupoid.

For small topological categories $\mathscr{A}$ and $\mathscr{B}$, let $\mathscr{C}at(\mathscr{A}, \mathscr{B})$ denote the category of all continuous functors $\mathscr{A} \to \mathscr{B}$ and all natural transformations. When $\mathscr{A}$ and $\mathscr{B}$ are G–categories, $\mathscr{C}at(\mathscr{A}, \mathscr{B})$ inherits an action of G, given by conjugation. We shall give more details in Section 1.1.

We assume that the reader is familiar with the classifying space functor B from categories to spaces, or more generally from topological categories to spaces. It works equally well to construct G–spaces from topological G–categories. It is the composite of the nerve functor N from topological categories to simplicial spaces (eg May [11, Section 7]) and geometric realization $|-|$ from simplicial spaces to spaces (eg May [10, Section 11]), both of which are product-preserving functors.

Theorem 0.4 *If G is discrete and Π is either discrete or a compact Lie group, then the canonical map*

$$B\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi) \to B\mathscr{C}at(\mathcal{E}G, \Pi)$$

is a universal principal (G, Π_G)–bundle.

Thus the classifying space of the G–category $\mathscr{C}at(\mathcal{E}G, \Pi)$ is a G–space that classifies (G, Π_G)–bundles.

Crossed homomorphisms, their automorphism groups, and the nonabelian cohomology group $H^1(G; \Pi_G)$ are defined in Definitions 4.1, 4.11, and 4.17.

Theorem 0.5 *The fixed point category $\mathscr{C}at(\mathcal{E}G, \Pi)^G$ is the disjoint union of the groups $\mathrm{Aut}\,\alpha$, where α runs over crossed homomorphisms representing the elements of $H^1(G; \Pi_G)$. Equivalently, $\mathscr{C}at(\mathcal{E}G, \Pi)^G$ is the disjoint union of the groups $\Pi \cap N_\Gamma \Lambda$, where Λ runs over the Π–conjugacy classes of subgroups Λ of Γ such that $\Lambda \cap \Pi = e$. Therefore $B\mathscr{C}at(\mathcal{E}G, \Pi)^G$ is the disjoint union of the classifying spaces $B(\Pi \cap N_\Gamma \Lambda)$.*

With more work, our hypotheses on G and Π could surely be weakened. We should admit that we are especially interested in discrete groups in many of our current applications. Since Π is the relevant structural group, we are then studying equivariant covering spaces. However, it is important for some applications to allow Π to have a topology. For example, in Merling [14], equivariant algebraic K–theory is related to equivariant topological K–theory and to Atiyah's real K–theory. There it is crucial that Π be allowed to be compact Lie in Theorem 0.4.

There is an earlier topological analogue of our categorical construction in terms of mapping spaces rather than mapping categories; see May [12]. It applies in considerably greater topological generality, but it does not generally start categorically. We compare the categorical and topological constructions in Section 5.

The choices of Π relevant to equivariant infinite loop space theory and equivariant algebraic K–theory, namely symmetric groups and the general linear groups of G–rings, have alternative categorical models, which play a key role. These alternative categorical models are given in Section 6, which is entirely algebraic, with all groups discrete. We call special attention to Section 6.2, where we relate crossed homomorphisms to skew group rings and their skew modules. The algebraic ideas here may not be as well known as they should be and deserve further study.

The letter B for the classifying space functor from categories to spaces would sometimes be awkward in our context, since the classifying space functor will also be used to construct universal bundles rather than classifying spaces for bundles, hence we agree to write out $|N-|$ rather than B whenever B seems likely to confuse.

This notation also displays a key technical problem that is sometimes overlooked in the literature. The functor $|-|$ is a left adjoint and therefore preserves all colimits, such as passage to orbits in the equivariant setting. The functor N is a right adjoint and it generally does not preserve colimits or passage to orbits, as we illustrate with elementary examples. This problem is the subject of Babson and Kozlov [1]. For topological categories, there is no discussion in the literature. Exceptionally, N does commute with passage to orbits in the key examples that appear in equivariant bundle theory. Clear understanding of passage to orbits is essential to our calculations of fixed point spaces.

Remark 0.6 The functor $\mathscr{C}at(\mathcal{E}G, -)$ from G–categories to G–categories plays a central role in our work. Its G–fixed category was introduced by Thomason [22, (2.1)], who called it the lax limit of the action of G on $\mathscr{C}$ and denoted it by $\mathscr{C}at_G(\underline{EG}, \underline{C})$. The relevance to equivariant bundle theory of the equivariant precursor $\mathscr{C}at(\mathcal{E}G, \mathscr{B})$ was first noticed by Murayama and Shimakawa [16] and Shimakawa [21].

Acknowledgements Merling thanks Matthew Morrow and Liang Xiao for answers to her questions that pointed out the striking relevance to our work of $H^1(G; \Pi)$ and Serre's general version of Hilbert's Theorem 90. We are grateful to an anonymous referee for a careful reading and suggestions for improving the notations and exposition. Guillou was partially supported by Simons Collaboration Grant No. 282316 and by NSF grant DMS-1710379 and Merling was partially supported by NSF grant DMS-1709461.

1 Preliminaries on chaotic and translation categories

The definitions we start with are familiar and elementary. However, to keep track of categorical data and group actions later, we shall be pedantically precise.

1.1 Preliminaries on topological G–categories

Let $\mathscr{C}at$ be the category of categories and functors. We may also view it as the 2–category of categories, with 0–cells, 1–cells, and 2–cells the categories, functors, and natural transformations. From that point of view, $\mathscr{C}at(\mathscr{A}, \mathscr{B})$ is the internal hom category whose objects are the functors $\mathscr{A} \to \mathscr{B}$ and whose morphisms are the natural transformations between them; they enrich $\mathscr{C}at$ over itself.

For a group G, a G–category $\mathscr{A}$ is a category with an action of G specified by a homomorphism from G to the automorphism group of $\mathscr{A}$. Regarding G as a groupoid with one object, the action is specified by a functor $G \to \mathscr{C}at$. We have the 2–category $G\mathscr{C}at$ of G–categories, G–functors, and G–natural transformations, where the latter notions are defined in the evident way: everything must be equivariant.

We may view $G\mathscr{C}at$ as the underlying 2–category of a category enriched over $G\mathscr{C}at$. The 0–cells are still G–categories, but now we have the G–category $\mathscr{C}at(\mathscr{A}, \mathscr{B})$ as the internal hom between them. Its underlying category is $\mathscr{C}at(\mathscr{A}, \mathscr{B})$, and G acts by conjugation on functors and natural transformations. Thus, for $F\colon \mathscr{A} \to \mathscr{B}$, $g \in G$, and A either an object or a morphism of $\mathscr{A}$, we have $(gF)(A) = gF(g^{-1}A)$. Similarly, for a natural transformation $\eta\colon E \to F$ and an object A of $\mathscr{A}$,

$$(g\eta)_A = g\eta_{g^{-1}A}\colon gE(g^{-1}A) \to gF(g^{-1}A).$$

The category $G\mathscr{C}at(\mathscr{A},\mathscr{B})$ is the same as the G–fixed category $\mathscr{C}at(\mathscr{A},\mathscr{B})^G$, and we sometimes vary the choice of notation.

We can topologize the definitions so far, starting with the 2–category of categories internal to the category $\mathscr{U}$ of (compactly generated) spaces, together with continuous functors and continuous natural transformations. Recall that a category $\mathscr{A}$ internal to a cartesian monoidal category $\mathscr{V}$ has object and morphism objects in $\mathscr{V}$ and structure maps source, target, identity and composition in $\mathscr{V}$. These maps are denoted by S, T, I and C, and the usual category axioms must hold. When $\mathscr{V}=\mathscr{U}$, we refer to internal categories as topological categories; we refer to them as topological G–categories when $\mathscr{V}=G\mathscr{U}$. These are more general than (small) topologically enriched categories, which have discrete sets of objects. We can now allow G to be a topological group in the equivariant picture. We continue to use the notations already given in the more general topological situation.

1.2 Chaotic topological G–categories

Definition 1.1 A small category $\mathscr{C}$ is *chaotic* if there is exactly one morphism from b to a for each pair of objects a and b. The unique morphism from a to b must then be inverse to the unique morphism from b to a. Thus $\mathscr{C}$ is a groupoid, and its classifying space is contractible since every object is initial and terminal; in fact, it is the unique contractible groupoid with the given object set. A topological category $\mathscr{C}$ is chaotic if its underlying category is chaotic. Its classifying space is again contractible (see Remark 2.11), but there are other topological groupoids with the given object space and contractible classifying spaces. Similarly, a topological G–category is chaotic if its underlying category is chaotic. It is then contractible but not usually G–contractible.

The senior author remembers hearing the name "chaotic" long ago, but we do not know its source. The idea is that everything is the same as everything else, which does seem rather chaotic.[2]

Lemma 1.2 *If $\mathscr{A}$ is any category and $\mathscr{B}$ is a chaotic category, then the category $\mathscr{C}at(\mathscr{A},\mathscr{B})$ is again chaotic.*

Proof The unique natural map $\zeta\colon E\to F$ between functors $E,F\colon \mathscr{A}\to\mathscr{B}$ is given on an object A of $\mathscr{A}$ by the unique map $\zeta_A\colon E(A)\to F(A)$ in $\mathscr{B}$. □

[2] Some category theorists suggest the name "indiscrete category", by formal analogy with indiscrete spaces in topology. The key difference is that indiscrete spaces are of no interest, whereas we hope to convince the reader that chaotic categories are of considerable interest.

Lemma 1.3 *If $\mathscr{A}$ is any topological G–category and $\mathscr{B}$ is a chaotic topological G–category, then the topological G–category $\mathscr{C}at(\mathscr{A},\mathscr{B})$ and its G–fixed category $G\mathscr{C}at(\mathscr{A},\mathscr{B})$ are again chaotic.*

Proof Since $\mathscr{C}at(\mathscr{A},\mathscr{B})$ is just the category $\mathscr{C}at(\mathscr{A},\mathscr{B})$ with its conjugation action by G, Lemma 1.2 implies the conclusion for $\mathscr{C}at(\mathscr{A},\mathscr{B})$. The conclusion is inherited by $G\mathscr{C}at(\mathscr{A},\mathscr{B}) = \mathscr{C}at(\mathscr{A},\mathscr{B})^G$ since the unique natural transformation between G–functors E and F is necessarily a G–natural transformation. □

Definition 1.4 The chaotic topological category $\mathcal{E}X$ generated by a space X is the topological category with object space X and morphism space $X \times X$. The source, target, identity, and composition maps are defined by

$$S = \pi_2\colon X \times X \to X, \quad T = \pi_1\colon X \times X \to X, \quad I = \Delta\colon X \to X \times X,$$
$$C = \mathrm{id} \times \varepsilon \times \mathrm{id}\colon (X \times X) \times_X (X \times X) \cong X \times X \times X \to X \times X,$$

where $\varepsilon\colon X \to *$ is the trivial map. On elements, $S(y,x) = x$, $T(y,x) = y$, $I(x) = (x,x)$ and $C(z,y,x) = (z,x)$. Forgetting the topology, the element (y,x) is the unique morphism $x \to y$. Reversing the order of source and target in the notation this way, so that $(z,y) \circ (y,x) = (z,x)$, will turn out to be helpful later.

A map $f\colon X \to Y$ induces the functor $\tilde{f}\colon \mathcal{E}X \to \tilde{Y}$ given by f on objects and $f \times f$ on morphisms. When X is a (left or right) G–space, we give $\mathcal{E}X$ the action specified by the given action on the object space X and the diagonal action on the morphism space $X \times X$; $\mathcal{E}X$ is then a chaotic topological G–category. Sending X to $\mathcal{E}X$ specifies a functor from the category $G\mathscr{U}$ of G–spaces to the category $G\mathscr{G}pd$ of topological G–groupoids (a full subcategory of $G\mathscr{C}at$).

1.3 The adjunction between G–spaces and topological G–groupoids

Sending a category to its set of objects restricts to an object functor $\mathscr{O}b\colon G\mathscr{G}pd \to G\mathscr{U}$.

Lemma 1.5 *The chaotic category functor is right adjoint to the object functor, so that*

$$G\mathscr{C}at(\mathscr{C}, \mathcal{E}X) \cong G\mathrm{Map}(\mathscr{O}b\mathscr{C}, X)$$

for a topological G–category $\mathscr{C}$ with object space $\mathscr{O}b\mathscr{C}$ and a topological G–space X. If $\mathscr{C}$ is chaotic with object G–space X, then the unit of the adjunction is an isomorphism of topological G–groupoids $\eta\colon \mathscr{C} \to \mathcal{E}X$.

Proof Let $\mathscr{M}or\mathscr{C}$ be the morphism G–space of $\mathscr{C}$. The functor $\mathscr{C} \to \mathcal{E}X$ determined by a continuous G–map $f\colon \mathscr{O}b\mathscr{C} \to X$ is f on object G–spaces and the composite

$$\mathscr{M}or\mathscr{C} \xrightarrow{(T,S)} \mathscr{O}b\mathscr{C} \times \mathscr{O}b\mathscr{C} \xrightarrow{f\times f} X \times X$$

on morphism G–spaces. The last statement rephrases the meaning of chaotic. □

1.4 Translation categories and chaotic categories

We use another simple definition to relate chaotic categories to other familiar categories. Let G be a topological group and Y be a left G–space. Generalizing how we think of G as a one object category, we can think of Y together with its action by G as the functor $Y\colon G \to \mathscr{U}$ that sends the single object $*$ to Y and is given on morphism spaces by the map $G \to \mathrm{Map}(Y, Y)$ adjoint to the action map $G \times Y \to Y$.

Definition 1.6 Let Y be a left G–space. Define the translation category $T(G, Y)$ as follows. The object space is Y and the morphism space is $G \times Y$. We think of (g, y) as a morphism $g\colon y \to gy$. The map $I\colon Y \to G \times Y$ sends y to (e, y). The maps S and T send (g, y) to y and gy, respectively. The domain of composition, $(G \times Y) \times_Y (G \times Y)$, can be identified with $(G \times G) \times Y$, and composition sends (h, g, y) to (hg, y). The construction is functorial in Y, for fixed G, and in the pair (G, Y) in general. If Y has a right action by G that commutes with the left action, then $T(Y, G)$ is a right G–category via the given right action on the object space Y and on the second coordinate of the morphism space $G \times Y$.

Remark 1.7 The definition makes sense when G is only a monoid, not necessarily a group. When Y is a point, $T(Y, G)$ is G regarded as a one object category. When G is a group, $T(Y, G)$ is the standard groupoid associated to a G–space, but it is not generally chaotic.

Proposition 1.8 *For left G–spaces Y, there is a natural comparison functor $\mu\colon T(G, Y) \to \widetilde{Y}$. If Y has a right action that commutes with its left action, then μ is a map of right G–categories. The functor $\mu\colon T(G, G) \to \mathcal{E}G$ is an isomorphism of right G–categories.*

Proof Define μ to be the identity map on object spaces and the map that sends (g, y) to (gy, y) on morphism spaces. Since $\widetilde{Y}$ is chaotic, this is the unique functor that is the identity on objects, and it is easy to check equivariance when Y has a right G–action. When $Y = G$ with left action and right action given by its product, μ is an isomorphism with $\mu^{-1}(h, g) = (hg^{-1}, g)$ on morphism spaces. □

In view of the differing group actions on the morphism spaces $G \times G$, namely action on the right coordinate in $T(G,G)$ and diagonal action in $\mathcal{E}G$, the isomorphism between $T(G,G)$ and $\mathcal{E}G$ must not be viewed as a tautology.

Remark 1.9 When we return to the split extension (0-1), the group Π there will play a role close to that of the group denoted G in Definition 1.6 and Proposition 1.8. When $G = e$, we would then specialize to $Y = \Pi$ with its natural left Π action and see the usual universal principal Π–bundle. When $G \neq e$, the relevant specialization is a little less obvious; see Lemma 3.4, which is a follow up of Proposition 1.8.

2 The category $\mathscr{C}at(\mathcal{E}X, \Pi)$

We let X be a space and Π be a topological group in this section. We regard Π as a category with one object without change of notation; it should be clear from the context when we mean the group Π and when we mean the category Π. From now on, functors and natural transformations are to be continuous (in the topological sense), even when we neglect to say so. We are especially interested in the functor categories $\mathscr{C}at(\mathcal{E}X, \mathcal{E}\Pi)$, which are chaotic by Lemma 1.2, and in the functor categories $\mathscr{C}at(\mathcal{E}X, \Pi)$, which are not. The right action of Π on $\mathcal{E}\Pi$ induces a right action of Π on $\mathscr{C}at(\mathcal{E}X, \mathcal{E}\Pi)$.

This section and the next give a pedantically explicit description of $\mathscr{C}at(\mathcal{E}X, \Pi)$ and of the induced map

$$\mathscr{C}at(\mathcal{E}X, \mathcal{E}\Pi) \to \mathscr{C}at(\mathcal{E}X, \Pi),$$

showing in particular that it is obtained by passage to orbits over Π. When $X = G$, this elementary analysis will be at the heart of all our proofs. We defer adding in the second group G that appears in the bundle theory until after we have this description in place since a group defined solely in terms of the diagonal on X and the product on Π plays a central role in the description.

2.1 An explicit description of $\mathscr{C}at(\mathcal{E}X, \Pi)$

By the adjunction given in Lemma 1.5 (with $G = e$), the object space of the chaotic category $\mathscr{C}at(\mathcal{E}X, \mathcal{E}\Pi)$ can be identified with the space $\mathrm{Map}(X, \Pi)$ of maps $X \to \Pi$ with its standard (compactly generated) function space topology. Therefore $\mathscr{C}at(\mathcal{E}X, \mathcal{E}\Pi)$ can be identified with the chaotic category $\mathcal{E}\mathrm{Map}(X, \Pi)$.

Definition 2.1 Define the pointwise product $*$ on $\mathrm{Map}(X, \Pi)$ by

$$(\alpha * \beta)(x) = \alpha(x)\beta(x)$$

for $\alpha, \beta\colon X \to \Pi$. The unit element ε is given by $\varepsilon(x) = e$ and inverses are given by $\alpha^{-1}(x) = \alpha(x)^{-1}$. The topological group $\mathrm{Map}(X, \Pi)$ contains Π as a (closed) subgroup, where we regard an element $\sigma \in \Pi$ as the constant map $\sigma\colon X \to \Pi$ at σ. The inclusion of Π in $\mathrm{Map}(X, \Pi)$ and composition give $\mathrm{Map}(X, \Pi)$ its right Π–action.

Definition 2.2 Choose a basepoint $x_0 \in X$. There is a unique representative map α such that $\alpha(x_0) = e$ in each orbit of $\mathrm{Map}(X, \Pi)$ under the right action by Π. Let $\mathscr{O}(X, \Pi) \subset \mathrm{Map}(X, \Pi)$ denote the subspace of such representative maps. It is a subgroup of $\mathrm{Map}(X, \Pi)$. The Π–action and the product $*$ on $\mathrm{Map}(X, \Pi)$ are related by $\alpha\sigma = \alpha * \sigma$ for $\sigma \in \Pi$, and $*$ restricts to a homeomorphism of Π–spaces $\mathscr{O}(X, \Pi) \times \Pi \to \mathrm{Map}(X, \Pi)$. Write elements of $\mathrm{Map}(X, \Pi)$ in the form $\alpha\sigma$, where $\alpha(x_0) = e$. Passage to orbits restricts to a homeomorphism $\mathscr{O}(X, \Pi) \cong \mathrm{Map}(X, \Pi)/\Pi$. Observe that the product $*$ on $\mathrm{Map}(X, \Pi)$ induces a left action of $\mathrm{Map}(X, \Pi)$ on $\mathscr{O}(X, \Pi)$ by sending (β, α) to the orbit representative of $\beta * \alpha$.

The proofs of the follow three lemmas are simple exercises from the fact that there is a unique morphism (y, x) from x to y in $\mathcal{E}X$; compare Lemma 1.2.

Lemma 2.3 *A functor $E\colon \mathcal{E}X \to \Pi$ is given by the trivial map $X \to *$ of object spaces and a map $E\colon X \times X \to \Pi$ of morphism spaces such that $E(x, x) = e$ and $E(z, y)E(y, x) = E(z, x)$. Define $\alpha \in \mathscr{O}(X, \Pi)$ by $\alpha(x) = E(x, x_0)$. Then α determines E by the formula*

$$E(y, x) = E(y, x_0)E(x_0, x) = \alpha(y)\alpha(x)^{-1}.$$

Writing $E = E_\alpha$, sending E_α to α specifies a homeomorphism from the space of functors $\mathcal{E}X \to \Pi$ to $\mathscr{O}(X, \Pi)$.

Lemma 2.4 *For $E_\alpha, E_\beta\colon \mathcal{E}X \to \Pi$, a natural transformation $\eta\colon E_\alpha \to E_\beta$ is given by a map $\eta\colon X \to \Pi$ such that $\eta(y)E_\alpha(y, x) = E_\beta(y, x)\eta(x)$ for $x, y \in X$. If $\sigma \in \Pi$ is defined by $\sigma = \eta(x_0)$, then the pair $(\beta\sigma, \alpha)$ determines η by the formula*

$$\eta(x) = E_\beta(x, x_0)\eta(x_0)E_\alpha(x, x_0)^{-1} = (\beta\sigma * \alpha^{-1})(x).$$

Writing $\eta = \eta_\sigma$, sending η_σ to $(\beta\sigma, \alpha)$ specifies a homeomorphism from the space of morphisms of $\mathscr{C}at(\mathcal{E}X, \Pi)$ to the space $\mathrm{Map}(X, \Pi) \times \mathscr{O}(X, \Pi)$.

Lemma 2.5 *Identify the object and morphism spaces of $\mathscr{C}at(\mathcal{E}X, \Pi)$ with*

$$\mathscr{O}(X, \Pi) \quad \text{and} \quad \mathscr{M}(X, \Pi) \equiv \mathrm{Map}(X, \Pi) \times \mathscr{O}(X, \Pi)$$

via the homeomorphisms of the previous two lemmas. Then the identity map I sends α to $(\alpha e, \alpha)$ and the source and target maps S and T send $(\beta\sigma, \alpha)$ to α and to β. The $S = T$ pullback

$$\mathscr{M}(X,\Pi) \times_{\mathscr{O}(X,\Pi)} \mathscr{M}(X,\Pi)$$

can be identified with $\mathrm{Map}(X,\Pi) \times \mathrm{Map}(X,\Pi) \times \mathscr{O}(X,\Pi)$ *via*

$$((\gamma\tau,\beta),(\beta\sigma,\alpha)) \leftrightarrow (\gamma\tau,\beta\sigma,\alpha)$$

and the composition map C sends $(\gamma\tau,\beta\sigma,\alpha)$ to $(\gamma\tau\sigma,\alpha)$.

Proof If we compose $\eta_\tau\colon E_\beta \to E_\gamma$ with $\eta_\sigma\colon E_\alpha \to E_\beta$, we obtain

$$\eta_\tau * \eta_\sigma = \gamma^{-1}\tau * \beta * \beta^{-1}\sigma * \alpha = \gamma^{-1}\tau\sigma * \alpha,$$

which corresponds to the given description. □

2.2 Two identifications of $\mathscr{C}at(\mathcal{E}X,\Pi)$

We show here that Proposition 1.8 leads to one identification of $\mathscr{C}at(\mathcal{E}X,\Pi)$, and the lemmas of the previous section lead to a closely related one. These elementary identifications commute passage to orbits with the functor $\mathscr{C}at(\mathcal{E}X,-)$, and that will be crucial to understanding $B\mathscr{C}at(\mathcal{E}G,\Pi)$ as an equivariant classifying space.

Notation 2.6 The category Π is isomorphic to the orbit category $\mathcal{E}\Pi/\Pi$. The quotient functor $p\colon \mathcal{E}\Pi \to \Pi$ is the trivial map $\Pi \to *$ on object spaces and is given on morphism spaces by the map $p\colon \Pi\times\Pi \to (\Pi\times\Pi)/\Pi \cong \Pi$ specified by $p(\tau,\sigma) = \tau\sigma^{-1}$. Let q denote the functor

$$\mathscr{C}at(\mathrm{id},p)\colon \mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi) \to \mathscr{C}at(\mathcal{E}X,\Pi).$$

We also let q denote the functor between translation categories

$$T(\mathrm{Map}(X,\Pi),\mathrm{Map}(X,\Pi)) \to T(\mathrm{Map}(X,\Pi),\mathscr{O}(X,\Pi))$$

that is induced by the quotient map $p\colon \mathrm{Map}(X,\Pi) \to \mathrm{Map}(X,\Pi)/\Pi \cong \mathscr{O}(X,\Pi)$.

Theorem 2.7 *There is a commutative diagram of topological categories in which μ, ν, and ξ are isomorphisms:*

$$\begin{array}{ccccc}
T(\mathrm{Map}(X,\Pi),\mathrm{Map}(X,\Pi)) & & \xrightarrow{\quad\mu\quad} & & \mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi) \\
\downarrow{\scriptstyle q} & & \swarrow{\scriptstyle p} & & \downarrow{\scriptstyle q} \\
T(\mathrm{Map}(X,\Pi),\mathscr{O}(X,\Pi)) & \xrightarrow[\nu]{} & \mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi)/\Pi & \xrightarrow[\xi]{} & \mathscr{C}at(\mathcal{E}X,\Pi)
\end{array}$$

Proof The map p is the quotient map given by passage to orbits over Π. Since q on the right is a Π–map with Π acting trivially on $\mathscr{C}at(\mathcal{E}X, \Pi)$, q factors through a map ξ that makes the triangle commute. Since $\mathscr{C}at(\mathcal{E}X, \mathcal{E}\Pi)$ is the chaotic category whose object space is the topological group $\mathrm{Map}(X, \Pi)$, Proposition 1.8 gives the isomorphism μ. Since q on the left is obtained by passage to orbits from the relevant action of Π, it is clear that μ induces an isomorphism ν making the left trapezoid commute.

All that remains is to prove that ξ is an isomorphism, and that follows from the results of Section 2.1. For a functor $E_\alpha\colon \mathcal{E}X \to \Pi$, the maps $\alpha\colon X \to \Pi$ and $\alpha\times\alpha\colon X\times X \to \Pi\times\Pi$ define the object and morphism maps of a functor $F\colon \mathcal{E}X \to \mathcal{E}\Pi$. The functoriality properties of E_α show that $p \circ F = E_\alpha$, so that q is surjective on objects. If we also have $p \circ F' = E_\alpha$, then a quick check shows that $F(x)^{-1}F'(x) = F(y)^{-1}F'(y)$ for all $x, y \in X$. If the common value is denoted by σ, then $F'(x) = F(x)\sigma$ for all x. In view of the specification of p and q in Notation 2.6, this implies that ξ is a homeomorphism on object spaces.

Let $E_\alpha, E_\beta\colon \mathcal{E}X \to \Pi$ be any two functors. For any chosen functors $F, F'\colon \mathcal{E}X \to \mathcal{E}\Pi$ such that $q\circ F = E_\alpha$ and $q \circ F' = E_\beta$, define $\zeta\colon X \to \Pi\times\Pi$ by $\zeta(x) = (F(x), F'(x))$. Then ζ is a map from the object space of $\mathcal{E}X$ to the morphism space of $\mathcal{E}\Pi$. A quick check shows that ζ is a natural transformation $F \to F'$ such that $\eta = q\circ\zeta$ is a natural transformation $E_\alpha \to E_\beta$ with $\eta_{x_0} = F'(x_0)F(x_0)^{-1}$. Via our enumeration of the possible choices, this implies that q restricted to the inverse image of the space of natural transformations $E_\alpha \to E_\beta$ can be identified with the quotient map $p\colon \Pi\times\Pi \to \Pi$ of Notation 2.6. It follows that ξ is a homeomorphism on morphism spaces. □

2.3 The nerve functor and classifying spaces

We recall the definition of the nerve functor N in more detail than might be thought warranted at this late date since, in the presence of the left-right action dichotomy of multiple group actions, the original definitions in category theory can cause real problems arising from categorical dyslexia. There are two standard conventions in the literature, and we must choose. Let $\mathscr{C}$ be a topological category with object space $\mathscr{O}$ and morphism space $\mathscr{M}$. Then $N_0\mathscr{C} = \mathscr{O}$ and, for $q > 0$,

$$N_q\mathscr{C} = \mathscr{M} \times_{\mathscr{O}} \cdots \times_{\mathscr{O}} \mathscr{M},$$

with q factors $\mathscr{M}$. The pullbacks are over pairs of maps (S, T). To avoid dyslexia, we remember that $g \circ f$ means first f and then g, and choose to forget the picture

$$\bullet \xrightarrow{f_1} \bullet \xrightarrow{f_2} \bullet \to \cdots \to \bullet \xrightarrow{f_{q-1}} \bullet \xrightarrow{f_q} \bullet$$

of q composable arrows and instead remember that the picture

$$x_0 \xleftarrow{f_1} x_1 \xleftarrow{f_2} x_2 \leftarrow \cdots \leftarrow x_{q-2} \xleftarrow{f_{q-1}} x_{q-1} \xleftarrow{f_q} x_q \tag{2-8}$$

corresponds to an element $[f_1, \dots, f_q]$ of $N_q\mathscr{C}$, so that $S(f_i) = T(f_{i+1})$. For $x \in \mathscr{O}$, we write $\mathrm{id} = I(x)$ generically. Then

$$d_0[f] = T(f), \quad d_1[f] = S(f), \quad \text{and} \quad s_0(x) = [\mathrm{id}_x].$$

For $q \geq 2$,

$$d_i[f_1, \dots, f_q] = \begin{cases} [f_2, \dots, f_q] & \text{if } i = 0, \\ [f_1, \dots, f_{i-1}, f_i \circ f_{i+1}, f_{i+2}, \dots, f_q] & \text{if } 0 < i < q, \\ [f_1, \dots, f_{q-1}] & \text{if } i = q, \end{cases}$$

and, for $q \geq 1$,

$$s_i[f_1, \dots, f_q] = [f_1, \dots, f_i, \mathrm{id}, f_{i+1}, \dots, f_q].$$

Of course, these can and should be expressed in terms of the maps S, T, I, and C, so as to remember the topology and check continuity.

Recall that a (right) action of a group G on a simplicial space Y_* is specified by levelwise group actions such that the d_i and s_i are G–maps; formally, Y_* is a simplicial object in the category of (right) G–spaces. Orbit and fixed point simplicial spaces are constructed levelwise, $(Y_*/G)_q = Y_q/G$ and $(Y_*)^G_q = Y^G_q$. For a G–category $\mathscr{C}$, $N(\mathscr{C}^G) \cong (N\mathscr{C})^G$ since N is a right adjoint, but it is rarely the case that $N(\mathscr{C}/G) \cong (N\mathscr{C})/G$, as the following counterexample should make clear.

Example 2.9 Let G be a group and let G act on itself by conjugation. Let A be the abelianization of G. Regarding G and A as categories with a single object, $G/G \cong A$, and NA is generally much smaller than NG/G. Here $[g_1, \dots, g_q]$ and $[h_1, \dots, h_q]$ are in the same orbit under the conjugation action if and only if there is a single g such that $gg_ig^{-1} = gh_ig^{-1}$ for all i. For example if G is a finite simple group of order n, then A is trivial but N_qG/G has at least n^{q-1} elements.

In this example, NG is the simplicial space, often denoted by B_*G, whose geometric realization is the classifying space BG. Parametrizing with a left G–space Y gives a familiar simplicial space $B_*(*, G, Y)$ (eg [11, Section 7]). Write $q\colon E_*G \to B_*G$ for the map

$$B_*(*, G, G) \to B_*(*, G, *) \cong B_*(*, G, G)/G$$

induced by $G \to *$. The isomorphism on the right is obvious, but it is in fact an example of an isomorphism of the form $N(\mathscr{C}/G) \cong (N\mathscr{C})/G$, as the following observations make clear. Recall the translation category from Definition 1.6.

Lemma 2.10 *The simplicial space $NT(G,Y)$ is isomorphic to $B_*(*,G,Y)$.*

Proof A typical q–tuple (2-8) in $N_qT(G,Y)$ has i^{th} term

$$f_i = (g_i, g_{i+1}\cdots g_q y)\colon g_{i+1}\cdots g_q y \to g_i g_{i+1}\cdots g_q y$$

for elements $g_i \in G$ and $y \in Y$. It corresponds to $[g_1,\dots,g_q]y$ in $B_q(*,G,Y)$. □

Remark 2.11 For any space X, $N\mathcal{E}X$ is the simplicial space denoted by D_*X in [10, page 97]. Our choice of S and T on $\mathcal{E}X$ is consistent with (2-8) and the usual notation $(x_0,\dots,x_q)$ for q–simplices. The claim in Definition 1.1 that $|N\mathcal{E}X|$ is contractible is immediate from [10, 10.4], which says that D_*X is simplicially contractible. The isomorphism $N\mu\colon NT(G,G)\to N\mathcal{E}G$ implied by Proposition 1.8 coincides with the isomorphism $\alpha_*\colon E_*G\to D_*G$ of [10, 10.4].

Applying geometric realization, write $B(*,G,Y) = |B_*(*,G,Y)|$, and similarly for EG and BG. Then $B(*,G,Y)\cong B(*,G,G)\times_G Y = EG\times_G Y$. By Lemma 2.10,

$$BT(Y,G) = EG\times_G Y.$$

A relevant example is $Y = G/H$ for a (closed) subgroup H of G. The space

$$BT(G,G/H) = EG\times_G (G/H)\cong (EG)/H$$

is a classifying space BH since EG is a free contractible H–space.

In particular, take $G = \mathrm{Map}(X,\Pi)$ and $H=\Pi$ for a space X and group Π, remembering that $\mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi)$ is the chaotic category with object space the group $\mathrm{Map}(X,\Pi)$. Applying the classifying space functor to the diagram of Theorem 2.7 and using Lemma 2.10, we obtain the following commutative diagram, in which the horizontal maps are homeomorphisms and, up to canonical homeomorphisms, the vertical maps are obtained by passage to orbits over Π:

$$\begin{array}{ccccc}
E(\mathrm{Map}(X,\Pi)) & \xrightarrow{\cong} & B\mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi) & \xrightarrow{=} & B\mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi) \\
\downarrow & & \downarrow & & \downarrow \\
(E\,\mathrm{Map}(X,\Pi))/\Pi & \xrightarrow[\cong]{} & B(\mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi)/\Pi) & \xrightarrow[\cong]{} & B\mathscr{C}at(\mathcal{E}X,\Pi)
\end{array}$$

Ignoring minor topological niceness conditions,[3] for any space X the diagram gives isomorphic categorical models for the universal principal Π–bundle $E\Pi\to B\Pi$.

[3] The identity element of the group $\mathrm{Map}(X,\Pi)$ should be a nondegenerate basepoint and the space $\mathrm{Map}(X,\Pi)$ should be paracompact; see [11, 9.10].

3 Categorical universal equivariant principal bundles

3.1 Preliminaries on actions by the semidirect product Γ

Now return to the split extension (0-1) of the introduction. For a Γ–category or Γ–space, passage to orbits with respect to Π gives a G–category or a G–space. It is standard in equivariant bundle theory to let G act from the left and Π act from the right. Thus suppose that X is a left G and right Π object in any category. Using elementwise notation, turn the right action of Π into a left action by setting $\sigma x = x\sigma^{-1}$.

By an action of Γ on X, we mean a left action that coincides with the given actions when restricted to the subgroups $G = e \times G$ and $\Pi = \Pi \times e$ of Γ. Since $(\sigma, g) = (\sigma, e)(e, g)$, the action must be defined by

$$(\sigma, g)x = (\sigma, e)(e, g)x = (\sigma, e)gx = \sigma g x = (gx)\sigma^{-1}. \tag{3-1}$$

For now, we will denote the action of G on Π by $\cdot$, but we just use juxtaposition for the prescribed actions of G and Π on X. Since the action by g on Π is a group homomorphism, $g \cdot (\sigma\tau) = (g \cdot \sigma)(g \cdot \tau)$ and $g \cdot \sigma^{-1} = (g \cdot \sigma)^{-1}$. The interaction of Π and G in Γ is given by the twisted commutation relation

$$(e, g)(\sigma, e) = (g \cdot \sigma, g) = (g \cdot \sigma, e)(e, g),$$

or the same relation with σ replaced by σ^{-1}. Therefore (3-1) gives an action of Γ if and only if the given actions of Π and G satisfy the twisted commutation relation

$$g(x\sigma) = (gx)(g \cdot \sigma). \tag{3-2}$$

The placement of parentheses is crucial: we are taking group actions in different orders. When the action of G on Π is trivial, $g \cdot \sigma = \sigma$, this is the familiar statement that commuting left and right actions define an action by the product $G \times \Pi$.

Lemma 3.3 *For a G–category $\mathscr{A}$, the left G and right Π–actions on $\mathscr{C}at(\mathscr{A}, \mathcal{E}\Pi)$ extend naturally to a Γ–action.*

Proof We must verify that $g(F\sigma) = (gF)(g \cdot \sigma)$ for $g \in G$, $\sigma \in \Pi$ and a functor $F\colon \mathscr{A} \to \Pi$. The unique natural transformation $E \to F$ between a pair of functors E and F will then necessarily be given by Γ–maps. The verification is formal from the fact that G acts by conjugation, so that the action of G on Π is part of the prescription of the action of G on F. Recall that the left action of G on $\mathscr{C}at(\mathscr{A}, \mathcal{E}\Pi)$ is given by conjugation, $(gF)(a) = g \cdot F(g^{-1}a)$ for $g \in G$ and an object or morphism $a \in \mathscr{A}$.

The right action of Π is given by $(F\sigma)(a) = F(a)\sigma$. Then

$$\begin{aligned}(g(F\sigma))(a) &= g\cdot(F\sigma)(g^{-1}a)\\ &= g\cdot(F(g^{-1}a)\sigma)\\ &= (g\cdot F(g^{-1}a))(g\cdot\sigma)\\ &= ((gF)(a))(g\cdot\sigma)\\ &= ((gF)(g\cdot\sigma))(a).\end{aligned}$$ □

In particular, let $\mathscr{A} = \mathcal{E}X$ for a left G–space X. Then the given action of G on the object space X and the diagonal action of G on the morphism space $X\times X$ give a left G–action on the category $\mathcal{E}X$. Lemma 3.3 shows that the left G and right Π–action on $\mathscr{C}at(\mathcal{E}X,\mathcal{E}\Pi)$ give it an action by Γ. Explicitly, the conjugation left action by G and the evident right action by Π on the object space $\mathrm{Map}(X,\Pi)$ induce diagonal actions on the morphism space $\mathrm{Map}(X,\Pi)\times\mathrm{Map}(X,\Pi)$, and these specify left G and right Π–actions on $\mathscr{C}at(\mathcal{E}X,\Pi)$ that satisfy the commutation relation required for a Γ–action.

Specializing further to $X = G$, we have the following equivariant elaboration of Proposition 1.8. We change the group G there to the group $\mathrm{Map}(G,\Pi)$ here and remember that the product on $\mathrm{Map}(G,\Pi)$ is just the pointwise product induced by the product on Π, with no dependence on the product of G. Ignoring the group action, we may identify the chaotic right $\mathrm{Map}(G,\Pi)$–category with object space $\mathrm{Map}(G,\Pi)$ with the category $\mathscr{C}at(\mathcal{E}G,\mathcal{E}\Pi)$. The following lemma identifies group actions. Remember that Π is a subgroup of $\mathrm{Map}(G,\Pi)$.

Lemma 3.4 *The isomorphism of right* $\mathrm{Map}(G,\Pi)$*–categories*

$$\mu\colon T(\mathrm{Map}(G,\Pi),\mathrm{Map}(G,\Pi)) \to \mathscr{C}at(\mathcal{E}G,\mathcal{E}\Pi)$$

is an isomorphism of Γ*–categories, where the* G*–action on both source and target categories is given by the conjugation action on the object space* $\mathrm{Map}(G,\Pi)$ *and the resulting diagonal action on the morphism space* $\mathrm{Map}(G,\Pi)\times\mathrm{Map}(G,\Pi)$.

Proof Since μ is an isomorphism and a Π–map, we can and must give the source category the unique G–action such that μ is a G–map. Since μ is the identity map on object spaces, the action must be the conjugation action on the object space. On an element (β,α) of the morphism space, we must define

$$g(\beta,\alpha) = \mu^{-1}(g\mu(\beta,\alpha)) = \mu^{-1}(g(\beta\alpha), g\alpha) = \mu^{-1}((g\beta)(g\alpha), g\alpha) = (g\beta, g\alpha).$$ □

Lemma 3.5 *With $X = G$, the diagram of Theorem 2.7 is a commutative diagram of Γ–categories and maps of Γ–categories, where Γ acts through the quotient homomorphism $\Gamma \to G$ on the three categories on the bottom row.*

Proof Since the trapezoid is obtained by passing to orbits under the action of Π, it is a diagram of Γ–categories by Lemma 3.4. The functor $p\colon \mathcal{E}\Pi \to \Pi$ of Notation 2.6 is a G–map since

$$g\cdot(\tau,\sigma) = g\cdot(\tau\sigma^{-1}) = (g\cdot\tau)(g\cdot\sigma)^{-1} = p(g\cdot\tau, g\cdot\sigma).$$

It follows that the right vertical arrow $q = \mathscr{C}at(\mathcal{E}G, p)$ is a map of Γ–categories. Letting $[F]$ denote the orbit of a functor $F\colon \mathcal{E}G \to \mathcal{E}\Pi$ under the right action of Π, the functor ξ is specified by $\xi[F] = p \circ F$, and it follows that ξ is Γ–equivariant. □

3.2 Universal principal (G, Π_G)–bundles

Observe that for any G–category $\mathscr{A}$, the corepresented functor $\mathscr{C}at(\mathscr{A}, -)$ from G–categories to G–categories is a right adjoint and therefore preserves all limits. We take $\mathscr{A}$ to be the G–category $\mathcal{E}G$ from now on, and we use the functor $\mathscr{C}at(\mathcal{E}G, -)$ to obtain a convenient categorical description of universal principal (G, Π_G)–bundles. Variants of the construction are given in [12; 16].

Definition 3.6 Let G and Π be topological groups and let G act on Π. Define $E(G, \Pi_G)$ to be the Γ–space $B\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi) = |N\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)|$ and define $B(G, \Pi_G)$ to be the orbit G–space $E(G, \Pi_G)/\Pi$. Let $p\colon E(G, \Pi_G) \to B(G, \Pi_G)$ be the quotient map.

We need a lemma in order to prove that p is a universal (G, Π_G)–bundle in favorable cases. We defer the proof to the next section. We believe that the result is true more generally, but there are point-set topological issues obstructing a proof. We shall not obscure the simplicity of our work by seeking maximum generality. As usual in equivariant bundle theory, we assume that all given subgroups are closed.

Lemma 3.7 *Let Λ be a subgroup of Γ. If $\Lambda \cap \Pi \neq e$, then the fixed point category $\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)^{\Lambda}$ is empty. At least if G is discrete, if $\Lambda \cap \Pi = e$, then $\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)^{\Lambda}$ is nonempty and chaotic.*

The following result is [7, Theorem 9], but the details of the proof are in [6, Section 2]. A principal (G, Π_G)–bundle is numerable if it is trivial over the subspaces of B in a numerable open cover.

Theorem 3.8 *A numerable principal (G, Π_G)–bundle $p\colon E \to B$ is universal if and only if E^Λ is contractible for all (closed) subgroups Λ of Γ such that $\Lambda \cap \Pi = \{e\}$.*

We comment on the hypotheses. Recall from point-set topology that a space X is completely regular if for every closed subspace C and every point x not in C, there is a continuous function $f\colon X \to [0,1]$ such that $f(x) = 0$ and $f(C) = 1$. This is a weak condition that is satisfied by reasonable spaces, such as CW complexes.

Remark 3.9 Specializing [7, Propositions 4 and 5], a principal (G, Π_G)–bundle with completely regular total space is locally trivial, and a locally trivial principal (G, Π_G)–bundle over a paracompact base space (such as a CW complex) is numerable. Therefore, modulo weak point-set topological conditions, the fixed point condition in Theorem 3.8 is the essential criterion for a universal bundle.

Therefore Lemma 3.7 has the following consequence. Its condition on Π serves only to ensure that p is a numerable principal (G, Π_G)–bundle.

Theorem 3.10 *If G is discrete and Π is either discrete or a compact Lie group, the map*

$$p\colon E(G, \Pi_G) \to B(G, \Pi_G)$$

obtained by passage to orbits over Π is a universal principal (G, Π_G)–bundle.

The classifying space $B(G, \Pi_G) = |N\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)|/\Pi$ is obtained by first applying the classifying space functor and then passing to orbits. On the other hand, the space $B\mathscr{C}at(\mathcal{E}G, \Pi) = |N\mathscr{C}at(\mathcal{E}G, \Pi)|$ is obtained by first passing to orbits on the categorical level and then applying the classifying space functor. The category $\mathscr{C}at(\mathcal{E}G, \Pi)$ is thoroughly understood, as explained in Section 2. The key virtue of our model for $B(G, \Pi_G)$ is that these two G–spaces can be identified, by Theorem 2.7.

Theorem 3.11 *The canonical map*

$$B(G, \Pi_G) = |N\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)|/\Pi \to |N\mathscr{C}at(\mathcal{E}G, \Pi)| = B\mathscr{C}at(\mathcal{E}G, \Pi)$$

is a homeomorphism of G–spaces. Therefore, if G is discrete and Π is either discrete or a compact Lie group, the map

$$Bq\colon B\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi) \to B\mathscr{C}at(\mathcal{E}G, \Pi)$$

is a universal principal (G, Π_G)–bundle.

Scholium 3.12 For finite groups G, this result is claimed in [16, page 1294]. For more general groups G, [16, 3.1] states an analogous result, but with $\mathcal{E}\Pi \to \Pi$ replaced by a functor defined in terms of the nonequivariant universal bundle $E\Pi \to B\Pi$, resulting in a much larger construction. The replacement is needed for the proof of their analogue [16, 3.3] of our Lemma 3.7. A commutation relation of the form $N(\mathscr{C}/\Pi) = (N\mathscr{C})/\Pi$ for their larger construction is stated (five lines above [16, 3.1]), but there is no hint of a proof or of the need for one. It is not altogether clear to us that the commutation relation stated there is true, and we view the commutation relation Theorem 2.7 as the main point of the proof of Theorem 3.11. Nevertheless, Murayama and Shimakawa [16] had the insightful right idea that led to our work.

4 Determination of fixed points

4.1 The fixed point spaces of $E(G, \Pi_G)$

We must prove Lemma 3.7, but we place no restrictions on G and Π until they are needed. Since Π acts freely on $\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)$, it is clear that $\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)^\Lambda$ is empty if $\Lambda \cap \Pi \neq e$. Thus assume that $\Lambda \cap \Pi = e$. By Lemma 1.3, the fixed point category $\mathscr{C}at(\mathcal{E}G, \mathcal{E}\Pi)^\Lambda$ is chaotic. It remains to prove that it is nonempty, and Lemma 1.5 implies that this is so if and only if the space $\mathrm{Map}(G, \Pi)^\Lambda$ is nonempty. Thus it suffices to show that $\mathrm{Map}(G, \Pi)$ has a Λ–fixed point, which means that there is a Λ–map $f\colon G \to \Pi$. We prove this using the following standard generalization of a homomorphism and a variant needed later.

Definition 4.1 A function $\alpha\colon G \to \Pi$ is a crossed homomorphism if

$$\alpha(gh) = \alpha(g)(g \cdot \alpha(h)) \tag{4-2}$$

for all $g, h \in G$. In particular,

$$\alpha(e) = e,\ \alpha(g)^{-1} = g \cdot \alpha(g^{-1}) \quad \text{and} \quad \alpha(g^{-1})^{-1} = g^{-1} \cdot \alpha(g). \tag{4-3}$$

A map $\alpha\colon G \to \Pi$ is a crossed antihomomorphism if

$$\alpha(gh) = (g \cdot \alpha(h))\alpha(g). \tag{4-4}$$

Note that we should require the function α to be continuous in our general topological context. However, the continuity is sometimes automatic, as indicated in the following lemma. Remember that we understand subgroups to be closed.

Lemma 4.5 *All subgroups Λ of Γ such that $\Lambda \cap \Pi = e$ are of the form*

$$\Lambda_\alpha = \{(\alpha(h), h) \mid h \in H\},$$

where H is a subgroup of G and $\alpha\colon H \to \Pi$ is a crossed homomorphism. At least if G is discrete or Γ is compact, α is continuous.

Proof Clearly Λ_α is a subgroup of Γ such that $\Lambda_\alpha \cap \Pi = e$. Conversely, let $\Lambda \cap \Pi = e$. Define H to be the image of the composite of the inclusion $\iota\colon \Lambda \subset \Gamma$ and the projection $\pi\colon \Gamma \to G$. Since $\Lambda \cap \Pi = e$, the composite $\pi \circ \iota$ is injective and so restricts to a continuous isomorphism $\nu\colon \Lambda \to H$. For $h \in H$, define $\alpha(h) = \sigma$, where σ is the unique element of Π such that $(\sigma, h) \in \Lambda$. Thus α is the composite of $\iota \circ \nu^{-1}\colon H \to \Gamma$ and the projection $\rho\colon \Gamma \to \Pi$. If G is discrete or if Γ and therefore Λ is compact, then ν is a homeomorphism and α is continuous. For $h, k \in H$,

$$(\alpha(h), h)(\alpha(k), k) = (\alpha(h)(h \cdot \alpha(k)), hk) \in \Lambda,$$

so $\alpha(hk) = \alpha(h)(h \cdot \alpha(k))$. Thus α is a crossed homomorphism and $\Lambda = \Lambda_\alpha$. □

Proof of Lemma 3.7 We must obtain a Λ–map $f\colon G \to \Pi$, where $\Lambda = \Lambda_\alpha$ for a crossed homomorphism α. By the definition of the action by Λ, this means that

$$f(g) = (h \cdot f(h^{-1}g))\alpha(h)^{-1}$$

or equivalently

$$h \cdot f(h^{-1}g) = f(g)\alpha(h)$$

for all $h \in H$ and $g \in G$. We choose right coset representatives $\{g_i\}$ to write G as a disjoint union of cosets Hg_i. We then define $f\colon G \to \Pi$ by

$$f(kg_i) = \alpha(k)^{-1}$$

for $k \in H$. By using (4-2), writing out the inverse of a product as the product of inverses, using that $h^{-1}\cdot$ and $h\cdot$ are group homomorphisms and that $\cdot$ is a group action, and finally using (4-3) and, again, that $\cdot$ is a group action, we see that

$$\begin{aligned} h \cdot f(h^{-1}kg_i) &= h \cdot \alpha(h^{-1}k)^{-1} \\ &= h \cdot \big(\alpha(h^{-1})(h^{-1} \cdot \alpha(k))\big)^{-1} \\ &= h \cdot \big((h^{-1} \cdot \alpha(k))^{-1}(\alpha(h^{-1}))^{-1}\big) \\ &= \big(h \cdot (h^{-1} \cdot \alpha(k)^{-1})\big)\big(h \cdot (\alpha(h^{-1})^{-1})\big) \\ &= \alpha(k)^{-1}\big(h \cdot (h^{-1} \cdot \alpha(h))\big) \\ &= f(kg_i)\alpha(h) \end{aligned}$$

for all $h \in H$. Thus f is a Λ–map. We have assumed that G is discrete in order to ensure that f is continuous. □

Remark 4.6 If we relax the condition that G is discrete, we do not see how to prove that f is continuous, as would be needed for a more general result.

4.2 The fixed point categories of $\mathscr{C}at(\mathcal{E}G, \Pi)$

For $H \subset G$, the structure of the fixed point space $B(G, \Pi_G)^H$ is known (up to homotopy), for example by specialization of more general results in [7]. We show here how to see that structure on the category level. In fact, we identify the fixed point categories $\mathscr{C}at(\mathcal{E}G, \Pi)^H$, with no restrictions on Π and G. However, the reader may prefer to assume that G is discrete for the rest of Section 4.

Since the functor B commutes with fixed points, this gives a categorically precise interpretation of the fixed point space $B(G, \Pi_G)^H$.

We return to Section 2, taking $X = G$ there. The H–fixed functors and H–natural transformations in $\mathscr{C}at(\mathcal{E}G, \Pi)$ are the H–equivariant functors and natural transformations, in accord with our notational convention $\mathscr{C}at(\mathcal{E}G, \Pi)^H = H\mathscr{C}at(\mathcal{E}G, \Pi)$. Since $\mathcal{E}G$ and $\widetilde{H}$ are both H–free contractible categories, they are equivalent as H–categories. Therefore,

$$\mathscr{C}at(\mathcal{E}G, \Pi)^H \simeq \mathscr{C}at(\widetilde{H}, \Pi)^H = H\mathscr{C}at(\widetilde{H}, \Pi). \tag{4-7}$$

This implies that we may restrict to the case $G = H$ and deduce conclusions in general. The objects and morphisms of $G\mathscr{C}at(\mathcal{E}G, \Pi)$ are the G–equivariant functors $E\colon \mathcal{E}G \to \Pi$ and the G–equivariant natural transformations η. In Lemma 2.3, we described a functor E in terms of the map $\alpha\colon G \to \Pi$ defined by $\alpha(h) = E(h, e)$.

Lemma 4.8 *The G–action on functors $E\colon \mathcal{E}G \to \Pi$ induces the G–action on maps $\alpha\colon G \to \Pi$ specified by*

$$(g\alpha)(h) = \big(g \cdot (\alpha(g^{-1}h))(g \cdot \alpha(g^{-1})^{-1})\big).$$

Proof $(gE)(h, e) = g \cdot E(g^{-1}h, g^{-1}) = g \cdot (E(g^{-1}h, e)E(e, g^{-1}))$. □

Lemma 4.9 *The space of objects of $G\mathscr{C}at(\mathcal{E}G, \Pi)$ can be identified with the subspace of* $\mathrm{Map}(G, \Pi)$ *consisting of the crossed antihomomorphisms $\alpha\colon G \to \Pi$.*

Proof Setting $g\alpha = \alpha$ and applying $g^{-1} \cdot (-)$ to the formula for the action of G on α, we obtain

$$g^{-1} \cdot \alpha(h) = \alpha(g^{-1}h)\alpha(g^{-1})^{-1}.$$

Replacing g^{-1} by g and multiplying on the right by $\alpha(g)$, this gives

$$\alpha(gh) = (g \cdot \alpha(h))\alpha(g)$$

for all $g, h \in G$, which says that α is a crossed antihomomorphism. □

Similarly, as in Lemma 2.4, a natural transformation $\eta\colon E_\alpha \to E_\beta$ is determined by $\sigma = \eta(e)$. Explicitly,

$$\eta(g) = E_\beta(g,e)\eta(e)E_\alpha(g,e)^{-1} = \beta(g)\sigma\alpha(g)^{-1}$$

for $g \in G$. Now a G–fixed natural transformation η satisfies $\eta(gh) = g \cdot \eta(h)$ for $g, h \in G$ and thus $\eta(g) = \eta(ge) = g \cdot \eta(e) = g \cdot \sigma$. Therefore the naturality square for G–fixed natural transformations translates into

$$g \cdot \sigma = \beta(g)\sigma\alpha(g)^{-1}$$

or, equivalently,

$$\beta(g)\sigma = (g \cdot \sigma)\alpha(g). \tag{4-10}$$

We use the following definitions and lemma to put things together.

Definition 4.11 Let G act on Π. Define the crossed functor category $\mathscr{C}at_\times(G, \Pi)$ to be the category whose objects are the crossed homomorphisms $G \to \Pi$ and whose morphisms $\sigma\colon \alpha \to \beta$ are the elements $\sigma \in \Pi$ such that $\beta(g)(g \cdot \sigma) = \sigma\alpha(g)$; they are called isomorphisms of crossed homomorphisms. The composite $\tau \circ \sigma$, where $\tau\colon \beta \to \gamma$, is given by $\tau\sigma$. Define the centralizer Π^α of a crossed homomorphism $\alpha\colon G \to \Pi$ to be the subgroup

$$\Pi^\alpha = \{\sigma \in \Pi \mid \alpha(g)(g \cdot \sigma) = \sigma\alpha(g) \text{ for all } g \in G\}$$

of Π. It is the automorphism group $\mathrm{Aut}(\alpha)$ of the object α in $\mathscr{C}at_\times(G, \Pi)$.

Definition 4.12 Define the anticrossed functor category $\mathscr{C}at_\times^-(G, \Pi)$ to have objects the crossed antihomomorphisms $\alpha\colon G \to \Pi$ and morphisms $\sigma\colon \alpha \to \beta$ the elements $\sigma \in \Pi$ such that $\beta(g)\sigma = (g \cdot \sigma)\alpha(g)$, with $\tau \circ \sigma = \tau\sigma$. The centralizer Π^α of a crossed antihomomorphism $\alpha\colon G \to \Pi$ is

$$\Pi^\alpha = \{\sigma \in \Pi \mid \alpha(g)\sigma = (g \cdot \sigma)\alpha(g) \text{ for all } g \in G\}.$$

Again, $\Pi^\alpha = \mathrm{Aut}(\alpha)$ in $\mathscr{C}at_\times^-(G, \Pi)$.

If the action of G on Π is trivial, then the crossed functor category is just the functor category $\mathscr{C}at(G,\Pi)$ since homomorphisms $\alpha\colon G\to\Pi$ correspond to functors $\alpha\colon G\to\Pi$ and elements $\sigma\in\Pi$ such that $\beta(g)\sigma=\sigma\alpha(g)$ for $g\in G$ correspond to natural transformations $\alpha\to\beta$. In that case,

$$\Pi^{\alpha}=\{\sigma\in\Pi\mid\sigma^{-1}\alpha(g)\sigma=\alpha(g)\ \text{ for all }\ g\in G\}$$

is the usual centralizer of α in Π, and then the following identification is obvious.

Lemma 4.13 *The categories $\mathscr{C}at_{\times}(G,\Pi)$ and $\mathscr{C}at_{\times}^{-}(G,\Pi)$ of crossed homomorphisms and crossed antihomomorphisms are canonically isomorphic.*

Proof For a crossed homomorphism $\alpha\colon G\to\Pi$, define $\overline{\alpha}\colon G\to\Pi$ by

$$\overline{\alpha}(g)=g\cdot\alpha(g^{-1}).$$

Then

$$\overline{\alpha}(gh)=(gh)\cdot\alpha(h^{-1}g^{-1})=g\cdot h\cdot(\alpha(h^{-1})(h^{-1}\cdot\alpha(g^{-1}))=(g\cdot\overline{\alpha}(h))(\overline{\alpha}(g)),$$

so that $\overline{\alpha}$ is a crossed antihomomorphism. If σ is a morphism $\alpha\to\beta$ in $\mathscr{C}at_{\times}(G,\Pi)$, then $\beta(g)(g\cdot\sigma)=\sigma\alpha(g)$. It follows that

$$\overline{\beta}(g)\sigma=(g\cdot\beta(g^{-1}))\sigma=g\cdot(\beta(g^{-1})(g^{-1}\cdot\sigma))=g\cdot(\sigma\alpha(g^{-1}))=(g\cdot\sigma)\overline{\alpha}(g),$$

so that σ is also a morphism $\overline{\alpha}\to\overline{\beta}$ in $\mathscr{C}at_{\times}^{-}(G,\Pi)$. The construction of the inverse isomorphism is similar. □

Returning to the G–fixed category of interest, we summarize our discussion in terms of these definitions and results.

Theorem 4.14 *The fixed point category $G\mathscr{C}at(\mathcal{E}G,\Pi)=\mathscr{C}at(\mathcal{E}G,\Pi)^{G}$ is isomorphic to the anticrossed functor category $\mathscr{C}at_{\times}^{-}(G,\Pi)$. Therefore it is also isomorphic to the crossed functor category $\mathscr{C}at_{\times}(G,\Pi)$.*

Corollary 4.15 *For $H\subset G$, the fixed point category $\mathscr{C}at(\mathcal{E}G,\Pi)^{H}$ is equivalent to the anticrossed functor category $\mathscr{C}at_{\times}^{-}(H,\Pi)$. Therefore it is also equivalent to the crossed functor category $\mathscr{C}at_{\times}(H,\Pi)$.*

Remark 4.16 The appearance of antihomomorphisms in this context is not new; see eg [23]. As we have seen, it is also innocuous. We have chosen not to introduce opposite groups, but the anti-isomorphism $(-)^{-1}\colon\Pi\to\Pi^{\mathrm{op}}$ is relevant.

4.3 Fixed point categories, $H^1(G;\Pi_G)$, and Hilbert's Theorem 90

Since $G\mathscr{C}at(\mathcal{E}G,\Pi)$ is a groupoid, it is equivalent to the coproduct of its subcategories $\mathrm{Aut}(\alpha)$, where we choose one α from each isomorphism class of objects. The following definition is standard when Π and G are discrete but makes sense in general.

Definition 4.17 The first nonabelian cohomology group $H^1(G;\Pi_G)$ is the pointed set of isomorphism classes of (continuous) crossed homomorphisms $G\to\Pi$. We write $[\alpha]$ for the isomorphism class of α. The basepoint of $H^1(G;\Pi_G)$ is $[\varepsilon]$, where ε is the trivial crossed homomorphism given by $\varepsilon(g)=e$ for $g\in G$.

With this language, (4-7) and Corollary 4.15 can be restated as follows.

Theorem 4.18 *For $H\subset G$, $\mathscr{C}at(\mathcal{E}G,\Pi)^H$ is equivalent to the coproduct of the categories* $\mathrm{Aut}(\alpha)$, *where the coproduct runs over* $[\alpha]\in H^1(H;\Pi_H)$.

Here $\mathrm{Aut}(\alpha)$ implicitly refers to the ambient group $\Pi\rtimes H$, not $\Gamma=\Pi\rtimes G$. By (4-7) or, more concretely, Lemma 4.22 below, we obtain the same group $\mathrm{Aut}(\alpha)$ for α considered as an object of $\mathscr{C}at(\widetilde{K},\Pi)^H$ for any $H\subset K\subset G$.

For any G–category $\mathscr{A}$, we have a natural map of G–categories

$$\iota\colon \mathscr{A}\to\mathscr{C}at(\mathcal{E}G,\mathscr{A}).$$

It is induced by the unique G–functor $\mathcal{E}G\to *$, where $*$ is the trivial G–category with one object and its identity morphism. The G–fixed point functor ι^G played a central role in Thomason [22]. When $\mathscr{A}=\Pi$ for a G–group Π, ι sends the unique object of Π to the basepoint $[\varepsilon]\in H^1(G;\Pi)$.

We shall describe the groups $\mathrm{Aut}(\alpha)$ in familiar group-theoretic terms in the next section. As a special case, $\mathrm{Aut}(\varepsilon)=\Pi^G$ and ι^G restricts to the identity functor from Π^G to $\mathrm{Aut}(\varepsilon)$. This implies the following result.

Proposition 4.19 *The functor $\iota^G\colon\Pi^G\to\mathscr{C}at(\mathcal{E}G,\Pi)^G$ is an equivalence of categories if and only if $H^1(G;\Pi_G)=[\varepsilon]$.*

Example 4.20 Let E be a Galois extension of a field F with Galois group G. Then G acts on E and $E^G=F$. Let G act entrywise on $\mathrm{GL}(n,E)$. Then Serre's general version of Hilbert's Theorem 90 [20, Chapter 10, Proposition 3] gives that $H^1(G;\mathrm{GL}(n,E)_G)=[\varepsilon]$. Since $\mathrm{GL}(n,E)^G=\mathrm{GL}(n,F)$, we conclude that ι^G is an equivalence of categories

$$\mathrm{GL}(n,F)\to\mathscr{C}at(\mathcal{E}G,\mathrm{GL}(n,E))^G.$$

More generally ι^H for $H \subset G$ is an equivalence of categories

$$\mathrm{GL}(n, E^H) \to \mathscr{C}at(\mathcal{E}G, \mathrm{GL}(n, E)^H.$$

As explained in [4] this gives precisely the information that ensures that the algebraic K–theory fixed point spectrum $\mathbb{K}_G(E)^H$ is equivalent to $\mathbb{K}(E^H)$. We shall return to consideration of G–rings such as E in Section 6.

We recall the easy calculation of $H^1(G;\Pi)$ in group-theoretic terms. Here we must restrict G since the proof depends on Lemma 3.7.

Lemma 4.21 *At least if G is discrete, the set $H^1(G;\Pi)$ is in bijective correspondence with the set of Π–conjugacy classes of subgroups Λ of Γ such that $\Lambda \cap \Pi = e$ and $q(\Lambda) = G$.*

Proof By Lemma 3.7, the subgroups Λ of Γ such that $\Lambda \cap \Pi = e$ are of the form

$$\Lambda_\alpha = \{(\alpha(h), h) \mid h \in H\}$$

for a crossed homomorphism $\alpha\colon H \to \Pi$. If $\sigma \in \Pi$, then $\sigma\Lambda_\alpha\sigma^{-1} \cap \Pi = e$ and therefore $\sigma\Lambda_\alpha\sigma^{-1} = \Lambda_\beta$ for some crossed homomorphism β. The equality forces β and α to be defined on the same subgroup H and to satisfy $\beta(g)(g\cdot\sigma) = \sigma\alpha(g)$. We are concerned only with the case $H = G$, and then this says that σ is a morphism and thus an isomorphism $\alpha \to \beta$ in $\mathscr{C}at_\times(G,\Pi)$. □

4.4 The fixed point spaces of $B(G, \Pi_G)$

We here identify the automorphism groups $\mathrm{Aut}(\alpha)$ group-theoretically and so complete the identification of $\mathscr{C}at(\mathcal{E}G,\Pi)^G$.

Lemma 4.22 *Let $\alpha\colon H \to \Pi$ be a crossed homomorphism and Π be a G–group, where $H \subset G$. Then the crossed centralizer Π^α is the intersection $\Pi \cap N_\Gamma\Lambda_\alpha$. Therefore this intersection is the same for all $\Gamma_K = \Pi \rtimes K$, $H \subset K \subset G$.*

Proof Let $(\pi, g) \in \Pi \rtimes G$ and $h \in H$. Calculating in $\Gamma = \Pi \rtimes G$, we have

$$\begin{aligned}(\sigma,g)^{-1}(\alpha(h),h)(\sigma,g) &= (g^{-1}\cdot\sigma^{-1}, g^{-1})(\alpha(h),h)(\sigma,g)\\ &= ((g^{-1}\cdot\sigma^{-1})(g^{-1}\cdot\alpha(h)), g^{-1}h)(\sigma,g)\\ &= \big((g^{-1}\cdot\sigma^{-1})(g^{-1}\cdot\alpha(h))((g^{-1}h)\cdot\sigma), g^{-1}hg\big).\end{aligned}$$

Therefore, (σ,g) is in $N_\Gamma\Lambda_\alpha$ if and only if g is in N_GH and

$$\alpha(g^{-1}hg) = (g^{-1}\cdot\sigma^{-1})(g^{-1}\cdot\alpha(h))((g^{-1}h)\cdot\sigma)$$

for all $h \in H$. When $g = e$, so that $\sigma = (\sigma, e)$ is a typical element of $\Pi \cap N_\Gamma \Lambda_\alpha$, this simplifies to

$$\alpha(h) = \sigma^{-1}\alpha(h)(h \cdot \sigma). \qquad \square$$

Passing to classifying spaces from Theorem 4.18 gives the following result.

Theorem 4.23 *For $H \subset G$,*

$$B(G, \Pi_G)^H = B\mathscr{C}at(\mathcal{E}G, \Pi)^H \simeq \coprod B\,\mathrm{Aut}(\alpha),$$

where the coproduct runs over $[\alpha] \in H^1(H; \Pi_H)$.

By Lemmas 4.21 and 4.22, at least when G is discrete we can restate Theorem 4.23 as follows.

Theorem 4.24 *Let $\Gamma = \Pi \rtimes G$, where G is discrete. For a subgroup H of G,*

$$B(G, \Pi_G)^H \simeq \coprod B(\Pi \cap N_\Gamma \Lambda),$$

where the coproduct runs over the Π–conjugacy classes of subgroups Λ of Γ such that $\Lambda \cap \Pi = e$ and $q(\Lambda) = H$.

Of course, we are only entitled to consider $B(G, \Pi_G)$ as a classifying space for principal Γ–bundles when Theorem 3.11 applies. The fixed point spaces $B(\Pi; \Gamma)^H$ of classifying spaces are studied more generally in [7] when Γ is given by a not necessarily split extension of compact Lie groups

$$1 \to \Pi \to \Gamma \xrightarrow{q} G \to 1. \tag{4-25}$$

For such groups Γ, Theorem 10 of [7] gives an entirely different bundle-theoretic proof that the conclusion of Theorem 4.24 still holds as stated, but without the restriction on G. However, when [7] was written, no particularly nice model for the homotopy type $B(\Pi; \Gamma)$ was known.

5 The comparison between $B\mathscr{C}at(\mathcal{E}G, \Pi)$ and $\mathrm{Map}(EG, B\Pi)$

A convenient model $p\colon E(\Pi; \Gamma) \to B(\Pi; \Gamma)$ for a universal principal $(\Pi; \Gamma)$–bundle was later given in terms of mapping spaces [12]. Here we assume given an extension (4-25), with no restrictions on our topological groups.[4] Start with the classical models

[4] We do assume their identity elements are nondegenerate basepoints.

in Section 2.3 for universal principal Π, G, and Γ–bundles and let $Eq\colon E\Gamma \to EG$ be the map induced by the quotient homomorphism $q\colon \Gamma \to G$. Let $\mathrm{Sec}(EG, E\Gamma)$ denote the Γ–space of sections $f\colon EG \to E\Gamma$, so that $Eq \circ f = \mathrm{id}$. The following result is part of [12, Theorem 5].

Theorem 5.1 *The quotient map* $p\colon \mathrm{Sec}(EG, E\Gamma) \to \mathrm{Sec}(EG, E\Gamma)/\Pi$ *is a universal principal* $(\Pi; \Gamma)$*–bundle.*

Now let the extension be split, so that $\Gamma = \Pi \rtimes G$. The given action of G induces a left action of G on $E\Pi$ that, together with the free right action by Π, makes it a Γ–space. Taking EG to be a left G–space and letting Γ act through q on EG, we have the product Γ–space $E\Pi \times EG$. It is free as a Γ–space because $E\Pi$ is free as a Π–space and EG is free as a G–space. Since it is contractible, we may as well take $E\Gamma = E\Pi \times EG$. Since the second coordinate of a section $f\colon EG \to E\Pi \times EG$ must be the identity, we then have

$$\mathrm{Sec}(EG, E\Gamma) = \mathrm{Map}(EG, E\Pi).$$

Its Γ–action is defined just as was the Γ–action on $\mathscr{C}at(\mathcal{E}G, \Pi)$ in Lemma 3.3. This gives the following specialization of Theorem 5.1, which is the space level forerunner of the categorical Theorem 3.10.

Theorem 5.2 *The quotient map* $p\colon \mathrm{Map}(EG, E\Pi) \to \mathrm{Map}(EG, E\Pi)/\Pi$ *is a universal principal* (G, Π_G)*–bundle.*

We also have the mapping space $\mathrm{Map}(EG, B\Pi)$. The canonical map $E\Pi \to B\Pi$ induces a map $q\colon \mathrm{Map}(EG, E\Pi) \to \mathrm{Map}(EG, B\Pi)$. Then there is an induced map ξ that makes the following diagram commute:

$$\begin{array}{ccc} & & \mathrm{Map}(EG, E\Pi) \\ & \swarrow^{p} & \downarrow q \\ \mathrm{Map}(EG, E\Pi)/\Pi & \xrightarrow[\xi]{} & \mathrm{Map}(EG, B\Pi) \end{array}$$

The analogy with the triangle in Theorem 2.7 should be evident. As observed in [12, Theorem 5], elementary covering space theory gives the following space level forerunner of the categorical Theorem 3.11.

Theorem 5.3 *If* Π *is discrete, then* $\xi\colon \mathrm{Map}(EG, E\Pi)/\Pi \to \mathrm{Map}(EG, B\Pi)$ *is a homeomorphism and therefore* $q\colon \mathrm{Map}(EG, E\Pi) \to \mathrm{Map}(EG, B\Pi)$ *is a universal principal* (G, Π_G)*–bundle.*

Note that G but not Π is required to be discrete in Theorem 3.11, whereas Π but not G is required to be discrete in Theorem 5.3.[5] There is an obvious comparison map relating the categorical and space level constructions. For any G–categories $\mathscr{A}$ and $\mathscr{B}$, we have the evaluation G–functor

$$\varepsilon\colon \mathscr{C}at(\mathscr{A},\mathscr{B})\times\mathscr{A}\to\mathscr{B}.$$

Applying the classifying space functor and taking adjoints, this gives a G–map

$$\xi\colon B\mathscr{C}at(\mathscr{A},\mathscr{B})\to \mathrm{Map}(B\mathscr{A},B\mathscr{B}). \tag{5-4}$$

When $\mathscr{A}$ and $\mathscr{B}$ are both discrete (in the topological sense), there is a simple analysis of this map in terms of the simplicial mapping space $\mathrm{Map}^{\Delta}(N\mathscr{A},N\mathscr{B})$. The following two lemmas are well-known nonequivariantly.

Lemma 5.5 *For discrete categories $\mathscr{A}$ and $\mathscr{B}$, there is a natural isomorphism*

$$\mu\colon N\mathscr{C}at(\mathscr{A},\mathscr{B})\cong \mathrm{Map}^{\Delta}(N\mathscr{A},N\mathscr{B}),$$

and this is an isomorphism of simplicial G–sets if $\mathscr{A}$ and $\mathscr{B}$ are G–categories.

Proof Let Δ_n be the poset $\{0,1,\dots,n\}$, viewed as a category. The n–simplices of $\mathscr{C}at(\mathscr{A},\mathscr{B})$ are the functors $\Delta_n\to\mathscr{C}at(\mathscr{A},\mathscr{B})$. By adjunction, they are the functors $\mathscr{A}\times\Delta_n\to\mathscr{B}$. Since N is full and faithful, these functors are the maps of simplicial sets

$$N\mathscr{A}\times N\Delta_n\cong N(\mathscr{A}\times\Delta_n)\to N\mathscr{B}.$$

By definition, these maps are the n–simplices of $\mathrm{Map}^{\Delta}(N\mathscr{A},N\mathscr{B})$. These identifications give the claimed isomorphism of simplicial sets. The compatibility with the actions of G when $\mathscr{A}$ and $\mathscr{B}$ are G–categories is clear. □

Lemma 5.6 *For simplicial sets K and L, there is a natural map*

$$\nu\colon |\mathrm{Map}^{\Delta}(K,L)|\to \mathrm{Map}(|K|,|L|).$$

If K and L are simplicial G–sets, ν is a map of G–spaces, and it is a weak equivalence of G–spaces when L is a Kan complex.

[5] When G is a compact Lie group acting trivially on a compact abelian Lie group Π, results of [8] imply that the map ξ is a weak G–equivalence; in [18], Charles Rezk proves that this remains true when Π is a finite extension of a torus (a compact Lie homotopy 1–type).

Proof The evaluation map $\mathrm{Map}^{\Delta}(K,L)\times K \to L$ induces a map

$$|\mathrm{Map}^{\Delta}(K,L)|\times|K| \cong |\mathrm{Map}^{\Delta}(K,L)\times K| \to |L|$$

whose adjoint is ν. When L is a Kan complex, so is $\mathrm{Map}^{\Delta}(K,L)$ (eg [9, 6.9]), and the natural maps $L\to S|L|$ and $\mathrm{Map}^{\Delta}(K,L)\to S|\mathrm{Map}^{\Delta}(K,L)|$ are homotopy equivalences, where S is the total singular complex functor. A diagram chase shows that ξ induces a bijection on homotopy classes of maps

$$\xi_*\colon [|J|,|\mathrm{Map}^{\Delta}(K,L)|]\to \big[|J|,\mathrm{Map}(|K|,|L|)\big]$$

for any simplicial set J. Letting G act trivially on J, all functors in sight commute with passage to H–fixed points, and the equivariant conclusions follow. □

Now the following result is immediate from the definitions and lemmas above.

Proposition 5.7 *For discrete G–categories $\mathscr{A}$ and $\mathscr{B}$, the map ξ of* (5-4) *is the composite $\nu\circ\mu$, and it is a weak G–equivalence if $\mathscr{B}$ is a groupoid.*

Returning to the topological setting, take $\mathscr{A}=\mathcal{E}G$ and write $EG=|N\mathcal{E}G|$, as we may. Recalling that $E\Pi\to B\Pi$ is obtained by applying B to the functor $\mathcal{E}\Pi\to\Pi$, we obtain the following commutative diagram:

$$\begin{array}{ccc}
B\mathscr{C}at(\mathcal{E}G,\mathcal{E}\Pi) & \longrightarrow & \mathrm{Map}(EG,E\Pi)\\
\downarrow & & \downarrow\\
B\mathscr{C}at(\mathcal{E}G,\mathcal{E}\Pi)/\Pi & \longrightarrow & \mathrm{Map}(EG,E\Pi)/\Pi\\
\downarrow & & \downarrow\\
B\mathscr{C}at(\mathcal{E}G,\Pi) & \longrightarrow & \mathrm{Map}(EG,B\Pi)
\end{array}$$

Theorems 3.10 and 5.2 say that the top two vertical arrows are often universal principal $(\Pi;\Gamma)$–bundles, in which case the top two horizontal arrows are equivalences. Theorems 3.11 and 5.3 say that the lower two vertical arrows and therefore also the bottom horizontal arrow are also often equivalences. When both Π and G are discrete, the equivalences are immediate from Proposition 5.7. More elaborate arguments might prove all of these results in greater topological generality.

6 Other categorical models for classifying spaces $B(G,\Pi_G)$

For particular G–groups Π, there are alternative categorical models for universal principal (G,Π_G)–bundles that are important in our applications in [4; 14]. They lead to equivalent, but more intuitive, constructions of categorical models for a number of interesting G–spectra, in particular suspension G–spectra and the equivariant K–theory spectra of rings with actions by G.

Perhaps surprisingly, the symmetric groups Σ_n with trivial G–action are of particular importance in equivariant infinite loop space theory. For a ring R with an action of a group G via ring maps, the general linear groups $\mathrm{GL}(n, R)$ with G–action on all matrix entries are of particular importance. We give alternative models for universal principal bundles applicable to these cases. We focus on the total spaces here and explain additional structure on the resulting classifying spaces in [4]. We assume that G is finite, although some of the definitions make sense and are interesting more generally.

6.1 A model $\widetilde{\mathscr{E}}_G(n)$ for $E(G, \Sigma_n)$

Definition 6.1 Let U be a countable ambient G–set that contains countably many copies of each orbit G/H. The action of G on U fixes bijections $g\colon A \to gA$ for all finite subsets A of U, denoted by $a \mapsto g \cdot a$.

Let $\boldsymbol{n} = \{1, \dots, n\}$ and view elements $\sigma \in \Sigma_n$ as functions $\boldsymbol{n} \to \boldsymbol{n}$, so that $\sigma(i) = \sigma \cdot i$ gives a left action of Σ_n on $\boldsymbol{n}$.

Definition 6.2 For $n \geq 0$, let $\widetilde{\mathscr{E}}_G(n)$ denote the chaotic $(\Sigma_n \times G)$-category whose set $\mathscr{O}b$ of objects is the set of pairs (A, α), where A is an n–element subset of U and $\alpha\colon \boldsymbol{n} \to A$ is a bijection. Let G act on $\mathscr{O}b$ on the left by postcomposition and let Σ_n act on the right by precomposition. Thus $g(A, \alpha) = (gA, g \circ \alpha)$ for $g \in G$, and $(A, \alpha)\sigma = (A, \alpha \circ \sigma)$ for $\sigma \in \Sigma$; of course,

$$(g \circ \alpha) \circ \sigma = g \circ \alpha \circ \sigma = g \circ (\alpha \circ \sigma).$$

The action of $\Sigma_n \times G$ is given by $(\sigma, g)(A, \alpha) = (gA, g \circ \alpha \circ \sigma^{-1})$. Since $\widetilde{\mathscr{E}}_G(n)$ is chaotic, this fixes the actions on the morphism set, which the map (S, T) identifies with $\mathscr{O}b \times \mathscr{O}b$ with $\Sigma_n \times G$ acting diagonally.

Proposition 6.3 *For each n, the classifying space $|N\widetilde{\mathscr{E}}_G(n)|$ is a universal principal (G, Σ_n)–bundle.*

Proof For each A, choose a base bijection $\eta_A\colon \boldsymbol{n} \to A$. The function sending σ to $(A, \eta_A \circ \sigma)$ is an isomorphism of right Σ_n–sets from Σ_n to the set of objects (A, α); its inverse sends (A, α) to $\eta_A^{-1} \circ \alpha$. Thus Σ_n acts freely on $\widetilde{\mathscr{E}}_G(n)$. Since $\widetilde{\mathscr{E}}_G(n)$ is chaotic, it suffices to show that the set of objects of $\widetilde{\mathscr{E}}_G(n)^\Lambda$ is nonempty if $\Lambda \cap \Sigma_n = \{e\}$. As usual, $\Lambda = \{(\rho(h), h) \mid h \in H\}$, where H is a subgroup of G and $\rho\colon H \to \Sigma_n$ is a homomorphism.

Let H act through ρ on $\boldsymbol{n}$, so that $h \cdot i = \rho(h)(i)$. Since U contains a copy of every finite G–set, there is a bijection of G–sets $\beta\colon G \times_H \boldsymbol{n} \to B \subset U$. Its restriction to $\boldsymbol{n}$ gives a bijection of H–sets $\alpha\colon \boldsymbol{n} \to A \subset B$. We claim that this (A, α) is a Λ–fixed object. Obviously $hA = A$ for $h \in H$. By Definition 6.2, we have $(\rho(h), h)(A, \alpha) = (A, h \circ \alpha \circ \rho(h)^{-1})$, where

$$(h \circ \alpha \circ \rho(h)^{-1})(i) = h \cdot \alpha(\rho(h)^{-1}(i)) = h \cdot h^{-1} \cdot \alpha(i) = \alpha(i). \qquad \square$$

Definition 6.4 Define $\mathscr{E}_G(n)$ to be the orbit G–category $\widetilde{\mathscr{E}}_G(n)/\Sigma_n$.

By Proposition 6.3 and Section 2.3, $B\mathscr{E}_G(n)$ is a classifying space $B(G, \Sigma_n)$. Up to isomorphism, the G–category $\mathscr{E}_G(n)$ admits the following more explicit description.

Lemma 6.5 *The objects of $\mathscr{E}_G(n)$ are the n–pointed subsets A of U. The morphisms are the bijections $\alpha\colon A \to B$, with the evident composition and identities. The group G acts by translation on objects and by conjugation on morphisms. That is, g sends A to gA and α to $g\alpha$, where $g\alpha = g \circ \alpha \circ g^{-1}$, so that $(g\alpha)(g \cdot a) = g \cdot \alpha(a)$.*

Proof The objects (A, α) are all in the same orbit, denoted by A, and the bijections η_A chosen in the proof of Proposition 6.3 give orbit representatives for the objects of $\mathscr{E}_G(n)$. In $\widetilde{\mathscr{E}}_G(n)$, we have a unique morphism $\iota_\beta\colon (A, \eta_A) \to (B, \beta)$ for each bijection $\beta\colon \boldsymbol{n} \to B$, and these morphisms give orbit representatives for the set of morphisms $A \to B$ in $\mathscr{E}_G(n)$. Letting the orbit of ι_β correspond to the bijection $\alpha = \beta \circ \eta_A^{-1}\colon A \to B$ and noting that $\alpha = \eta_B \circ \sigma \circ \eta_A^{-1}$ for a unique $\sigma \in \Sigma_n$, we obtain the claimed description of $\mathscr{E}_G(n)$. Since η_A specifies an ordering on A, η_{gA} is fixed as $g \circ \eta_A$. Then, if $\alpha = \beta \circ \eta_A^{-1}$,

$$g \circ \alpha \circ g^{-1} = g \circ (\beta \circ \eta_A^{-1}) \circ (\eta_A \circ \eta_{gA}^{-1}) = g \circ \beta \circ \eta_{gA}^{-1}\colon gA \to gB. \qquad \square$$

6.2 G–rings, G–ring modules, and crossed homomorphisms

By a G–ring we understand a ring R with a left action of G on R through ring automorphisms. We do not assume that R is commutative, although that is the case of greatest interest to us. Following the literature, we write $g(r) = r^g$ for the automorphism $g\colon R \to R$ determined by $g \in G$. Then $r^{gh} = g(h(r)) = (r^h)^g$.

When R is a subquotient of $\mathbb{Q}$, the only automorphism of R is the identity and the action of G must be trivial, but nontrivial examples abound. One important example is the action of the Galois group on a Galois extension E of a field F.

In the next section we will give an analogue of $\widetilde{\mathscr{E}}_G(n)$ but with $\Pi = \Sigma_n$ replaced by $\Pi = \mathrm{GL}(n, R)$ with the entrywise action of G. We will need a tiny bit of what appears to us to be a relatively undeveloped part of representation theory.

For a G–ring R, there are standard notions of a "crossed product" ring, a "group-graded ring", and, as a special case of both, a "skew group ring", variously denoted $R \rtimes G$ or $R * G$. We shall use the notation $R_G[G]$ for the last of these notions. If the action of G on R is given by the homomorphism $\theta\colon G \to \mathrm{Aut}(R)$, a more precise notation would be $R_\theta[G]$. Observe that R is a k–algebra, where k denotes the intersection of the center of R with R^G.

Definition 6.6 As an R–module, $R_G[G]$ is the same as the group ring $R[G]$, which is the case when G acts trivially on R. We define the product on $R_G[G]$ by k–linear (not R–linear) extension of the relation

$$(rg)\,(sh) = rs^g\, gh$$

for $r, s \in R$ and $g, h \in G$. Thus R and $k[G]$ are subrings of $R_G[G]$ and

$$g\, r = r^g\, g.$$

Definition 6.7 We call (left) $R_G[G]$–modules "G–ring modules" or "skew G–modules". Such an M is a left R–module and a left $k[G]$–module such that $g(rm) = r^g(gm)$ for $m \in M$. If M is R–free, we call M a skew representation of G over R.

Although special cases have appeared and there is a substantial literature on crossed products, group-graded rings, and skew group rings (for example [2; 15; 17]), we have not found a systematic study of these representations in the literature. Kawakubo [5] gives a convenient starting point. The following relationship with crossed homomorphisms is [5, 5.1].

Theorem 6.8 *Let R be a G–ring. Then the set of isomorphism classes of $R_G[G]$–module structures on the R–module R^n is in canonical bijective correspondence with $H^1(G; \mathrm{GL}(n, R))$. In detail, let $\{e_i\}$ be the standard basis for R^n. Then the formula*

$$ge_i = \rho(g)(e_i)$$

establishes a bijection between $R_G[G]$–module structures on R^n and crossed homomorphisms $\rho\colon G \to \mathrm{GL}(n, R)$. Moreover, two $R_G[G]$–modules with underlying R–module R^n are isomorphic if and only if their corresponding crossed homomorphisms are isomorphic.

Proof Given an $R_G[G]$–module structure on R^n, define the matrix $\rho(g)$ in $\mathrm{GL}(n,R)$ by letting its i^{th} column be $(s_{i,j})$, where

$$ge_i = \sum_j s_{i,j} e_j.$$

Conversely, given ρ, write $\rho(g) = (s_{i,j})$ and define ge_i by the same formula. From either starting point, we have $ge_i = \rho(g)(e_i)$. For a second element $h \in G$, write $\rho(h) = (t_{i,j})$, where $\rho(h)$ is either determined by an $R_G[G]$–module structure or is given by a crossed homomorphism ρ. Since $g\,r = r^g\,g$ in $R_G[G]$ and $g\,(r_{i,j}) = (r^g_{i,j})$ in $\mathrm{GL}(n,R)$, the relation $(gh)e_i = g(he_i)$ required of an $R_G[G]$–module is the same as the relation $\rho(g)\rho(h)(e_i) = \rho(g)(g\rho(h))(e_i)$ required of a crossed homomorphism. Indeed, $(gh)e_i = \rho(gh)(e_i)$ and

$$\begin{aligned} g(he_i) = g\rho(h)(e_i) &= \sum_j g(t_{i,j}e_j) = \sum_j t^g_{i,j} g e_j = \sum_j \sum_k t^g_{i,j} s_{j,k} e_k \\ &= \rho(g)\Bigl(\sum_j t^g_{i,j} e_j\Bigr) = \rho(g)(g\rho(h)(e_i)). \end{aligned}$$

The remaining compatibilities, in particular for the transitivity relation required of a module, are equally straightforward verifications, as is the verification of the statement about isomorphisms. □

The following easy observation specifies the permutation skew representations. For a set A, let $R[A]$ denote the free R–module on the basis A.

Proposition 6.9 *Let A be a G–set and define*

$$g\Bigl(\sum_a r_a a\Bigr) = \sum_a r^g_a\, ga$$

for $g \in G$, $r_a \in R$ and $a \in A$. Then $R[A]$ is an $R_G[G]$–module.

In view of Theorem 6.8, this has the following immediate consequence.

Corollary 6.10 *For a G–ring R, any n–pointed G–set A canonically gives rise to a crossed homomorphism $\rho_A\colon G \to \mathrm{GL}(n,R)$.*

We shall need to embed skew representations in permutation skew representations to apply these notions in equivariant bundle (or covering space) theory. Of course, in classical representation theory over $\mathbb{C}$, every representation embeds in a permutation representation. We need an analogue for skew representations.

Definition 6.11 A G–ring R is amenable if there is a monomorphism of $R_G[G]$–modules that embeds any finite-dimensional skew representation of G over R into a finite-dimensional permutation skew representation.

Example 6.12 Let G act trivially on $\boldsymbol{n} = \{1, \dots, n\}$. The trivial permutation skew representation $R[\boldsymbol{n}]$ is the $R_G[G]$–module corresponding to the trivial crossed homomorphism $\varepsilon\colon G \to \mathrm{GL}(n, R)$. Thus, when $H^1(G; \mathrm{GL}(n, R)) = [\varepsilon]$ for all n, every skew representation of G over R is isomorphic to a permutation skew representation and R is amenable. This holds, for example, when G is the Galois group of a Galois extension $R = K$ over a field k.

More generally, we have the following analogue of the situation in classical representation theory, which shows that amenability is not an unduly restrictive condition. It is proven in Passman [17, 4.1 in Chapter 1]. Even in this generality, he ascribes it to Maschke.

Lemma 6.13 *Let $N \subset M$ be $R_G[G]$–modules with no $|G|$–torsion. If $M = N \oplus V$ as an R–module, then there is an $R_G[G]$–submodule $P \subset M$ such that $|G|M \subset N \oplus P$.*

An irreducible skew representation is one that has no nontrivial proper skew subrepresentations.

Theorem 6.14 *Suppose that R is semisimple and $|G|^{-1} \in R$. Then every $R_G[G]$–module is completely reducible and R is amenable.*

Proof By the lemma, if $N \subset M$, then $M = N \oplus P$. That is, the complete reducibility of R–modules implies the complete reducibility of $R_G[G]$–modules. If N is an irreducible $R_G[G]$–module, then any choice of an element $n \neq 0$ determines a map of $R_G[G]$–modules $f\colon R_G[G] \to N$ such that $f(1) = n$. The image of f is a submodule of N, and it is all of N since N is irreducible. By complete reducibility, $\mathrm{Ker}(f)$ has a complement in $R_G[G]$, and that complement must be isomorphic to N. Thus N is a direct summand of the permutation skew representation $R_G[G]$. Therefore, by complete reducibility, all skew representations are direct summands of permutation skew representations. □

6.3 A model $\widetilde{\mathscr{GL}}_G(n, R)$ for $E(G, \mathrm{GL}(n, R)_G)$

Again let R be a G–ring, and assume that R is amenable. We have the entrywise left action of G on $\mathrm{GL}(n, R)$, and we have the right action of $\mathrm{GL}(n, R)$ on $\mathrm{GL}(n, R)$ given by matrix multiplication.

Lemma 6.15 *The left action of G and the right action of $\mathrm{GL}(n,R)$ on $\mathrm{GL}(n,R)$ specify an action of $\mathrm{GL}(n,R)\rtimes G$ on $\mathrm{GL}(n,R)$ via $(\tau,g)(x)=(gx)\tau^{-1}$ for $g\in G$, $x\in\mathrm{GL}(n,R)$, and $\tau\in\mathrm{GL}(n,R)$.*

Proof The required relation $g\cdot(x\tau)=(g\cdot x)(g\cdot\tau)$ is immediate from the fact that $g\colon R\to R$ is an automorphism of rings. □

Recall the G–set U from Definition 6.1. By Proposition 6.9, $R[U]$ is an $R_G[G]$–module with

(6-16) $$g\cdot(ru)=r^g gu \quad \text{for } g\in G,\ r\in R \text{ and } u\in U.$$

Similarly, we have the entrywise (equivalently, diagonal) left action of g on R^n, $g\cdot(re_i)=r^g e_i$, where we think of G as acting trivially on the set $\{e_i\}$. Regard elements $\tau\in\mathrm{GL}(n,R)$ as homomorphisms $\tau\colon R^n\to R^n$. That fixes the left action of $\mathrm{GL}(n,R)$ on R^n given by matrix multiplication, where elements of R^n are thought of as row matrices.

Definition 6.17 We define the chaotic general linear category $\widetilde{\mathscr{GL}}_G(n,R)$. The objects of $\widetilde{\mathscr{GL}}_G(n,R)$ are the monomorphisms of left R–modules $\alpha\colon R^n\to R[U]$. Let G act from the left on objects by $g\alpha=g\circ\alpha\circ g^{-1}$. By (6-16), we have

$$\begin{aligned}(g\circ\alpha\circ g^{-1})\Big(\sum r_ie_i\Big)&=\sum_i(g\circ\alpha)(r_i^{g^{-1}}e_i))=\sum_i g(r_i^{g^{-1}})\alpha(e_i)\\&=\sum_i r_i^{g^{-1}g}(g\cdot\alpha(e_i))=\sum_i r_i(g\cdot\alpha(e_i)).\end{aligned}$$

In particular, $(g\alpha)(e_i)=g\cdot\alpha(e_i)$. Let $\mathrm{GL}(n,R)$ act from the right on objects by $\alpha\tau=\alpha\circ\tau\colon R^n\to R[U]$; this uses the left, not the right, action of $\mathrm{GL}(n,R)$ on R^n. Since $\widetilde{\mathscr{GL}}_G(n,R)$ is chaotic, this fixes the actions on the morphism set, which the map (S,T) identifies with the product of two copies of the object set.

Proposition 6.18 *The actions of G and $\mathrm{GL}(n,R)$ on $\widetilde{\mathscr{GL}}_G(n,R)$ determine a left action of $\mathrm{GL}(n,R)\rtimes G$ via*

$$(\tau,g)\alpha=(g\alpha)\tau^{-1}.$$

The classifying space $|N\widetilde{\mathscr{GL}}_G(n,R)|$ is a universal principal $(G,\mathrm{GL}(n,R)_G)$–bundle.

Proof For the first claim, we must show that $g(\alpha\tau)=(g\alpha)(g\cdot\tau)\colon R^n\to R[U]$ for $\alpha\colon R^n\to R[U]$, $g\in G$, and $\tau=(t_{i,j})\in\mathrm{GL}(n,R)$. On elements e_i,

$$g(\alpha\tau)(e_i) = g\cdot(\alpha\tau)(e_i) = g\cdot\left(\alpha\left(\sum_j t_{i,j}e_j\right)\right)$$
$$= g\cdot\sum_j (t_{i,j}\alpha(e_j)) = \sum_j t^g_{i,j}(g\cdot\alpha(e_j))$$
$$= (g\alpha)\left(\sum_j t^g_{i,j}\, e_j\right) = (g\alpha)(g\cdot\tau)(e_i).$$

For each free R–module $M \subset R[U]$, choose an R–linear isomorphism $\eta_M\colon R^n \to M$. Sending $\alpha\colon R^n \to M$ to $\eta_M^{-1}\circ\alpha$ specifies an isomorphism of right $\mathrm{GL}(n,R)$–sets from the set of objects α with image M to $\mathrm{GL}(n,R)$; the inverse sends $\tau\in\mathrm{GL}(n,R)$ to $\eta_M\circ\tau$. Therefore $\mathrm{GL}(n,R)$ acts freely on $\widetilde{\mathscr{GL}}_G(n,R)$. Since $\widetilde{\mathscr{GL}}_G(n,R)$ is chaotic, it only remains to show that the set of objects of $\widetilde{\mathscr{GL}}_G(n,R)^\Lambda$ is nonempty if $\Lambda\cap\mathrm{GL}(n,R)=\{e\}$. By Lemma 4.5, $\Lambda=\{(\rho(h),h)\mid h\in H\}$, where H is a subgroup of G and $\rho\colon H\to\mathrm{GL}(n,R)$ is a crossed homomorphism.

By Theorem 6.8, we may use ρ to endow R^n with a structure of left $R_H[H]$–module. By the assumed amenability of R, there is a monomorphism of left $R_H[H]$–modules $R^n \to R[A]$ for some finite H–set A. We can embed A in the finite G–set $B = G\times_H A$ and then B is isomorphic to a sub-G–set of U. This fixes a monomorphism $\alpha\colon R^n\to R[U]$ of left $R_H[H]$–modules. Writing $\rho(h)=(s_{i,j})$ and $\rho(h)^{-1}=(t_{i,j})$, we have

$$h\alpha(e_j) = \alpha(\rho(h)(e_j)) = \alpha\left(\sum_k s_{j,k}e_k\right) = \sum_k s_{j,k}\alpha(e_k)$$

and therefore, using the display in Definition 6.17,

$$((h\alpha)\rho(h)^{-1})(e_i) = (h\alpha)\left(\sum_j t_{i,j}e_j\right) = \sum_j t_{i,j}h\cdot\alpha(e_j)$$
$$= \sum_j\sum_k t_{i,j}s_{j,k}\alpha(e_k) = \alpha(e_i). \qquad \square$$

Definition 6.19 Let $\mathscr{GL}_G(n,R)$ be the orbit G–category $\widetilde{\mathscr{GL}}_G(n,R)/\mathrm{GL}(n,R)$.

The classifying space $|N\mathscr{GL}_G(n,R)|$ is a model for $B(G,\mathrm{GL}(n,R)_G)$. Up to isomorphism, the G–category $\mathscr{GL}_G(n,R)$ admits the following explicit description.

Lemma 6.20 *The objects of $\mathscr{GL}_G(n,R)$ are the n–dimensional free R–submodules M of $R[U]$. The morphisms $\alpha\colon M\to N$ are the isomorphisms of R–modules. The group G acts by translation on objects, so that $gM=\{gm\mid m\in M\}$, and by conjugation on morphisms, so that $(g\alpha)(gm)=\alpha(m)$ for $m\in M$ and $g\in G$.*

Proof The objects α of $\widetilde{\mathscr{GL}}_G(n,R)$ with a fixed image M are all in the same orbit. Choose $\eta_M\colon R^n \to M$ to fix an orbit representative. In $\widetilde{\mathscr{GL}}_G(n,R)$, we have a unique morphism $\iota\colon \eta \to \beta$ for each object $\beta\colon R^n \to N$. We define $\alpha\colon M \to N$ to be the composite $\beta \circ \eta_M^{-1}$. The α are isomorphisms of R–modules that give orbit representatives specifying the morphisms of $\mathscr{GL}_G(n,R)$. As in the proof of Lemma 6.5, the description of the action of G follows. □

References

[1] **E Babson**, **D N Kozlov**, *Group actions on posets*, J. Algebra 285 (2005) 439–450 MR

[2] **E C Dade**, *Group-graded rings and modules*, Math. Z. 174 (1980) 241–262 MR

[3] **T tom Dieck**, *Faserbündel mit Gruppenoperation*, Arch. Math. (Basel) 20 (1969) 136–143 MR

[4] **B Guillou**, **J P May**, *Equivariant iterated loop space theory and permutative G–categories*, preprint (2012) To appear in Algebr. Geom. Topol.

[5] **K Kawakubo**, *Induction theorems for equivariant K–theory and J–theory*, J. Math. Soc. Japan 38 (1986) 173–198 MR

[6] **R K Lashof**, *Equivariant bundles*, Illinois J. Math. 26 (1982) 257–271 MR

[7] **R K Lashof**, **J P May**, *Generalized equivariant bundles*, Bull. Soc. Math. Belg. Sér. A 38 (1986) 265–271 MR

[8] **R K Lashof**, **J P May**, **G B Segal**, *Equivariant bundles with abelian structural group*, from "Proceedings of the Northwestern Homotopy Theory Conference" (H R Miller, S B Priddy, editors), Contemp. Math. 19, Amer. Math. Soc., Providence, RI (1983) 167–176 MR

[9] **J P May**, *Simplicial objects in algebraic topology*, Van Nostrand Mathematical Studies 11, D. Van Nostrand, Princeton, NJ (1967) MR

[10] **J P May**, *The geometry of iterated loop spaces*, Lectures Notes in Mathematics 271, Springer (1972) MR

[11] **J P May**, *Classifying spaces and fibrations*, Mem. Amer. Math. Soc. 155, Amer. Math. Soc., Providence, RI (1975) MR

[12] **J P May**, *Some remarks on equivariant bundles and classifying spaces*, from "International Conference on Homotopy Theory", Astérisque 191, Soc. Math. France, Paris (1990) 239–253 MR

[13] **J P May**, *Equivariant homotopy and cohomology theory*, CBMS Regional Conference Series in Mathematics 91, Amer. Math. Soc. (1996) MR

[14] **M Merling**, *Equivariant algebraic K–theory of G–rings*, Math. Z. 285 (2017) 1205–1248 MR

[15] **S Montgomery**, *Fixed rings of finite automorphism groups of associative rings*, Lecture Notes in Mathematics 818, Springer (1980) MR

[16] **M Murayama**, **K Shimakawa**, *Universal equivariant bundles*, Proc. Amer. Math. Soc. 123 (1995) 1289–1295 MR

[17] **D S Passman**, *Infinite crossed products*, Pure and Applied Mathematics 135, Academic Press, Boston (1989) MR

[18] **C Rezk**, *Classifying spaces for* 1*–truncated compact Lie groups*, preprint (2016) arXiv

[19] **G Segal**, *Classifying spaces and spectral sequences*, Inst. Hautes Études Sci. Publ. Math. 34 (1968) 105–112 MR

[20] **J-P Serre**, *Local fields*, Graduate Texts in Mathematics 67, Springer (1979) MR

[21] **K Shimakawa**, *Infinite loop* G*–spaces associated to monoidal* G*–graded categories*, Publ. Res. Inst. Math. Sci. 25 (1989) 239–262 MR

[22] **R W Thomason**, *The homotopy limit problem*, from "Proceedings of the Northwestern Homotopy Theory Conference" (H R Miller, S B Priddy, editors), Contemp. Math. 19, Amer. Math. Soc., Providence, R.I. (1983) 407–419 MR

[23] **V M Usenko**, *Subgroups of semidirect products*, Ukrain. Mat. Zh. 43 (1991) 1048–1055 MR In Russian; translated in Ukrainian Math. J. 43 (1991) 982–988

Department of Mathematics, The University of Kentucky
Lexington, KY, United States

Department of Mathematics, The University of Chicago
Chicago, IL, United States

Department of Mathematics, Johns Hopkins University
Baltimore, MD, United States

bertguillou@uky.edu, may@math.uchicago.edu, mmerling@math.jhu.edu

Received: 14 July 2012 Revised: 25 September 2016

msp

Algebraic & Geometric Topology 17 (2017) 2603–2634

Bounds on alternating surgery slopes

DUNCAN MCCOY

We show that if p/q–surgery on a nontrivial knot K yields the branched double cover of an alternating knot, then $|p/q| \leq 4g(K)+3$. This generalises a bound for lens space surgeries first established by Rasmussen. We also show that all surgery coefficients yielding the double branched covers of alternating knots must be contained in an interval of width two and this full range can be realised only if the knot is a cable knot. The work of Greene and Gibbons shows that if $S^3_{p/q}(K)$ bounds a sharp 4–manifold X, then the intersection form of X takes the form of a changemaker lattice. We extend this to show that the intersection form is determined uniquely by the knot K, the slope p/q and the Betti number $b_2(X)$.

57M12, 57M25; 57M27

1 Introduction

For a knot $K \subset S^3$ and $p/q \in \mathbb{Q}$ we say that $S^3_{p/q}(K)$ is an *alternating surgery* if it is the double branched cover of an alternating knot or link. In this paper, we will prove some bounds on the slopes of alternating surgeries. The first of these generalises a bound for lens space surgeries originally due to Rasmussen [27].

Theorem 1.1 *If K is a nontrivial knot with an alternating surgery $S^3_{p/q}(K)$, then the slope p/q satisfies the inequality $|p/q| \leq 4g(K)+3$.*

The bound in Theorem 1.1 is sharp with equality being attained by the $T_{2,n}$ torus knots. It turns out that whenever this bound is realised, the resulting alternating surgery yields a lens space. Hence, work of Baker [1, Theorem 1.2] shows that the $T_{2,n}$ torus knots are the only knots achieving equality in Theorem 1.1.

We can also obtain a bound on the range of slopes yielding alternating surgeries.

Theorem 1.2 *If K is a nontrivial knot admitting an alternating surgery, then there is an integer N such that for any alternating surgery $S^3_{p/q}(K)$, the coefficient p/q lies in the interval*

$$N-1 \leq \frac{p}{q} \leq N+1.$$

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2603

The definition of N is given Section 4.3. Theorem 1.2 shows that the range of slopes which yield alternating surgeries is contained in an interval with integer endpoints of width two. When every slope in this interval yields an alternating surgery, then we will show that the knot must be a cable knot. For the purposes of this paper, we consider torus knots to be cable knots.

Theorem 1.3 *Suppose that K is a nontrivial knot admitting alternating surgeries $S^3_r(K)$ for each of the slopes $r \in \{r_1, r_2, N\}$, where N is the integer appearing in Theorem 1.2. If r_1 and r_2 satisfy*

$$N-1 \leq r_1 < N < r_2 < N+1,$$

then $S^3_N(K)$ is a reducible surgery and K is a cable knot.

Remark 1.4 It can be shown that Theorem 1.3 still holds under the slightly weaker condition that $r_2 \leq N+1$. However, this relatively minor extension requires a substantial amount of work so we will not prove it here.

The starting point for the proof of these results is the work of Gibbons [6], which generalizes the work of Greene [9; 10; 11]. It provides strong restrictions on the intersection form of a negative-definite sharp 4–manifold X bounding $S^3_{p/q}(K)$ for $p/q > 0$, which must take the form of a changemaker lattice. In order to prove Theorems 1.2 and 1.3, we are required to determine the extent to which this intersection form depends on the knot K and the surgery slope p/q. This leads us to define the stable coefficients of a changemaker lattice. The definition of a changemaker lattice and its stable coefficients are given in Section 2.1. Let p/q have continued fraction expansion $p/q = [a_0, \dots, a_l]^-$, where $a_i \geq 2$ for $1 \leq i \leq l$ and $a_0 \geq 1$. Here $[a_0, \dots, a_l]^-$ denotes the Hirzebruch–Jung continued fraction

$$[a_0, \dots, a_l]^- = a_0 - \cfrac{1}{a_1 - \cfrac{1}{\ddots - \cfrac{1}{a_l}}}.$$

A p/q–changemaker lattice takes the form of an orthogonal complement

$$L = \langle w_0, \dots, w_l \rangle^{\perp} \subseteq \mathbb{Z}^{t+s+1} = \langle f_1, \dots, f_t, e_0, \dots, e_s \rangle,$$

where the f_i and e_j form an orthonormal basis for $\mathbb{Z}^{r+s+1}$, and the w_i have the properties that

$$w_i \cdot w_j = \begin{cases} a_i & \text{if } i = j, \\ -1 & \text{if } |i-j| = 1, \\ 0 & \text{if } |i-j| \geq 2, \end{cases}$$

and

$$\begin{aligned} w_0 \cdot e_0 &= 1, \\ w_0 \cdot e_i &= 0 \quad \text{for } 1 \le i \le s, \\ w_0 \cdot f_i &\ge 0 \quad \text{for } 1 \le i \le t, \\ w_j \cdot f_i &= 0 \quad \text{for } 1 \le i \le t \text{ and } 1 \le j \le l. \end{aligned}$$

The stable coefficients of L are defined to be the values of $w_0 \cdot f_i$ satisfying $w_0 \cdot f_i > 1$.

Theorem 1.5 *Let $K \subset S^3$ be a knot and suppose that $S^3_{p/q}(K)$ bounds a negative-definite sharp 4–manifold X with intersection form Q_X for some $p/q > 0$. Then the positive-definite lattice $-Q_X$ embeds into $\mathbb{Z}^{b_2(X)+l+1}$ as a p/q–changemaker lattice, where the stable coefficients are determined by K.*

The stable coefficients in Theorem 1.5 form an invariant of the knot K that can be calculated from the knot Floer homology of K. Section 2.3 provides an algorithm for this calculation. When K is an L–space knot, the stable coefficients can be computed directly from its Alexander polynomial. The integer N appearing in Theorems 1.2 and 1.3 is defined in terms of stable coefficients and hence is an invariant of K and can be calculated from the Alexander polynomial.

Remark 1.6 In addition to being a lower bound for alternating surgeries, the integer $N-1$ appearing in Theorem 1.2 also has the property that if $S^3_{p/q}(K)$ bounds a negative-definite sharp 4–manifold then $p/q \ge N-1$. We explain this observation after the proof of Theorem 1.2.

Given one negative-definite sharp 4–manifold, bounding a 3–manifold Y we can obtain another by taking a connected sum with $\overline{\mathbb{CP}}^2$. It follows from Theorem 1.5 that if $Y = S^3_{p/q}(K)$, then at the level of intersection forms this is the only possibility.

Corollary 1.7 *Let $K \subset S^3$ be a knot such that for some $p/q > 0$, the 3–manifold $S^3_{p/q}(K)$ bounds negative-definite sharp 4–manifolds X and X', with $b_2(X') = b_2(X) + k$ for $k \ge 0$. Then*

$$Q_{X'} \cong Q_X \oplus (-\mathbb{Z}^k) \cong Q_{X \#_k \overline{\mathbb{CP}}^2}.$$

Acknowledgements The author would like to thank his supervisor, Brendan Owens, for his guidance and careful reading of this paper. He is grateful to Liam Watson for helpful conversations about quasi-alternating links and many other things. He also wishes to thank the anonymous referee for their feedback.

2 Changemaker lattices and sharp 4–manifolds

The aim of this section is to prove Theorem 1.5. We begin by defining changemaker lattices and recalling the necessary definitions and properties from Heegaard Floer homology. We finish the section by stating the properties of L–space surgeries that we will require to prove the results on alternating surgeries.

2.1 Changemaker lattices

We will define p/q–changemaker lattices for any $p/q > 0$. Changemaker lattices corresponding to the case $q = 1$ were defined by Greene in his solution to the lens space realisation problem [9] and work on the cabling conjecture [11]. The case $q = 2$ arose in his work on unknotting numbers [10]. The more general definition we state here is the one which arises in Gibbons' work [6].

Definition 2.1 We say $(\sigma_1, \dots, \sigma_t)$ satisfies the *changemaker condition* if the following conditions hold:

$$0 \le \sigma_1 \le 1 \qquad \text{and} \qquad \sigma_{i-1} \le \sigma_i \le \sigma_1 + \cdots + \sigma_{i-1} + 1 \quad \text{for } 1 < i \le t.$$

The changemaker condition is equivalent to the following combinatorial result.

Proposition 2.2 (Brown [2]) *Let $\sigma = (\sigma_1, \dots, \sigma_t)$, with $\sigma_1 \le \cdots \le \sigma_t$. There is $A \subseteq \{1, \dots, t\}$ such that $k = \sum_{i \in A} \sigma_i$ for every integer k with $0 \le k \le \sigma_1 + \cdots + \sigma_t$ if and only if σ satisfies the changemaker condition.*

Now we are ready to define changemaker lattices. It is convenient to define integer and noninteger changemaker lattices separately, although the two are clearly similar.

Definition 2.3 (integral changemaker lattice) First suppose that $q = 1$, so that $p/q > 0$ is an integer. Let $f_0, \dots, f_t$ be an orthonormal basis for $\mathbb{Z}^t$. Let $w_0 = \sigma_1 f_1 + \cdots + \sigma_t f_t$ be a vector such that $\|w_0\|^2 = p$ and $(\sigma_1, \cdots, \sigma_t)$ satisfies the changemaker condition. Then

$$L = \langle w_0 \rangle^{\perp} \subseteq \mathbb{Z}^{t+1}$$

is a *p/q–changemaker lattice*. Let m be minimal such that $\sigma_m > 1$. We define the *stable coefficients* of L to be the tuple $(\sigma_m, \dots, \sigma_t)$. If no such m exists, then we take the stable coefficients to be the empty tuple.

Definition 2.4 (nonintegral changemaker lattice) Now suppose that $q \geq 2$, so that $p/q > 0$ is not an integer. This has continued fraction expansion of the form $p/q = [a_0, a_1, \dots, a_l]^-$, where $a_k \geq 2$ for $1 \leq k \leq l$ and $a_0 = \lceil p/q \rceil \geq 1$. Now define

$$m_0 = 0 \qquad \text{and} \qquad m_k = \sum_{i=1}^{k} a_i - k \quad \text{for } 1 \leq k \leq l.$$

Set $s = m_l$ and let $f_1, \dots, f_t, e_0, \dots, e_s$ be an orthonormal basis for the lattice $\mathbb{Z}^{t+s+1}$. Let $w_0 = e_0 + \sigma_1 f_1 + \cdots + \sigma_t f_t$ be a vector such that $(\sigma_1, \dots, \sigma_t)$ satisfies the changemaker condition and $\|w_0\|^2 = a_0$. For $1 \leq k \leq l$, define

$$w_k = -e_{m_{k-1}} + e_{m_{k-1}+1} + \cdots + e_{m_k}.$$

We say that

$$L = \langle w_0, \dots, w_l \rangle^\perp \subseteq \mathbb{Z}^{t+s+1}$$

is a *p/q–changemaker lattice*. Let m be minimal such that $\sigma_m > 1$. We define the *stable coefficients* of L to be the tuple $(\sigma_m, \dots, \sigma_t)$. If no such m exists, then we take the stable coefficients to be the empty tuple.

Remark 2.5 Since $m_k - m_{k-1} = a_k - 1$, the vectors $w_0, \dots, w_l$ constructed in Definition 2.4 satisfy

$$w_i . w_j = \begin{cases} a_j & \text{if } i = j, \\ -1 & \text{if } |i-j| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 2.6 Let L be a p/q–changemaker lattice

$$L = \langle w_0 = e_0 + \sigma_1 f_1 + \cdots + \sigma_t f_t, w_1, \dots, w_l \rangle^\perp \subseteq \mathbb{Z}^{t+s+1}.$$

By definition, the stable coefficients determine the values of the σ_i satisfying $\sigma_i > 1$. Since $\|w_0\|^2 = \lceil p/q \rceil$, the stable coefficients fix the number of σ_i equal to 1 and this accounts for all nonzero σ_i. It follows that the number of σ_i equal to zero can be deduced from the rank of L. Thus we see that the value p/q, the stable coefficients and the rank determine L uniquely. Since we have $f_i \in L$ if and only if $\sigma_i = 0$, any two p/q–changemaker lattices L and L' with the same stable coefficients and $\operatorname{rk}(L') = \operatorname{rk}(L) + k$ satisfy $L' \cong L \oplus \mathbb{Z}^k$

2.2 Sharp 4–manifolds

Now we will give a summary of the necessary background on Heegaard Floer homology and its d–invariants. Let Y be a rational homology 3–sphere. Its Heegaard Floer homology, $\widehat{\mathrm{HF}}(Y)$, when defined with coefficients in $\mathbb{Z}/2\mathbb{Z}$, takes the form of a finite-dimensional vector space over $\mathbb{Z}/2\mathbb{Z}$. The group $\widehat{\mathrm{HF}}(Y)$ splits as a direct sum over spin^c–structures:

$$\widehat{\mathrm{HF}}(Y) \cong \bigoplus_{\mathfrak{s}\in \mathrm{Spin}^c(Y)} \widehat{\mathrm{HF}}(Y,\mathfrak{s}),$$

where $\widehat{\mathrm{HF}}(Y,\mathfrak{s}) \neq 0$ for all $\mathfrak{s} \in \mathrm{Spin}^c(Y)$. We say that Y is an *L–space* if $\widehat{\mathrm{HF}}(Y)$ is as small as possible:

$$\dim_{\mathbb{F}_2} \widehat{\mathrm{HF}}(Y) = |H^2(Y;\mathbb{Z})| = |\mathrm{Spin}^c(Y)|.$$

Associated to each summand there is a numerical invariant $d(Y,\mathfrak{s}) \in \mathbb{Q}$, called the *$d$–invariant* [22]. If Y is the boundary of a smooth negative-definite 4–manifold X, then for any $\mathfrak{t} \in \mathrm{Spin}^c(X)$ which restricts to $\mathfrak{s} \in \mathrm{Spin}^c(Y)$ there is a bound on the corresponding d–invariant,

$$c_1(\mathfrak{t})^2 + b_2(X) \leq 4d(Y,\mathfrak{s}). \tag{2-1}$$

We say that X is *sharp* if for every $\mathfrak{s} \in \mathrm{Spin}^c(Y)$ there is some $\mathfrak{t} \in \mathrm{Spin}^c(X)$ which restricts to $\mathfrak{s}$ and attains equality in (2-1).

We will be interested in the case where Y arises as surgery on a knot in S^3. Let $K \subset S^3$ be a knot. For fixed $p/q \in \mathbb{Q}\setminus\{0\}$, there are canonical identifications [26]

$$\mathrm{Spin}^c(S^3_{p/q}(K)) \leftrightarrow \mathbb{Z}/p\mathbb{Z} \leftrightarrow \mathrm{Spin}^c(S^3_{p/q}(U)).$$

Using these identifications we are able to define

$$D_{p/q}(i) := d(S^3_{p/q}(K),i) - d(S^3_{p/q}(U),i)$$

for each $i \in \mathbb{Z}/p\mathbb{Z}$.

The work of Ni and Wu shows that for $0 \leq i \leq p-1$ these values may be calculated by the formula [20, Proposition 1.6]

$$D_{p/q}(i) = -2\max\{V_{\lfloor i/q\rfloor}, H_{\lfloor (i-p)/q\rfloor}\}, \tag{2-2}$$

where V_j and H_j are sequences of positive integers, depending only on K, which are nonincreasing and nondecreasing, respectively. These further satisfy $H_{-j} = V_j = 0$ for $j \geq g(K)$, where $g(K)$ is the genus of K. In fact, it can be shown that $V_j = H_{-j}$ for all j [21, Proof of Theorem 3]. Using these properties of the V_j and H_j, (2-2) can be rewritten as

$$D_{p/q}(i) = -2V_{\min\{\lfloor i/q\rfloor, \lceil (p-i)/q\rceil\}}. \tag{2-3}$$

Let $p/q = [a_0,\dots,a_l]^-$ be the continued fraction of p/q with $a_0 \geq 1$ and $a_i \geq 2$ for $i \geq 1$. The changemaker theorem we will use is the following.

Theorem 2.7 (Gibbons [6]) *Let $K \subset S^3$ be a knot and suppose that $S^3_{p/q}(K)$ bounds a smooth, negative-definite 4–manifold X with intersection form Q_X for some $p/q > 0$. If the manifold X is sharp, then $-Q_X$ embeds into $\mathbb{Z}^{b_2(X)+l+1}$ as a p/q–changemaker lattice,*

$$-Q_X \cong L = \langle w_0, \dots, w_l \rangle^{\perp} \subseteq \mathbb{Z}^{t+s+1},$$

where w_0 satisfies the formula

$$8V_{|i|} = \min_{\substack{c\cdot w_0 \equiv a_0+2i \bmod 2a_0 \\ c\in \mathrm{Char}(\mathbb{Z}^{t+1})}} \|c\|^2 - t - 1 \tag{2-4}$$

for $|i| \le \frac{1}{2}a_0$.

Here $\mathrm{Char}(\mathbb{Z}^{t+1})$ denotes the set of all characteristic vectors in $\mathbb{Z}^{t+1}$, where a *characteristic vector $x \in \mathbb{Z}^{t+1}$* is one with odd coefficients with respect to any orthonormal basis for $\mathbb{Z}^{t+1}$.

The equation (2-4) is not explicitly stated by Gibbons. However, Greene shows that it holds in the case of integer surgeries [11] and it follows from Gibbons' proof that it must also hold for noninteger surgeries. Further discussion of this can be found in [15].

2.3 Calculating stable coefficients

We will deduce Theorem 1.5 from Theorem 2.7 by showing that (2-4) determines the stable coefficients uniquely. The argument is entirely combinatorial and uses only the properties of the V_i stated in Section 2.2.

Let $(V_i)_{i\ge 0}$ be the nonincreasing, nonnegative sequence

$$V_0 \ge V_1 \ge \cdots \ge V_{\widetilde{g}-1} > V_{\widetilde{g}} = V_{\widetilde{g}+1} = \cdots = 0,$$

for which $V_i = 0$ if and only if $i \ge \widetilde{g}$ and $V_i \le V_{i+1} + 1$ for all i. Suppose that there is $\rho = (\rho_0, \dots, \rho_t) \in \mathbb{Z}^{t+1}$, with $\|\rho\|^2 = n \ge 2\widetilde{g}$, such that

$$8V_{|k|} = \min_{\substack{c\cdot \rho \equiv n+2k \bmod 2n \\ c\in \mathrm{Char}(\mathbb{Z}^{t+1})}} \|c\|^2 - t - 1 \tag{2-5}$$

for $|k| \le \frac{1}{2}n$. Possibly after an automorphism of $\mathbb{Z}^{t+1}$, we may assume that $\rho_i \ge 0$ for all i and that the ρ_i form a decreasing sequence

$$\rho_0 \ge \rho_1 \ge \cdots \ge \rho_t \ge 0.$$

Observe that (2-5) has three pieces of input data: the sequence $(V_i)_{i\ge 0}$ and the integers n and t. Given some choice of $(V_i)_{i\ge 0}$, n and t, there is no guarantee that there is

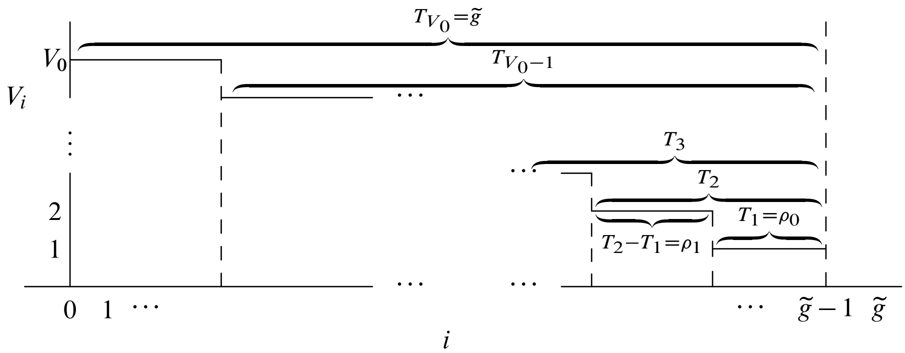


Figure 1: A graph to show the relationship between the V_i and the T_i. We have also shown how ρ_0 and ρ_1 occur as the number of V_i equal to one and two, respectively.

ρ satisfying (2-5). However, we will show that when there is such a ρ, it is unique. Moreover we will see that the coefficients of ρ satisfying $\rho_i > 1$ are determined by the sequence $(V_i)_{i\geq 0}$.

Remark 2.8 If $\rho_t = 0$, then any minimiser in the right-hand side of (2-5) must have $c_t = \pm 1$. So we see that $\rho' = (\rho_0, \dots, \rho_{t-1})$ satisfies

$$8V_{|k|} = \min_{\substack{c\cdot\rho' \equiv n+2k \bmod 2n \\ c\in \mathrm{Char}(\mathbb{Z}^t)}} \|c\|^2 - t$$

for all $0 \leq |k| \leq \frac{1}{2}n$. This allows us to assume that $\rho_i \geq 1$ for all i.

If we restrict our attention to $0 \leq k \leq \frac{1}{2}n$, we find that (2-5) simplifies as follows.

Lemma 2.9 *For* $0 \leq k \leq \frac{1}{2}n$,

$$8V_k = \min_{\substack{c\cdot\rho = 2k-n \\ c\in \mathrm{Char}(\mathbb{Z}^{t+1})}} \|c\|^2 - t - 1.$$

Proof Suppose $c \in \mathrm{Char}(\mathbb{Z}^{t+1})$ satisfies $c\cdot\rho = 2mn - n + 2k$ for some $m \in \mathbb{Z}$. Consider the vector $c' = c - 2m\rho$. This satisfies

$$c'\cdot\rho = 2k - n \equiv c\cdot\rho \mod 2n$$

and

$$\|c'\|^2 = \|c\|^2 - 4mc\cdot\rho + 4m^2 n = \|c\|^2 - 4m(nm - n + 2k).$$

Since we are assuming $-n \leq 2k - n \leq 0$, we have $m(nm + 2k - n) \geq 0$ for all $m \in \mathbb{Z}$. Therefore, we have $\|c'\|^2 \leq \|c\|^2$. This shows that if c is a minimiser in (2-5) we can assume it satisfies $c\cdot\rho = 2k - n$. □

For $m \geq 1$, it will be convenient to consider the quantities

$$T_m = \left|\{0 \leq i < \widetilde{g} \mid 0 < V_i \leq m\}\right|.$$

These are illustrated in Figure 1. We will show how to calculate these in terms of ρ. First we need to define the following collection of tuples for each $m \geq 0$:

$$S_m = \left\{\alpha \in \mathbb{Z}^{r+1} : \alpha_i \geq 0,\ 2m = \sum \alpha_i(\alpha_i + 1)\right\}.$$

Lemma 2.10 *For $0 \leq m < V_0$, we can calculate T_m by*

$$T_m = \max_{\alpha \in S_m} \rho \cdot \alpha$$

and T_{V_0} satisfies

$$T_{V_0} = \widetilde{g} = \frac{1}{2}\sum_{i=0}^{t} \rho_i^2 - \rho_i \quad \textit{and} \quad T_{V_0} \leq \max_{\alpha \in S_{V_0}} \rho \cdot \alpha.$$

Proof Since the V_k form a decreasing sequence with $V_k = 0$ if and only if $k \geq \widetilde{g}$, we necessarily have $T_{V_0} = \widetilde{g}$. Using Lemma 2.9, we that $V_k = 0$ if and only if there is $c \in \{\pm 1\}^{t+1}$ with $c \cdot \rho = 2k - n$. The smallest of value k for which this is true is $k = \frac{1}{2}\left(n - \sum_{i=0}^{t} \rho_i\right)$, which is obtained by taking $c = \{-1\}^{t+1}$. Thus $2\widetilde{g} = \sum_{i=0}^{t} \rho_i^2 - \rho_i$, as required (see [11, Proposition 3.1]).

Now observe that for $0 \leq m < V_0$, we have

$$T_m = \widetilde{g} - \min\{k : V_k = m\}.$$

By Lemma 2.9, $V_k = m$ and $0 \leq k < \frac{1}{2}n$ implies there is $c \in \mathrm{Char}(\mathbb{Z}^{t+1})$ such that $\|c\|^2 - t - 1 = 8m$ and $c \cdot \rho = 2k - n$. If we write the coefficients of c in the form $c_i = -(2\alpha_i + 1)$, then $\sum_{i=0}^{t} \alpha_i(\alpha_i + 1) = 2m$ and

$$2k = n - \sum_{i=0}^{t} \rho_j - 2\alpha \cdot \rho = 2\widetilde{g} - 2\alpha \cdot \rho. \tag{2-6}$$

We see that for any α minimising (2-6), we must have $\alpha \in S_m$, since it must satisfy $\alpha_i \geq 0$ for all i. Thus we see that

$$T_m = \max_{\alpha \in S_m} \rho \cdot \alpha$$

for $0 \leq m < V_0$. The equation (2-6) also shows that there must exist α satisfying $\sum_{i=0}^{t} \alpha_i(\alpha_i + 1) = 2V_0$ and $\alpha \cdot \rho = \widetilde{g}$. This implies the inequality

$$T_{V_0} \leq \max_{\alpha \in S_{V_0}} \rho \cdot \alpha,$$

which completes the proof. □

Remark 2.11 It follows from this lemma that $T_1=\rho_0$ and $T_2=\rho_0+\rho_1$. In particular this implies that $\rho_1 = T_2 - T_1$. This is illustrated in Figure 1.

We now begin the process of showing how the remaining ρ_i can be recovered from the sequence $(V_i)_{i\geq 0}$. We begin with the simplest case, which is when $V_0\leq 1$.

Lemma 2.12 *If $V_0\leq 1$, then $\widetilde{g}\leq 3$ and ρ takes the form*

$$\rho=\begin{cases}(1,1,\dots,1) & \text{if } \widetilde{g}=0,\\ (2,1,\dots,1) & \text{if } \widetilde{g}=1,\\ (2,2,1,\dots,1) & \text{if } \widetilde{g}=2,\\ (3,1,\dots,1) & \text{if } \widetilde{g}=3.\end{cases}$$

Proof If $V_0=0$, then $\widetilde{g}=0$ and Lemma 2.10 implies that $\sum_{i=0}^t \rho_i^2-\rho_i=0$. This shows that we have $\rho_i=1$ for all $0\leq i\leq t$.

Suppose now that $V_0=1$. By Lemma 2.10, we have

$$0<T_1=\widetilde{g}=\frac{1}{2}\sum_{i=0}^t \rho_i^2-\rho_i\leq \max_{\alpha\in S_1}\rho\cdot\alpha.$$

Since S_1 consists of vectors with a single nonzero coordinate, which equals one, we have $\max_{\alpha\in S_1}\rho\cdot\alpha=\rho_0$. Thus we must have $\rho_0^2-\rho_0\leq 2\rho_0$, and hence $\rho_0\leq 3$. If $\rho_0=3$, then we have

$$\widetilde{g}=3+\frac{1}{2}\sum_{i=1}^t \rho_i(\rho_i-1)\leq\rho_0=3,$$

which implies that $\rho_i=1$ for $1\leq i\leq t$ and $\widetilde{g}=3$. If $\rho_0=2$, then $\widetilde{g}\leq 2$ implies that $\rho_1\in\{1,2\}$, giving the other two possibilities in the statement of the lemma. □

From now on we will suppose that $V_0>1$. This allows us to define the quantity

$$\mu=\min_{1\leq i<V_0}\{T_i-T_{i-1}\}.$$

Since $T_1=\rho_0$ and $T_0=0$, we must have $\mu\leq\rho_0$.

Lemma 2.13 *If $\rho_0\geq 5$ or $\sum_{\rho_i\,\text{even}}\rho_i\geq 6$, then $\mu\leq 2$.*

Proof For $m<V_0$, Lemma 2.10 shows that there is $\alpha\in S_m$ such that $\rho\cdot\alpha=T_m$. If $\alpha_l>0$, then consider α' defined by

$$\alpha_i'=\begin{cases}\alpha_i & \text{if } i\neq l,\\ \alpha_i-1 & \text{if } i=l.\end{cases}$$

By construction, we have $\alpha' \in S_{m-\alpha_l}$ and $\alpha' \cdot \rho = \alpha \cdot \rho - \rho_l = T_m - \rho_l$. As $\alpha' \cdot \rho \le T_{m-\alpha_l}$, we get

$$\rho_l \ge T_m - T_{m-\alpha_l} \ge \alpha_l \mu. \tag{2-7}$$

If we have a maximiser $\alpha \in S_m$ such that $\rho \cdot \alpha = T_m$ and α does not satisfy

$$\alpha_i \le \begin{cases} \frac{1}{2}(\rho_i - 2) & \text{if } \rho_i \text{ is even,} \\ \frac{1}{2}(\rho_i - 3) & \text{if } \rho_i > 3 \text{ is odd,} \\ \frac{1}{2}(\rho_i - 1) & \text{if } \rho_i \in \{1, 3\}, \end{cases} \tag{2-8}$$

for all i, then there is l such that $\rho_l/\alpha_l < 3$. So, by (2-7), we see that $\mu \le 2$. We will show that if ρ satisfies the hypotheses of the lemma, then such a maximiser must exist.

Let $c \in \mathrm{Char}(\mathbb{Z}^{t+1})$ be such that $c \cdot \rho = n$. By (2-5), we have

$$8V_0 \ge \|c\|^2 - t - 1.$$

On the other hand, the Cauchy–Schwarz inequality implies that

$$|c \cdot \rho|^2 = n^2 \le \|\rho\|^2 \|c\|^2 = n\|c\|^2,$$

showing that $\|c\|^2 \ge n$ with equality if and only if $c = \rho$. Altogether, this yields

$$V_0 \ge \frac{1}{8}(\|\rho\|^2 - t - 1) = \frac{1}{8}\sum_{i=0}^{t}(\rho_i^2 - 1),$$

with equality if and only if $\rho \in \mathrm{Char}(\mathbb{Z}^{t+1})$. We will let N denote the quantity

$$N = \left\lfloor \frac{1}{8}\sum_{i=0}^{t}(\rho_i^2 - 1) \right\rfloor \le V_0.$$

Now take $\alpha \in S_m$, which satisfies the conditions given by (2-8). It follows that

$$\begin{aligned} m &= \frac{1}{2}\sum_{i=0}^{t}\alpha_i(\alpha_i + 1) \\ &\le \sum_{\rho_i > 3 \text{ odd}} \frac{(\rho_i - 3)(\rho_i - 1)}{8} + \sum_{\rho_i \text{ even}} \frac{\rho_i(\rho_i - 2)}{8} + \sum_{\rho_i \in \{1,3\}} \frac{\rho_i^2 - 1}{8} \\ &= \sum_{i=0}^{t} \frac{\rho_i^2 - 1}{8} + \sum_{\rho_i > 3 \text{ odd}} \frac{1 - \rho_i}{2} + \sum_{\rho_i \text{ even}} \frac{1 - 2\rho_i}{8}. \end{aligned} \tag{2-9}$$

If ρ_0 is odd and $\rho_0 \geq 5$, then (2-9) shows that

$$m \leq \sum_{i=0}^{t} \frac{\rho_i^2 - 1}{8} - 2 < N - 1$$

In particular, there is no $\beta \in S_{N-1}$ satisfying (2-8). Since $N - 1 < V_0$, there is $\beta \in S_{N-1}$ with $\beta \cdot \rho = T_{N-1}$ and so (2-7) implies that $\mu \leq 2$. If $\sum_{\rho_i \text{ even}} \rho_i \geq 6$, then we must have $\sum_{\rho_i \text{ even}} (2\rho_i - 1) \geq \frac{3}{2} \sum_{\rho_i \text{ even}} \rho_i \geq 9$. Therefore, (2-9) shows that

$$m < \sum_{i=0}^{t} \frac{\rho_i^2 - 1}{8} - 1 < N.$$

In particular, there is no $\beta \in S_N$ satisfying (2-7). Since we are assuming there is an even ρ_i, we have $N < V_0$ and so there exists $\beta \in S_N$ such that $\beta \cdot \rho = T_N$ and so (2-7) implies that $\mu \leq 2$. □

If $\mu > 2$, then ρ must fall into one of a small number of cases.

Lemma 2.14 *If $\mu > 2$ then either $T_1 = 3$ or $T_1 = 4$. If $T_1 = 3$, then ρ takes the form*

$$\rho = \begin{cases} (\underbrace{3, \dots, 3}_{d}, 1, \dots, 1) & \text{if } \widetilde{g} = 3d, \\ (\underbrace{3, \dots, 3}_{d}, 2, 1, \dots, 1) & \text{if } \widetilde{g} = 3d + 1, \\ (\underbrace{3, \dots, 3}_{d}, 2, 2, 1, \dots, 1) & \text{if } \widetilde{g} = 3d + 2. \end{cases}$$

If $T_1 = 4$, then ρ must take the form

$$\rho = (4, \underbrace{3, \dots, 3}_{d}, 1, \dots, 1), \quad \text{where } \widetilde{g} = 3d + 6.$$

Proof If $\mu > 2$, then Lemma 2.13 and the observation that $\mu \leq T_1 = \rho_0$, we must have $\rho_0 \in \{3, 4\}$. If $\rho_0 = 3$, then Lemma 2.13 implies that we have $\rho_i = 2$ for at most two values i. If $\rho_0 = 4$, then Lemma 2.13 implies that ρ_i is odd for all $i \geq 1$. It is then easy to deduce that ρ must take the required form by using the formula $\widetilde{g} = \frac{1}{2} \sum_{i=0}^{t} \rho_i^2 - \rho_i$. □

Remark 2.15 Although it suffices for our purposes, Lemma 2.14 does not quite tell the full story. If $\rho = (4, 3, \dots, 3, 1, \dots, 1)$, then one can show that we have $\mu = 1$. This shows that the only cases with $\mu > 2$ are those given in Lemma 2.14 with $\rho_0 = 3$. For these examples we do have $\mu = 3$.

Now we show that the sequence $(V_i)_{i\geq 0}$ determines ρ when $\mu \leq 2$.

Lemma 2.16 *If $\mu \leq 2$, then the vector ρ satisfying* (2-5) *is unique.*

Proof We will show that can calculate the coefficients of ρ iteratively from the values $T_0 < T_1 < \dots < T_{V_0} = \widetilde{g}$. Using the T_i, we will construct a sequence $s^{(0)}, s^{(1)}, \dots, s^{(N)}$, which we will show to satisfy

$$s^{(k)} = (\rho_0, \dots, \rho_k, 0, \dots, 0)$$

for each $k \leq N$. The integer N will be large enough that $S^{(N)}$ satisfies

$$\max_{\alpha \in S_t} s^{(N)} \cdot \alpha = T_t$$

for all $t < V_0$. We will show we can deduce ρ_i for any $i > N$ by considering $T_{V_0} = \widetilde{g}$.

Start by setting

$$s^{(0)} = (T_1, 0, \dots, 0) = (\rho_0, 0, \dots, 0).$$

Now suppose that for $l \geq 0$ we have $s_i^{(l)} = \rho_i$ for all $i \leq l$. Suppose there is $t < V_0 - 1$ minimal such that $M = \max_{\alpha \in S_t} s^{(l)} \cdot \alpha < T_t$.

Claim 1 *We have $\rho_{l+1} = T_t - T_{t-1}$.*

Proof of Claim 1 Let $\alpha \in S_{t-1}$ be such that $s^{(l)} \cdot \alpha = T_{t-1}$. Such an α must also satisfy $\rho \cdot \alpha = T_{t-1}$. In particular, $\alpha_i = 0$ for $i > l$.

Now we consider $\alpha' \in S_t$ defined by

$$\alpha'_i = \begin{cases} \alpha_i & \text{if } i \neq l+1, \\ 1 & \text{if } i = l+1. \end{cases}$$

We have $\alpha' \cdot \rho = T_{t-1} + \rho_{l+1} \leq T_t$. This implies that

$$\rho_{l+1} \leq T_t - T_{t-1}. \tag{2-10}$$

Let $\beta \in S_t$ be such that $\rho \cdot \beta = T_t$. Since $M < T_t$, we may assume $\beta_{l+1} > 0$. Thus we can define β' by

$$\beta'_i = \begin{cases} \beta_i & \text{if } i \neq l, \\ \beta_i - 1 & \text{if } i = l+1. \end{cases}$$

We have $\beta' \in S_{t-\beta_l}$. Therefore we obtain

$$T_{t-1} \geq T_{t-\beta_l} \geq \rho \cdot \beta' = T_t - \rho_{l+1}. \tag{2-11}$$

Combining (2-10) and (2-11) gives $\rho_{l+1} = T_t - T_{t-1}$, as claimed. □

Thus if we define $s^{(l+1)}$ by

$$s_i^{(l+1)} = \begin{cases} s_i^{(l)} & \text{if } i \neq l+1, \\ T_t - T_{t-1} & \text{if } i = l+1, \end{cases}$$

we see that $s^{(l+1)}$ satisfies

$$s^{(l+1)} = (\rho_0, \dots, \rho_{l+1}, 0, \dots, 0)$$

and

$$s^{(l+1)} \cdot \alpha' = \max_{\alpha \in S_t} \alpha \cdot s^{(l+1)} = T_t,$$

where $\alpha' \in S_t$ is as defined in the proof of Claim 1

Proceeding in this way, we eventually obtain $s^{(N)}$ such that $T_t = \max_{\alpha \in S_t} \alpha \cdot s^{(N)}$ for all $0 \le t < V_0$ and

$$s^{(N)} = (\rho_0, \dots, \rho_N, 0, \dots, 0).$$

Claim 2 *We have $\rho_l \le \mu \le 2$ for all $l > N$.*

Proof of Claim 2 Let $\tau < V_0 - 1$ be such that $T_{\tau+1} - T_\tau = \mu$. There is $\alpha \in S_\tau$ such that $\alpha \cdot \rho = \alpha \cdot s^{(N)} = T_\tau$. Such an α must satisfy $\alpha_l = 0$ for $l > N$. Let $\alpha' \in S_{\tau+1}$ be defined by

$$\alpha_i' = \begin{cases} \alpha_i & \text{if } i \neq l, \\ 1 & \text{if } i = l. \end{cases}$$

We have

$$\rho_l = \alpha' \cdot \rho - T_\tau \le T_{\tau+1} - T_\tau = \mu \le 2,$$

as required. □

It remains to determine how many values of $i > N$ satisfy $\rho_i = 2$. Since we have the formula $T_{V_0} = \frac{1}{2} \sum_{i=0}^{t} \rho_i (\rho_i - 1)$, we see that there are

$$T_{V_0} - \frac{1}{2} \sum_{i=0}^{t} s_i^{(N)} (s_i^{(N)} - 1)$$

values of $i > N$ with $\rho_i = 2$. Since $\rho_i = 1$ for all remaining values of i, this shows that ρ is determined by the T_i. □

The proof of Lemma 2.16 combined with Lemmas 2.12 and 2.14 provides an algorithm for calculating ρ. This shows that ρ is the unique vector with $\rho_0 \ge \rho_1 \ge \dots \ge \rho_t > 0$ and $\|\rho\|^2 = n$ satisfying (2-5). Moreover, if we take m to be maximal such that $\rho_m > 1$, then this algorithm calculates the tuple $(\rho_0, \dots, \rho_m)$ using only the sequence $(V_i)_{i \ge 0}$.

This allows us to deduce Theorem 1.5 and Corollary 1.7 from Theorem 2.7.

Proof of Theorem 1.5 Theorem 2.7 shows that the intersection form Q_X takes the form of a p/q–changemaker lattice,

$$-Q_X \cong L = \langle w_0, \dots, w_l \rangle^\perp \subseteq \mathbb{Z}^{t+s+1},$$

where the sequence $(V_i)_{i\geq 0}$, which is an invariant of K, can be calculated from $w_0 = \sigma_t f_t + \dots + \rho_1 f_1 + e_0$ by the formula (2-4). Thus, w_0 satisfies (2-5) and using the algorithm provided by Lemma 2.12, Lemma 2.14 and the proof of Lemma 2.16, we see that the tuple $(\sigma_m, \dots, \sigma_t)$, where m is minimal such that $\sigma_m > 1$, is independent of t and $\|w_0\|^2 = \lceil p/q \rceil$. By definition, $(\sigma_m, \dots, \sigma_t)$ are the stable coefficients of L and it follows that they are independent of $b_2(X)$ and p/q. □

Proof of Corollary 1.7 This follows combining Theorem 1.5 with Remark 2.6. Theorem 1.5 shows that $-Q_X$ and $-Q_{X'}$ are both p/q–changemaker lattices with the same stable coefficients. Remark 2.6 then shows that $Q_{X'} \cong Q_X \oplus (-\mathbb{Z}^k)$. The isomorphism of intersection forms $Q_X \oplus (-\mathbb{Z}^k) \cong Q_{X \#_k \overline{\mathbb{CP}}^2}$ is clear. □

2.4 L–space knots

Now we specialise to the case of L–space surgeries. A knot K is said to be an *L–space knot* if $S^3_{p/q}(K)$ is an L–space for some $p/q \in \mathbb{Q}$. The knot Floer homology of an L–space knot is known to be determined by its Alexander polynomial, which can be written in the form

$$\Delta_K(t) = a_0 \sum_{i=1}^{g} a_i (t^i + t^{-i}),$$

where $g = g(K)$, $a_g = 1$ and the nonzero values of a_i alternate in sign and assume values in $\{\pm 1\}$ [23; 24]. Given an Alexander polynomial in this form, we can compute its *torsion coefficients* by the formula

$$t_i(K) = \sum_{j\geq 1} j a_{|i|+j}.$$

When K is an L–space knot, the V_i appearing in (2-3) satisfy $V_i = t_i(K)$ for $i \geq 0$ [26]. Thus if $S^3_{p/q}(K)$ is an L–space bounding a negative-definite sharp 4–manifold X, then Theorem 1.5 shows that the intersection form is isomorphic to a p/q–changemaker lattice L, where the stable coefficients, $(\sigma_r, \dots, \sigma_m)$, are determined by the torsion coefficients. Since $t_i(K) = 0$ if and only if $i \geq g(K)$, Lemma 2.10 shows that the genus can be computed by the formula

$$g(K) = \frac{1}{2} \sum_{i=m}^{r} \sigma_i(\sigma_i - 1), \tag{2-12}$$

which was first proven by Greene [11, Proposition 3.1].

Remark 2.17 Lemma 2.10 shows that σ_r and σ_{r-1} have particularly simple interpretations in terms of torsion coefficients:

$$\sigma_r = \#\{0 \le i < g \mid t_i(K) = 1\} \quad \text{and} \quad \sigma_{r-1} = \#\{0 \le i < g \mid t_i(K) = 2\}.$$

As in the proof of Lemma 2.16, the remaining stable coefficients can be also be computed from the torsion coefficients. However, the relationship is more complicated.

3 Graph lattices and obtuse superbases

In this section, we gather together some lattice-theoretic concepts and properties that we will need.

3.1 Graph lattices

We recall the definition of a graph lattice and state the results that we will require for this paper. All statements in this section can be found with proof in [17].

Let $G = (V, E)$ be a finite, connected, undirected graph with no self-loops. For a pair of disjoint subsets $R, S \subset V$, let $E(R, S)$ be the set of edges between R and S. Define $e(R, S) = |E(R, S)|$. We will use the notation $d(R) = e(R, V \setminus R)$.

Let $\overline{\Lambda}(G)$ be the free abelian group generated by $v \in V$. Define a symmetric bilinear form on $\overline{\Lambda}(G)$ by

$$v \cdot w = \begin{cases} d(v) & \text{if } v = w, \\ -e(v, w) & \text{if } v \neq w. \end{cases}$$

In this section we will use the notation $[R] = \sum_{v \in R} v$, for $R \subseteq V$. The above definition gives

$$v \cdot [R] = \begin{cases} -e(v, R) & \text{if } v \notin R, \\ e(v, V \setminus R) & \text{if } v \in R. \end{cases} \tag{3-1}$$

From this it follows that $[V] \cdot x = 0$ for all $x \in \overline{\Lambda}(G)$. We define the *graph lattice* of G to be

$$\Lambda(G) := \frac{\overline{\Lambda}(G)}{\mathbb{Z}[V]}.$$

The bilinear form on $\overline{\Lambda}(G)$ descends to $\Lambda(G)$. Since we have assumed that G is connected, the pairing on $\Lambda(G)$ is positive-definite. This makes $\Lambda(G)$ into an integral lattice. Henceforth, we will abuse notation by using v to denote its image in $\Lambda(G)$.

Recall that a vector z in a lattice is *irreducible* if it cannot be written in the form $z = x + y$ for nonzero x and y with $x \cdot y \ge 0$. The irreducible vectors in $\Lambda(G)$ can be characterised in terms of the graph G.

Lemma 3.1 *The vector $x \in \Lambda(G) \setminus \{0\}$ is irreducible if and only if $x = [R]$ for some $R \subseteq V$ such that R and $V \setminus R$ induce connected subgraphs of G.* □

A connected graph is said to be 2*–connected* if it cannot be disconnected by deleting a vertex. This property is equivalent to $\Lambda(G)$ being *indecomposable*, that is, $\Lambda(G)$ cannot be written as the orthogonal direct sum $\Lambda(G) = L_1 \oplus L_2$ with L_1 and L_2 nonzero sublattices.

Lemma 3.2 *The following are equivalent:*

(i) *The graph G is 2–connected.*

(ii) *Every vertex $v \in V$ is irreducible.*

(iii) *The lattice $\Lambda(G)$ is indecomposable.* □

Given a graph lattice of some graph G, the following lemma will be useful for identifying other graphs with isomorphic graph lattices.

Lemma 3.3 *Suppose that G is 2–connected. Let v be a vertex such that we can find $x, y \in \Lambda(G)$, with $v = x + y$ and $x \cdot y = -1$. Then there is a cut edge e in $G \setminus \{v\}$ and, if R and S are the vertices of the two components of $(G \setminus \{v\}) \setminus \{e\}$, then $\{x, y\} = \{[R] + v, [S] + v\}$. Let u_1 and u_2 be the endpoints of e. These are the unique vertices $u_1, u_2 \neq v$, with $x \cdot u_1 = y \cdot u_2 = 1$. Furthermore, any vertex $w \notin \{v, u_1, u_2\}$ satisfies $w \cdot x, w \cdot y \leq 0$.* □

3.2 Obtuse superbases

Given a positive definite integral lattice L of rank r, we say that L *admits an obtuse superbase* if it contains a set $B = \{v_0, \dots, v_r\}$ such that $v_1, \dots, v_r$ form a basis for L, $v_0 + \dots + v_r = 0$ and $v_i \cdot v_j \leq 0$ for all $0 \leq i \neq j \leq r$. We will call the set B a *an obtuse superbase* for L. This terminology is taken from the work of Conway and Sloane [3].

Given an obtuse superbase $B = \{v_0, \dots, v_r\}$ for L, we can construct a graph G_B by taking vertex set B with $|v_i \cdot v_j|$ edges between vertices v_i and v_j for $i \neq j$. With this construction in mind, we will frequently refer to elements of a given obtuse superbase as vertices of L.

Proposition 3.4 *The graph G_B is connected and L is isomorphic to $\Lambda(G_B)$.*

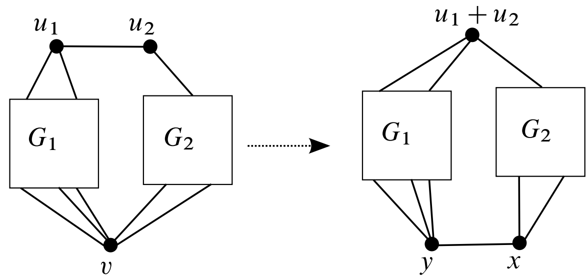


Figure 2: The graphs G_B and $G_{B'}$ corresponding to the obtuse superbases appearing in Lemma 3.5

Proof First we show that G_B is connected. Let $R \subseteq B$ be the vertices of a nonempty connected component of G_B. We see that the vector $[R] = \sum_{x\in R} x$ satisfies $[R]\cdot v_i = 0$ for all $0 \le i \le r$ (see (3-1)). Since L is positive-definite, this implies that $[R] = 0$. By definition, $v_1, \dots, v_r$ must be linearly independent. It follows that $R = B$ and hence G_B is connected, as required.

To show that $\Lambda(G_B)$ is isomorphic to L, take the linear map which takes vertices to the corresponding vectors in L. Since $v_0 + \dots + v_r = 0$, we have

$$d(v_k) = -\sum_{i\neq k} v_k \cdot v_i = \|v_k\|^2,$$

and by construction we have $e(v_i, v_j) = -v_i \cdot v_j$ for $i \neq j$. This shows that this map is the required isomorphism. □

For any given lattice there may be many choices of obtuse superbase. The following lemma shows one way to convert one obtuse superbase into another.

Lemma 3.5 *Let L be an indecomposable lattice with an obtuse superbase B. Suppose that we have $v \in B$ which can be written as $v = x + y$, where $x, y \in L$ and $x \cdot y = -1$. There are unique $u_1, u_2 \in B$ with $u_1 \cdot x > 0$ and $u_2 \cdot y > 0$ and the set $B' = (B \setminus \{v, u_1, u_2\}) \cup \{x, y, u_1 + u_2\}$ is also an obtuse superbase for L.*

Proof Since L is indecomposable, Lemma 3.2 shows that the graph G_B is 2–connected. Thus we may apply Lemma 3.3, which shows that there are disjoint connected subgraphs G_1 and G_2 of G_B and vertices u_1 and u_2 such that $x = v + u_1 + \sum_{z\in G_1} z$ and $y = v + u_2 + \sum_{z\in G_2}$, with a unique edge between u_1 and u_2 which is a cut-edge in $G_B \setminus \{v\}$. It is straightforward to verify that $B' = (B \setminus \{v, u_1, u_2\}) \cup \{x, y, u_1 + u_2\}$ is an obtuse superbase for L. An illustration of how the graph $G_{B'}$ is obtained from G_B is given in Figure 2. □

4 Alternating surgeries

In this section, we will prove our main results.

4.1 The Goeritz form

A diagram D of a link L divides the plane into connected regions. We may colour these regions black and white in a chessboard fashion. This colouring can be done in two different ways. Each of the possible colourings gives an incidence number, $\mu(c) \in \{\pm 1\}$, at each crossing c of D, as shown in Figure 3. We construct a planar graph, Γ_D, by drawing a vertex in each white region and an edge e for every crossing c between the two white regions it joins. We define an incidence number on each edge by $\mu(e) = \mu(c)$. We call this the *white graph* corresponding to D. This gives rise to a *Goeritz matrix*, $G_D = (G_{ij})$, defined by labelling the vertices of Γ_D by $v_1, \dots, v_{r+1}$ and, for $1 \le i, j \le r$, setting

$$g_{ij} = \sum_{e \in E(v_i, v_j)} \mu(e)$$

for $i \neq j$ and

$$g_{ii} = -\sum_{e \in E(v_i, \Gamma_D \backslash v_i)} \mu(e)$$

otherwise [13, Chapter 9].

Now suppose that L is an alternating, nonsplit link. If D is any alternating diagram, then we may fix the colouring so that $\mu(c) = -1$ for all crossings. In this case, G_D defines a positive-definite bilinear form. This in turn gives a lattice, Λ_D, which we will refer to as the *white lattice* of D. Observe that if D is reduced (ie contains no nugatory crossings), then Γ_D contains no self-loops or cut-edges and Λ_D is isomorphic to the graph lattice $\Lambda(\Gamma_D)$.

Ozsváth and Szabó have shown that the Heegaard Floer homology d–invariants of the branched double cover $\Sigma(L)$ are determined by Λ_D [25].

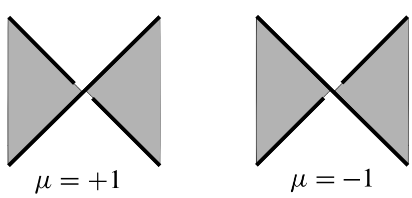


Figure 3: The incidence number of a crossing

Theorem 4.1 [25] *Let L be a nonsplit alternating link with a reduced alternating diagram D. The double branched cover $\Sigma(L)$ is an L–space which bounds a simply connected negative-definite sharp 4–manifold with intersection form isomorphic to $-\Lambda_D$.* □

4.2 Changemaker lattices admitting obtuse superbases

We will establish some restrictions on changemaker lattice which admits an obtuse superbase. The following proposition, which combines results from [17; 16], will allow us to restrict our attention to integer changemaker lattices.

Proposition 4.2 *Suppose that for some $p/q = n - r/q$ with $q > r \geq 1$, the changemaker lattice*

$$L_{p/q} = \langle w_0, \dots, w_l \rangle^{\perp} \subseteq \mathbb{Z}^{t+s+1},$$

where $w_0 = e_0 + \sigma_1 f_1 + \cdots + \sigma_t f_t$, admits an obtuse superbase. Then the changemaker lattices

$$L_n = \langle w_0 \rangle^{\perp} \subseteq \mathbb{Z}^{t+1} = \langle e_0, f_1, \dots, f_t \rangle,$$
$$L_{n-1} = \langle w_0 - e_0 \rangle^{\perp} \subseteq \mathbb{Z}^{t} = \langle f_1, \dots, f_t \rangle,$$

both admit obtuse superbases. Furthermore, if $\sigma_t > 1$, then we can assume the obtuse superbase for L_{n-1} contains a vector x with $x \cdot f_1 = -2$.

Proof Since $L_{p/q}$ admits an obtuse superbase, it follows from [16, Proposition 7.7] that the lattice

$$L_{n-1/2} = \langle w_0, e_1 - e_0 \rangle^{\perp} \subseteq \mathbb{Z}^{t+2} = \langle e_1, e_0, f_1, \dots, f_t \rangle$$

also admits an obtuse superbase, which we will call B. The results of [17] show that there are precisely two vertices v and w in B with $v \cdot e_0,\ w \cdot e_0 \neq 0$ and they satisfy $v \cdot w \leq -1$. Moreover, the results of the same paper show that we can assume that $v = -f_1 + e_0 + e_1$ and $w \cdot e_0 = w \cdot e_1 = -1$, and if there is k such that $\sigma_k > 1$, then we can assume that $w \cdot f_1 = -1$.

Consider the set $B' = B \setminus \{v, w\} \cup \{v + w\}$. Since $(v + w) \cdot e_0 = (v + w) \cdot e_1 = 0$, we have $B' \subset L_{n-1}$. Since B spans $L_{n-1/2}$, we see that B' must span L_{n-1}. Since B is an obtuse superbase for $L_{n-1/2}$, it follows that B' is an obtuse superbase for L_{n-1}, where the graph $G_{B'}$ is obtained from G_B by contracting the edge between v and w. Furthermore, if there is $\sigma_k > 1$, then $x = v + w$ is the required vector with $x \cdot f_1 = -2$.

Now consider the set $B'' = B \setminus \{v, w\} \cup \{v - e_1, w + e_1\}$. Since every element $x \in B \setminus \{v, w\}$ has $x \cdot e_1 = 0$, we see that every $x \in B''$ satisfies $x \cdot e_1 = 0$, so we have $B'' \subseteq \langle e_0, f_1, \dots, f_t\rangle$ and hence $B'' \subseteq L_{n+1}$. Since B is an obtuse superbase for $L_{n-1/2}$ and $v \cdot w \le -1$, it follows that B'' is an obtuse superbase for L_{n+1}, where the graph $G_{B''}$ is obtained from G_B by deleting an edge between v and w. □

The next lemma gives bounds on when a changemaker lattice can be decomposable.

Lemma 4.3 [9, Lemma 5.1] *Suppose that $L = \langle w_0\rangle^\perp \subseteq \mathbb{Z}^t$ is a changemaker lattice, where $w_0 = \sigma_1 f_1 + \dots + \sigma_t f_t$ with $\sigma_i \ge 1$ for all i and $\sigma_t > 1$. Let $m \le t$ be minimal such that $\sigma_m > 1$. If L is decomposable, then $\sigma_m = m - 1$.* □

We get a similar bound on a changemaker lattice admitting an obtuse superbase in terms of its stable coefficients. This will allow us to prove the upper bound in Theorem 1.2.

Lemma 4.4 *Suppose that $L = \langle w_0\rangle^\perp \subseteq \mathbb{Z}^t$ is a changemaker lattice, where $w_0 = \sigma_1 f_1 + \dots + \sigma_t f_t$ with $\sigma_i \ge 1$ for all i. If the stable coefficients $(\sigma_m, \dots, \sigma_t)$ are a nonempty tuple and L admits an obtuse superbase, then $\sigma_m \ge m - 2$ and*

$$\|w_0\|^2 \le 1 + \sigma_m + \sum_{i=m}^{t} \sigma_i^2.$$

Proof If L is decomposable, then Lemma 4.3 shows that the bound is automatically satisfied. We will assume from now on that L is indecomposable.

For $2 \le i \le m-1$, let v_i be the vector $v_i = e_i - e_{i-1}$. Since $\sigma_i = \sigma_{i-1} = 1$ for i in this range, we have $v_i \in L$. We will use Lemma 3.5 to show that L admits an obtuse superbase containing the vectors $v_2, \dots, v_{m-1}$.

Let B be an obtuse superbase and let $k \le m-1$, be minimal such that v_k is not in B. Suppose first that $k = 2$. Since v_2 is irreducible, Lemma 3.1 implies that it can be written as a sum of elements of B. Hence, there is a vector $u \in B$ with $u \cdot v_2 > 0$. By Lemma 3.2, the indecomposability of L implies that u is irreducible. In turn, this implies that $(u - v_2) \cdot v_2 = u \cdot v_2 - 2 = -1$. Therefore by applying Lemma 3.5, we see that there is an obtuse superbase containing v_2.

Now we suppose that $k > 2$. Since v_k is irreducible, Lemma 3.1 shows that it can be written as a sum of elements of B. Since v_{k-1} is a vertex of B and $v_k \cdot v_{k-1} = -1$, there is $u \in B$ with $u \cdot v_k = -u \cdot v_{k-1} = 1$. This must satisfy $(u - v_k) \cdot v_k = -1$. By Lemma 3.5, this implies we can find an obtuse superbase containing v_k. Moreover, since $(u - v_k) \cdot v_j \le 0$ and $v_k \cdot v_j \le 0$ for all $2 \le j < k$, we can assume that $v_2, \dots, v_{k-1}$ are also in this obtuse superbase. Thus, proceeding inductively, we see that we can assume that $v_2, \dots, v_{m-1}$ are all contained in the obtuse superbase B.

Suppose that $\sigma_m = m - b$ for some $m-2 \geq b \geq 2$. Consider the vector $v_m = -e_m + e_{m-1} + \cdots + e_b \in L$. Since this is irreducible, Lemma 3.1 shows that we may write it as a sum of vertices $v_m = \sum_{x \in R} x$ for some subset $R \subseteq B$. Since $v_m \cdot v_b = -1$, we have $v_b \notin R$ and there must exist $u \in R$ with $u \cdot v_b = -1$ and $u \cdot v_m = 1$. However, as $\|v_b\|^2 = 2$, there are at most two vectors in B which pair nontrivially with v_b. If $b \geq 3$ then we have $v_{b-1} \cdot v_b = v_{b+1} \cdot v_b = -1$ and $v_{b-1} \cdot v_m = v_{b+1} \cdot v_m = 0$. This implies that the required $u \in B$ cannot exist if $b \geq 3$. Thus we must have $b = 2$. This shows that $\sigma_m \geq m-2$, as required. Since $\sigma_i = 1$ for $i < m$, we have

$$\|w_0\|^2 = m - 1 + \sum_{i=m}^{t} \sigma_i^2 \leq 1 + \sigma_m + \sum_{i=m}^{t} \sigma_i^2,$$

which is the required bound. This completes the proof. □

This allows us to prove the inequality which will give Theorem 1.1.

Lemma 4.5 *Suppose that $L = \langle \sigma_1 f_1 + \cdots + \sigma_t f_t \rangle^{\perp} \subseteq \mathbb{Z}^t$ is a changemaker lattice which admits an obtuse superbase and $\sigma_t > 1$. Then*

$$\sum_{i=1}^{t} \sigma_i^2 \leq 2 \sum_{i=1}^{t} \sigma_i(\sigma_i - 1) + 3.$$

Proof Let m be minimal such that $\sigma_m > 1$. Since L admits an obtuse superbase, Lemma 4.4 shows that we have

$$\sum_{i=1}^{t} \sigma_i^2 \leq \sum_{i=m}^{t} \sigma_i^2 + \sigma_m + 1.$$

Observe that if $\sigma_i \geq 2$, then $\sigma_i^2 \leq 2\sigma_i(\sigma_i - 1)$. Since $\sigma_m \geq 2$, we also have $\sigma_m^2 + \sigma_m \leq 2\sigma_m(\sigma_m - 1) + 2$. Combining these inequalities, we obtain

$$\sum_{i=1}^{t} \sigma_i^2 \leq \sum_{i=m}^{t} \sigma_i^2 + \sigma_m + 1 \leq 2 \sum_{i=m}^{t} \sigma_i(\sigma_i - 1) + 3 = 2 \sum_{i=1}^{t} \sigma_i(\sigma_i - 1) + 3,$$

which is the required inequality. □

4.3 The main results

Suppose that K is an nontrivial knot such that $S^3_{p/q}(K)$ is an alternating surgery, that is, $S^3_{p/q}(K) = \Sigma(L)$ for an alternating knot or link L. Since a nontrivial L–space knot cannot admit both positive and negative L–space surgeries and

$$-S^3_r(K) = S^3_{-r}(\overline{K}) = \Sigma(\overline{L}),$$

we may assume that $p/q > 0$ and that all other alternating surgeries on K arise from positive slopes.

Let D be a reduced alternating diagram of L. By Theorems 1.5 and 4.1, the lattice Λ_D is isomorphic to a p/q–changemaker lattice,

$$\Lambda_{p/q} = \langle w_0, \dots, w_l \rangle^{\perp} \subseteq \mathbb{Z}^{\operatorname{rk}\Lambda_D + l + 1},$$

whose stable coefficients are determined by the Alexander polynomial of K.

Since Λ_D is the graph lattice associated to the white graph of D, the lattice $\Lambda_{p/q}$ admits an obtuse superbase. We write w_0 in the form

$$w_0 = \begin{cases} e_0 + \sigma_1 f_1 + \dots + \sigma_t f_t & \text{if } q > 1, \\ \sigma_1 f_1 + \dots + \sigma_t f_t & \text{if } q = 1. \end{cases}$$

Since D is reduced, Γ_D contains no cut-edges. This implies that Λ_D contains no vectors of norm 1 and so $\sigma_i \geq 1$ for all i. As we are assuming that $g(K) > 0$, (2-12) implies that $\sigma_t > 1$. So the stable coefficients form a nonempty tuple, $(\sigma_m, \dots, \sigma_t)$. This allows us to define

$$N = \sigma_m + \sum_{i=m}^{t} \sigma_i^2,$$

which will be the integer appearing in the statement of Theorem 1.2.

Proof of Theorems 1.1 and 1.2 Proposition 4.2 implies that the $\lceil p/q \rceil$–changemaker lattice

$$\Lambda' = \langle w_0 \rangle^{\perp} \subseteq \begin{cases} \mathbb{Z}^{t+1} & \text{if } q > 1, \\ \mathbb{Z}^{t} & \text{if } q = 1, \end{cases}$$

also admits an obtuse superbase. As shown by (2-12), we have

$$2g(K) = \sum_{i=1}^{t} \sigma_i(\sigma_i - 1).$$

Therefore, Lemma 4.5 gives the bound

$$\left\lceil \frac{p}{q} \right\rceil = \|w_0\|^2 \leq 4g(K) + 3.$$

This proves Theorem 1.1. From Lemma 4.4, we get the upper bound

$$\frac{p}{q} \leq \|w_0\|^2 \leq 1 + \sigma_m + \sum_{i=m}^{t} \sigma_i^2 = N + 1.$$

Since $(\sigma_1, \ldots, \sigma_t)$ satisfies the changemaker condition, we must have

$$\sigma_m \leq 1 + \sum_{i=1}^{m-1} \sigma_i = 1 + \sum_{i=1}^{m-1} \sigma_i^2,$$

where the second inequality holds since $\sigma_i = 1$ for $1 \leq i < m$. Thus we obtain

$$\frac{p}{q} \geq \sum_{i=1}^{t} \sigma_i^2 \geq \sum_{i=m}^{t} \sigma_i^2 + \sigma_m - 1 = N - 1.$$

This completes the proof of Theorem 1.2. □

The lower bound $N-1$ appearing in this proof arises from the fact that there can be no r–changemaker lattice for any $r < N-1$ with stable coefficients $(\sigma_m, \ldots, \sigma_t)$. Thus it follows from Theorem 1.5 that if $S^3_r(K)$ bounds a negative-definite sharp manifold for $r > 0$, then $r \geq N-1$. This justifies the claim made in Remark 1.6.

Now it remains to prove Theorem 1.3.

Proof of Theorem 1.3 Assume that $S^3_r(K)$ is an alternating surgery for $r \in \{r_1, N, r_2\}$ with $N-1 \leq r_1 < N < r_2 < N+1$. Let $S^3_{r_i}(K) = \Sigma(L_i)$ for $i = 1, 2$ and $S^3_N(K) = \Sigma(L)$ for L and L_i alternating. For $i = 1, 2$, let D_i be a reduced alternating diagram for L_i and let D be a reduced alternating diagram for L. Theorem 1.5 shows that there is $w_0 = \sigma_t f_t + \cdots + \sigma_2 f_2$ such that Λ_{D_1} is isomorphic to the r_1–changemaker lattice

$$\Lambda_{r_1} = \langle w_0 + e_0, w_1, \ldots, w_{l_1} \rangle^{\perp} \subseteq \langle f_2, \cdots, f_t, e_0, \ldots, e_{s_1} \rangle,$$

Λ_{D_2} is isomorphic to the r_2–changemaker lattice

$$\Lambda_{r_2} = \langle w_0 + f_1 + e_0, w_1, \ldots, w_{l_2} \rangle^{\perp} \subseteq \langle f_1, f_2, \cdots, f_t, e_0, \ldots, e_{s_2} \rangle,$$

and Λ_D is isomorphic the N–changemaker lattice

$$\Lambda_N = \langle w_0 \rangle^{\perp} \subseteq \langle f_1, \ldots, f_t \rangle.$$

Since Λ_{r_2} admits an obtuse superbase, Proposition 4.2 implies that Λ_N admits an obtuse superbase containing a vertex v with $v \cdot f_1 = -2$. Since Λ_{r_1} is a changemaker lattice, $(\sigma_2, \ldots, \sigma_t)$ must satisfy the changemaker condition. Therefore, if $g > 1$ is minimal such that $v \cdot f_g \geq 0$, then Proposition 2.2 implies that there is $A \subseteq \{2, \ldots, g-1\}$ with $\sigma_g - 1 = \sum_{i \in A} \sigma_i$. If we set $z = f_g - f_1 - \sum_{i \in A} f_i$, we have $z \in \Lambda_N$ and we can compute

$$(v - z) \cdot z = v \cdot f_g - 1 + -(v \cdot f_1 + 1) - \sum_{i \in A} (v \cdot f_i + 1)$$
$$\geq v \cdot f_g - v \cdot f_1 - 2 = v \cdot f_g \geq 0.$$

Since $z \neq v$, this shows that v is reducible. Thus Lemma 3.2 implies that Λ_N is decomposable and that if Λ_N is isomorphic to a graph lattice $\Lambda(G)$ for any connected graph G, then G contains a cut vertex. This shows that the white graph Γ_D contains a cut vertex. Since, we have assumed that D is reduced, this implies that $L = L_1 \# L_2$ for nontrivial L_1 and L_2. Therefore $S^3_N(K) = \Sigma(L_1) \# \Sigma(L_2)$ is reducible. Using work of Hoffman [12], Matignon and Sayari [14] showed that if $S^3_N(K)$ is a reducible surgery, then either $N \leq 2g(K) - 1$ or K is a cable knot. Since we have

$$N > 2g(K) = \sum_{i=1}^{t} \sigma_i(\sigma_i - 1),$$

it follows that K is cable knot. This completes the proof of Theorem 1.3. □

5 Examples and questions

We give some examples relating to alternating surgeries and sharp 4–manifolds to illustrate the results of this paper. We then conclude the paper by discussing some questions that arise naturally from this work.

5.1 Alternating surgeries via the Montesinos trick

We will now describe a construction for building knots admitting alternating surgeries. As far as the author is aware, this construction accounts for all known examples of alternating surgeries.

An *almost-alternating diagram* D is one which can be obtained by a crossing change from an alternating diagram. We call a crossing which can be changed to obtain an alternating diagram a *dealternating crossing*. Now let D be an almost-alternating diagram of the unknot with a dealternating crossing c and let B be a small ball containing c. Since the double cover of S^3 branched over the unknot is S^3, the ball B lifts to a solid torus $T \subseteq S^3$ when we take the double cover of S^3 branched over D. Let $K \subseteq S^3$ be the knot given by the core of T. If D' is obtained from D by replacing c with some other rational tangle, then the Montesinos trick shows that $\Sigma(D')$ is obtained by surgery on K [18]. Since we may perform tangle replacements such that the resulting diagram is alternating, we see that K admits alternating surgeries. If we take D' to be the alternating diagram obtained by changing c, then the resulting surgery is half-integral:

$$S^3_{n+1/2}(K) = \Sigma(D')$$

for some $n \in \mathbb{Z}$. By reflecting D, if necessary, we may assume that n is positive. It can be shown (eg [16, Proposition 5.4]) there are tangle replacements showing that $S^3_r(K)$ is an alternating surgery for all r in the range $n \le r \le n+1$.

Remark 5.1 It follows from the work of Watson that for all $r \ge n$, the manifold $S^3_r(K)$ is the double branched cover of a quasi-alternating link L [28]. However, Theorem 1.2 shows that when K is nontrivial L can only be alternating for $r \le n+2$. Thus we see that almost-alternating diagrams of the unknot gives rise to infinite families of nonalternating quasi-alternating knots and links.

Remark 5.2 It follows from Theorem 1.3 that if K is not a cable knot or the unknot, then K can admit at most one other alternating surgery with $r = n+2$ or $r = n-1$. If one uses the generalisation of Theorem 1.3 asserted in Remark 1.4, then we see that actually neither of these possibilities can arise and that $S^3_r(K)$ is an alternating surgery if and only if $n \le r \le n+1$.

As an example, we see what the results of this paper say about alternating surgeries on the $(-2,3,7)$–pretzel knot and describe how they arise through the construction given in this section.

Example 5.3 Let K denote the $(-2,3,7)$–pretzel knot. It is well known that K admits two lens space surgeries [5]. This implies that K is an L–space knot and in particular that it has alternating surgeries. The Alexander polynomial is

$$\Delta_K(t) = t^5 + t^{-5} - (t^4 + t^{-4}) + t^2 + t^{-2} - (t^1 + t^{-1}) + 1.$$

The corresponding nonzero torsion coefficients are $t_0 = t_1 = 2$ and $t_2 = t_3 = t_4 = 1$. From Lemma 2.13 we can deduce that the stable coefficients of the corresponding changemaker vector are $(2,2,3)$. If we apply Theorem 1.2 to K, then integer N we obtain is $N = 3^2 + 2^2 + 2^2 + 2 = 19$. Therefore, if $S^3_r(K)$ is an alternating surgery, then $18 \le r \le 20$.

Since the changemaker lattice

$$L = \langle 3f_6 + 2f_5 + 2f_4 + f_3 + f_2 + f_1 \rangle^{\perp}$$

does not admit an obtuse superbase, we see that $S^3_r(K)$ cannot be an alternating surgery for $19 < r \le 20$.

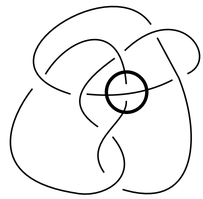


Figure 4: A diagram of 8_{17} with its unknotting crossing circled. For each r in the range $18 \le r \le 19$, r–surgery on the $(-2,3,7)$–pretzel knot yields the branched double cover of an alternating knot or link obtained by replacing the unknotting crossing in 8_{17} by some rational tangle. Note that both resolutions of the unknotting crossing give a 2–bridge knot or link. The two resolutions correspond to the cases $r = 18$ and 19.

In fact, K arises through the construction given in Section 5.1, and for each r in $18 \le r \le 19$, $S^3_r(K)$ branches over an alternating knot or link obtained by tangle replacement on the knot 8_{17}, as shown in Figure 4.

5.2 Some knots with no alternating surgeries

We use the results of this paper to exhibit two examples of L–space knots which do not admit any alternating surgeries. Although both are cables of the trefoil, they do not admit alternating surgeries for different reasons: in one case, the cabling slope is "too large" and in the other it is "too small".

Example 5.4 Let K be the $(2,15)$–cable of $T_{2,3}$. Since

$$S^3_{30}(K) = S^3_{15/2}(T_{3,2}) \# L(2,1)$$

is an L–space, K is an L–space knot. We will show that this does not admit any alternating surgeries. The Alexander polynomial of K is given by

$$\Delta_K(t) = t^9 + t^{-9} - (t^8 + t^{-8}) + t^5 + t^{-5} - (t^4 + t^{-4}) + t^3 + t^{-3} - (t^2 + t^{-2}) + t + t^{-1} - 1.$$

By the observations of Remark 2.17 and (2-12), we see that the stable coefficients given by K must be $(2,2,2,4)$. Thus the quantity N in Theorem 1.2 is given by $N = 30$. Combining this with Proposition 4.2, we see that to verify that K has no alternating surgeries we need only check that none of the three changemaker lattices

$$L_{29} = \langle 4e_0 + 2e_1 + 2e_2 + 2e_3 + e_4 \rangle^{\perp},$$
$$L_{30} = \langle 4e_0 + 2e_1 + 2e_2 + 2e_3 + e_4 + e_5 \rangle^{\perp},$$
$$L_{31} = \langle 4e_0 + 2e_1 + 2e_2 + 2e_3 + e_4 + e_5 + e_6 \rangle^{\perp},$$

admit obtuse superbases. Since this can be verified relatively easily, for example by using that in each case there are only a small number of irreducible vectors v with $v \cdot e_0 \neq 0$, we see that K does not admit any alternating surgeries.

Example 5.5 Let K be the $(2,3)$–cable of $T_{2,3}$. We will show that this is an L–space knot not admitting any alternating surgeries. The Alexander polynomial of K is

$$\Delta_K(t) = t^3 - t^2 + 1 - t^{-2} + t^{-3}.$$

Observe that this is the same as the Alexander polynomial for the torus knot $T_{3,4}$. If $S^3_r(K) = \Sigma(L)$ were an alternating surgery, then for any reduced alternating diagram D of L, the white lattice Λ_D would be isomorphic to an r–changemaker lattice with stable coefficients the tuple (3). It follows that we must have $11 \leq r \leq 13$. Since $S^3_r(T_{4,3})$ is an alternating surgery for any r in this range, we must have $\Lambda_D \cong \Lambda_{D'}$, where D' and is any reduced alternating diagram for an alternating knot or link L' such that $\Sigma(L') = S^3_r(T_{4,3})$. Since L and L' are alternating, this isomorphism of white lattices implies that L and L' must be mutants of one another and that $\Sigma(L) = \Sigma(L') = S^3_r(T_{4,3})$ [8]. Surgery on a torus knot is always a small Seifert fibred space [19], but $S^3_r(K)$ is a small Seifert fibred space only if r takes the form $r = 6 \pm 1/q$ [7]. Thus K admits no alternating surgeries.

5.3 Surgeries bounding sharp 4–manifold

It seems natural to wonder what we can say about the set of positive surgery slopes for which a given knot bounds a negative-definite sharp manifold. It can be shown that if it is nonempty then this set is an unbounded interval.

Theorem 5.6 [15, Theorem 1.2] *Let K be a knot in S^3. If $S^3_{p/q}(K)$ bounds a sharp negative-definite 4–manifold for some $p/q > 0$, then $S^3_{p'/q'}(K)$ bounds a sharp negative-definite 4–manifold for all $p'/q' \geq p/q$.* □

This allows us to characterise the set of all such slopes for torus knots admitting positive L–space surgeries.

Proposition 5.7 *For $r, s > 1$ and $p/q > 0$, the manifold $S^3_{p/q}(T_{r,s})$ bounds a negative-definite sharp 4–manifold if and only if $p/q \geq rs - 1$.*

Proof Since $S^3_{rs-1}(T_{r,s})$ is a lens space [19], Theorem 5.6 shows that $S^3_{p/q}(T_{r,s})$ bounds a negative-definite sharp 4–manifold for any $p/q \geq rs - 1$. To obtain the converse, observe that $S^3_{rs+1}(T_{r,s})$ is also a lens space and hence also an alternating surgery. Thus, for $K = T_{r,s}$, we see that the integer N in Theorem 1.2 is $N = rs$. Thus Remark 1.6 gives the desired lower bound. □

There are also examples of L–space knots for which no such slopes exist.

Example 5.8 Let K be the $(2,5)$–cable of $T_{2,3}$. We will show that K is an L–space knot such that $S^3_r(K)$ cannot bound a sharp negative-definite 4–manifold for any $r>0$. Since $S^3_{10}(K)=S^3_{5/2}(T_{3,2})\,\#\,L(2,1)$ is an L–space, K is an L–space knot. To show that $S^3_r(K)$ cannot bound a sharp 4–manifold, we show there is no vector satisfying (2-4). The Alexander polynomial of K is

$$\Delta_K(t)=t^4-t^3+1-t^{-3}+t^{-4},$$

which has nonzero torsion coefficients $t_0(K)=t_1(K)=t_2(K)=t_3(K)=1$. Thus, by Remark 2.17, we can assume that the first coordinate of any vector satisfying (2-4) is $\sigma_0=4$. However this contradicts (2-12), which implies that we must have $\sigma_0(\sigma_0-1)\le 2g(K)=8$.

5.4 Further questions

Given the results of this paper, it is natural to wonder how the set of knots admitting alternating surgeries are contained within the set of all L–space knots. For the purposes of this discussion we define several classes of L–space knots. We will restrict our attention to those admitting positive L–space surgeries. We say that $S^3_r(K)$ is a *quasi-alternating surgery* if it is the double branched cover of a quasi-alternating knot or link.

$$\begin{aligned}
\mathcal{L}&=\{K : S^3_r(K) \text{ is an } L\text{–space for some } r>0\},\\
\mathcal{A}&=\{K : S^3_r(K) \text{ is an alternating surgery for some } r>0\},\\
\mathcal{D}&=\{K : K \text{ is the double branched cover of an unknotting arc in an}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{alternating diagram}\},\\
\mathcal{QA}&=\{K : S^3_r(K) \text{ is a quasi-alternating surgery for some } r>0\}.
\end{aligned}$$

Since the double branched cover of a quasi-alternating knot is an L–space and any alternating link is quasi-alternating, these sets satisfy the inclusions

$$\mathcal{D}\subseteq\mathcal{A}\subseteq\mathcal{QA}\subseteq\mathcal{L}.$$

Watson has shown that any sufficiently large cable of a torus knot is in $\mathcal{QA}$ [28]. In particular, the $(2,15)$–cable of $T(2,3)$ is in $\mathcal{QA}$. As we have shown that it is not in $\mathcal{A}$, this shows that $\mathcal{A}\subsetneq\mathcal{QA}$.

Remark 5.9 It seems probable that there are L–space knots which do not admit quasi-alternating surgeries. The $(2,3)$–cable of $T_{2,3}$ and the $(2,5)$–cable of $T_{2,3}$ seem to be potential candidates for knots in $\mathcal{L}\setminus\mathcal{QA}$.

As far as the author is aware, all known examples of knots in $\mathcal{A}$ are also in $\mathcal{D}$, ie they arise through the construction in Section 5.1. Moreover, it is known that for every noninteger alternating surgeries, there is a knot in $\mathcal{D}$ with the same surgery.

Theorem 5.10 [16, Theorem 1.2] *If $S^3_{p/q}(K)$ is an alternating surgery with $q > 1$, then there is $K' \in \mathcal{D}$ with $S^3_{p/q}(K) = S^3_{p/q}(K')$.* □

This suggests the following conjecture.

Conjecture 1 *Every alternating surgery arises as tangle replacement on an almost-alternating diagram of the unknot, that is, we have $\mathcal{A} = \mathcal{D}$.*

Since lens spaces arise as the double branched covers of alternating links, one can ask how this conjecture agrees with results and conjectures on lens space surgeries. The cyclic surgery theorem of Culler, Gordon, Luecke and Shalen shows that only torus knots admit noninteger lens space surgeries [4]. Since torus knots are in $\mathcal{D}$, this verifies Conjecture 1 in certain cases.

Short of attacking Conjecture 1 in full, there are various related questions we can ask.

Question 2 Does Theorem 5.10 extend to the case of integer alternating surgeries?

It follows from their construction that every knot in $\mathcal{D}$ admits a strong inversion.

Question 3 Is every knot in $\mathcal{A}$ strongly invertible?

It seems likely that any progress on Conjecture 1 would require an alternative description of the class $\mathcal{D}$.

Question 4 Is there a characterisation of $\mathcal{D}$ which does not refer to almost-alternating diagrams of the unknot?

Finally, as we demonstrated with the $(2,5)$–cable of $T_{2,3}$, (2-4) can be used to show that for some knots no manifold obtained by positive surgery can bound a negative-definite sharp manifold. As we saw in Example 5.5, the $(2,3)$–cable of $T_{2,3}$ passes this obstruction as it has the same Alexander polynomial as $T_{3,4}$. However, it seems unlikely that any positive surgery on the $(2,3)$–cable of $T_{2,3}$ bounds a sharp manifold.

Question 5 Can one find alternative ways to show that surgery on a knot does not bound a sharp 4–manifold? In particular, is it possible to show that no positive surgery on the $(2,3)$–cable of $T_{2,3}$ bounds a sharp manifold?

References

[1] **K L Baker**, *Small genus knots in lens spaces have small bridge number*, Algebr. Geom. Topol. 6 (2006) 1519–1621 MR

[2] **J L Brown, Jr**, *Note on complete sequences of integers*, Amer. Math. Monthly 68 (1961) 557–560 MR

[3] **J H Conway**, **N J A Sloane**, *Low-dimensional lattices, VI: Voronoĭ reduction of three-dimensional lattices*, Proc. Roy. Soc. London Ser. A 436 (1992) 55–68 MR

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

[5] **R Fintushel**, **R J Stern**, *Constructing lens spaces by surgery on knots*, Math. Z. 175 (1980) 33–51 MR

[6] **J Gibbons**, *Deficiency symmetries of surgeries in S^3*, Int. Math. Res. Not. 2015 (2015) 12126–12151 MR

[7] **C M Gordon**, *Dehn surgery and satellite knots*, Trans. Amer. Math. Soc. 275 (1983) 687–708 MR

[8] **J E Greene**, *Lattices, graphs, and Conway mutation*, Invent. Math. 192 (2013) 717–750 MR

[9] **J E Greene**, *The lens space realization problem*, Ann. of Math. 177 (2013) 449–511 MR

[10] **J E Greene**, *Donaldson's theorem, Heegaard Floer homology, and knots with unknotting number one*, Adv. Math. 255 (2014) 672–705 MR

[11] **J E Greene**, *L-space surgeries, genus bounds, and the cabling conjecture*, J. Differential Geom. 100 (2015) 491–506 MR

[12] **J A Hoffman**, *There are no strict great x–cycles after a reducing or P^2 surgery on a knot*, J. Knot Theory Ramifications 7 (1998) 549–569 MR

[13] **W B R Lickorish**, *An introduction to knot theory*, Graduate Texts in Mathematics 175, Springer (1997) MR

[14] **D Matignon**, **N Sayari**, *Longitudinal slope and Dehn fillings*, Hiroshima Math. J. 33 (2003) 127–136 MR

[15] **D McCoy**, *Surgeries, sharp 4–manifolds and the Alexander polynomial*, preprint (2014) arXiv

[16] **D McCoy**, *Non-integer surgery and branched double covers of alternating knots*, J. Lond. Math. Soc. 92 (2015) 311–337 MR

[17] **D McCoy**, *Alternating knots with unknotting number one*, Adv. Math. 305 (2017) 757–802 MR

[18] **J M Montesinos**, *Seifert manifolds that are ramified two-sheeted cyclic coverings*, Bol. Soc. Mat. Mexicana 18 (1973) 1–32 MR

[19] **L Moser**, *Elementary surgery along a torus knot*, Pacific J. Math. 38 (1971) 737–745 MR

[20] **Y Ni**, **Z Wu**, *Cosmetic surgeries on knots in* S^3, J. Reine Angew. Math. 706 (2015) 1–17 MR

[21] **B Owens**, **S Strle**, *Immersed disks, slicing numbers and concordance unknotting numbers*, Comm. Anal. Geom. 24 (2016) 1107–1138 MR

[22] **P Ozsváth**, **Z Szabó**, *Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary*, Adv. Math. 173 (2003) 179–261 MR

[23] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and genus bounds*, Geom. Topol. 8 (2004) 311–334 MR

[24] **P Ozsváth**, **Z Szabó**, *On knot Floer homology and lens space surgeries*, Topology 44 (2005) 1281–1300 MR

[25] **P Ozsváth**, **Z Szabó**, *On the Heegaard Floer homology of branched double-covers*, Adv. Math. 194 (2005) 1–33 MR

[26] **P S Ozsváth**, **Z Szabó**, *Knot Floer homology and rational surgeries*, Algebr. Geom. Topol. 11 (2011) 1–68 MR

[27] **J Rasmussen**, *Lens space surgeries and a conjecture of Goda and Teragaito*, Geom. Topol. 8 (2004) 1013–1031 MR

[28] **L Watson**, *A surgical perspective on quasi-alternating links*, from "Low-dimensional and symplectic topology" (M Usher, editor), Proc. Sympos. Pure Math. 82, Amer. Math. Soc., Providence, RI (2011) 39–51 MR

Department of Mathematics, University of Texas
Austin, TX, United States

d.mccoy@math.utexas.edu

http://www.ma.utexas.edu/users/d.mccoy/

Received: 15 December 2014 Revised: 27 February 2017

msp

Algebraic & Geometric Topology 17 (2017) 2635–2685

Link homology and equivariant gauge theory

PRAYAT POUDEL
NIKOLAI SAVELIEV

Singular instanton Floer homology was defined by Kronheimer and Mrowka in connection with their proof that Khovanov homology is an unknot detector. We study this theory for knots and two-component links using equivariant gauge theory on their double branched covers. We show that the special generator in the singular instanton Floer homology of a knot is graded by the knot signature mod 4, thereby providing a purely topological way of fixing the absolute grading in the theory. Our approach also results in explicit computations of the generators and gradings of the singular instanton Floer chain complex for several classes of knots with simple double branched covers, such as two-bridge knots, some torus knots, and Montesinos knots, as well as for several families of two-component links.

57M27; 57R58

1 Introduction

This paper studies the Floer homology $I_*(\Sigma, \mathcal{L})$ of two-component links $\mathcal{L} \subset \Sigma$ in homology 3–spheres defined by Kronheimer and Mrowka [24] using singular SO(3) instantons. An important special case of this theory is the singular instanton knot Floer homology $I^\natural(k)$ for knots $k \subset S^3$ obtained by applying $I_*(S^3, \mathcal{L})$ to the link $\mathcal{L}$ which is a connected sum of k with the Hopf link. The Floer homology $I_*(\Sigma, \mathcal{L})$ has a relative $\mathbb{Z}/4$ grading, which can be upgraded to an absolute $\mathbb{Z}/4$ grading in the special case of $I^\natural(k)$. Kronheimer and Mrowka [24] used $I^\natural(k)$ and its close cousin $I^\sharp(k)$ to prove that the reduced Khovanov homology is an unknot-detector.

The definition of the groups $I_*(\Sigma, \mathcal{L})$ uses singular gauge theory, which makes them difficult to compute. We propose a new approach to these computations which uses equivariant gauge theory in place of the singular one. Given a two-component link $\mathcal{L}$ in an integral homology sphere Σ, we pass to the double branched cover $M \to \Sigma$ with branch set $\mathcal{L}$ and observe that the singular connections on Σ used in the definition of $I_*(\Sigma, \mathcal{L})$ pull back to equivariant smooth connections on M. The generators of the Floer chain complex $IC_*(\Sigma, \mathcal{L})$, whose homology is $I_*(\Sigma, \mathcal{L})$, are then derived from

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2635

the equivariant representations $\pi_1 M \to \mathrm{SO}(3)$, and their Floer gradings are computed using equivariant rather than singular index theory.[1]

As our first application of this approach, we determine the grading of the special generator in the Floer chain complex $IC^\natural(k)$ of a knot $k \subset S^3$; see Section 5. This fixes the absolute $\mathbb{Z}/4$ grading on $I^\natural(k)$ and confirms the conjecture of Hedden, Herald and Kirk [20, Section 12.6].

Theorem *For any knot $k \subset S^3$, the grading of the special generator in the Floer chain complex $IC^\natural(k)$ equals* $\operatorname{sign} k \pmod 4$.

We also achieve significant simplifications in computing the Floer chain complexes $IC^\natural(k)$ and $IC_*(\Sigma, \mathcal{L})$ for knots and links with simple double branched covers, such as torus and Montesinos knots and links, whose double branched covers are Seifert fibered manifolds. Explicit calculations for these knots and links are possible because the gauge theory on Seifert fibered manifolds is sufficiently well developed; see Fintushel and Stern [15] and, in the equivariant setting, Collin and Saveliev [11] and Saveliev [36]. Here are sample results of our calculations:

(1) The Floer chain complex $IC^\natural(k)$ of a two-bridge knot k is calculated in Section 7.1. For example, the Floer chain complex of the figure-eight knot consists of free abelian groups of ranks $(1,1,2,1)$. In fact, the Kronheimer–Mrowka spectral sequence [24] is known to collapse for all two-bridge knots k, which implies that $IC^\natural(k) = I^\natural(k)$ for all such knots.

(2) The Floer chain complex $IC^\natural(k)$ of a Montesinos knot $k = k(p,q,r)$ whose double branched cover is a Brieskorn homology sphere $\Sigma(p,q,r)$ consists of free abelian groups of ranks $(1+b, b, b, b)$, where b equals -2 times the Casson invariant of $\Sigma(p,q,r)$; see Section 7.2. General Montesinos knots are discussed in Section 7.3.

(3) The Floer chain complex $IC_*(S^3, \mathcal{L})$ of two-component Montesinos links $\mathcal{L} = K((a_1,b_1),\dots,(a_n,b_n))$ whose double branched cover is a homology $S^1 \times S^2$ is calculated in Section 8.3. For example, the chain complex of the pretzel link $\mathcal{L} = P(2,-3,-6)$ consists of free abelian groups of ranks $(2,0,2,0)$ up to cyclic permutation; see Section 8.2. It has zero differential, hence $IC_*(S^3,\mathcal{L}) = I_*(S^3,\mathcal{L})$.

(4) Our calculations for torus knots are less satisfactory because the equivariant index theory in this setting is less well developed. For instance, we prove that the Floer chain complex $IC^\natural(k)$ of a torus knot $k = T_{p,q}$ with odd coprime integers p and q has rank $1 + 4a$, where $a = -\operatorname{sign}(T_{p,q})/4$, and we conjecture that the Floer chain groups

[1] The theory $I_*(\Sigma, \mathcal{L})$ is different from $I^\natural(\Sigma, \mathcal{L})$ studied in [24]: the latter is a Floer homology of a three-component link obtained by summing $\mathcal{L}$ with the Hopf link.

have ranks $(1+a, a, a, a)$; see Section 7.4.[2] A complete calculation of the Floer chain complex of the torus knot $T_{3,4}$ can be found in Example 7.9.

Some of the above results concerning two-bridge and torus knots were obtained earlier by Hedden, Herald, and Kirk [20] using pillowcase techniques, which are completely different from our equivariant methods. We do not discuss the more difficult problem of computing the boundary operators in the Floer chain complexes $IC^{\natural}(k)$ and $IC_*(\Sigma, \mathcal{L})$. Such calculations are still out of reach except in a few special cases. However, it may be worth investigating whether our equivariant techniques can shed some light on this problem.

Here is an outline of the paper. It begins with a sketch of the definition of $I_*(\Sigma, \mathcal{L})$ mainly following Kronheimer and Mrowka [24] but using the language of projective representations developed in Ruberman and Saveliev [33]; see also Dostoglou and Salamon [13]. We obtain a purely algebraic description of the generators in $IC_*(\Sigma, \mathcal{L})$ as well as of a certain natural $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ action on them, which is crucial to the rest of the paper.

Equivariant gauge theory is developed in Section 3. The section begins with a computation of $\mathbb{Z}/2$ cohomology rings of double branched covers $M \to \Sigma$ of two-component links, followed by a computation of the characteristic classes of SO(3)–bundles on M pulled back from orbifold bundles on Σ. The results are used to establish a bijective correspondence between equivariant SO(3) representations of $\pi_1 M$ and orbifold SO(3) representations of $\pi_1 \Sigma$. In the rest of the section, we discuss equivariant index theory which is used later in the paper to compute Floer gradings of the generators in $IC_*(\Sigma, \mathcal{L})$. Our equivariant index theory approach is also used to recover the Kronheimer–Mrowka singular index formulas [24, Lemma 2.11] along the lines of Wang's paper [42].

The next five sections are dedicated to the singular knot Floer homology $I^{\natural}(k)$ for knots $k \subset S^3$. Section 4 describes generators in the chain complex $IC^{\natural}(k)$ in terms of equivariant representations $\pi_1 Y \to \mathrm{SO}(3)$ on the double branched cover $Y \to S^3$ with branch set the knot k. These representations fall into three categories: trivial, reducible nontrivial, and irreducible.

The trivial representation $\theta\colon \pi_1 Y \to \mathrm{SO}(3)$ gives rise to a special generator $\alpha \in IC^{\natural}(k)$ which is used in [24] to fix an absolute grading on $I^{\natural}(k)$. This generator is dealt with in Section 5. We pass to the double branched cover and use Taubes index theory [40] on manifolds with periodic ends to show that the Floer grading of α equals $\operatorname{sign} k \pmod 4$.

[2] Extensive calculations for torus knots have recently been done by Anvari [2] using similar equivariant techniques.

Having computed the absolute grading of α, we only need to compute the relative gradings of the remaining generators. We derive formulas for these gradings in Section 6 using equivariant index calculations on double branched covers, and apply these formulas to Montesinos and torus knots in Section 7.

Section 8 contains calculations of $IC_*(\Sigma, \mathcal{L})$ for several two-component links $\mathcal{L}$ not of the form $k^\natural$. For the pretzel link $\mathcal{L} = P(2,-3,-6)$ in the 3–sphere we obtain a complete calculation of the Floer homology groups of $P(2,-3,-6)$ and not just of the Floer chain complex. The same answer is independently confirmed by computing the Floer homology of Harper and Saveliev [19] for this two-component link: the latter theory is isomorphic to $I_*(\Sigma, \mathcal{L})$ but does not use singular connections in its definition.

Finally, Section 8.3 contains proofs of some topological results, which were postponed earlier in the paper for the sake of exposition.

Acknowledgements We are thankful to Ken Baker, Paul Kirk, and Daniel Ruberman for useful discussions. Both authors were partially supported by NSF Grant 1065905.

2 Link homology

In this section, we will sketch the definition of the singular instanton Floer homology $I_*(\Sigma, \mathcal{L})$ of a two-component link $\mathcal{L} \subset \Sigma$ in an integral homology 3–sphere. We will follow Kronheimer and Mrowka [24] closely, deviating in just two respects: we will use the language of projective representations to describe the generators in the Floer chain complex, and will introduce a canonical $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ action on these generators.

2.1 The Chern–Simons functional

Given a two-component link $\mathcal{L}$ in an integral homology sphere Σ, the second homology of its exterior $X = \Sigma - \operatorname{int} N(\mathcal{L})$ is isomorphic to a copy of $\mathbb{Z}$ spanned by either one of the boundary tori of X. Let $P \to X$ be the unique SO(3)–bundle with a nontrivial second Stiefel–Whitney class $w_2(P) \in H^2(X; \mathbb{Z}/2) = \mathbb{Z}/2$. The flat connections in this bundle serve as the starting point for building $I_*(\Sigma, \mathcal{L})$. Since $w_2(P)$ evaluates nontrivially on the boundary tori, these connections are necessarily irreducible and have order-2 holonomy along the meridians of the link components. Therefore, they give rise to flat connections in an orbifold SO(3)–bundle on Σ, which we again call P. The homology sphere Σ itself is viewed as an orbifold with the orbifold singularity $\mathcal{L}$, equipped with a Riemannian metric with cone angle π along the singular set.

Kronheimer and Mrowka [24] interpreted the gauge equivalence classes of the orbifold flat connections in P as the critical points of an orbifold Chern–Simons functional

$$\mathbf{cs}\colon \mathcal{B}(\Sigma, \mathcal{L}) \to \mathbb{R}/\mathbb{Z}, \tag{1}$$

and defined $I_*(\Sigma, \mathcal{L})$ as its Morse homology. An important feature of this construction is the use of the restricted orbifold gauge group $\mathcal{G}_S$ in the definition of the configuration space,

$$\mathcal{B}(\Sigma, \mathcal{L}) = \mathcal{A}(\Sigma, \mathcal{L})/\mathcal{G}_S,$$

where $\mathcal{A}(\Sigma, \mathcal{L})$ is an affine space of orbifold connections and $\mathcal{G}_S$ is the quotient of the determinant-1 orbifold gauge group $\mathcal{G}(\check{P})$ of Kronheimer and Mrowka [24, Section 2.6] by its center $\{\pm 1\}$. The group $\mathcal{G}_S$ is a normal subgroup of the full orbifold gauge group $\mathcal{G}$ with the quotient $\mathcal{G}/\mathcal{G}_S = H^1(X; \mathbb{Z}/2) = \mathbb{Z}/2 \oplus \mathbb{Z}/2$. The full gauge group $\mathcal{G}$ acts on $\mathcal{A}(\Sigma, \mathcal{L})$ preserving the gradient of **cs**, thereby giving rise to the residual action of $H^1(X; \mathbb{Z}/2)$ on the configuration space $\mathcal{B}(\Sigma, \mathcal{L})$ and on the critical point set of the Chern–Simons functional.

We will next describe the critical points of the functional (1) algebraically using the holonomy correspondence between flat connections and representations of the fundamental group. A variant of this classical correspondence which applies to the situation at hand was described in [33, Section 3.2] using projective SU(2) representations. We will review these first; see [33, Section 3.1] for details.

2.2 Projective representations

Let G be a finitely presented group and view the center of SU(2) as $\mathbb{Z}/2 = \{\pm 1\}$. A map $\rho\colon G \to \mathrm{SU}(2)$ is called a projective representation if

$$c(g,h) = \rho(gh)\rho(h)^{-1}\rho(g)^{-1} \in \mathbb{Z}/2 \quad \text{for all } g, h \in G.$$

The function $c\colon G \times G \to \mathbb{Z}/2$ is a 2–cocycle on G defining a cohomology class $[c] \in H^2(G; \mathbb{Z}/2)$. This class has the following interpretation. The composition of $\rho\colon G \to \mathrm{SU}(2)$ with $\mathrm{Ad}\colon \mathrm{SU}(2) \to \mathrm{SO}(3)$ is a representation $\mathrm{Ad}\,\rho\colon G \to \mathrm{SO}(3)$. As such, it induces a continuous map $BG \to B\,\mathrm{SO}(3)$ which is unique up to homotopy. The pullback of the universal Stiefel–Whitney class $w_2 \in H^2(B\,\mathrm{SO}(3); \mathbb{Z}/2)$ via this map is our class $[c] = w_2(\mathrm{Ad}\,\rho) \in H^2(G; \mathbb{Z}/2)$. It serves as an obstruction to lifting $\mathrm{Ad}\,\rho\colon G \to \mathrm{SO}(3)$ to an SU(2) representation.

Let $\mathcal{PR}_c(G; \mathrm{SU}(2))$ be the space of conjugacy classes of projective representations $\rho\colon G \to \mathrm{SU}(2)$ whose associated cocycle is c. The topology on $\mathcal{PR}_c(G; \mathrm{SU}(2))$ is supplied by the algebraic set structure. One can easily see that $\mathcal{PR}_c(G; \mathrm{SU}(2))$ is determined uniquely up to homeomorphism by the cohomology class of c. The group $H^1(G; \mathbb{Z}/2) = \mathrm{Hom}(G, \mathbb{Z}/2)$ acts on $\mathcal{PR}_c(G; \mathrm{SU}(2))$ by sending ρ to $\chi \cdot \rho$ for any $\chi \in \mathrm{Hom}(G, \mathbb{Z}/2)$. The orbits of this action are in a bijective correspondence with the conjugacy classes of representations $G \to \mathrm{SO}(3)$ whose second Stiefel–Whitney class equals $[c]$. The bijection is given by taking the adjoint representation.

Projective representations $\rho\colon G \to \mathrm{SU}(2)$ can also be described in terms of a presentation $G = F/R$. Consider a homomorphism $\gamma\colon R \to \mathbb{Z}/2$ defined by its values $\gamma(r) = \pm 1$ on the relators $r \in R$ and by the condition that it is constant on the orbits of the adjoint action of F on R. Also, choose a set-theoretic section $s\colon G \to F$ in the exact sequence

$$1 \to R \xrightarrow{i} F \xrightarrow{\pi} G \to 1$$

and let $r\colon G \times G \to R$ be the function defined by the formula $s(gh) = r(g,h)s(g)s(h)$.

Proposition 2.1 *A choice of a section $s\colon G \to F$ establishes a bijective correspondence between the conjugacy classes of projective representations $\rho\colon G \to \mathrm{SU}(2)$ with the cocycle $c(g,h) = \gamma(r(g,h))$ and the conjugacy classes of homomorphisms $\sigma\colon F \to \mathrm{SU}(2)$ such that $i^*\sigma = \gamma$. A different choice of s results in a cohomologous cocycle.*

Proof We begin by checking that $c(g,h) = \gamma(r(g,h))$ is a cocycle. For any $g, h, k \in G$, we have

$$\begin{aligned} s(ghk) &= r(gh,k)s(gh)s(k) = r(gh,k)r(g,h)s(g)s(h)s(k), \\ s(ghk) &= r(g,hk)s(g)s(hk) = r(g,hk)s(g)r(h,k)s(h)s(k), \end{aligned}$$

which results in $r(gh,k)r(g,h) = r(g,hk)s(g)r(h,k)s(g)^{-1}$. Since the homomorphism γ is constant on the orbits of the adjoint action of F on R, its application to the above equality gives the cocycle condition $c(gh,k)c(g,h) = c(g,hk)c(h,k)$ as desired.

Now, given a homomorphism $\sigma\colon F \to \mathrm{SU}(2)$ such that $i^*\sigma = \gamma$, define $\rho\colon G \to \mathrm{SU}(2)$ by the formula $\rho(g) = \sigma(s(g))$. Then

$$\begin{aligned} \rho(gh) &= \sigma(s(gh)) = \sigma(r(g,h)s(g)s(h)) \\ &= \gamma(r(g,h))\sigma(s(g))\sigma(s(h)) = c(g,h)\rho(g)\rho(h), \end{aligned}$$

hence ρ is a projective representation with cocycle c. It is clear that conjugate representations σ define conjugate projective representations ρ, and that a different choice of s leads to a cohomologous cocycle c.

The inverse correspondence is defined as follows. Given a projective representation $\rho\colon G \to \mathrm{SU}(2)$, write elements of F in the form $r \cdot s(g)$ with $r \in R$ and $g \in G$, and define $\sigma\colon F \to \mathrm{SU}(2)$ by the formula $\sigma(r \cdot s(g)) = \gamma(r)\rho(g)$. That σ is a homomorphism can be checked by a straightforward calculation using the fact that $c(g,h) = \gamma(r(g,h))$. □

Example 2.2 Let $G = \pi_1 M$ be the fundamental group of a manifold M obtained by 0–surgery on a knot k in an integral homology sphere Σ. The group $\pi_1 M$ is obtained from $\pi_1 K$ by imposing the relation $\lambda = 1$, where λ is a canonical longitude of k. Therefore, $\pi_1 M$ admits a presentation $\pi_1 M = F/R$ with λ being one of the relators. Let $\gamma(\lambda) = -1$ and $\gamma(r) = 1$ for the rest of the relators $r \in R$. It has been known since Floer [16] that the action of $H^1(M;\mathbb{Z}/2) = \mathbb{Z}/2$ on the set of conjugacy classes of projective representations $\sigma\colon F \to \mathrm{SU}(2)$ with $i^*\sigma = \gamma$ is free, providing a two-to-one correspondence between this set and the set of the conjugacy classes of representations $\pi_1 M \to \mathrm{SO}(3)$ with nontrivial $w_2 \in H^2(M;\mathbb{Z}/2) = \mathbb{Z}/2$.

2.3 Holonomy correspondence

We will now apply the general theory of Section 2.2 to the group $G = \pi_1 X$, where X is the exterior of a two-component link $\mathcal{L}$ in an integral homology sphere Σ. We begin with the following simple observation.

Lemma 2.3 *Unless the link $\mathcal{L}$ is split, $H^2(X;\mathbb{Z}/2) = H^2(\pi_1 X;\mathbb{Z}/2) = \mathbb{Z}/2$. For split links, $I_*(\Sigma,\mathcal{L}) = 0$.*

Proof For a split link $\mathcal{L}$, the splitting sphere generates the group $H_2(X;\mathbb{Z}) = \mathbb{Z}$. Since there are no flat connections on this sphere with nontrivial $w_2(P)$, the group $I_*(\Sigma,\mathcal{L})$ must vanish. For a nonsplit link, the claimed equality follows from the Hopf exact sequence

$$\pi_2(X) \to H_2(X) \to H_2(\pi_1 X) \to 0$$

and the vanishing of the Hurewicz homomorphism $\pi_2(X) \to H_2(X)$. □

From now on, we will assume that the link $\mathcal{L} \subset \Sigma$ is not split. The holonomy correspondence of [33, Section 3.1] identifies the critical point set of the functional (1) with the set $\mathcal{PR}_c(X,\mathrm{SU}(2))$ of conjugacy classes of projective representations $\rho\colon \pi_1 X \to \mathrm{SU}(2)$, for any choice of cocycle c such that $0 \neq [c] = w_2(P) \in H^2(X;\mathbb{Z}/2) = \mathbb{Z}/2$. Note that this identification commutes with the $H^1(X;\mathbb{Z}/2)$ action, and that the orbits of this action on $\mathcal{PR}_c(X,\mathrm{SU}(2))$ are in bijective correspondence with the conjugacy classes of representations $\mathrm{Ad}\,\rho\colon \pi_1 X \to \mathrm{SO}(3)$ having $w_2(\mathrm{Ad}\,\rho) \neq 0$.

Lemma 2.4 *Any representation $\mathrm{Ad}\,\rho\colon \pi_1 X \to \mathrm{SO}(3)$ with $w_2(\mathrm{Ad}\,\rho) \neq 0$ is irreducible, that is, its image is not contained in a copy of $\mathrm{SO}(2) \subset \mathrm{SO}(3)$.*

Proof The restriction of ρ to either boundary torus of X has nontrivial second Stiefel–Whitney class, which implies that it does not lift to an $\mathrm{SU}(2)$ representation. However, any reducible representation $\pi_1 T^2 \to \mathrm{SO}(3)$ admits an $\mathrm{SU}(2)$ lift, therefore, the image of ρ cannot be contained in a copy of $\mathrm{SO}(2) \subset \mathrm{SO}(3)$. It is essential here that $H_1(T^2)$ has no 2–torsion: a nontrivial $\mathrm{SO}(3)$ representation of $\mathbb{Z}/2$ is reducible but does not admit an $\mathrm{SU}(2)$ lift. □

2.4 Floer gradings

Given flat orbifold connections ρ and σ in the orbifold bundle $P \to \Sigma$, consider an arbitrary orbifold connection A in the pullback bundle on the product $\mathbb{R} \times \Sigma$ matching ρ and σ near the negative and positive ends, respectively. Equip $\mathbb{R} \times \Sigma$ with the orbifold product metric and consider the ASD operator

$$\mathcal{D}_A(\rho,\sigma) = -d_A^* \oplus d_A^+ \colon \Omega^1(\mathbb{R} \times \Sigma, \operatorname{ad} P) \to (\Omega^0 \oplus \Omega^2_+)(\mathbb{R} \times \Sigma, \operatorname{ad} P) \tag{2}$$

completed in the orbifold Sobolev L^2 norms as in [24, Section 3.1]. Since ρ and σ are irreducible, this operator will be Fredholm if we further assume that ρ and σ are nondegenerate as the critical points of the Chern–Simons functional (1). Define the relative Floer grading as

$$\operatorname{gr}(\rho,\sigma) = \operatorname{ind} \mathcal{D}_A(\rho,\sigma) \pmod 4. \tag{3}$$

This grading is well defined because replacing either ρ or σ by its gauge equivalent within the restricted gauge group $\mathcal{G}_S$ results in adding a multiple of four to the index of $\mathcal{D}_A$, see [24, Section 2.5]. This is no longer true if we use the full gauge group. The following lemma makes it precise; it will be proved in Section 3.7.

Lemma 2.5 *Let χ_1 and χ_2 be the generators of $H^1(X;\mathbb{Z}/2) = \mathbb{Z}/2 \oplus \mathbb{Z}/2$ dual to the meridians of the link $\mathcal{L} = \ell_1 \cup \ell_2$. Then*

$$\operatorname{gr}(\chi_1 \cdot \rho, \sigma) = \operatorname{gr}(\chi_2 \cdot \rho, \sigma) = \operatorname{gr}(\rho,\sigma) + 2 \cdot \delta \pmod 4,$$

and similarly for the action on σ, where

$$\delta = \begin{cases} 0 & \text{if } \ell k(\ell_1,\ell_2) \text{ is odd,} \\ 1 & \text{if } \ell k(\ell_1,\ell_2) \text{ is even.} \end{cases}$$

2.5 Perturbations

The critical points of the Chern–Simons functional need not be nondegenerate, therefore we may have to perturb it to define $I_*(\Sigma,\mathcal{L})$. The perturbations used in [24, Section 3.4] are the standard Wilson loop perturbations along loops in Σ disjoint from the link $\mathcal{L}$. There are sufficiently many such perturbations to guarantee the nondegeneracy of the critical points of the perturbed Chern–Simons functional as well as the transversality properties for the moduli spaces of trajectories of its gradient flow. This allows us to define the boundary operator and to complete the definition of $I_*(\Sigma, \mathcal{L})$.

3 Equivariant gauge theory

In this section, we survey some equivariant gauge theory on the double branched cover $M \to \Sigma$ of a homology sphere Σ with branch set a two-component link $\mathcal{L}$. It will be used in the forthcoming sections to make headway in computing the link homology $I_*(\Sigma, \mathcal{L})$.

3.1 Topological preliminaries

Let Σ be an integral homology 3–sphere and $\mathcal{L} = \ell_1 \cup \ell_2$ a link of two components in Σ. The link exterior $X = \Sigma - \operatorname{int} N(\mathcal{L})$ is a manifold whose boundary consists of two tori, with $H_1(X;\mathbb{Z}) = \mathbb{Z}^2$ spanned by the meridians μ_1 and μ_2 of the link components. The homomorphism $\pi_1 X \to \mathbb{Z}/2$ sending μ_1 and μ_2 to the generator of $\mathbb{Z}/2$ gives rise to a regular double cover $\widetilde{X} \to X$, and also to a double branched cover $\pi\colon M \to \Sigma$ with branching set $\mathcal{L}$ and the covering translation $\tau\colon M \to M$. Denote by $\Delta(t)$ the one-variable Alexander polynomial of $\mathcal{L}$.

Proposition 3.1 *The first Betti number of M is 1 if $\Delta(-1) = 0$ and 0 otherwise. In the latter case, $H_1(M;\mathbb{Z})$ is a finite group of order $|\Delta(-1)|$. The induced involution $\tau_*\colon H_1(M) \to H_1(M)$ is multiplication by -1.*

Proof This is essentially proved in Kawauchi [21, Section 5.5]. The statement about τ_* follows from an isomorphism of $\mathbb{Z}[t, t^{-1}]$ modules $H_1(M) = H_1(E)/(1+t)H_1(E)$, where E is the infinite cyclic cover of X, established in [21, Theorem 5.5.1]. A completely different proof for the special case of double branched covers of S^3 with branch set a knot can be found in Ruberman [31, Lemma 5.5]. □

Proposition 3.2 *Let M be the double branched cover of an integral homology sphere with branch set a two-component link. Then $H_i(M;\mathbb{Z}/2) = H^i(M;\mathbb{Z}/2)$ is isomorphic to $\mathbb{Z}/2$ if $i = 0, 1, 2, 3$, and is zero otherwise. The cup product*

$$H^1(M;\mathbb{Z}/2) \times H^1(M;\mathbb{Z}/2) \to H^2(M;\mathbb{Z}/2)$$

is given by the linking number $\ell k(\ell_1, \ell_2) \pmod 2$.

The proof of Proposition 3.2 will be postponed until Section 8.3 for the sake of exposition.

An important example of $\mathcal{L}$ to consider is the two-component link $k^\natural$ obtained as the connected sum of a knot $k \subset S^3$ with the Hopf link. The double branched cover $M \to S^3$ in this case is the connected sum $M = Y \# \mathbb{R}\mathrm{P}^3$, where Y is the double branched cover of k. Proposition 3.2 easily follows because $H_*(Y;\mathbb{Z}/2) = H_*(S^3;\mathbb{Z}/2)$.

3.2 The orbifold exact sequence

We will view $\Sigma = M/\tau$ as an orbifold with the singular set $\mathcal{L}$. To be precise, the regular double cover $\widetilde{X} \to X$ is a 3–manifold whose boundary consists of two tori, and

$$M = \widetilde{X} \cup_h N(\mathcal{L}),$$

where the gluing homeomorphism $h\colon \partial\widetilde{X} \to \partial N(\mathcal{L})$ identifies $\pi^{-1}(\mu_i)$ with the meridian μ_i for $i = 1, 2$. The involution $\tau\colon M \to M$ acts by meridional rotation on $N(\mathcal{L})$, thereby fixing the link $\mathcal{L}$, and by covering translation on $\widetilde{X}$. Define the orbifold fundamental group

$$\pi_1^V(\Sigma,\mathcal{L}) = \pi_1 X/\langle {\mu_1}^2 = {\mu_2}^2 = 1\rangle.$$

Then the homotopy exact sequence of the covering $\widetilde{X} \to X$ gives rise to a split short exact sequence, called the orbifold exact sequence,

$$1 \to \pi_1 M \xrightarrow{\pi_*} \pi_1^V(\Sigma,\mathcal{L}) \xrightarrow{j} \mathbb{Z}/2 \to 1. \tag{4}$$

The homomorphism j maps the meridians μ_1, μ_2 to the generator of $\mathbb{Z}/2$ and one obtains a splitting by sending this generator to either μ_1 or μ_2.

It follows from the definition of the orbifold fundamental group $\pi_1^V(\Sigma,\mathcal{L})$ that its abelianization is given by

$$H_1(X)/\langle {\mu_1}^2 = {\mu_2}^2 = 1\rangle = H_1(X;\mathbb{Z}/2) = \mathbb{Z}/2 \oplus \mathbb{Z}/2,$$

with the canonical generators μ_1 and μ_2. The homomorphism π_* of the orbifold exact sequence (4) then induces a map $\pi_*\colon H_1(M;\mathbb{Z}/2) \to H_1(X;\mathbb{Z}/2)$, which can be described as follows.

Lemma 3.3 *The homomorphism $\pi_*\colon H_1(M;\mathbb{Z}/2) \to H_1(X;\mathbb{Z}/2)$ sends the generator of $H_1(M;\mathbb{Z}/2) = \mathbb{Z}/2$ to the sum of the meridians $\mu_1 + \mu_2 \in H_1(X;\mathbb{Z}/2)$.*

Proof That $H_1(M;\mathbb{Z}/2) = \mathbb{Z}/2$ follows from Proposition 3.2. An explicit generator of this group is described in the proof of Proposition A.2 as the circle $\pi^{-1}(w)$, where w is an embedded arc in Σ with endpoints on the two different components of $\mathcal{L}$. The commutative diagram

$$\begin{array}{ccc} \pi_1 \widetilde{X} & \xrightarrow{\pi_*} & \pi_1 X \\ \downarrow & & \downarrow \\ \pi_1 M & \xrightarrow{\pi_*} & \pi_1^V(\Sigma, \mathcal{L}) \end{array}$$

gives rise to the commutative diagram in homology

$$\begin{array}{ccc} H_1(\widetilde{X}; \mathbb{Z}/2) & \xrightarrow{\pi_*} & H_1(X; \mathbb{Z}/2) \\ \downarrow & & \downarrow \\ H_1(M; \mathbb{Z}/2) & \xrightarrow{\pi_*} & H_1(X; \mathbb{Z}/2) \end{array}$$

The cycle $\pi^{-1}(w)$ in M is homologous to a cycle in $\widetilde{X}$ which consists of the two arcs $\pi^{-1}(w) \cap \widetilde{X}$ whose endpoints on each of the tori in $\partial\widetilde{X}$ are connected by an arc. The map $\pi_*\colon H_1(\widetilde{X}; \mathbb{Z}/2) \to H_1(X; \mathbb{Z}/2)$ takes the homology class of this cycle to $\mu_1 + \mu_2$, and the result follows. □

3.3 Pulled-back bundles

Let $P \to \Sigma$ be the orbifold SO(3)–bundle used in the definition of $I_*(\Sigma, \mathcal{L})$ in Section 2. It pulls back to an orbifold SO(3)–bundle $Q \to M$ because the projection map $\pi\colon M \to \Sigma$ is regular in the sense of Chen and Ruan [10]. The bundle Q is in fact smooth because orbifold connections on P with order-2 holonomy along the meridians of $\mathcal{L}$ lift to connections in Q with trivial holonomy along the meridians of the two-component link $\widetilde{\mathcal{L}} = \pi^{-1}(\mathcal{L})$.

Proposition 3.4 *The bundle $Q \to M$ is nontrivial.*

The rest of this section is dedicated to the proof of this proposition. We will accomplish it by showing that $w_2(Q) \in H^2(M; \mathbb{Z}/2) = \mathbb{Z}/2$ is nonvanishing. Our argument will split into two cases, corresponding to the parity of the linking number between the components of $\mathcal{L}$.

Suppose that $\ell k(\ell_1, \ell_2)$ is even and consider the regular double cover $\pi\colon M - \widetilde{\mathcal{L}} \to \Sigma - \mathcal{L}$. It gives rise to the Gysin exact sequence

$$\cdots \to H^1(\Sigma - \mathcal{L}; \mathbb{Z}/2) \xrightarrow{\cup w_1} H^2(\Sigma - \mathcal{L}; \mathbb{Z}/2) \xrightarrow{\pi^*} H^2(M - \widetilde{\mathcal{L}}; \mathbb{Z}/2)$$
$$\longrightarrow H^2(\Sigma - \mathcal{L}; \mathbb{Z}/2) \xrightarrow{\cup w_1} H^3(\Sigma - \mathcal{L}; \mathbb{Z}/2) \longrightarrow \cdots$$

where $\cup w_1$ means taking the cup product with the first Stiefel–Whitney class of the cover. The cup product on $H^*(\Sigma - \mathcal{L}; \mathbb{Z}/2)$ can be determined from the following commutative diagram:

$$\begin{array}{ccc} H_2(\Sigma,\mathcal{L};\mathbb{Z}/2)\times H_2(\Sigma,\mathcal{L};\mathbb{Z}/2) & \xrightarrow{\ \cdot\ } & H_1(\Sigma,\mathcal{L};\mathbb{Z}/2) \\ \uparrow \text{PD} & & \uparrow \text{PD} \\ H^1(\Sigma-\mathcal{L};\mathbb{Z}/2)\times H^1(\Sigma-\mathcal{L};\mathbb{Z}/2) & \xrightarrow{\ \cup\ } & H^2(\Sigma-\mathcal{L};\mathbb{Z}/2) \end{array}$$

where PD stands for the Poincaré duality isomorphism and the dot in the upper row for the intersection product. Note that Seifert surfaces of knots ℓ_1 and ℓ_2 generate $H_2(\Sigma,\mathcal{L};\mathbb{Z}/2)=\mathbb{Z}/2\oplus\mathbb{Z}/2$, and any arc in Σ with one endpoint on ℓ_1 and the other on ℓ_2 generates $H_1(\Sigma,\mathcal{L};\mathbb{Z}/2)=\mathbb{Z}/2$. An easy calculation shows that, with respect to these generators, the intersection product is given by the matrix

$$\begin{pmatrix} 0 & \ell k(\ell_1,\ell_2) \\ \ell k(\ell_1,\ell_2) & 0 \end{pmatrix}.$$

Since $\ell k(\ell_1,\ell_2)$ is even, this gives a trivial cup product structure on the link complement $\Sigma-\mathcal{L}$. Therefore, the map $\cup w_1$ in the Gysin sequence is zero and the map $\pi^*\colon H^2(\Sigma-\mathcal{L};\mathbb{Z}/2)\to H^2(M-\widetilde{\mathcal{L}};\mathbb{Z}/2)$ is injective. Since $w_2(P)\in H^2(\Sigma-\mathcal{L};\mathbb{Z}/2)$ is nonzero we conclude that $\pi^*(w_2(P))\neq 0$. This implies that $w_2(Q)\neq 0$ because $Q=\pi^*P$ over $M-\widetilde{\mathcal{L}}$.

Now suppose that $\ell k(\ell_1,\ell_2)$ is odd. The above calculation implies that the second Stiefel–Whitney class of π^*P vanishes in $H^2(M-\widetilde{\mathcal{L}};\mathbb{Z}/2)$. We will prove, however, that $w_2(Q)\in H^2(M;\mathbb{Z}/2)$ is nonzero, by showing that Q carries a flat connection with nonzero w_2.

Note that the orbifold bundle P carries a flat SO(3) connection whose holonomy is a representation $\alpha\colon \pi_1^V(\Sigma,\mathcal{L})\to \mathrm{SO}(3)$ of the orbifold fundamental group $\pi_1^V(\Sigma,\mathcal{L})=\pi_1X/\langle {\mu_1}^2={\mu_2}^2=1\rangle$ sending the two meridians to $\mathrm{Ad}\,i$ and $\mathrm{Ad}\,j$. This flat connection pulls back to a flat connection on Q with holonomy $\pi^*\alpha\colon \pi_1M\to\mathrm{SO}(3)$. We wish to compute the second Stiefel–Whitney class of $\pi^*\alpha$.

Lemma 3.5 *The representation $\pi^*\alpha\colon \pi_1M\to\mathbb{Z}/2\oplus\mathbb{Z}/2$ is nontrivial.*

Proof Our proof will rely on the orbifold exact sequence (4). Assume that $\pi^*\alpha$ is trivial. Then $\pi_1M\subset\ker(\pi^*\alpha)$, hence α factors through $\pi_1^V(\Sigma,\mathcal{L})/\pi_*(\pi_1M)\to\mathbb{Z}/2\oplus\mathbb{Z}/2$. Since $\pi_1^V(\Sigma,\mathcal{L})/\pi_*(\pi_1M)=\mathbb{Z}/2$, we obtain a contradiction with the surjectivity of α. □

Since the group $\mathbb{Z}/2\oplus\mathbb{Z}/2$ is abelian, the representation $\pi^*\alpha\colon\pi_1M\to\mathbb{Z}/2\oplus\mathbb{Z}/2$ factors through a homomorphism $H_1(M)\to\mathbb{Z}/2\oplus\mathbb{Z}/2$ which is uniquely determined by its two components $\xi,\ \eta\in\mathrm{Hom}(H_1(M),\mathbb{Z}/2)=H^1(M;\mathbb{Z}/2)=\mathbb{Z}/2$;

see Proposition 3.2. A calculation identical to that in [33, Proposition 4.3] shows that $w_2(\pi^*\alpha) = \xi \cup \xi + \xi \cup \eta + \eta \cup \eta$ (note that, unlike in [33], the classes $\xi \cup \xi$ and $\eta \cup \eta$ need not vanish). Since ξ and η cannot both be trivial by Lemma 3.5, we may assume without loss of generality that $\xi \neq 0$. If $\eta = 0$ then $w_2(\pi^*\alpha) = \xi \cup \xi$. If $\eta \neq 0$ then $\xi = \eta$ due to the fact that $H^1(M;\mathbb{Z}/2) = \mathbb{Z}/2$, and therefore again $w_2(\pi^*\alpha) = \xi \cup \xi$. Since $\ell k(\ell_1, \ell_2)$ is odd, it follows from Proposition 3.2 that $w_2(\pi^*\alpha) \neq 0$.

3.4 Pulled-back representations

Assuming that $\mathcal{L} \subset \Sigma$ is nonsplit, we identified in Section 2.3 the critical point set of the Chern–Simons functional (1) with the space $\mathcal{PR}_c(X, \mathrm{SU}(2))$ of the conjugacy classes of projective representations $\pi_1 X \to \mathrm{SU}(2)$ on the link exterior, for any choice of cocycle c not cohomologous to zero. We further identified the quotient of $\mathcal{PR}_c(X, \mathrm{SU}(2))$ by the natural $H^1(X;\mathbb{Z}/2)$ action with the subspace $\mathcal{R}_w(X;\mathrm{SO}(3))$ of the $\mathrm{SO}(3)$ character variety of $\pi_1 X$ cut out by the condition $w_2 \neq 0$. The latter condition implies that both meridians μ_1 and μ_2 are represented by $\mathrm{SO}(3)$ matrices of order 2, which leads to a natural identification of this subspace with

$$\mathcal{R}_\omega(\Sigma, \mathcal{L}; \mathrm{SO}(3)) = \{\rho\colon \pi_1^V(\Sigma, \mathcal{L}) \to \mathrm{SO}(3) \mid w_2(\rho) \neq 0\}/\operatorname{Ad}\mathrm{SO}(3),$$

where the condition $w_2(\rho) \neq 0$ applies to the representation ρ restricted to X. To summarize, the group $H^1(X;\mathbb{Z}/2)$ acts on $\mathcal{PR}_c(X, \mathrm{SU}(2))$ with the quotient map

$$\mathcal{PR}_c(X, \mathrm{SU}(2)) \to \mathcal{R}_\omega(\Sigma, \mathcal{L}; \mathrm{SO}(3)).$$

We now wish to study $\mathcal{R}_\omega(\Sigma, \mathcal{L}; \mathrm{SO}(3))$ using representations on the double branched cover $M \to \Sigma$ equivariant with respect to the covering translation $\tau\colon M \to M$.

Lemma 3.6 *Let $\rho\colon \pi_1^V(\Sigma, \mathcal{L}) \to \mathrm{SO}(3)$ be a representation with $w_2(\rho) \neq 0$ and $\pi^*\rho\colon \pi_1 M \to \mathrm{SO}(3)$ its pullback via the homomorphism π_* of the orbifold exact sequence* (4). *Then there exists an element $u \in \mathrm{SO}(3)$ of order 2 such that $\tau^*(\pi^*\rho) = u \cdot (\pi^*\rho) \cdot u^{-1}$.*

Proof Let $\widetilde{X} \to X$ be the regular double cover as in Section 3.2. Choose a basepoint b in one of the boundary tori of $\widetilde{X}$ and consider the commutative diagram

$$\begin{array}{ccccc}
\pi_1(\widetilde{X}, b) & \xrightarrow{\ \tau_*\ } & \pi_1(\widetilde{X}, \tau(b)) & \xrightarrow{\ \psi_f\ } & \pi_1(\widetilde{X}, b) \\
{\scriptstyle \pi_*}\searrow & & \swarrow{\scriptstyle \pi_*} & & \downarrow{\scriptstyle \pi_*} \\
& \pi_1(X, \pi(b)) & \xrightarrow{\qquad \varphi \qquad} & & \pi_1(X, \pi(b))
\end{array}$$

whose maps ψ_f and φ are defined as follows. Given a path $f\colon [0,1] \to X$ from b to $\tau(b)$, take its inverse $\overline{f}(s) = f(1-s)$ and define the map ψ_f by the formula $\psi_f(\beta) = f \cdot \beta \cdot \overline{f}$. Since $\pi(b) = \pi(\tau(b))$, the path f projects to a loop in X based at $\pi(b)$, and the map φ is the conjugation by that loop. In fact, one can choose the path f to project onto the meridian μ_i of the boundary torus on which $\pi(b)$ lies so that $\varphi(x) = \mu_i \cdot x \cdot \mu_i^{-1}$. After filling in the solid tori, we obtain the commutative diagram

$$\begin{array}{ccc} \pi_1 M & \xrightarrow{\ \tau_*\ } & \pi_1 M \\ {\scriptstyle \pi_*}\downarrow & & \downarrow{\scriptstyle \pi_*} \\ \pi_1^V(\Sigma, \mathcal{L}) & \xrightarrow{\ \varphi\ } & \pi_1^V(\Sigma, \mathcal{L}) \end{array}$$

which tells us that, for any $\rho\colon \pi_1^V(\Sigma, \mathcal{L}) \to \mathrm{SO}(3)$, the pullback representation $\pi^*\rho$ has the property that $\tau^*(\pi^*\rho) = u \cdot (\pi^*\rho) \cdot u^{-1}$ with $u = \rho(\mu_i)$ of order 2. □

Example 3.7 Let $\mathcal{L} \subset S^3$ be the Hopf link. Then $M = \mathbb{RP}^3$ and the orbifold exact sequence (4) takes the form

$$1 \to \mathbb{Z}/2 \xrightarrow{\pi_*} \mathbb{Z}/2 \oplus \mathbb{Z}/2 \xrightarrow{j} \mathbb{Z}/2 \to 1$$

with the two copies of $\mathbb{Z}/2$ in the middle group generated by the meridians μ_1 and μ_2. Define $\rho\colon \mathbb{Z}/2 \oplus \mathbb{Z}/2 \to \mathrm{SO}(3)$ on the generators by $\rho(\mu_1) = \mathrm{Ad}\, i$ and $\rho(\mu_2) = \mathrm{Ad}\, j$; up to conjugation, this is the only representation $\mathbb{Z}/2 \to \mathrm{SO}(3)$ with $w_2(\rho) \neq 0$. The pullback representation $\pi^*\rho\colon \mathbb{Z}/2 \to \mathrm{SO}(3)$ sends the generator to $\mathrm{Ad}\, i \cdot \mathrm{Ad}\, j = \mathrm{Ad}\, k$. Since $\tau^*(\pi^*\rho) = \pi^*\rho$, the identity $\tau^*(\pi^*\rho) = u \cdot (\pi^*\rho) \cdot u^{-1}$ holds for multiple choices of u, including the second-order u of the form $u = \mathrm{Ad}\, q$, where q is any unit quaternion such that $-qk = kq$.

Given a double branched cover $\pi\colon M \to \Sigma$ with branch set $\mathcal{L}$ and the covering translation $\tau\colon M \to M$, define

$$\mathcal{R}_\omega(M; \mathrm{SO}(3)) = \{\beta\colon \pi_1 M \to \mathrm{SO}(3) \mid w_2(\beta) \neq 0\} / \mathrm{Ad}\, \mathrm{SO}(3).$$

Since $w_2(\tau^*\beta) = w_2(\beta) \in H^2(M; \mathbb{Z}/2) = \mathbb{Z}/2$, the pullback of representations via τ gives rise to a well defined involution

$$\tau^*\colon \mathcal{R}_\omega(M; \mathrm{SO}(3)) \to \mathcal{R}_\omega(M; \mathrm{SO}(3)). \tag{5}$$

Its fixed point set $\mathrm{Fix}(\tau^*)$ consists of those conjugacy classes of representations $\beta\colon \pi_1 M \to \mathrm{SO}(3)$ such that $w_2(\beta) \neq 0$ and there exists an element $u \in \mathrm{SO}(3)$ having the property that $\tau^*\beta = u \cdot \beta \cdot u^{-1}$. Consider the subvariety

$$\mathcal{R}^\tau_w(M; \mathrm{SO}(3)) \subset \mathrm{Fix}(\tau^*) \tag{6}$$

defined by the condition that the conjugating element u can be chosen to be of order 2. This subvariety is well defined because all elements of order 2 in $\mathrm{SO}(3)$ are conjugate to each other. The following proposition is the main result of this section.

Proposition 3.8 *The homomorphism $\pi_*\colon \pi_1 M \to \pi_1^V(\Sigma,\mathcal{L})$ of the orbifold exact sequence* (4) *induces via the pullback a homeomorphism*

$$\pi^*\colon \mathcal{R}_\omega(\Sigma,\mathcal{L};\mathrm{SO}(3)) \to \mathcal{R}_\omega^\tau(M;\mathrm{SO}(3)).$$

Proof Orbifold representations $\pi_1^V(\Sigma,\mathcal{L}) \to \mathrm{SO}(3)$ with nontrivial w_2 pull back to representations $\pi_1 M \to \mathrm{SO}(3)$ with nontrivial w_2; see Section 3.3. In addition, these pullback representations are equivariant in the sense of Lemma 3.6. Therefore, the map

$$\pi^*\colon \mathcal{R}_\omega(\Sigma,\mathcal{L};\mathrm{SO}(3)) \to \mathcal{R}_\omega^\tau(M;\mathrm{SO}(3))$$

is well defined. To finish the proof, we will construct an inverse of π^*. Given $\beta\colon \pi_1 M \to \mathrm{SO}(3)$ whose conjugacy class belongs to $\mathcal{R}_\omega^\tau(M;\mathrm{SO}(3))$, there exists an element $u\in\mathrm{SO}(3)$ of order 2 such that $\tau^*\beta = u\cdot\beta\cdot u^{-1}$. The pair (β,u) then defines an $\mathrm{SO}(3)$ representation of $\pi_1^V(\Sigma,\mathcal{L}) = \pi_1 M \rtimes \mathbb{Z}/2$ by the formula $\rho(x,t^\ell) = \beta(x)\cdot u^\ell$, where $x\in\pi_1 M$ and t is the generator of $\mathbb{Z}/2$. □

3.5 Equivariant index

All orbifolds we encounter in this paper are obtained by taking the quotient of a smooth manifold by an orientation-preserving involution. The orbifold elliptic theory on such global quotient orbifolds is equivalent to the equivariant elliptic theory on their branched covers; see for instance [42]. In particular, the orbifold index of the ASD operator (2) can be computed as an equivariant index as explained below.

Let X be a smooth oriented Riemannian 4–manifold without boundary, which may or may not be compact. If X is not compact, we assume that its only noncompactness comes from a product end $(0,\infty)\times Y$ equipped with a product metric. Let $\tau\colon X\to X$ be a smooth orientation-preserving isometry of order 2 with nonempty fixed point set F making X into a double branched cover over X' with branch set F'. Let $P\to X$ be an $\mathrm{SO}(3)$–bundle to which τ lifts so that its action on the fibers over the fixed point set of τ has order 2. This lift will be denoted by $\tilde\tau\colon P\to P$. The quotient of P by the involution $\tilde\tau$ is naturally an orbifold $\mathrm{SO}(3)$–bundle $P'\to X'$, and any equivariant connection A in P gives rise to an orbifold connection A' in P'. The ASD operator

$$\mathcal{D}_A(X) = -d_A^*\oplus d_A^+\colon \Omega^1(X,\operatorname{ad}P)\to(\Omega^0\oplus\Omega^2_+)(X,\operatorname{ad}P)$$

associated with A is equivariant in that the diagram

$$\begin{array}{ccc} \Omega^1(X, \operatorname{ad} P) & \xrightarrow{\mathcal{D}_A(X)} & (\Omega^0 \oplus \Omega^2_+)(X, \operatorname{ad} P) \\ \tilde{\tau}^* \big\downarrow & & \big\downarrow \tilde{\tau}^* \\ \Omega^1(X, \operatorname{ad} P) & \xrightarrow{\mathcal{D}_A(X)} & (\Omega^0 \oplus \Omega^2_+)(X, \operatorname{ad} P) \end{array}$$

commutes, giving rise to the orbifold operator

$$\mathcal{D}_{A'}(X')\colon \Omega^1(X', \operatorname{ad} P') \to (\Omega^0 \oplus \Omega^2_+)(X', \operatorname{ad} P').$$

From this we immediately conclude that

$$\operatorname{ind} \mathcal{D}_{A'}(X') = \operatorname{ind} \mathcal{D}^{\tau}_A(X), \tag{7}$$

where $\mathcal{D}^{\tau}_A(X)$ is the operator $\mathcal{D}_A(X)$ restricted to the $(+1)$–eigenspaces of the involution $\tilde{\tau}^*$. If X is closed, the operators in (7) are automatically Fredholm. If X has a product end, we ensure Fredholmness by completing with respect to the weighted Sobolev norms

$$\|\varphi\|_{L^2_{k,\delta}(X)} = \|h \cdot \varphi\|_{L^2_k(X)},$$

where $h\colon X \to \mathbb{R}$ is a smooth function which is τ–invariant and which, over the end, takes the form $h(t, y) = e^{\delta t}$ for a sufficiently small positive δ. We choose to work with these particular norms to match the global boundary conditions of Atiyah, Patodi and Singer [4].

In particular, if ρ and σ are nondegenerate critical points of the orbifold Chern–Simons functional on Σ, they pull back to the flat connections $\pi^*\rho$ and $\pi^*\sigma$ on the double branched cover $M \to \Sigma$. The formula (3) for the relative Floer grading can then be written as

$$\operatorname{gr}(\rho, \sigma) = \operatorname{ind} \mathcal{D}^{\tau}_A(\pi^*\rho, \pi^*\sigma) \pmod 4,$$

where A is an equivariant connection on $\mathbb{R} \times Y$ whose limits at the negative and positive ends are $\pi^*\rho$ and $\pi^*\sigma$, respectively. The index in the above formula can be understood as the L^2_δ index for any sufficiently small $\delta \geq 0$ because the operator $\mathcal{D}^{\tau}_A(\pi^*\rho, \pi^*\sigma)$ is Fredholm in the usual L^2 Sobolev completion.

3.6 Index formulas

Let us continue with the setup of the previous subsection. One can easily see that

$$\operatorname{ind} \mathcal{D}^{\tau}_A(X) = \tfrac{1}{2} \operatorname{ind} \mathcal{D}_A(X) + \tfrac{1}{2} \operatorname{ind}(\tau, \mathcal{D}_A)(X),$$

where

$$\operatorname{ind}(\tau, \mathcal{D}_A)(X) = \operatorname{tr}(\tilde{\tau}^* \,|\, \ker \mathcal{D}_A(X)) - \operatorname{tr}(\tilde{\tau}^* \,|\, \operatorname{coker} \mathcal{D}_A(X)).$$

We will use this observation together with the standard index theorems to obtain explicit formulas for the index of the operators in question.

Proposition 3.9 *Let X be a closed manifold. Then*

$$\operatorname{ind}\mathcal{D}^{\tau}_{A}(X) = -p_1(P) - \tfrac{3}{4}(\sigma(X)+\chi(X)) + \tfrac{1}{4}(\chi(F)+F\cdot F).$$

Proof The index of $\mathcal{D}_A(X)$ can be expressed topologically using the Atiyah–Singer index theorem [6]. Since the operator $\mathcal{D}_A$ has the same symbol as the positive chiral Dirac operator twisted by $S^+ \otimes (\operatorname{ad} P)_{\mathbb{C}}$ (see [3]), we obtain

$$\begin{aligned}\operatorname{ind}\mathcal{D}_A(X) &= \int_X \widehat{A}(X)\operatorname{ch}(S^+)\operatorname{ch}(\operatorname{ad} P)_{\mathbb{C}}\\ &= \int_X -2p_1(A) - \tfrac{1}{2}p_1(TX) - \tfrac{3}{2}e(TX)\\ &= -2p_1(P) - \tfrac{3}{2}(\sigma(X)+\chi(X)),\end{aligned}$$

using the Hirzebruch signature theorem in the last line. A similar expression for $\operatorname{ind}(\tau,\mathcal{D}_A)(X)$ is obtained using the G–index theorem of Atiyah and Singer [6]. For the twisted Dirac operator in question, an explicit calculation in Shanahan [39, Section 19] leads us to the formula

$$\operatorname{ind}(\tau,\mathcal{D}_A)(X) = -\frac{1}{2}\int_F (e(TF)+e(NF))\operatorname{ch}_g(\operatorname{ad} P)_{\mathbb{C}} = \tfrac{1}{2}(\chi(F)+F\cdot F).$$

Here TF and NF are the tangent and the normal bundle of the fixed point set $F \subset X$, and the zero-order term in $\operatorname{ch}_g(\operatorname{ad} P)_{\mathbb{C}}$ equals -1 because this is the trace of the second-order SO(3) operator acting on the fiber. Adding these formulas together, we obtain the desired formula. □

Remark 3.10 Our formula matches the formulas for $\operatorname{ind}\mathcal{D}_{A'}(X')$ of Kronheimer and Mrowka [24, Lemma 2.11] and Wang [42, Theorem 18],

$$\operatorname{ind}\mathcal{D}_{A'}(X') = -p_1(P) - \tfrac{3}{2}(\sigma(X')+\chi(X')) + \chi(F') + \tfrac{1}{2}F'\cdot F',$$

after taking into account that $F'\cdot F' = 2(F\cdot F)$, $\chi(F)=\chi(F')$, $2\chi(X')=\chi(X)+\chi(F)$, and $2\sigma(X') = \sigma(X) + F\cdot F$; see for instance Viro [41].

Next, let X be a manifold with a product end $(0,\infty)\times Y$, where Y need not be connected, and work with the L^2_δ norms for sufficiently small $\delta > 0$. In a temporal gauge over the end, the operator $\mathcal{D}_A(X)$ takes the form $\mathcal{D}_A(X) = \partial/\partial t + K_{A(t)}$.

Proposition 3.11 *Let X be a manifold with product end as above, and A an equivariant connection whose limit over the end is a flat connection β. Then*

$$\operatorname{ind}\mathcal{D}_A^\tau(X) = \frac{1}{2}\int_X \hat{A}(X)\operatorname{ch}(S^+)\operatorname{ch}(\operatorname{ad}P)_{\mathbb{C}} + \tfrac{1}{4}(\chi(F) + F\cdot F) - \tfrac{1}{4}(h_\beta - \eta_\beta(0)) - \tfrac{1}{4}(h_\beta^\tau - \eta_\beta^\tau(0)).$$

The notation here is as follows:

- h_β is the dimension of $H^0(Y;\operatorname{ad}\beta)\oplus H^1(Y;\operatorname{ad}\beta)$;
- h_β^τ is the trace of the map induced by $\tilde{\tau}^*$ on $H^0(Y;\operatorname{ad}\beta)\oplus H^1(Y;\operatorname{ad}\beta)$;
- $\eta_\beta(0)$ is the Atiyah–Patodi–Singer spectral asymmetry of K_β; and
- $\eta_\beta^\tau(0)$ its equivariant version, defined as follows. For any eigenvalue λ of the operator K_β, the λ–eigenspace W_λ^β is acted upon by $\tilde{\tau}^*$ with trace $\operatorname{tr}(\tilde{\tau}^*|W_\lambda^\beta)$. The infinite series
$$\eta_\beta^\tau(s) = \sum_{\lambda\neq 0}\operatorname{sign}\lambda\cdot\operatorname{tr}(\tilde{\tau}^*|W_\lambda^\beta)|\lambda|^{-s}$$
converges for $\operatorname{Re}(s)$ large enough and has a meromorphic continuation to the entire complex s–plane with no pole at $s=0$; see Donnelly [12]. This makes $\eta_\beta^\tau(0)$ a well-defined real number.

Proof of Proposition 3.11 The index $\operatorname{ind}\mathcal{D}_A(X)$ can be computed using the index theorem of Atiyah, Patodi and Singer [4] as

$$\operatorname{ind}\mathcal{D}_A(X) = \int_X \hat{A}(X)\operatorname{ch}(S^+)\operatorname{ch}(\operatorname{ad}P)_{\mathbb{C}} - \tfrac{1}{2}(h_\beta - \eta_\beta(0))(Y),$$

and $\operatorname{ind}(\tau,\mathcal{D}_A)(X)$ using its equivariant counterpart, the G–index theorem of Donnelly [12], as

$$\operatorname{ind}(\tau,\mathcal{D}_A)(X) = \frac{1}{2}\int_F (e(TF) + e(NF)) - \tfrac{1}{2}(h_\beta^\tau - \eta_\beta^\tau(0))(Y).$$

The desired formula now follows because, according to the Gauss–Bonnet theorem,

$$\int_F e(TF) = \chi(F) \quad\text{and}\quad \int_F e(NF) = F\cdot F. \qquad \square$$

Example 3.12 Let $P\to Y$ be a trivial $\mathrm{SO}(3)$–bundle with an involution $\tilde{\tau}$ acting as a second-order operator on the fibers. Application of Proposition 3.11 to the product connection A on the manifold $X=\mathbb{R}\times Y$ results in the formula $\operatorname{ind}\mathcal{D}_\theta^\tau(X)=-1$, which corresponds to the fact that the $(+1)$–eigenspace of the involution $\tilde{\tau}^*\colon H^0(X;\operatorname{ad}\theta)\to H^0(X;\operatorname{ad}\theta)$ is 1–dimensional.

3.7 Proof of Lemma 2.5

Since both ρ and σ are irreducible and nondegenerate, we have $\mathrm{gr}(\chi_1 \cdot \rho, \sigma) = \mathrm{gr}(\chi_1 \cdot \rho, \rho) + \mathrm{gr}(\rho, \sigma)$. Therefore, we only need to compute $\mathrm{gr}(\chi_1 \cdot \rho, \rho) \pmod 4$.

Let $g \in \mathcal{G}$ be a gauge transformation matching ρ and $\chi_1 \cdot \rho$. The mapping torus of g is an orbifold bundle P_0 over $S^1 \times \Sigma$, and

$$\mathrm{gr}(\chi_1 \cdot \rho, \rho) = \mathrm{ind}\, \mathcal{D}_A\, (S^1 \times \Sigma) \pmod 4$$

for any choice of orbifold connection A in P_0. Let M be the double branched cover of Σ with branch set $\mathcal{L}$. Then the index in the above formula, treated as an equivariant index on $S^1 \times M$, equals $-p_1(Q_0)$ by the formula of Proposition 3.9 applied to the pullback bundle $Q_0 = \pi^* P_0$. This reduces the above formula to

$$\mathrm{gr}(\chi_1 \cdot \rho, \rho) = -p_1(Q_0) \pmod 4.$$

To compute the Pontryagin number $p_1(Q_0)$ we observe that the bundle Q_0 on $S^1 \times M$ can be obtained from the bundle $Q = \pi^* P$ on M as the mapping torus of a gauge transformation matching $\pi^* \rho$ with $\pi^*(\chi_1 \cdot \rho) = \eta \cdot \pi^* \rho$, where $\eta = \pi^* \chi_1 \in H^1(M; \mathbb{Z}/2)$. According to Braam and Donaldson [8, Part II, Propositions 1.9 and 1.13],

$$p_1(Q_0) = 2 \cdot (\eta \cup w_2(Q) + \eta \cup \eta \cup \eta)[M] \pmod 4.$$

We already know that $w_2(Q)$ is a generator of $H^2(M; \mathbb{Z}/2) = \mathbb{Z}/2$; see Proposition 3.4. It follows from Lemma 3.3 that the class η is a generator of $H^1(M; \mathbb{Z}/2) = \mathbb{Z}/2$. The desired formula now follows from the calculation of the cohomology ring $H^*(M; \mathbb{Z}/2)$ in Proposition 3.2.

4 Knot homology: the generators

We will now use the equivariant theory of Section 3 to better understand the chain complex $IC^\natural(k)$ which computes the singular instanton knot homology $I^\natural(k) = I_*(S^3, k^\natural)$ of Kronheimer and Mrowka [24]. In this section, we will describe the conjugacy classes of projective SU(2) representations on the exterior of $k^\natural$ with nontrivial $[c]$ and separate them into the orbits of the canonical $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ action. The next two sections will be dedicated to computing Floer gradings.

4.1 Projective representations

Given a knot $k \subset S^3$, denote by $K = S^3 - \mathrm{int}\, N(k)$ its exterior and by $K^\natural = S^3 - \mathrm{int}\, N(k^\natural)$ the exterior of the two-component link $k^\natural = k \cup \ell$ obtained as the connected sum of k with the Hopf link. The Wirtinger presentation

$$\pi_1 K = \langle a_1, a_2, \dots, a_n \mid r_1, \dots, r_m \rangle$$

with meridians a_i and relators r_j gives rise to the Wirtinger presentation

$$\pi_1 K^\natural = \langle a_1, a_2, \dots, a_n, b \mid r_1, \dots, r_m,\ [a_1, b] = 1\rangle,$$

where b stands for the meridian of the component ℓ. Since the link $k^\natural$ is not split, it follows from Lemma 2.3 that $H^2(\pi_1 K^\natural; \mathbb{Z}/2) = H^2(K^\natural; \mathbb{Z}/2) = \mathbb{Z}/2$. The generator of the latter group evaluates nontrivially on both boundary components of $K^\natural$, which makes it Poincaré dual to any arc connecting these two boundary components. It follows from Proposition 2.1 that the projective representations with nontrivial $[c]$ which we are interested in are precisely the homomorphisms $\rho\colon F \to \mathrm{SU}(2)$ of the free group F generated by the meridians $a_1, \dots, a_n, b$ such that

$$\rho(r_1) = \cdots = \rho(r_n) = 1 \quad \text{and} \quad \rho([a_1, b]) = -1.$$

Representations ρ are uniquely determined by the SU(2) matrices $A_i = \rho(a_i)$ and $B = \rho(b)$ subject to the above relations, and the space $\mathcal{PR}_c(K^\natural, \mathrm{SU}(2))$ consists of all such tuples $(A_1, \dots, A_n; B)$ up to conjugation.

The relation $A_1 B = -B A_1$ implies that, up to conjugation, $A_1 = i$ and $B = j$. Since the Wirtinger relations $r_1 = 1, \dots, r_m = 1$ are of the form $a_i a_j a_i^{-1} = a_k$, all the matrices A_i must have zero trace. In particular, the matrices $A_1 = \cdots = A_n = i$ and $B = j$ satisfy all of the relations, thereby giving rise to the special projective representation $\alpha = (i, i, \dots, i; j)$. On the other hand, if we assume that not all A_i commute with each other, we have an entire circle of projective representations,

$$(i, e^{i\varphi} A_2 e^{-i\varphi}, \dots, e^{i\varphi} A_n e^{-i\varphi}; j). \tag{8}$$

It is parametrized by $e^{2i\varphi} \in S^1$ because the center of SU(2) is the stabilizer of the adjoint action of SU(2) on itself. Note that two tuples like (8) are conjugate if and only if they are equal to each other. One can easily see that the formula $\psi(A_1, \dots, A_n; B) = (A_1, \dots, A_n)$ defines a surjective map

$$\psi\colon \mathcal{PR}_c(K^\natural, \mathrm{SU}(2)) \to \mathcal{R}_0(K, \mathrm{SU}(2)), \tag{9}$$

where $\mathcal{R}_0(K, \mathrm{SU}(2))$ is the space of the conjugacy classes of traceless representations $\rho_0\colon \pi_1 K \to \mathrm{SU}(2)$. If ρ_0 is irreducible, the fiber $C(\rho_0) = \psi^{-1}([\rho_0])$ is a circle of the form (8). The special projective representation α is a fiber of (9) in its own right over the unique (up to conjugation) reducible traceless representation $\pi_1 K \to H_1(K) \to \mathrm{SU}(2)$ sending all the meridians to the same traceless matrix i. Therefore, assuming that $\mathcal{R}_0(K, \mathrm{SU}(2))$ is nondegenerate, the space $\mathcal{PR}_c(K^\natural, \mathrm{SU}(2))$ consists of an isolated point and finitely many circles, one for each conjugacy class of irreducible representations in $\mathcal{R}_0(K, \mathrm{SU}(2))$. The same result holds in general after perturbation.

4.2 The action of $H^1(K^\natural; \mathbb{Z}/2)$

The group $H^1(K^\natural; \mathbb{Z}/2) = \mathbb{Z}/2 \oplus \mathbb{Z}/2$ generated by the duals χ_k and χ_ℓ of the meridians of the link $k^\natural = k \cup \ell$ acts on the space of projective representations $\mathcal{PR}_c(K^\natural, \mathrm{SU}(2))$ as explained in Section 2.2. In terms of the tuples (8), the generators χ_k and χ_ℓ send $(i, e^{i\varphi} A_2 e^{-i\varphi}, \dots, e^{i\varphi} A_n e^{-i\varphi}; j)$ to

$$(-i, -e^{i\varphi} A_2 e^{-i\varphi}, \dots, -e^{i\varphi} A_n e^{-i\varphi}; j) \quad \text{and} \quad (i, e^{i\varphi} A_2 e^{-i\varphi}, \dots, e^{i\varphi} A_n e^{-i\varphi}; -j),$$

respectively. The isolated point $\alpha = (i, i, \dots, i; j)$ is a fixed point of this action since

$$(-i, -i, \dots, -i; j) = j \cdot (i, i, \dots, i; j) \cdot j^{-1}, \quad (i, i, \dots, i; -j) = i \cdot (i, i, \dots, i; j) \cdot i^{-1}.$$

To describe the action of χ_ℓ on the circle $C(\rho_0)$ for an irreducible ρ_0, conjugate $(i, e^{i\varphi} A_2 e^{-i\varphi}, \dots, e^{i\varphi} A_n e^{-i\varphi}; -j)$ by i to obtain

$$(i, e^{i(\varphi+\pi/2)} A_2 e^{-i(\varphi+\pi/2)}, \dots, e^{i(\varphi+\pi/2)} A_n e^{-i(\varphi+\pi/2)}; j).$$

Since the circle $C(\rho_0)$ is parametrized by $e^{2i\varphi}$, we conclude that the involution χ_ℓ acts on $C(\rho_0)$ via the antipodal map.

The action of χ_k on the circle $C(\rho_0)$ for an irreducible ρ_0 will depend on whether ρ_0 is a binary dihedral representation or not. Recall that a representation $\rho_0\colon \pi_1 K \to \mathrm{SU}(2)$ is called *binary dihedral* if it factors through a copy of the binary dihedral subgroup $S^1 \cup j \cdot S^1 \subset \mathrm{SU}(2)$, where S^1 stands for the circle of unit complex numbers. Equivalently, ρ_0 is binary dihedral if its adjoint representation $\mathrm{Ad}(\rho_0)\colon \pi_1 K \to \mathrm{SO}(3)$ is *dihedral* in that it factors through a copy of $O(2)$ embedded into $\mathrm{SO}(3)$ via the map $A \to (A, \det A)$.

One can show that a representation ρ_0 is binary dihedral if and only if $\chi \cdot \rho_0$ is conjugate to ρ_0, where $\chi\colon \pi_1 K \to \mathbb{Z}/2$ is the generator of $H^1(K; \mathbb{Z}/2) = \mathbb{Z}/2$. Note that χ defines an involution on $\mathcal{R}_0(K, \mathrm{SU}(2))$ which makes the following diagram commute:

$$\begin{array}{ccc} \mathcal{PR}_c(K^\natural, \mathrm{SU}(2)) & \xrightarrow{\pi} & \mathcal{R}_0(K, \mathrm{SU}(2)) \\ \downarrow{\scriptstyle \chi_k} & & \downarrow{\scriptstyle \chi} \\ \mathcal{PR}_c(K^\natural, \mathrm{SU}(2)) & \xrightarrow{\pi} & \mathcal{R}_0(K, \mathrm{SU}(2)) \end{array}$$

The action of χ_k can now be described as follows. If an irreducible $\rho_0\colon \pi_1 K \to \mathrm{SU}(2)$ is not binary dihedral, the involution χ_k takes the circle $C(\rho_0)$ to the circle $C(\chi \cdot \rho_0)$. Since $\chi \cdot \rho_0$ is not conjugate to ρ_0, these two circles are disjoint from each other, and χ_k permutes them. If an irreducible $\rho_0\colon \pi_1 K \to \mathrm{SU}(2)$ is binary dihedral, there exists $u \in \mathrm{SU}(2)$ such that $uiu^{-1} = -i$ and $uA_iu^{-1} = -A_i$ for $i = 2, \dots, n$. The irreducibility of ρ_0 also implies that $u^2 = -1$, so after conjugation we may assume that $u = k$. Now conjugate

$$\chi_k \cdot (i, e^{i\varphi} A_2 e^{-i\varphi}, \dots, e^{i\varphi} A_n e^{-i\varphi}; j) = (-i, -e^{i\varphi} A_2 e^{-i\varphi}, \dots, -e^{i\varphi} A_n e^{-i\varphi}; j)$$

by j to obtain

$$\begin{aligned}(i, j(-e^{i\varphi} A_2 e^{-i\varphi}) j^{-1}, \dots, j(-e^{i\varphi} A_n e^{-i\varphi}) j^{-1}; j) \\ &= (i, -e^{-i\varphi} j A_2 j^{-1} e^{i\varphi}, \dots, -e^{-i\varphi} j A_n j^{-1} e^{i\varphi}; j) \\ &= (i, -(i e^{-i\varphi}) k A_2 k^{-1} (i^{-1} e^{i\varphi}), \dots, -(i e^{-i\varphi}) k A_n k^{-1} (i^{-1} e^{i\varphi}); j) \\ &= (i, e^{i(\pi/2-\varphi)} A_2 e^{-i(\pi/2-\varphi)}, \dots, e^{i(\pi/2-\varphi)} A_n e^{-i(\pi/2-\varphi)}; j).\end{aligned}$$

Therefore, χ_k acts on $C(\rho_0)$ by sending $e^{2i\varphi}$ to $-e^{-2i\varphi}$, which is an involution on the complex unit circle with two fixed points, i and $-i$.

Finally, observe that the quotient of $\mathcal{R}_0(K, \mathrm{SU}(2))$ by the involution χ is precisely the space $\mathcal{R}_0(K, \mathrm{SO}(3))$ of the conjugacy classes of representations $\operatorname{Ad}\rho_0\colon \pi_1 K \to \mathrm{SO}(3)$. Since $H^2(K; \mathbb{Z}/2) = 0$, every $\mathrm{SO}(3)$ representation lifts to an $\mathrm{SU}(2)$ representation, hence $\mathcal{R}_0(K, \mathrm{SO}(3))$ can also be described as the space of the conjugacy classes of representations $\pi_1 K \to \mathrm{SO}(3)$ sending the meridians to $\mathrm{SO}(3)$ matrices of trace -1. Compose (9) with the projection $\mathcal{R}_0(K, \mathrm{SU}(2)) \to \mathcal{R}_0(K, \mathrm{SO}(3))$ to obtain a surjective map $\psi\colon \mathcal{PR}_c(K^\natural, \mathrm{SU}(2)) \to \mathcal{R}_0(K, \mathrm{SO}(3))$. The above discussion can now be summarized as follows.

Proposition 4.1 *The group $H^1(K^\natural; \mathbb{Z}/2) = \mathbb{Z}/2 \oplus \mathbb{Z}/2$ acts on $\mathcal{PR}_c(K^\natural, \mathrm{SU}(2))$ preserving the fibers of the map $\psi\colon \mathcal{PR}_c(K^\natural, \mathrm{SU}(2)) \to \mathcal{R}_0(K, \mathrm{SO}(3))$. Furthermore:*

(a) *For the unique reducible in $\mathcal{R}_0(K, \mathrm{SO}(3))$, the fiber of ψ consists of just one point, which is the conjugacy class of the special projective representation α. This point is fixed by both χ_k and χ_ℓ.*

(b) *For any dihedral representation in $\mathcal{R}_0(K, \mathrm{SO}(3))$, the fiber of ψ is a circle. The involution χ_k is a reflection of this circle with two fixed points, while χ_ℓ is the antipodal map.*

(c) *Otherwise, the fiber of ψ consists of two circles. The involution χ_k permutes these circles, while χ_ℓ acts as the antipodal map on both.*

It should be noted that perturbing the Chern–Simons functional (1) may easily break the $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ symmetry. Finding a perturbation which preserves this symmetry runs as usual into the equivariant transversality problem, which we do not try to address here. It should be noted, however, that such a problem was successfully solved in [33] in a similar setting.

4.3 Double branched covers

Next, we would like to describe the space $\mathcal{PR}_c(K^\natural, \mathrm{SU}(2))$ using the equivariant theory of Section 3. We could proceed as in that section, by passing to the double branched cover $M \to S^3$ with branch set the link $k^\natural$ and working with the equivariant representations $\pi_1 M \to \mathrm{SO}(3)$. However, in the special case at hand, one can observe that M is simply the connected sum $Y \# \mathbb{R}\mathrm{P}^3$, where Y is the double branched cover of S^3 with branch set the knot k, hence the same information about $\mathcal{PR}_c(K^\natural, \mathrm{SU}(2))$ can be extracted more easily by working directly with Y and using Proposition 4.1. The only missing step in this program is a description of $\mathcal{R}_0(K, \mathrm{SO}(3))$ in terms of equivariant representations $\pi_1 Y \to \mathrm{SO}(3)$, which we will take up next.

Every representation $\rho\colon \pi_1 K \to \mathrm{SO}(3)$ gives rise to a representation of the orbifold fundamental group $\pi_1^V(S^3, k) = \pi_1 K/\langle \mu^2 = 1\rangle$, where we choose $\mu = a_1$ to be our meridian. The latter group can be included into the split orbifold exact sequence

$$1 \to \pi_1 Y \xrightarrow{\pi_*} \pi_1^V(S^3, k) \xrightarrow{j} \mathbb{Z}/2 \to 1.$$

Proposition 4.2 *Let Y be the double branched cover of S^3 with branch set a knot k and let $\tau\colon Y \to Y$ be the covering translation. The pullback of representations via the map π_* in the orbifold exact sequence establishes a homeomorphism*

$$\pi^*\colon \mathcal{R}_0(K, \mathrm{SO}(3)) \to \mathcal{R}^\tau(Y, \mathrm{SO}(3)),$$

where $\mathcal{R}^\tau(Y)$ is the fixed point set of the involution $\tau^*\colon \mathcal{R}(Y, \mathrm{SO}(3)) \to \mathcal{R}(Y, \mathrm{SO}(3))$. The unique reducible representation in $\mathcal{R}_0(K, \mathrm{SO}(3))$ pulls back to the trivial representation of $\pi_1 Y$, and the dihedral representations in $\mathcal{R}_0(K, \mathrm{SO}(3))$ are precisely those that pull back to reducible representations of $\pi_1 Y$.

Proof A slight modification of the argument of Proposition 3.8 (see also [11, Proposition 3.3]), establishes a homeomorphism between $\mathcal{R}_0(K, \mathrm{SO}(3))$ and the subspace of $\mathcal{R}^\tau(Y, \mathrm{SO}(3))$ consisting of the conjugacy classes of representations $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ such that $\tau^*\beta = u \cdot \beta \cdot u^{-1}$ for some $u \in \mathrm{SO}(3)$ of order 2. The proof of the first statement of the proposition will be complete after we show that this subspace in fact comprises the entire space $\mathcal{R}^\tau(Y, \mathrm{SO}(3))$.

If $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ is reducible, it factors through a representation $H_1(Y) \to \mathrm{SO}(2)$. According to Proposition 3.1, the involution τ_* acts on $H_1(Y)$ as multiplication by -1. Therefore, $\tau^*\beta = \beta^{-1}$, and the latter representation can obviously be conjugated to β by an element $u \in \mathrm{SO}(3)$ of order 2. If $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ is irreducible, the condition $\beta \in \mathrm{Fix}(\tau^*)$ implies that there exists a unique $u \in \mathrm{SO}(3)$ such that $\tau^*\beta = u \cdot \beta \cdot u^{-1}$ and $u^2 = 1$. If $u = 1$, then $\tau^*\beta = \beta$, which implies that β is the pullback of a representation of $\pi_1^V(S^3,k)$ which sends the meridian μ to the identity matrix and hence factors through $\pi_1 S^3 = 1$. This contradicts the irreducibility of β.

To prove the second statement of the proposition, observe that the homomorphism j in the above orbifold exact sequence sending μ to the generator of $\mathbb{Z}/2$ is in fact the abelianization homomorphism. This implies that the unique reducible representation in $\mathcal{R}_0(K,\mathrm{SO}(3))$ pulls back to the trivial representation of $\pi_1 Y$. Since $\pi_1 Y$ is the commutator subgroup of $\pi_1^V(S^3,k)$, any dihedral representation $\rho\colon \pi_1^V(S^3,k)\to O(2)$ must map $\pi_1 Y$ to the commutator subgroup of $O(2)$, which happens to be $\mathrm{SO}(2)$. This ensures that the pullback of ρ is reducible. Conversely, if the pullback of ρ is reducible, its image is contained in a copy of $\mathrm{SO}(2)$, and the image of ρ itself in its 2–prime extension. The latter group is of course just a copy of $O(2)\subset\mathrm{SO}(3)$. □

Remark 4.3 For future use note that, for any projective representation $\rho\colon \pi_1 K^\natural \to \mathrm{SU}(2)$ in $C(\rho_0)$ described by a tuple (8), the adjoint representation $\mathrm{Ad}\,\rho\colon \pi_1 K^\natural\to \mathrm{SO}(3)$ pulls back to an $\mathrm{SO}(3)$ representation of $\pi_1(Y\,\#\,\mathbb{RP}^3) = \pi_1 Y * \mathbb{Z}/2$ of the form $\beta*\gamma\colon \pi_1 Y*\mathbb{Z}/2\to\mathrm{SO}(3)$, where $\beta=\pi^*\,\mathrm{Ad}\,\rho_0$ and $\gamma\colon \mathbb{Z}/2\to\mathrm{SO}(3)$ sends the generator of $\mathbb{Z}/2$ to $\mathrm{Ad}\,i\cdot\mathrm{Ad}\,j=\mathrm{Ad}\,k$. The representation $\beta*\gamma$ is equivariant, $\tau^*(\beta*\gamma)=u\cdot(\beta*\gamma)\cdot u^{-1}$, with the conjugating element u given by $\mathrm{Ad}\,\rho_0(a_1)=\mathrm{Ad}\,i$.

5 Knot homology: grading of the special generator

Given a knot $k\subset S^3$, we will continue using the notation K for its exterior and $K^\natural$ for the exterior of the two-component link $k^\natural = k\cup\ell$ obtained as the connected sum of k with the Hopf link h. The special projective representation $\alpha\colon \pi_1 K^\natural\to\mathrm{SU}(2)$, which sends all the meridians of k to i and the meridian of ℓ to j, is a generator in the chain complex $IC^\natural(k)$. In this section, we compute its absolute Floer grading.

Theorem 5.1 *For any knot k in S^3, we have* $\mathrm{gr}(\alpha)=\operatorname{sign}k \pmod 4$.

Before we go on to prove this theorem, recall the definition of $\mathrm{gr}(\alpha) \pmod 4$. Let (W',S) be a cobordism of pairs (S^3,u) and (S^3,k), where u is an unknot in S^3. The manifold W' is required to be oriented but the surface S is not. Construct a new cobordism (W',S') of the pairs (S^3,h) and $(S^3,k^\natural)$ by letting S' be the disjoint union of S with the normal circle bundle along a path in S connecting the two boundary components (the surface S' is called $S^\natural$ in [24, Section 4.3]). According to [24, Proposition 4.4], the generator α has grading

$$\mathrm{gr}(\alpha) = -\operatorname{ind}\mathcal{D}_{A'}(\alpha_u,\alpha)-\tfrac{3}{2}(\chi(W')+\sigma(W'))-\chi(S') \pmod 4, \tag{10}$$

where α_u stands for the special generator in the Floer chain complex of u, and we use the fact that $\chi(S)=\chi(S')$. The operator $\mathcal{D}_{A'}(\alpha_u,\alpha)$ refers to the ASD operator on the noncompact manifold obtained from W' by attaching cylindrical ends to the two boundary components; this manifold is again called W'. The connection A' can be any connection on W' which is singular along the surface S' and whose limits on the two ends are flat connections with holonomies α_u and α. The index of $\mathcal{D}_{A'}(\alpha_u, \alpha)$ is understood as the L^2_δ index for a small positive δ.

5.1 Constructing the cobordism

Our calculation of the Floer index $\mathrm{gr}(\alpha)$ will use a specific cobordism (W', S') constructed as follows.

Let Σ be the double branched cover of S^3 with branch set the knot k. Choose a Seifert surface F' of k and push its interior slightly into the ball D^4 so that the resulting surface, which we still call F', is transverse to $\partial D^4 = S^3$. Let V be the double branched cover of D^4 with branch set the surface F'. Then V is a smooth simply connected spin 4–manifold with boundary Σ, which admits a handle decomposition with only 0– and 2–handles; see Akbulut and Kirby [1, page 113].

Next, choose a point in the interior of the surface $F' \subset D^4$. Excising a small open 4–ball containing that point from (D^4, F') results in a manifold W_1' diffeomorphic to $I \times S^3$ together with the surface $F_1' = F' - \operatorname{int} D^2$ properly embedded into it, thereby providing a cobordism (W_1', F_1') from an unknot to the knot k. The double branched cover $W_1 \to W_1'$ with branch set F_1' is a cobordism from S^3 to Σ. The manifold W_1 is simply connected because it can be obtained from the simply connected manifold V by excising an open 4–ball.

Similarly, consider the manifold $W_2' = I \times S^3$ and surface $F_2' = I \times h \subset W_2'$ providing a product cobordism from the Hopf link h to itself. The double branched cover $W_2 \to W_2'$ with branch set F_2' is then a cobordism $W_2 = I \times \mathbb{R}\mathrm{P}^3$ from $\mathbb{R}\mathrm{P}^3$ to itself.

As the final step of the construction, consider a path γ_1' in the surface F_1' connecting its two boundary components. Similarly, consider a path γ_2' of the form $I \times \{p\}$ in the surface $F_2' = I \times H$. Remove tubular neighborhoods of these two paths and glue the resulting manifolds and surfaces together using an orientation-reversing diffeomorphism $1 \times h\colon I \times S^2 \to I \times S^2$. The resulting pair (W', S') is the desired cobordism of the pairs (S^3, h) and $(S^3, k^\natural)$. One can easily see that

$$\chi(W') = \sigma(W') = 0 \quad \text{and} \quad \chi(S') = \chi(F') - 1. \tag{11}$$

Note that the double branched cover $W \to W'$ with branch set S' is a cobordism from $\mathbb{R}\mathrm{P}^3$ to $\Sigma \# \mathbb{R}\mathrm{P}^3$ which can be obtained from the cobordisms W_1 and W_2 by taking a connected sum along the paths $\gamma_1 \subset W_1$ and $\gamma_2 \subset W_2$ lifting, respectively, the paths γ_1' and γ_2'. To be precise,

$$W = W_1^\circ \cup W_2^\circ, \tag{12}$$

where W_1° and W_2° are obtained from W_1 and W_2 by removing tubular neighborhoods of γ_1 and γ_2. The identification in (12) is done along a copy of $I \times S^2$. In particular, we see that $\pi_1 W = \mathbb{Z}/2$.

5.2 L^2–index

We will rely on Ruberman [32] and Taubes [40] in our index calculations.

Let $\pi\colon W \to W'$ be the double branched cover with branch set S' constructed in the previous section, and $\tau\colon W \to W$ the covering translation. Let us consider a representation $\rho\colon \pi_1^V(W', S') \to \mathrm{SO}(3)$ sending the two sets of meridians of S' to $\mathrm{Ad}\, i$ and $\mathrm{Ad}\, j$. Then the representation $\pi^*\rho\colon \pi_1 W \to \mathrm{SO}(3)$ sends the generator of $\pi_1 W$ to $\mathrm{Ad}\, k$ and it is equivariant in that $\tau^*(\pi^*\rho) = u \cdot \pi^*\rho \cdot u^{-1}$ with $u = \mathrm{Ad}\, i$; compare with Example 3.7. The representation ρ restricts to α_u and α over the two ends of W', therefore $\pi^*\rho$ makes W into a flat cobordism between $\gamma\colon \pi_1(\mathbb{R}\mathrm{P}^3) \to \mathrm{SO}(3)$ and $\theta * \gamma\colon \pi_1 \Sigma * \pi_1(\mathbb{R}\mathrm{P}^3) \to \mathrm{SO}(3)$, where γ is the representation of Example 3.7.

Let A and A' be flat connections on W and W' whose holonomies are, respectively, $\pi^*\rho$ and ρ. We will use A' as the twisting connection of the operator $\mathcal{D}_{A'}(\alpha_u, \alpha)$. Instead of computing the index of this operator, we will compute the equivariant index $\mathrm{ind}\, \mathcal{D}_A^\tau(\gamma, \theta * \gamma)$ of its pullback to W. The latter index equals minus the equivariant index of the elliptic complex

$$\Omega^0(W, \mathrm{ad}\, P) \xrightarrow{-d_A} \Omega^1(W, \mathrm{ad}\, P) \xrightarrow{d_A^+} \Omega^2_+(W, \mathrm{ad}\, P). \tag{13}$$

The equivariance here is understood with respect to a lift of $\tau\colon W \to W$ to the bundle $\mathrm{ad}\, P$ which has second order on the fibers over the fixed point set. The connection A is equivariant with respect to this lift, hence it splits the coefficient bundle $\mathrm{ad}\, P$ into a sum of three real line bundles corresponding to $\mathrm{Ad}\, k = \mathrm{diag}(-1, -1, 1)$. Accordingly, the complex (13) splits into a sum of three elliptic complexes, one with the trivial real coefficients and two with the twisted coefficients. Application of [32, Proposition 4.1] to the former complex and of [32, Corollary 4.2] to the latter two reduces the index problem to computing the singular cohomology

$$H^k(W; \mathrm{ad}\, \pi^*\rho) = H^k(W; \mathbb{R}) \oplus H^k(W; \mathbb{R}_-) \oplus H^k(W; \mathbb{R}_-) \quad \text{for } k = 0, 1, 2,$$

where $\mathbb{R}_-$ stands for the real line coefficients on which $\mathbb{Z}/2$ acts as multiplication by -1, and their equivariant versions.

The zeroth equivariant cohomology of the complex (13) vanishes since $H^0(W;\mathbb{R}_-)=0$ and the lift of τ acts as minus identity on the remaining group $H^0(W;\mathbb{R})=\mathbb{R}$. This vanishing result could also be derived directly from the irreducibility of the singular connection A'. The next two subsections are dedicated to computing the first and second cohomology of (13).

5.3 Trivial coefficients

Our computation will be based on the Mayer–Vietoris exact sequence applied twice, first to compute cohomology of W_1° and W_2°, and then to compute cohomology of $W=W_1^\circ\cup W_2^\circ$.

The cohomology groups of W_1° and $W_1=W_1^\circ\cup(I\times D^3)$ are related by the following Mayer–Vietoris exact sequence:

$$\begin{aligned}0\longrightarrow H^1(W_1;\mathbb{R})&\longrightarrow H^1(W_1^\circ;\mathbb{R})\longrightarrow 0\\ \longrightarrow H^2(W_1;\mathbb{R})&\longrightarrow H^2(W_1^\circ;\mathbb{R})\longrightarrow H^2(I\times S^2;\mathbb{R})\\ \xrightarrow{\delta} H^3(W_1;\mathbb{R})&\longrightarrow H^3(W_1^\circ;\mathbb{R})\longrightarrow 0\end{aligned}$$

Since W_1 and therefore W_1° are simply connected, both $H^1(W_1;\mathbb{R})$ and $H^1(W_1^\circ;\mathbb{R})$ vanish. Applying the Poincaré–Lefschetz duality to the manifold W_1 and using the long exact sequence of the pair $(W_1,\partial W_1)$, we obtain

$$H^3(W_1;\mathbb{R})=H_1(W_1,\partial W_1;\mathbb{R})=\widetilde{H}_0(\partial W_1;\mathbb{R})=\mathbb{R}.$$

Similarly, viewing W_1° as a manifold whose boundary is a connected sum of the two boundary components of W_1, we obtain

$$H^3(W_1^\circ;\mathbb{R})=H_1(W_1^\circ,\partial W_1^\circ;\mathbb{R})=\widetilde{H}_0(\partial W_1^\circ;\mathbb{R})=0.$$

Therefore, the connecting homomorphism δ in the above exact sequence must be an isomorphism, which leads to the isomorphisms

$$H^2(W_1^\circ;\mathbb{R})=H^2(W_1;\mathbb{R})=H^2(V;\mathbb{R}).$$

A similar long exact sequence relates the cohomology of W_2° and $W_2=W_2^\circ\cup(I\times D^3)$, implying that

$$H^2(W_2^\circ;\mathbb{R})=H^2(W_2;\mathbb{R})=H^2(\mathbb{R}\mathrm{P}^3;\mathbb{R})=0.$$

Since $\pi_1 W_2 = \pi_1 W_2^\circ = \mathbb{Z}/2$, both $H^1(W_2;\mathbb{R})$ and $H^1(W_2^\circ;\mathbb{R})$ vanish. The Mayer–Vietoris exact sequence of the splitting $W = W_1^\circ \cup W_2^\circ$,

$$\begin{aligned} 0 \longrightarrow H^1(W;\mathbb{R}) &\longrightarrow H^1(W_1^\circ;\mathbb{R}) \oplus H^1(W_2^\circ;\mathbb{R}) \longrightarrow 0 \\ \longrightarrow H^2(W;\mathbb{R}) &\longrightarrow H^2(W_1^\circ;\mathbb{R}) \oplus H^2(W_2^\circ;\mathbb{R}) \to H^2(I\times S^2;\mathbb{R}) \\ \longrightarrow H^3(W;\mathbb{R}) &\longrightarrow H^3(W_1^\circ;\mathbb{R}) \oplus H^3(W_2^\circ;\mathbb{R}) \longrightarrow 0 \end{aligned}$$

together with the isomorphisms $H^3(W;\mathbb{R}) = H_1(W,\partial W;\mathbb{R}) = \widetilde{H}_0(\partial W;\mathbb{R}) = \mathbb{R}$ and $\pi_1 W = \mathbb{Z}/2$, implies that

$$H^1(W;\mathbb{R}) = 0 \quad \text{and} \quad H^2(W;\mathbb{R}) = H^2(V;\mathbb{R}).$$

5.4 Twisted coefficients

We will now do a similar calculation using the Mayer–Vietoris sequence of $W = W_1^\circ \cup W_2^\circ$ with twisted coefficients. Since W_1° is simply connected, the twisted coefficients $\mathbb{R}_-$ pull back to the trivial $\mathbb{R}$–coefficients over W_1° and the cohomology calculations from the previous section are unchanged. A direct calculation using homotopy equivalences $W_2 \simeq \mathbb{R}\mathrm{P}^3$ and $W_2^\circ \simeq \mathbb{R}\mathrm{P}^2$ shows that

$$H^1(W_2^\circ;\mathbb{R}_-) = 0 \quad \text{and} \quad H^2(W_2^\circ;\mathbb{R}_-) = \mathbb{R}.$$

The latter isomorphism is induced by the inclusion $I \times S^2 \to W_2^\circ$, which can be easily seen from the Mayer–Vietoris exact sequence of $W_2 = W_2^\circ \cup (I \times D^3)$. Now, consider the Mayer–Vietoris exact sequence of the splitting $W = W_1^\circ \cup W_2^\circ$ with twisted $\mathbb{R}$–coefficients:

$$\begin{aligned} 0 \longrightarrow H^1(W;\mathbb{R}_-) &\longrightarrow H^1(W_1^\circ;\mathbb{R}) \oplus H^1(W_2^\circ;\mathbb{R}_-) \longrightarrow 0 \\ \longrightarrow H^2(W;\mathbb{R}_-) &\longrightarrow H^2(W_1^\circ;\mathbb{R}) \oplus H^2(W_2^\circ;\mathbb{R}_-) \to H^2(I\times S^2;\mathbb{R}) \\ \longrightarrow H^3(W;\mathbb{R}_-) &\longrightarrow H^3(W_1^\circ;\mathbb{R}) \oplus H^3(W_2^\circ;\mathbb{R}_-) \longrightarrow 0 \end{aligned}$$

Keeping in mind that the map $H^2(W_1^\circ;\mathbb{R}) \to H^2(I\times S^2;\mathbb{R})$ in this sequence is zero and the map $H^2(W_2^\circ;\mathbb{R}_-) \to H^2(I\times S^2;\mathbb{R})$ is an isomorphism $\mathbb{R}\to\mathbb{R}$, we conclude that

$$H^1(W;\mathbb{R}_-) = 0 \quad \text{and} \quad H^2(W;\mathbb{R}_-) = H^2(V;\mathbb{R}).$$

5.5 Equivariant cohomology

Combining results of the previous two sections, we obtain $H^1(W;\operatorname{ad} P) = 0$ and $H^2(W;\operatorname{ad} P) = H^2(V;\mathbb{R}^3)$. The action of τ is compatible with these isomorphisms, from which we immediately conclude that

$$H^1_\tau(W;\operatorname{ad} P) = 0$$

and $H^2_\tau(W;\operatorname{ad} P)$ is the fixed point set of the map $H^2(V;\mathbb{R}^3) \to H^2(V;\mathbb{R}^3)$ obtained by twisting $\tau^*\colon H^2(V;\mathbb{R}) \to H^2(V;\mathbb{R})$ by the action on the coefficients $\mathbb{R}^3 \to \mathbb{R}^3$. The involution τ^* is minus the identity, which follows from the usual transfer argument applied to the covering $V \to D^4$, while the action on the coefficients is given by an SO(3) operator of second order. Such an operator must have a single eigenvalue 1 and a double eigenvalue -1, which leads us to the conclusion that $\operatorname{rk} H^2_\tau(W;\operatorname{ad} P) = 2 \cdot b_2(V)$. Similarly,

$$\operatorname{rk} H^2_{\tau,+}(W;\operatorname{ad} P) = 2 \cdot b_2^+(V).$$

5.6 Proof of Theorem 5.1

It follows from the discussion in Section 5.2 and the calculation in Section 5.5 that

$$\operatorname{ind} \mathcal{D}_{A'}(\alpha_u, \alpha) = \operatorname{rk} H^1_\tau(W;\operatorname{ad} P) - \operatorname{rk} H^2_{+,\tau}(W;\operatorname{ad} P) = -2 \cdot b_2^+(V).$$

Taking into account (10) and (11), we obtain the formula

$$\operatorname{gr}(\alpha) = 2 \cdot b_2^+(V) - \chi(F') + 1 \pmod 4.$$

To simplify it, let us compute $\chi(V)$ in two different ways: $\chi(V) = 1 + b_2^+(V) + b_2^-(V)$ by definition, and $\chi(V) = 2\chi(D^4) - \chi(F') = 2 - \chi(F')$ using the fact that V is a double branched cover of D^4 with branch set F'. Combining these formulas with the knot signature formula of Viro [41], we obtain the desired result (remember that $\operatorname{sign} k$ is always even):

$$\operatorname{gr}(\alpha) = -\operatorname{sign} V = -\operatorname{sign} k = \operatorname{sign} k \pmod 4.$$

6 Knot homology: gradings of other generators

Proposition 4.1 identified the critical points of the Chern–Simons functional with the fibers of the map $\psi\colon \mathcal{PR}_c(K^\natural, \mathrm{SU}(2)) \to \mathcal{R}_0(K, \mathrm{SO}(3))$. Assuming that the space $\mathcal{R}_0(K, \mathrm{SO}(3))$ is nondegenerate, all of these fibers (with the exception of the special generator α) are Morse–Bott circles. In this section, we will compute their Floer gradings using the equivariant index theory of Section 3.5. The actual generators of the chain complex $IC^\natural(k)$ are then obtained by perturbing each Morse–Bott circle of index μ into two points of indices μ and $\mu + 1$, as in [20]. Our index calculation will depend on whether an irreducible trace-free representation $\rho_0\colon \pi_1 K \to \mathrm{SO}(3)$ giving rise to the Morse–Bott circle $C(\rho_0)$ is dihedral or not. The two cases will be considered separately, starting with the case when ρ_0 is dihedral. If $\mathcal{R}_0(K, \mathrm{SO}(3))$ fails to be nondegenerate, similar results hold after additional perturbations.

6.1 Dihedral representations

Let $\rho_0\colon \pi_1 K \to \mathrm{SO}(3)$ be an irreducible trace-free dihedral representation. The pullback via $\pi\colon M \to \Sigma$ identifies the Morse–Bott circle $C(\rho_0)$ with the circle of the conjugacy classes of equivariant representations of the form $\beta * \gamma\colon \pi_1 Y * \mathbb{Z}/2 \to \mathrm{SO}(3)$, where β is a nontrivial reducible representation of $\pi_1 Y$ and γ is the representation of $\mathbb{Z}/2$ sending the generator to $\mathrm{Ad}\, k$. These representations are equivariant in that $\tau^*(\beta * \gamma) = u \cdot (\beta * \gamma) \cdot u^{-1}$ with $u = \mathrm{Ad}\, i$; see Remark 4.3.

We wish to compute the equivariant index $\operatorname{ind} \mathcal{D}^{\tau}_A(\beta * \gamma, \theta * \gamma)$, where A is any equivariant connection on the cylinder $\mathbb{R} \times (Y \# \mathbb{R}\mathrm{P}^3)$ whose limits are the flat connections $\beta * \gamma$ and $\theta * \gamma$ over the negative and positive ends, respectively. The Morse–Bott index of the circle corresponding to $\beta * \gamma$ will then equal

$$\mu = \operatorname{ind} \mathcal{D}^{\tau}_A(\beta * \gamma, \theta * \gamma) + \operatorname{sign} k \pmod 4. \tag{14}$$

Proposition 6.1 *Let $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ be a nontrivial equivariant reducible representation. Then for any equivariant connection B on the cylinder $\mathbb{R} \times Y$ whose limits are the flat connections β and θ over the negative and positive ends,*

$$\operatorname{ind} \mathcal{D}^{\tau}_A(\beta * \gamma, \theta * \gamma) = \operatorname{ind} \mathcal{D}^{\tau}_B(\beta, \theta) \pmod 4.$$

Proof To compute the index on the left-hand side of this formula, we will apply the formula of Proposition 3.11 to the manifold $X = \mathbb{R} \times (Y \# \mathbb{R}\mathrm{P}^3)$ with two product ends. Since the metric on X is a product metric, the terms $p_1(TX)$ and $e(TX)$ in the integrand

$$\widehat{A}(X)\, \mathrm{ch}(S^+)\, \mathrm{ch}(\mathrm{ad}\, P)_{\mathbb{C}} = -2 p_1(A) - \tfrac{1}{2} p_1(TX) - \tfrac{3}{2} e(TX)$$

will vanish, as will the topological terms $\chi(F)$ and $F \cdot F$, leading to the formula

$$\begin{aligned} \operatorname{ind} \mathcal{D}^{\tau}_A(\beta * \gamma, \theta * \gamma) = -\int_X p_1(A) &- \tfrac{1}{4}(h_{\theta*\gamma} - \rho_{\theta*\gamma}) - \tfrac{1}{4}(h_{\beta*\gamma} + \rho_{\beta*\gamma}) \\ &- \tfrac{1}{4}(h^{\tau}_{\theta*\gamma} - \rho^{\tau}_{\theta*\gamma}) - \tfrac{1}{4}(h^{\tau}_{\beta*\gamma} + \rho^{\tau}_{\beta*\gamma}), \end{aligned}$$

where $\rho_{\beta*\gamma} = \eta_{\beta*\gamma}(0) - \eta_\theta(0)$ and $\rho^{\tau}_{\beta*\gamma} = \eta^{\tau}_{\beta*\gamma}(0) - \eta^{\tau}_\theta(0)$ are ρ–invariants of the manifold $Y \# \mathbb{R}\mathrm{P}^3$.

The connection A in this formula is any equivariant connection whose limits are the flat connections $\beta * \gamma$ and $\theta * \gamma$ at the two ends of X, hence we are free to choose A

to equal γ over $\mathbb{R}\times(\mathbb{RP}^3 - D^3)$ and to be trivial in the gluing region. This determines the integral term in the above formula as follows:

$$\int_X p_1(A) = \int_{\mathbb{R}\times Y} p_1(A).$$

To evaluate the ρ–invariants, build a cobordism W from the disjoint union $Y \cup \mathbb{RP}^3$ to the connected sum $Y \,\#\, \mathbb{RP}^3$ by attaching a 1–handle to $[0,1]\times(Y\cup\mathbb{RP}^3)$. The flat connection $\beta*\gamma$ extends to W making it into a flat cobordism from $(Y,\beta)\cup(\mathbb{RP}^3,\gamma)$ to $(Y\,\#\,\mathbb{RP}^3, \beta*\gamma)$. It then follows from [5, Theorem 2.4] that

$$\rho_{\beta*\gamma} - \rho_\beta - \rho_\gamma = \operatorname{sign}_{\beta*\gamma} W - 3\operatorname{sign} W,$$

where ρ_β and ρ_γ are ρ–invariants of the manifolds Y and $\mathbb{RP}^3$, respectively. One can easily see from the description of W that both signature terms in the above formula vanish, implying that $\rho_{\beta*\gamma} = \rho_\beta + \rho_\gamma$. Since the involution τ extends to W, a similar argument using the index theorem of Donnelly [12] instead of [5, Theorem 2.4] shows that $\rho^\tau_{\beta*\gamma} = \rho^\tau_\beta + \rho^\tau_\gamma$. Similar formulas also hold with $\theta*\gamma$ in place of $\beta*\gamma$.

Plugging all of this back into the above index formula and keeping in mind that $\rho_\theta = \rho^\tau_\theta = 0$, we obtain

$$\begin{aligned}\operatorname{ind}\mathcal{D}^\tau_A(\beta*\gamma,\theta*\gamma) = -\int_{\mathbb{R}\times Y} & p_1(A)\\ & -\tfrac14(h_{\beta*\gamma}+\rho_\beta) - \tfrac14 h_{\theta*\gamma} - \tfrac14(h^\tau_{\beta*\gamma}+\rho^\tau_\beta) - \tfrac14 h^\tau_{\theta*\gamma}.\end{aligned}$$

On the other hand, one can apply the formula of Proposition 3.11 to the manifold $X = \mathbb{R}\times Y$ to obtain

$$\operatorname{ind}\mathcal{D}^\tau_A(\beta,\theta) = -\int_{\mathbb{R}\times Y} p_1(A) - \tfrac14(h_\beta+\rho_\beta) - \tfrac14 h_\theta - \tfrac14(h^\tau_\beta+\rho^\tau_\beta) - \tfrac14 h^\tau_\theta.$$

Therefore,

$$\begin{aligned}\operatorname{ind}\mathcal{D}^\tau_A(\beta*\gamma,\theta*\gamma) - \operatorname{ind}\mathcal{D}^\tau_A(\beta,\theta) &= -\tfrac14(h_{\beta*\gamma}-h_\beta) - \tfrac14(h_{\theta*\gamma}-h_\theta)\\ &\qquad - \tfrac14(h^\tau_{\beta*\gamma}-h^\tau_\beta) - \tfrac14(h^\tau_{\theta*\gamma}-h^\tau_\theta),\end{aligned}$$

and the proof of the proposition reduces to a calculation with twisted cohomology.

Since Y is a rational homology sphere, $H^1(Y;\operatorname{ad}\theta)=0$, which implies that

$$h_\theta = \dim H^0(Y;\operatorname{ad}\theta) = 3 \quad\text{and}\quad h^\tau_\theta = \operatorname{tr}(\operatorname{Ad} u) = -1.$$

It follows from a calculation in Section 5 that $H^1(Y\,\#\,\mathbb{RP}^3;\operatorname{ad}(\theta*\gamma))=0$. Therefore, $h_{\theta*\gamma} = \dim H^0(Y;\operatorname{ad}(\theta*\gamma)) = 1$ because $H^0(Y;\operatorname{ad}(\theta*\gamma))$ is the $(+1)$–eigenspace of $\mathrm{Ad}(k)\colon \mathfrak{so}(3) \to \mathfrak{so}(3)$. The operator $\mathrm{Ad}\, i$ acts as minus identity on the $(+1)$–eigenspace of $\mathrm{Ad}\, k$, making $h^{\tau}_{\theta * \gamma} = -1$.

The calculation with $\beta * \gamma$ will rely on the Mayer–Vietoris exact sequence of the splitting $Y \# \mathbb{R}\mathrm{P}^3 = Y_0 \cup \mathbb{R}\mathrm{P}^3_0$ with twisted coefficients:

$$0 \to H^0(Y \# \mathbb{R}\mathrm{P}^3; \mathrm{ad}(\beta * \gamma)) \to H^0(Y; \mathrm{ad}\, \beta) \oplus H^0(\mathbb{R}\mathrm{P}^3; \mathrm{ad}\, \gamma) \\ \to H^0(S^2; \mathrm{ad}\, \theta) \to H^1(Y \# \mathbb{R}\mathrm{P}^3; \mathrm{ad}(\beta * \gamma)) \\ \to H^1(Y; \mathrm{ad}\, \beta) \oplus H^1(\mathbb{R}\mathrm{P}^3; \mathrm{ad}\, \gamma) \to 0.$$

Since β is reducible but nontrivial, $H^0(Y; \mathrm{ad}\, \beta) = \mathbb{R}$. Therefore, keeping in mind that $H^0(S^2; \mathrm{ad}\, \theta) = \mathbb{R}^3$, $H^0(\mathbb{R}\mathrm{P}^3; \mathrm{ad}\, \gamma) = \mathbb{R}$, and $H^1(\mathbb{R}\mathrm{P}^3; \mathrm{ad}\, \gamma) = 0$, we obtain

$$h_{\beta * \gamma} - h_{\beta} = 2 \cdot \dim H^0(Y \# \mathbb{R}\mathrm{P}^3; \mathrm{ad}(\beta * \gamma)).$$

The involution τ induces involutions $\tilde{\tau}^*$ on each of the groups in the Mayer–Vietoris exact sequence comprising a chain map. Keeping in mind that the traces of $\tilde{\tau}^*$ are equal to -1 on both $H^0(S^2; \mathrm{ad}\, \theta) = \mathbb{R}^3$ and $H^0(\mathbb{R}\mathrm{P}^3; \mathrm{ad}\, \gamma) = \mathbb{R}$, we obtain

$$h^{\tau}_{\beta * \gamma} - h^{\tau}_{\beta} = 2\, \mathrm{tr}\big(\tilde{\tau}^* | H^0(Y \# \mathbb{R}\mathrm{P}^3; \mathrm{ad}(\beta * \gamma))\big) - 2\, \mathrm{tr}\big(\tilde{\tau}^* | H^0(Y; \mathrm{ad}\, \beta)\big).$$

Even though both β and γ are reducible, the representation $\beta * \gamma$ may be either reducible or irreducible. In the former case, $H^0(Y \# \mathbb{R}\mathrm{P}^3; \mathrm{ad}(\beta * \gamma)) = \mathbb{R}$ is the $(+1)$–eigenspace of the operator $\mathrm{Ad}\, k\colon \mathfrak{so}(3) \to \mathfrak{so}(3)$ on which $\tilde{\tau}^*$ acts as minus identity, therefore

$$h_{\beta * \gamma} - h_{\beta} = 2 \quad \text{and} \quad h^{\tau}_{\beta * \gamma} - h^{\tau}_{\beta} = 0.$$

In the latter case, $H^0(Y \# \mathbb{R}\mathrm{P}^3; \mathrm{ad}(\beta * \gamma)) = 0$, therefore

$$h_{\beta * \gamma} - h_{\beta} = 0 \quad \text{and} \quad h^{\tau}_{\beta * \gamma} - h^{\tau}_{\beta} = 2.$$

In both cases, we conclude that

$$\mathrm{ind}\, \mathcal{D}^{\tau}_{A}(\beta * \gamma, \theta * \gamma) = \mathrm{ind}\, \mathcal{D}^{\tau}_{A}(\beta, \theta).$$

The result now follows from the fact that $\mathrm{ind}\, \mathcal{D}^{\tau}_{A}(\beta, \theta) = \mathrm{ind}\, \mathcal{D}^{\tau}_{B}(\beta, \theta) \pmod 4$ for any choice of connections A and B on the cylinder $\mathbb{R} \times Y$ whose limits are β and θ over the negative and positive ends. □

Remark 6.2 The formula of Proposition 6.1 holds as well for equivariant irreducible representations β, the proof requiring just minor adjustments.

Combining Proposition 6.1 with formula (14), we obtain the following formula for the Floer grading.

Corollary 6.3 *Let $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ be a nontrivial equivariant reducible representation. Then the Floer grading of the Morse–Bott circle arising from $\beta * \gamma$ is given by*

$$\mu = \operatorname{ind} \mathcal{D}^{\tau}_{B}(\beta, \theta) + \operatorname{sign} k \pmod 4,$$

where B is an arbitrary equivariant connection on the infinite cylinder $\mathbb{R} \times Y$ whose limits are β and θ over the negative and positive ends.

The index $\operatorname{ind} \mathcal{D}^{\tau}_{B}(\beta, \theta)$ in the above corollary can be computed using the formula

$$\operatorname{ind} \mathcal{D}^{\tau}_{B}(\beta, \theta) = \tfrac{1}{2} \operatorname{ind} \mathcal{D}_{B}(\beta, \theta) + \tfrac{1}{2} \operatorname{ind}(\tau, \mathcal{D}_{B})(\beta, \theta). \tag{15}$$

According to Donnelly [12],

$$\operatorname{ind}(\tau, \mathcal{D}_{B})(\beta, \theta) = \frac{1}{2} \int_F (e(TF) + e(NF)) - \tfrac{1}{2}(h^{\tau}_{\theta} - \eta^{\tau}_{\theta}(0))(Y) - \tfrac{1}{2}(h^{\tau}_{\beta} + \eta^{\tau}_{\beta}(0))(Y),$$

where the integral term vanishes and $h^{\tau}_{\beta} = h^{\tau}_{\theta} = -1$ as in the proof of Proposition 6.1. Therefore,

$$\operatorname{ind}(\tau, \mathcal{D}_{B})(\beta, \theta) = 1 - \tfrac{1}{2} \cdot \rho^{\tau}_{\beta}(Y). \tag{16}$$

The ρ–invariants in this formula are difficult to compute in general but they can be shown to vanish in several special cases, for example for two-bridge knots, as discussed in Section 7.1.

6.2 Nondihedral representations

Let $\rho_0\colon \pi_1 K \to \mathrm{SO}(3)$ be an irreducible trace-free representation which is not dihedral, and assume that it is nondegenerate. Proposition 4.1(c) then tells us that the fiber $C(\rho_0)$ consists of two circles which are permuted by the involution χ_k.

Lemma 6.4 *The involution χ_k permuting the two circles in $C(\rho_0)$ has degree zero mod 4.*

Proof This follows as in Lemma 2.5 whose proof in Section 3.7 needs to be amended to allow for the 1–dimensional critical point sets $C(\rho_0)$. This is easily accomplished by replacing $\operatorname{gr}(\chi_1 \cdot \rho, \rho)$ with $\operatorname{gr}(\chi_1 \cdot \rho, \rho) + 1$ in the first two displayed formulas. □

Therefore, the two circles in $C(\rho_0)$ have the same Morse–Bott index μ. Perturbing both of them, we obtain four generators, two of grading μ and two of grading $\mu + 1$. The calculation of the previous section leading up to the formula of Corollary 6.3 can be easily amended to work in the current situation, producing the following result.

Proposition 6.5 *Let $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ be an irreducible representation. Then the Floer grading of the two Morse–Bott circles arising from $\beta * \gamma$ is*

$$\mu = \operatorname{ind} \mathcal{D}^{\tau}_{\boldsymbol{B}}(\beta,\theta) + \operatorname{sign} k \pmod 4,$$

where B is an arbitrary equivariant connection on the infinite cylinder $\mathbb{R}\times Y$ whose limits are β and θ over the negative and positive ends.

The index $\operatorname{ind} \mathcal{D}^{\tau}_{\boldsymbol{B}}(\beta,\theta)$ in this proposition can be computed using the formula (15). Since h^{τ}_{β} now vanishes, the formula (16) takes the form

$$\operatorname{ind}(\tau,\mathcal{D}_{\boldsymbol{B}})(\beta,\theta) = \tfrac{1}{2} - \tfrac{1}{2}\cdot \rho^{\tau}_{\beta}(Y). \tag{17}$$

Remark 6.6 Let $\Delta(t)$ be the Alexander polynomial of a knot $k \subset S^3$ normalized so that $\Delta(t) = \Delta(t^{-1})$ and $\Delta(1) = 1$. The knots k with $\Delta(-1) = 1$ are precisely the knots whose double branched covers Y are integral homology spheres, and which are known to have no dihedral representations in $\mathcal{R}_0(K,\mathrm{SO}(3))$; see [23, Theorem 10] or [11, Proposition 3.4]. Therefore, all the generators in $IC^{\natural}(k)$ are of the nondihedral type studied in this section. In addition, $\operatorname{sign} k = 0 \pmod 8$ because $1 = \Delta(-1) = \det(i\cdot Q)$, where Q is the (even) quadratic form of the knot.

7 Knot homology: explicit calculations

The equivariant techniques work particularly well for Montesinos knots, including two-bridge and pretzel knots, as we will demonstrate in this section. We begin with two-bridge knots, then discuss the Montesinos knots whose double branched covers are integral homology spheres, and then move on to the general Montesinos knots. We finish with a short section on torus knots.

7.1 Two-bridge knots

Let p be an odd positive integer and k a two-bridge knot of type $-p/q$ in the 3–sphere. Its double branched cover Y is the lens space $L(p,q)$ oriented as the $(-p/q)$–surgery on an unknot in S^3. One can use Proposition 3.1 to show that all representations $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ are equivariant. The invariant $\rho^{\tau}_{\beta}(Y)$ of formula (16) has been shown to vanish in [36, Proposition 27]. Therefore, $\operatorname{ind}(\tau,\mathcal{D}_{\boldsymbol{B}})(\beta,\theta) = 1$ and formula (15) reduces to

$$\operatorname{ind} \mathcal{D}^{\tau}_{\boldsymbol{B}}(\beta,\theta) = \tfrac{1}{2}(\operatorname{ind} \mathcal{D}_{\boldsymbol{B}}(\beta,\theta) + 1) \pmod 4.$$

Let $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ be a representation sending the canonical generator of $\pi_1 Y$ to the adjoint of $\exp(2\pi i\ell/p)$. The quantity $\operatorname{ind}\mathcal{D}_{\boldsymbol{B}}(\beta,\theta)+1 \pmod 8$ was shown by Sasahira [34, Corollary 4.3] (see also Austin [7]) to equal

$$2N_1(k_1,k_2)+N_2(k_1,k_2) \pmod 8,$$

where the integers $0<k_1<p$ and $0<k_2<p$ are uniquely determined by the equations

$$k_1=\ell \pmod p, \quad k_2=-r\ell \pmod p, \quad qr=1 \pmod p,$$

and

$$N_1(k_1,k_2)=\#\big\{(i,j)\in\mathbb{Z}^2 \mid i+qj=0 \pmod p, |i|<k_1, |j|<k_2\big\},$$

$$N_2(k_1,k_2)=\#\big\{(i,j)\in\mathbb{Z}^2 \mid i+qj=0 \pmod p \text{ and } \text{either } |i|=k_1, |j|<k_2 \text{ or } |i|<k_1, |j|=k_2\big\}.$$

Example 7.1 The figure-eight knot k is the two-bridge knot of type $-\frac{5}{3}$. Its double branched cover is the lens space $L(5,3)$, whose fundamental group has no irreducible representations and has two nontrivial reducible representations, up to conjugacy. For these two representations, ℓ equals 1 and 2 and, by Sasahira's formula, $\operatorname{ind}\mathcal{D}_{\boldsymbol{B}}(\beta,\theta)+1$ equals 2 and 4 mod 8, respectively. Since $\operatorname{sign} k=0$, the corresponding Morse–Bott circles have indices $\mu=1$ and 2 mod 4 by Corollary 6.3. After perturbation, they contribute the generators of Floer indices 1, 2 and 2, 3 mod 4, respectively. The ranks of the chain groups $IC^\natural(k)$ are then equal to $(1,0,0,0)+(0,1,1,0)+(0,0,1,1)=(1,1,2,1)$. This equals the Khovanov homology of (the mirror image of) k, hence we conclude from the Kronheimer–Mrowka spectral sequence that the ranks of $I^\natural(k)$ also equal $(1,1,2,1)$.

7.2 Special Montesinos knots

Let p, q, and r be pairwise relatively prime positive integers, and view the Brieskorn homology sphere $\Sigma(p,q,r)$ as the link of the singularity at zero of the complex polynomial $x^p+y^q+z^r$. The involution τ induced by complex conjugation on the link makes $\Sigma(p,q,r)$ into a double branched cover of S^3 with branch set a Montesinos knot which will be called $k(p,q,r)$; see for instance [36, Section 7].

Since $\Sigma(p,q,r)$ is an integral homology sphere, apart from the trivial one, all representations $\beta\colon \pi_1(\Sigma(p,q,r)) \to \mathrm{SO}(3)$ are irreducible. Fintushel and Stern [15] showed that all irreducible representations β are nondegenerate and, up to conjugation, there are $-2\lambda(\Sigma(p,q,r))$ of them, where $\lambda(\Sigma(p,q,r))$ is the Casson invariant of $\Sigma(p,q,r)$. The representations β are also equivariant (see [36, Proposition 8]), hence each conjugacy class of them contributes four generators to the chain complex $IC^\natural(k(p,q,r))$, two of grading $\mu(\beta)$ and two of grading $\mu(\beta)+1$.

Theorem 7.2 *The ranks of the chain groups $IC^\natural(k(p,q,r))$ are $(1+b,b,b,b)$, where $b=-2\lambda(\Sigma(p,q,r))$.*

Proof Our proof will use the flat cobordism of Fintushel and Stern [15], which is constructed as follows. The mapping torus of the Seifert fibration $\Sigma(p,q,r)\to S^2$ is an orbifold with three singular points whose neighborhoods are open cones over lens spaces. The compact manifold obtained from W by excising these cones is an equivariant flat cobordism W_0 between $\Sigma(p,q,r)$ and the lens spaces. One can easily see that the intersection form on $H^2(W_0;\mathbb{R})=\mathbb{R}$ is negative definite.

An equivariant version of [5, Theorem 2.4] together with the vanishing of the ρ^τ–invariants of lens spaces [36, Proposition 27] imply that

$$\rho^\tau_\beta(\Sigma(p,q,r)) = \operatorname{sign}_\beta(\tau,W_0)-\operatorname{sign}_\theta(\tau,W_0),$$

where

$$\operatorname{sign}_\beta(\tau,W_0)=\operatorname{tr}(\tilde{\tau}^*\,|\,H^2_+(W_0;\operatorname{ad}\beta))-\operatorname{tr}(\tilde{\tau}^*\,|\,H^2_-(W_0;\operatorname{ad}\beta)),$$

and similarly for $\operatorname{sign}_\theta(\tau,W_0)$. It follows from [15, Proposition 2.5 and Lemma 2.6] that $H^2(W_0;\operatorname{ad}\beta)=0$, hence $\rho^\tau_\beta(\Sigma(p,q,r))=\operatorname{tr}(\operatorname{Ad}u)=-1$ and $\operatorname{ind}(\tau,\mathcal{D}_{\boldsymbol{B}})(\beta,\theta)=1$ by formula (17). Proposition 6.5 and formula (15) now imply that

$$\mu(\beta)=\tfrac{1}{2}(\operatorname{ind}\mathcal{D}_{\boldsymbol{B}}(\beta,\theta)+1).$$

The index $\operatorname{ind}\mathcal{D}_{\boldsymbol{B}}(\beta,\theta)$ in this formula can be computed explicitly using either [15] or Corollary 7.7, however, this alone will not lead us to the closed-form formula of Theorem 7.2.

Instead, we will use the 4–periodicity in the instanton Floer homology due to Frøyshov [17, Theorem 2]. In the case at hand, the Floer homology of $\Sigma(p,q,r)$ equals its Floer chain complex, whose generators are the conjugacy classes of irreducible representations β, hence the 4–periodicity simply means that there is a (noncanonical) free involution of degree 4 on these generators. For any pair of generators β_1 and β_2,

$$\mu(\beta_2)-\mu(\beta_1)=\tfrac{1}{2}(\operatorname{ind}\mathcal{D}_{\boldsymbol{B}}(\beta_2,\theta)-\operatorname{ind}\mathcal{D}_{\boldsymbol{B}}(\beta_1,\theta)) \pmod 4,$$

which is exactly half the relative grading of the generators β_1 and β_2 in the Floer chain complex of $\Sigma(p,q,r)$. For any involutive pair (β_1,β_2), we have

$$\mu(\beta_2)-\mu(\beta_1)=2 \pmod 4,$$

therefore, each such pair contributes $(2,2,2,2)$ to the chain complex $IC^\natural(k(p,q,r))$. The special generator α resides in degree zero so the result follows. □

Example 7.3 $\Sigma(2,3,7)$ is a double branched cover of S^3 whose branch set $k(2,3,7)$ is the pretzel knot $P(-2,3,7)$. Since $\lambda(\Sigma(2,3,7)) = -1$, we conclude that the ranks of the chain groups $IC^{\natural}(P(-2,3,7))$ are $(3,2,2,2)$. This is consistent with the calculation in [18, Section 5].

We expect that the formula of Theorem 7.2 can be proved for all Seifert fibered homology spheres $\Sigma(a_1,\dots,a_n)$ and the corresponding Montesinos knots $k(a_1,\dots,a_n)$ using τ–equivariant perturbations of [38] modeled after the perturbations of Kirk and Klassen [22]. Note that the action of $H^1(K;\mathbb{Z}/2)$ on the conjugacy classes of projective representations is free hence it causes no equivariant transversality issues.

7.3 General Montesinos knots

Let $(a_1,b_1),\dots,(a_n,b_n)$ be pairs of integers such that, for each i, the integers a_i and b_i are relatively prime and a_i is positive. Burde and Zieschang [9, Chapter 7] associated with these pairs a Montesinos link $K((a_1,b_1),\dots,(a_n,b_n))$ and showed that its double branched cover is a Seifert fibered manifold Y with unnormalized Seifert invariants $(a_1,b_1),\dots,(a_n,b_n)$. In particular,

$$\pi_1 Y = \langle x_1,\dots x_n, h \mid h \text{ central},\ x_i^{a_i} = h^{-b_i},\ x_1\cdots x_n = 1\rangle,$$

with the covering translation $\tau\colon Y \to Y$ acting on the fundamental group by the rule

$$\tau_*(h) = h^{-1}, \qquad \tau_*(x_i) = x_1\cdots x_{i-1}x_i^{-1}x_{i-1}^{-1}\cdots x_1^{-1} \quad \text{for } i = 1,\dots,n;$$

see Burde and Zieschang [9, Proposition 12.30]. The knots $k(a_1,\dots,a_n)$ of the previous section are of the type $K((a_1,b_1),\dots,(a_n,b_n))$; we omitted the parameters $(b_1,\dots,b_n)$ from the notation because they can be uniquely recovered from the pairwise relatively prime $a_1,\dots,a_n$ up to isotopy of the knot. All two-bridge and pretzel knots and links are special cases of Montesinos knots and links. In this section, we will only be interested in Montesinos knots; the case of Montesinos links of two components will be addressed in Section 8.3.

Let k be a Montesinos knot $K((a_1,b_1),\dots,(a_n,b_n))$ and Y the double branched cover of S^3 with branch set k. The manifold Y need not be an integral homology sphere; in fact, one can easily see that its first homology is a finite abelian group of order

$$|H_1(Y;\mathbb{Z})| = \left(\sum_{i=1}^n b_i/a_i\right)\cdot a_1\cdots a_n.$$

Note that this integer is always odd because Y is a $\mathbb{Z}/2$ homology sphere.

All reducible representations $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ are equivariant because the involution $\tau_*\colon H_1(Y) \to H_1(Y)$ acts as multiplication by -1; see Proposition 3.1. There are no irreducible representations for $n \le 2$. If $n = 3$, all irreducible representations are nondegenerate and equivariant, which can be shown using a minor modification of the arguments of [15, Proposition 2.5] and [36, Proposition 30]. For $n \ge 4$, one encounters positive-dimensional manifolds of representations; the action of τ^* on these manifolds was described in [38], together with equivariant perturbations making them nondegenerate. This discussion, together with Propositions 4.1 and 4.2, identifies the generators of the chain complex $IC^\natural(k)$ for all Montesinos knots in terms of representations for Seifert fibered manifolds, which are well known. An independent calculation of the generators of $IC^\natural(k)$ for pretzel knots k with $n = 3$ can be found in Zentner [43].

Let W_0 be the mapping cylinder of the Seifert fibration $Y \to S^2$ with excised open cones around its singular points. Then W_0 is a cobordism from a disjoint union of the lens spaces $L(a_i, -b_i)$ to Y.

Lemma 7.4 *W_0 is a flat cobordism provided $a_1 \cdots a_n = \mathrm{lcm}(a_1, \dots, a_n)\cdot |H_1(Y;\mathbb{Z})|$.*

Proof The fundamental group $\pi_1 W_0$ is obtained from $\pi_1 Y$ by setting the homotopy class $h \in \pi_1 Y$ of the circle fiber equal to one. Since h is a central element in $\pi_1 Y$, every irreducible representation $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ has the property that $\beta(h) = 1$. This property need not hold for reducible representations but it does if $h = 1$ in the first homology group $H_1(Y)$. The algebraic condition of the lemma ensures exactly that; see Lee and Raymond [26, page 331]. □

To avoid dealing with perturbations, we will assume from now on that our knot k is a Montesinos knot of type $K((a_1, b_1)(a_2, b_2), (a_3, b_3))$ and that W_0 is a flat cobordism. We wish to calculate Floer gradings of the generators in the chain complex $IC^\natural(k)$. Recall that every conjugacy class of nontrivial reducible representations $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ gives rise to two generators of gradings $\mu(\beta)$ and $\mu(\beta)+1$, and every conjugacy class of irreducible representations to four generators, two of grading $\mu(\beta)$ and two of grading $\mu(\beta) + 1$. The trivial representation as usual gives rise to just one generator α of grading $\operatorname{sign} k$.

Lemma 7.5 *For any nontrivial representation $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$, we have*

$$\mu(\beta) = \operatorname{sign} k + \tfrac{1}{2}(\operatorname{ind} \mathcal{D}_{\boldsymbol{B}}(\beta, \theta) + 1) \pmod 4.$$

Proof This formula holds for all irreducible representations β by the same argument as in the proof of Theorem 7.2. That argument can be easily amended for nontrivial reducible representations $\beta\colon \pi_1 Y \to \mathrm{SO}(3)$ by using (16) in place of (17). The ρ–invariant in formula (16) is given by the formula

$$\rho^\tau_\beta(Y) = \mathrm{sign}_\beta(\tau, W_0) - \mathrm{sign}_\theta(\tau, W_0),$$

with $\mathrm{sign}_\theta(\tau, W_0) = 1$. To compute the cohomology of W_0 with coefficients in $\mathrm{ad}\,\beta$, write $\mathrm{ad}\, P = \mathbb{R} \oplus L$, where L is a line bundle with a nontrivial flat connection. Then $H^2(W_0; L) = 0$ by the argument of [15, Lemma 2.6] and $H^2(W_0; \mathbb{R}) = \mathbb{R}$. Since the manifold W_0 is negative definite, we easily conclude that $\mathrm{sign}_\beta(\tau, W_0) = 1$. Therefore, $\rho^\tau_\beta(Y) = 0$, and the result follows. □

To complete the calculation of Floer gradings, we only need to compute the index $\mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(\beta, \theta)$. This can be done by extending the formulas of Fintushel and Stern [15] from integral homology spheres to the more general situation at hand. We will restrict ourselves to the relatively easy case of odd a_i and leave the case of even a_i open because it would require passing to a double branched cover as in the proof of [15, Theorem 3.7].

Given a flat cobordism W_0, any representation $\beta\colon \pi_1(Y) \to \mathrm{SO}(3)$ gives rise to a representation $\pi_1(W_0) \to \mathrm{SO}(3)$ and to representations $\beta_i\colon \pi_1(L(a_i, -b_i)) \to \mathrm{SO}(3)$. Let us assume that a_i are odd and $\beta_i \neq \theta$ for $i = 1, \dots, m$, and that $\beta_i = \theta$ for $i = m+1, \dots, 3$. Applying the excision principle for the ASD operator twice, first to $\mathbb{R} \times L(a_i, -b_i)$ with $i = 1, \dots, m$, and then to W_0 with the attached product ends, we obtain

$$-3 = \mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(\theta, \theta) = \mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(\theta, \beta_i) + 1 + \mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(\beta_i, \theta)$$

$$-3 = \mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(W_0, \theta, \theta) = \sum_{i=1}^{m} (\mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(\theta, \beta_i) + 1) + \mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(W_0) + 1 + \mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(\beta, \theta),$$

where $\mathcal{D}_{\boldsymbol{B}}(W_0)$ stands for the ASD operator on W_0 twisted by a flat connection B whose holonomy is the representation $\pi_1(W_0) \to \mathrm{SO}(3)$. A similar argument with even a_i does not work because representations β_i and θ may end up living in different SO(3)–bundles.

Lemma 7.6 *Let $\beta\colon \pi_1(Y) \to \mathrm{SO}(3)$ be a nontrivial representation. Then $\mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(W_0) = -1$ if β is reducible, and $\mathrm{ind}\,\mathcal{D}_{\boldsymbol{B}}(W_0) = 0$ if β is irreducible.*

Proof The proof of [15, Proposition 3.3] implies the formula for irreducible β immediately, and for reducible β after a minor modification. To be precise, let us assume that β is reducible. The index at hand equals $h^1 - h^0 - h^2$, where h^0, h^1, and h^2 are the Betti numbers of the elliptic complex

$$0 \to \Omega^0(W_0, \operatorname{ad} P) \xrightarrow{-d_B} \Omega^1(W_0, \operatorname{ad} P) \xrightarrow{d_B^+} \Omega^2_+(W_0, \operatorname{ad} P).$$

Since B has 1–dimensional stabilizer we immediately conclude that $h^0 = 1$. To compute the remaining Betti numbers, write $\operatorname{ad} P = \mathbb{R} \oplus L$, where L is a line bundle with a nontrivial flat connection. The argument of [15, Lemma 2.6] can be used to show that the homomorphisms $H^1(W_0; L) \to H^1(Y; L)$ and $H^2(W_0; L) \to H^2(Y; L)$ induced by the inclusion $Y \to W_0$ are injective. Both $H^1(W_0; \mathbb{R})$ and $H^1(Y; \mathbb{R})$ vanish, and the long exact sequence of the pair (W_0, Y) shows that the kernel of the map $H^2(W_0; \mathbb{R}) \to H^2(Y; \mathbb{R})$ is 1–dimensional. Keeping in mind that the manifold W_0 is negative definite, we conclude as in the proof of [15, Proposition 3.3] that $h^1 = h^2 = 0$. □

Corollary 7.7 *Let $\beta\colon \pi_1(Y) \to \mathrm{SO}(3)$ be a nontrivial representation such that a_i is odd and $\beta_i \neq \theta$ for $i = 1, \dots, m$, and $\beta_i = \theta$ for $i = m+1, \dots, 3$. Then*

$$\mu(\beta) = \operatorname{sign} k - 1 + \frac{1}{2} \sum_{i=1}^{m} (\operatorname{ind} \mathcal{D}_B(\beta_i, \theta) + 3) \pmod 4,$$

where the index $\operatorname{ind} \mathcal{D}_B(\beta_i, \theta)$ on the infinite cylinder $\mathbb{R} \times L(a_i, -b_i)$ can be computed as in Section 7.1.

Example 7.8 Let us view the pretzel knot $k = P(-2, 3, 3)$ as the Montesinos knot $K((2, -1), (3, 1), (3, 1))$. It obviously satisfies the condition of Lemma 7.4. Its double branched cover is a Seifert fibered manifold Y whose fundamental group has presentation

$$\langle x_1, x_2, x_3, x_4, h \mid h \text{ central}, \ x_1^2 = h, \ x_2^3 = h^{-1}, \ x_3^3 = h^{-1}, \ x_1 x_2 x_3 = 1 \rangle.$$

This group admits one nontrivial reducible representation β with $\beta(x_1) = 1$, $\beta(x_2) = \operatorname{Ad}(\exp(2\pi i/3))$ and $\beta(x_3) = \operatorname{Ad}(\exp(-2\pi i/3))$ contributing generators of gradings μ and $\mu + 1$ to the chain complex $IC^\natural(k)$. To compute μ, we apply the formulas of Section 7.1 to the lens space $L(3, -1) = L(3, 2)$ twice to obtain $\operatorname{ind} \mathcal{D}_B(\beta_2, \theta) = \operatorname{ind} \mathcal{D}_B(\beta_3, \theta) = 1 \pmod 8$. Since $\operatorname{sign} k = 2 \pmod 4$, it follows from Corollary 7.7 that $\mu = 1 \pmod 4$ hence the contribution of β to the chain complex is $(0, 1, 1, 0)$. The special generator α contributes $(0, 0, 1, 0)$.

The group $\pi_1 Y$ also admits one irreducible representation β such that all of the induced representations $\beta_1\colon \pi_1(L(2, 1)) \to \mathrm{SO}(3)$ and $\beta_2, \beta_3\colon \pi_1(L(3, 2)) \to \mathrm{SO}(3)$

are nontrivial. Corollary 7.7 no longer applies, hence we can only conclude that the contribution of this representation to $IC^{\natural}(k)$ is $(2,2,0,0)$ up to cyclic permutation.

This information can be combined with the fact that the Kronheimer–Mrowka spectral sequence of the knot $k = P(-2,3,3)$ is trivial and that the Khovanov homology groups of k have ranks $(2,1,1,1)$; see Lobb and Zentner [28]. It then follows that the ranks of the chain groups $IC^{\natural}(k)$ must be $(2,1,2,2)$, with the contribution of the irreducible being $(2,0,0,2)$, and that the boundary operator $IC^{\natural}_2(k) \to IC^{\natural}_3(k)$ must be nontrivial.

A similar calculation can be done for all Montesinos knots $K((a_1,b_1),\dots,(a_n,b_n))$ satisfying the condition of Lemma 7.4 with the help of the equivariant perturbations of [38].

7.4 Torus knots

Let p and q be positive integers which are odd and relatively prime. The double branched cover of the right-handed (p,q)–torus knot $T_{p,q}$ is the Brieskorn homology sphere $\Sigma(2,p,q)$. According to Fintushel and Stern [15], all irreducible SO(3) representations of the fundamental group of $\Sigma(2,p,q)$ are nondegenerate and, up to conjugacy, there are $-\operatorname{sign}(T_{p,q})/4$ of them. All of these representations are equivariant [11, Section 4.2], hence each of them contributes four generators to the chain complex of $I^{\natural}(T_{p,q})$, two of index μ and two of index $\mu+1$. Calculating μ would require equivariant index theory on the double branched cover of $T_{p,q}$ which is currently not sufficiently well developed. We know that the special generator resides in degree zero because $\operatorname{sign} T_{p,q} = 0 \pmod 8$, and we conjecture that the ranks of the chain groups $IC^{\natural}(T_{p,q})$ are

$$(1+a,a,a,a), \quad \text{where } a = -\operatorname{sign}(T_{p,q})/4.$$

This conjecture is consistent with the calculations for torus knots by Hedden, Herald and Kirk [20].

Let us now assume that p and q are relatively prime positive integers such that p is odd and $q = 2r$ is even. The double branched cover Y, which is no longer an integral homology sphere, is the link of the singularity at zero of the complex polynomial $x^2 + y^p + z^{2r} = 0$, with the covering translation given by the formula $\tau(x,y,z) = (-x,y,z)$. Neumann and Raymond [30] showed that Y admits a fixed-point-free circle action making it into a Seifert fibration over S^2 with the Seifert invariants

$$\{(a_1,b_1),\dots,(a_n,b_n)\} = \{(1,b_1),(p,b_2),(p,b_2),(r,b_3)\},$$

where $b_1 \cdot pr + 2b_2 \cdot r + b_3 \cdot p = 1$. In principle, this allows for calculation of the generators in the Floer chain complex $IC^\natural(T_{p,q})$.

Example 7.9 Let us consider the torus knot $T_{3,4}$. The Seifert invariants of the manifold Y are $\{(1,-1),(2,1),(3,1),(3,1)\}$, while those of the manifold in Example 7.8 are $\{(2,-1),(3,1),(3,1)\}$. These match for the good reason that $P(-2,3,3)$ and $T_{3,4}$ are the same knot. The calculation of Example 7.8 then tells us that the ranks of the chain groups $IC^\natural(T_{3,4})$ are $(2,1,2,2)$, with a nontrivial boundary operator $IC_2^\natural(T_{3,4}) \to IC_3^\natural(T_{3,4})$. This is consistent with [20].

8 Link homology of general two-component links

This section deals with general two-component links $\mathcal{L} = \ell_1 \cup \ell_2$ and not just the links $\mathcal{L} = k^\natural$ used in the definition of the knot Floer homology $I^\natural(k)$. After computing the Euler characteristic of $I_*(\Sigma,\mathcal{L})$, we explicitly compute the Floer chain groups for some links $\mathcal{L}$ with particularly simple double branched covers.

8.1 Euler characteristic

Let $\mathcal{L} = \ell_1 \cup \ell_2$ be a two-component link in an integral homology sphere Σ. The linking number $\ell k(\ell_1,\ell_2)$ is well defined up to a sign for nonoriented links $\mathcal{L}$.

Theorem 8.1 *The Euler characteristic of the Floer homology $I_*(\Sigma,\mathcal{L})$ of a two-component link $\mathcal{L} = \ell_1 \cup \ell_2$ equals $\pm \ell k(\ell_1,\ell_2)$.*

Proof The Floer excision principle can be used as in [24] to establish an isomorphism between $I_*(\Sigma,\mathcal{L})$ and the sutured Floer homology of $\mathcal{L}$. The latter is the Floer homology of the 3–manifold X_φ obtained by identifying the two boundary components of $S^3 - \operatorname{int} N(\mathcal{L})$ via an orientation-reversing homeomorphism $\varphi\colon T^2 \to T^2$. According to [19, Lemma 2.1], the homeomorphism φ can be chosen so that X_φ has integral homology of $S^1 \times S^2$. The result then follows from [19, Theorem 2.3], which asserts that the Euler characteristic of the sutured Floer homology of $\mathcal{L}$ equals $\pm \ell k(\ell_1,\ell_2)$. □

Theorem 8.1 implies in particular that the Euler characteristic of $I^\natural(k)$ equals ± 1, which is the linking number of the two components of the link $k^\natural$. This also follows from the fact that the critical point set of the orbifold Chern–Simons functional used to define $I^\natural(k)$ consists of an isolated point and finitely many isolated circles, possibly after a perturbation. An absolute grading on $I^\natural(k)$ was fixed in [24] so that the grading of the isolated point is even; this is consistent with our Theorem 5.1 because $\operatorname{sign} k$ is always even. The Euler characteristic of $I^\natural(k)$ then equals $+1$. We do not know how to fix an absolute grading on $I_*(\Sigma,\mathcal{L})$ for a general two-component link $\mathcal{L}$.

8.2 Pretzel link $P(2,-3,-6)$

This is the two-component link $\mathcal{L}$ whose double branched cover is the Seifert fibered manifold M with unnormalized Seifert invariants $(2,1)$, $(3,-1)$, and $(6,-1)$; see for instance [37, Section 4]. In particular,

$$\pi_1 M = \langle x, y, z, h \mid h \text{ central},\ x^2 = h^{-1},\ y^3 = h,\ z^6 = h,\ xyz = 1\rangle,$$

with the covering translation $\tau\colon M \to M$ acting on the fundamental group by the rule

$$\tau_*(h) = h^{-1}, \quad \tau_*(x) = x^{-1}, \quad \tau_*(y) = xy^{-1}x^{-1}, \quad \tau_*(z) = xyz^{-1}y^{-1}x^{-1};$$

see Burde and Zieschang [9, Proposition 12.30]. The manifold M has integral homology of $S^1 \times S^2$. In fact, it can be obtained by 0–surgery on the right-handed trefoil, so that $\pi_1 M = \pi_1 K/\langle\lambda\rangle$, where K is the exterior of the trefoil and λ is its longitude. The relation $\lambda = 1$ shows up as the relation $z^6 = h$ in the above presentation of $\pi_1 M$.

We will use this surgery presentation of M to describe representations $\pi_1 M \to \mathrm{SO}(3)$ with nontrivial $w_2 \in H^2(M;\mathbb{Z}/2) = \mathbb{Z}/2$. According to Example 2.2, the conjugacy classes of such representations are in one-to-two correspondence with the conjugacy classes of representations $\rho\colon \pi_1 K \to \mathrm{SU}(2)$ such that $\rho(\lambda) = -1$. In the terminology of Section 2.2, these ρ are projective representations $\rho\colon \pi_1 M \to \mathrm{SU}(2)$, and the group $H^1(M;\mathbb{Z}/2) = \mathbb{Z}/2$ acts on them freely, providing the claimed one-to-two correspondence. Therefore, we wish to find all the $\mathrm{SU}(2)$ matrices $\rho(h)$, $\rho(x)$, $\rho(y)$, and $\rho(z)$ such that

$$\rho(x)^2 = \rho(h)^{-1}, \quad \rho(y)^3 = \rho(h), \quad \rho(z)^6 = -\rho(h), \quad \rho(x)\rho(y)\rho(z) = 1$$

and such that $\rho(h)$ commutes with $\rho(x)$, $\rho(y)$, and $\rho(z)$. Since ρ is irreducible, we conclude as in Fintushel and Stern [15, Section 2] that $\rho(h) = -1$ and that $\rho(x)$ is conjugate to i, $\rho(y)$ is conjugate to $e^{\pi i/3}$, and $\rho(z)$ is conjugate to either $e^{\pi i/3}$ or $e^{2\pi i/3}$. These give rise to two conjugacy classes of projective representations $\rho\colon \pi_1 M \to \mathrm{SU}(2)$ corresponding to a single conjugacy class of representations $\mathrm{Ad}\,\rho\colon \pi_1 M \to \mathrm{SO}(3)$.

The arguments of [15, Proposition 2.5] and [36, Proposition 8] can be easily adapted to conclude that the representation $\mathrm{Ad}\,\rho$ is nondegenerate and equivariant. It gives rise to a single $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ orbit of generators in $IC_*(S^3, \mathcal{L})$. Since the linking number between the components of $\mathcal{L}$ is even, Lemma 2.5 tells us that the (relative) Floer indices of these four generators are $0, 0, 2, 2 \pmod 4$. The boundary operators then must vanish, and we conclude that the Floer homology groups $I_k(S^3, \mathcal{L})$ are free abelian groups of ranks $(2, 0, 2, 0)$, up to cyclic permutation.

Remark 8.2 The same result can be obtained independently using the isomorphism between $I_*(S^3, \mathcal{L})$ and the sutured Floer homology of $\mathcal{L}$ defined in [25]. The latter is the Floer homology of the manifold X_φ obtained by identifying the two boundary components of $X = S^3 - \operatorname{int} N(\mathcal{L})$ via an orientation-reversing homeomorphism $\varphi\colon T^2 \to T^2$. A surgery description of X_φ can be found in [19]; computing its Floer homology is then an exercise in applying the Floer exact triangle to this surgery description.

8.3 Montesinos links

Let $(a_1, b_1), \dots, (a_n, b_n)$ be pairs of integers such that, for each i, the integers a_i and b_i are relatively prime and a_i is positive. Associated with these pairs is the Montesinos link $K((a_1, b_1), \dots, (a_n, b_n))$, whose definition can be found for instance in [9, Chapter 7]. All two-bridge and pretzel links are Montesinos links; for example, the link $P(2, -3, -6)$ considered in the previous section is the Montesinos link with the parameters $(2, 1)$, $(3, -1)$ and $(6, -1)$. The double branched covers M of Montesinos links were described in Section 7.3. In this section, we will only be interested in Montesinos links whose double branched covers have integral homology of $S^1 \times S^2$, a condition that is easily checked by abelianizing $\pi_1 M$. This condition guarantees that the unique SO(3)–bundle $P \to M$ with nontrivial $w_2(P) \in H^2(M; \mathbb{Z}/2) = \mathbb{Z}/2$ does not carry any reducible connections.

The generators of Floer chain complex of the link $K((a_1, b_1), \dots, (a_n, b_n))$ and their gradings can be computed explicitly using the equivariant theory developed in this paper; here is a brief outline.

Since M is Seifert fibered, the representations $\pi_1 M \to \mathrm{SO}(3)$ with nontrivial w_2 can be described in terms of their rotation numbers using a slight modification of the Fintushel–Stern algorithm [15]; complete details can be found in [35]. If $n = 3$, there are finitely many conjugacy classes of such representations, all of which are nondegenerate and equivariant with the conjugating element of order 2. If $n \geq 4$, the same conclusion holds after using τ–equivariant perturbations similar to those described in [38]. Note that no equivariant transversality issues are caused by the action of $H^1(M; \mathbb{Z}/2)$ or $H^1(X; \mathbb{Z}/2)$ because both actions are free. In what follows, we will restrict ourselves to the case when $n = 3$; however, we expect that the same results will hold for all n.

The relative indices of the operator $\mathcal{D}_A$ on $\mathbb{R} \times M$ were computed explicitly in [35] and shown to be even. The relative Floer gradings of the generators in the Floer chain complex of the link $K((a_1, b_1), (a_2, b_2), (a_3, b_3))$ are equal to one half times those indices, by the argument of [36, Section 5.2] modified to take into account the nontriviality of the bundle $P \to M$.

The final outcome of this calculation can be stated in terms of the Floer homology groups $I_*(M,P)$ of the unique admissible bundle $P \to M$ as follows. The groups $I_*(M,P)$ are free abelian of ranks (n_0,n_1,n_2,n_3), up to cyclic permutation, with either $n_0 = n_2 = 0$ or $n_1 = n_3 = 0$. Assume for the sake of concreteness that $n_0 = n_2 = 0$. Then the Floer chain groups of $K((a_1,b_1),\dots,(a_n,b_n))$, up to cyclic permutation, have the ranks

$$(2n_1, 2n_3, 2n_1, 2n_3). \tag{18}$$

Example 8.3 The double branched cover M of the Montesinos link

$$\mathcal{L} = K((2,1),(5,-2),(10,-1))$$

can be obtained by 0–surgery on the right-handed torus knot $T_{2,5}$. Applying the Floer exact triangle to this surgery, we see that $I_*(M,P) \oplus I_{*+4}(M,P) = I_*(\Sigma(2,15,11))$, where we use the mod 8 grading in both groups. Fintushel and Stern [15] showed[3] that the groups $I_k(\Sigma(2,5,11))$ are free abelian of the ranks $(0,1,0,2,0,1,0,2)$. Therefore $n_1 = 1$, $n_3 = 2$, and the Floer chain groups of the link $\mathcal{L}$ have the ranks $(2,4,2,4)$.

In fact, the integers n_1 and n_3 in the formula (18) can be computed much more easily in terms of classical knot invariants without any reference to the Floer homology. They are known to satisfy the equations

$$-n_1 - n_3 = \lambda'(M) \quad \text{and} \quad -n_1 + n_3 = \bar{\mu}'(M),$$

where $\lambda'(M)$ is the Casson invariant of M and $\bar{\mu}'(M)$ its Neumann invariant [29]. The former equation follows from the Casson surgery formula and the latter from [37]. The Casson and Neumann invariants can then be computed explicitly using the formulas

$$\lambda'(M) = -\tfrac{1}{2}\cdot\Delta''_M(1) \quad \text{and} \quad \bar{\mu}'(M) = \pm\ell k(\ell_1,\ell_2),$$

where $\Delta_M(t)$ is the Alexander polynomial of M normalized so that $\Delta_M(1) = 1$ and $\Delta(t) = \Delta(t^{-1})$, and $\ell k(\ell_1,\ell_2)$ is the linking number between the components of the link $\mathcal{L}$. Note that there is no need to fix the sign in the above formula because switching that sign preserves the answer (18) up to cyclic permutation.

Appendix: Homology of double branched covers

This section contains a proof of Proposition 3.2 which was postponed until later in Section 3.1.

[3] We adjusted the formulas of [15] to take into account that Fintushel and Stern work with SD rather than ASD equations.

A.1 Computing $H_*(M;\mathbb{Z}/2)$

In this section, we will compute the groups $H_*(M;\mathbb{Z}/2)$ using the transfer homomorphism approach of [27].

The transfer homomorphisms can be defined in the following two equivalent ways; see for instance [14, Section 3]. For each singular simplex $\sigma\colon \Delta \to \Sigma$, choose a lift $\widetilde{\sigma}\colon \Delta \to M$ and define the chain map $\pi_!\colon C_*(\Sigma) \to C_*(M)$ by the formula $\pi_!(\sigma) = \widetilde{\sigma} + \tau \circ \widetilde{\sigma}$. This map is obviously independent of the choice of $\widetilde{\sigma}$, and it induces homomorphisms $\pi_!\colon H_*(\Sigma) \to H_*(M)$ and $\pi^!\colon H^*(M) \to H^*(\Sigma)$ in homology and cohomology with arbitrary coefficients, called transfer homomorphisms. Another way to define $\pi_!$ is as the map that makes the diagram

$$\begin{array}{ccc} H_*(M) & \xleftarrow{\text{PD}} & H^*(M) \\ \uparrow{\scriptstyle \pi_!} & & \uparrow{\scriptstyle \pi^*} \\ H_*(\Sigma) & \xleftarrow{\text{PD}} & H^*(\Sigma) \end{array}$$

commute, where PD stands for the Poincaré duality isomorphism, and similarly for $\pi^!$.

From now on, all chain complexes and (co)homology will be assumed to have $\mathbb{Z}/2$ coefficients. It is then immediate from the definition of $\pi_!\colon C_*(\Sigma) \to C_*(M)$ that $\ker \pi_! = C_*(\mathcal{L})$ and that we have a short exact sequence of chain complexes

$$0 \longrightarrow C_*(\Sigma,\mathcal{L}) \xrightarrow{\pi_!} C_*(M) \xrightarrow{\pi_*} C_*(\Sigma) \longrightarrow 0.$$

This exact sequence induces long exact sequences in homology

$$\begin{aligned} 0 \longrightarrow H_3(\Sigma,\mathcal{L}) &\xrightarrow{\pi_!} H_3(M) \longrightarrow H_3(\Sigma) \\ \longrightarrow H_2(\Sigma,\mathcal{L}) &\xrightarrow{\pi_!} H_2(M) \longrightarrow H_2(\Sigma) \\ \longrightarrow H_1(\Sigma,\mathcal{L}) &\xrightarrow{\pi_!} H_1(M) \longrightarrow H_1(\Sigma) \longrightarrow 0 \end{aligned}$$

and in cohomology

$$\begin{aligned} 0 \longrightarrow H^1(\Sigma) \longrightarrow H^1(M) &\xrightarrow{\pi^!} H^1(\Sigma,\mathcal{L}) \\ \longrightarrow H^2(\Sigma) \longrightarrow H^2(M) &\xrightarrow{\pi^!} H^2(\Sigma,\mathcal{L}) \\ \longrightarrow H^3(\Sigma) \longrightarrow H^3(M) &\xrightarrow{\pi^!} H^3(\Sigma,\mathcal{L}) \to 0. \end{aligned}$$

Combining these with the long exact sequence of the pair $(\Sigma,\mathcal{L})$, we obtain the following result.

Proposition A.1 *Let $\pi\colon M \to \Sigma$ be a double branched cover over an integral homology sphere Σ with branching set a two-component link $\mathcal{L}$. Then $H_i(M;\mathbb{Z}/2) = H^i(M;\mathbb{Z}/2)$ is isomorphic to $\mathbb{Z}/2$ if $i = 0, 1, 2, 3$, and is zero otherwise.*

A.2 The cup product on $H^*(M;\mathbb{Z}/2)$

This section is devoted to the proof of the following result. We continue working with $\mathbb{Z}/2$ coefficients.

Proposition A.2 *The cup product $H^1(M) \times H^1(M) \to H^2(M)$ is the bilinear form $\mathbb{Z}/2 \times \mathbb{Z}/2 \to \mathbb{Z}/2$ with the matrix $\ell k(\ell_1, \ell_2)$ (mod 2).*

Proof We will reduce the cup product calculation to intersection theory using the commutative diagram

$$\begin{array}{ccc} H_2(M) \times H_2(M) & \xrightarrow{\;\cdot\;} & H_1(M) \\ \uparrow{\scriptstyle \mathrm{PD}} & & \uparrow{\scriptstyle \mathrm{PD}} \\ H^1(M) \times H^1(M) & \xrightarrow{\;\cup\;} & H^2(M) \end{array}$$

where PD stands for the Poincaré duality isomorphisms and $\cdot$ for the intersection product. The transfer homomorphism $\pi_!\colon H_*(\Sigma, \mathcal{L}) \to H_*(M)$ will give us explicit generators of $H_1(M)$ and $H_2(M)$ that we need to proceed with this approach.

We begin with the group $H_1(M)$. Note that $H_1(\Sigma, \mathcal{L}) = \mathbb{Z}/2$ is generated by the homology class $[w]$ of any embedded arc $w \subset \Sigma$ whose endpoints belong to two different components of $\mathcal{L}$. The transfer homomorphism $\pi_!\colon H_1(\Sigma, \mathcal{L}) \to H_1(M)$ maps the homology class of w to that of the circle $\pi^{-1}(w)$. Since $\pi_!$ is an isomorphism, we conclude that the circle $\pi^{-1}(w)$ represents a generator of $H_1(M)$.

To describe a generator of $H_2(M)$, observe that $H_2(\Sigma, \mathcal{L}) = \mathbb{Z}/2 \oplus \mathbb{Z}/2$ is generated by the homology classes of Seifert surfaces S_1 and S_2 of the knots ℓ_1 and ℓ_2. We will assume that S_1 and S_2 intersect transversely in a finite number of circles and arcs, and note that $S_1 \cap S_2$ is homologous to $\ell k(\ell_1, \ell_2) \cdot w$. We claim that the closed orientable surfaces $\pi^{-1}(S_1)$ and $\pi^{-1}(S_2)$, representing the homology classes $\pi_!([S_1])$ and $\pi_!([S_2])$, are homologous to each other and generate $H_2(M)$. To see this, we will appeal to Theorem 2 of [27], which supplies us with the commutative diagram with an exact row

$$\begin{array}{ccccccccc} 0 & \longrightarrow & H_3(\Sigma) & \xrightarrow{d_*} & H_2(\Sigma, \mathcal{L}) & \xrightarrow{\pi_!} & H_2(M) & \longrightarrow & 0 \\ & & & \searrow^{f} & \downarrow{\scriptstyle \partial_*} & & & & \\ & & & & H_1(\mathcal{L}) & & & & \end{array}$$

where $f([\Sigma]) = [\ell_1] + [\ell_2]$ and ∂_* is the connecting homomorphism in the long exact sequence of the pair $(\Sigma, \mathcal{L})$. One can easily see that ∂_* is an isomorphism. Since $\partial_*([S_1] + [S_2]) = [\ell_1] + [\ell_2] = f([\Sigma])$, we conclude that $[S_1] + [S_2] \in \operatorname{im} d_* = \ker \pi_!$ and hence $\pi_!([S_1]) = \pi_!([S_2])$ is a generator of $H_2(M)$.

The calculation of the intersection form $H_2(M) \times H_2(M) \to H_1(M)$ is now completed as follows:

$$\begin{aligned}[\pi^{-1}(S_1)] \cdot [\pi^{-1}(S_2)] &= [\pi^{-1}(S_1) \cap \pi^{-1}(S_2)] \\ &= [\pi^{-1}(S_1 \cap S_2)] = \ell k(\ell_1, \ell_2) \cdot [\pi^{-1}(w)]. \qquad \square\end{aligned}$$

Remark A.3 Let $\beta \in H^1(M) = \mathbb{Z}/2$ be a generator and assume that $\ell k(\ell_1, \ell_2)$ is odd. Proposition A.2 implies that $\beta \cup \beta \in H^2(M)$ is nontrivial, and a straightforward argument with Poincaré duality shows that $\beta \cup \beta \cup \beta$ generates $H^3(M)$. If $\ell k(\ell_1, \ell_2)$ is even then $\beta \cup \beta = 0$, and the cup product of β with a generator of $H^2(M)$ generates $H^3(M)$. This gives a complete description of the cohomology ring $H^*(M)$.

Example A.4 The real projective space $\mathbb{R}\mathrm{P}^3$ is a double branched cover over the Hopf link in S^3 with linking number ± 1. Choose Seifert surfaces S_1 and S_2 to be the obvious disks intersecting in a single interval w. Then $\pi^{-1}(S_1)$ and $\pi^{-1}(S_2)$ are two copies of $\mathbb{R}\mathrm{P}^2$, each represented as a double branched cover of a disk with branching set a disjoint union of a circle and a point. These two copies of $\mathbb{R}\mathrm{P}^2$ intersect in the circle $\pi^{-1}(w)$, thereby recovering the familiar cup product structure on $H^*(\mathbb{R}\mathrm{P}^3; \mathbb{Z}/2)$.

References

[1] **S Akbulut**, **R Kirby**, *Branched covers of surfaces in 4–manifolds*, Math. Ann. 252 (1979/80) 111–131 MR

[2] **N Anvari**, *Equivariant rho-invariants and instanton homology of torus knots*, preprint (2016) arXiv To appear in Comm. Anal. Geom.

[3] **M F Atiyah**, **N J Hitchin**, **I M Singer**, *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. London Ser. A 362 (1978) 425–461 MR

[4] **M F Atiyah**, **V K Patodi**, **I M Singer**, *Spectral asymmetry and Riemannian geometry, I*, Math. Proc. Cambridge Philos. Soc. 77 (1975) 43–69 MR

[5] **M F Atiyah**, **V K Patodi**, **I M Singer**, *Spectral asymmetry and Riemannian geometry, II*, Math. Proc. Cambridge Philos. Soc. 78 (1975) 405–432 MR

[6] **M F Atiyah**, **I M Singer**, *The index of elliptic operators, III*, Ann. of Math. 87 (1968) 546–604 MR

[7] **D M Austin**, $\mathrm{SO}(3)$*–instantons on* $L(p,q)\times\mathbf{R}$, J. Differential Geom. 32 (1990) 383–413 MR

[8] **P J Braam**, **S K Donaldson**, *Floer's work on instanton homology, knots and surgery*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 195–256 MR

[9] **G Burde**, **H Zieschang**, *Knots*, de Gruyter Studies in Mathematics 5, de Gruyter, Berlin (1985) MR

[10] **W Chen**, **Y Ruan**, *Orbifold Gromov–Witten theory*, from "Orbifolds in mathematics and physics" (A Adem, J Morava, Y Ruan, editors), Contemp. Math. 310, Amer. Math. Soc., Providence, RI (2002) 25–85 MR

[11] **O Collin**, **N Saveliev**, *Equivariant Casson invariants via gauge theory*, J. Reine Angew. Math. 541 (2001) 143–169 MR

[12] **H Donnelly**, *Eta invariants for* G*–spaces*, Indiana Univ. Math. J. 27 (1978) 889–918 MR

[13] **S Dostoglou**, **D Salamon**, *Instanton homology and symplectic fixed points*, from "Symplectic geometry" (D Salamon, editor), London Math. Soc. Lecture Note Ser. 192, Cambridge Univ. Press (1993) 57–93 MR

[14] **A Durfee**, **L Kauffman**, *Periodicity of branched cyclic covers*, Math. Ann. 218 (1975) 157–174 MR

[15] **R Fintushel**, **R J Stern**, *Instanton homology of Seifert fibred homology three spheres*, Proc. London Math. Soc. 61 (1990) 109–137 MR

[16] **A Floer**, *Instanton homology and Dehn surgery*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 77–97 MR

[17] **K A Frøyshov**, *Equivariant aspects of Yang–Mills Floer theory*, Topology 41 (2002) 525–552 MR

[18] **Y Fukumoto**, **P Kirk**, **J Pinzón-Caicedo**, *Traceless* $SU(2)$ *representations of* 2*–stranded tangles*, Math. Proc. Cambridge Philos. Soc. 162 (2017) 101–129 MR

[19] **E Harper**, **N Saveliev**, *Instanton Floer homology for two-component links*, J. Knot Theory Ramifications 21 (2012) art. id. 1250054, 8 pages MR

[20] **M Hedden**, **C M Herald**, **P Kirk**, *The pillowcase and perturbations of traceless representations of knot groups*, Geom. Topol. 18 (2014) 211–287 MR

[21] **A Kawauchi**, *A survey of knot theory*, Birkhäuser, Basel (1996) MR

[22] **P A Kirk**, **E P Klassen**, *Representation spaces of Seifert fibered homology spheres*, Topology 30 (1991) 77–95 MR

[23] **E P Klassen**, *Representations of knot groups in* $\mathrm{SU}(2)$, Trans. Amer. Math. Soc. 326 (1991) 795–828 MR

[24] **P B Kronheimer**, **T S Mrowka**, *Khovanov homology is an unknot-detector*, Publ. Math. Inst. Hautes Études Sci. 113 (2011) 97–208 MR

[25] **P B Kronheimer**, **T S Mrowka**, *Knot homology groups from instantons*, J. Topol. 4 (2011) 835–918 MR

[26] **K B Lee**, **F Raymond**, *Seifert fiberings*, Mathematical Surveys and Monographs 166, Amer. Math. Soc., Providence, RI (2010) MR

[27] **R Lee**, **S H Weintraub**, *On the homology of double branched covers*, Proc. Amer. Math. Soc. 123 (1995) 1263–1266 MR

[28] **A Lobb**, **R Zentner**, *On spectral sequences from Khovanov homology*, preprint (2013) arXiv

[29] **W D Neumann**, *An invariant of plumbed homology spheres*, from "Topology Symposium Siegen 1979" (U Koschorke, W D Neumann, editors), Lecture Notes in Math. 788, Springer (1980) 125–144 MR

[30] **W D Neumann**, **F Raymond**, *Seifert manifolds, plumbing, μ–invariant and orientation reversing maps*, from "Algebraic and geometric topology" (K C Millett, editor), Lecture Notes in Math. 664, Springer (1978) 163–196 MR

[31] **D Ruberman**, *Doubly slice knots and the Casson–Gordon invariants*, Trans. Amer. Math. Soc. 279 (1983) 569–588 MR

[32] **D Ruberman**, *Rational homology cobordisms of rational space forms*, Topology 27 (1988) 401–414 MR

[33] **D Ruberman**, **N Saveliev**, *Rohlin's invariant and gauge theory, I: Homology 3–tori*, Comment. Math. Helv. 79 (2004) 618–646 MR

[34] **H Sasahira**, *Instanton Floer homology for lens spaces*, Math. Z. 273 (2013) 237–281 MR

[35] **N N Savel'ev**, *Adding relations to instanton homology groups of Seifert fibered homology spheres*, Mat. Sb. 183 (1992) 125–140 MR In Russian; translated in Russian Acad. Sci. Sb. Math 77 (1994) 497–510

[36] **N Saveliev**, *Floer homology of Brieskorn homology spheres*, J. Differential Geom. 53 (1999) 15–87 MR

[37] **N Saveliev**, *A surgery formula for the $\overline{\mu}$–invariant*, Topology Appl. 106 (2000) 91–102 MR

[38] **N Saveliev**, *Representation spaces of Seifert fibered homology spheres*, Topology Appl. 126 (2002) 49–61 MR

[39] **P Shanahan**, *The Atiyah–Singer index theorem*, Lecture Notes in Math. 638, Springer (1978) MR

[40] **C H Taubes**, *Gauge theory on asymptotically periodic 4–manifolds*, J. Differential Geom. 25 (1987) 363–430 MR

[41] **O J Viro**, *Branched coverings of manifolds with boundary and link invariants, I*, Izv. Akad. Nauk SSSR Ser. Mat. 37 (1973) 1241–1258 MR In Russian; translated in Math. USSR-Izv. 7 (1973) 1239–1256

[42] **S Wang**, *Moduli spaces over manifolds with involutions*, Math. Ann. 296 (1993) 119–138 MR

[43] **R Zentner**, *Representation spaces of pretzel knots*, Algebr. Geom. Topol. 11 (2011) 2941–2970 MR

Department of Mathematics & Statistics, McMaster University
Hamilton, ON, Canada

Department of Mathematics, University of Miami
Coral Gables, FL, United States

ppoudel@math.mcmaster.ca, saveliev@math.miami.edu

http://www.ms.mcmaster.ca/~ppoudel,
http://www.math.miami.edu/~saveliev

Received: 15 June 2015

msp

Algebraic & Geometric Topology 17 (2017) 2687–2735

Higher Toda brackets and the Adams spectral sequence in triangulated categories

J Daniel Christensen
Martin Frankland

The Adams spectral sequence is available in any triangulated category equipped with a projective or injective class. Higher Toda brackets can also be defined in a triangulated category, as observed by B Shipley based on J Cohen's approach for spectra. We provide a family of definitions of higher Toda brackets, show that they are equivalent to Shipley's and show that they are self-dual. Our main result is that the Adams differential d_r in any Adams spectral sequence can be expressed as an $(r+1)$–fold Toda bracket and as an r^{th} order cohomology operation. We also show how the result simplifies under a sparseness assumption, discuss several examples and give an elementary proof of a result of Heller, which implies that the 3–fold Toda brackets in principle determine the higher Toda brackets.

55T15; 18E30

1 Introduction

The Adams spectral sequence is an important tool in stable homotopy theory. Given finite spectra X and Y, the classical Adams spectral sequence is

$$E_2^{s,t} = \operatorname{Ext}_{\mathcal{A}}^{s,t}(H^*Y, H^*X) \implies [\Sigma^{t-s}X, Y_p^\wedge],$$

where $H^*X := H^*(X;\mathbb{F}_p)$ denotes mod p cohomology and $\mathcal{A} = H\mathbb{F}_p^* H\mathbb{F}_p$ denotes the mod p Steenrod algebra. Determining the differentials in the Adams spectral sequence generally requires a combination of techniques and much ingenuity. The approach that provides a basis for our work is found in [28], where Maunder showed that the differential d_r in this spectral sequence is determined by r^{th} order cohomology operations, which we now review.

A primary cohomology operation in this context is simply an element of the Steenrod algebra, and it is immediate from the construction of the Adams spectral sequence that the differential d_1 is given by primary cohomology operations. A secondary cohomology operation corresponds to a relation among primary operations, and is partially defined and multivalued: it is defined on the kernel of a primary operation

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2687

and takes values in the cokernel of another primary operation. Tertiary operations correspond to relations between relations, and have correspondingly more complicated domains and codomains. The pattern continues for higher operations. Using that cohomology classes are representable, secondary cohomology operations can also be expressed using 3–fold Toda brackets involving the cohomology class and two operations whose composite is null. However, what one obtains in general is a *subset* of the Toda bracket with less indeterminacy. This observation will be the key to our generalization of Maunder's result to other Adams spectral sequences in other categories.

The starting point of this paper is the following observation. On the one hand, the Adams spectral sequence can be constructed in any triangulated category equipped with a projective class or an injective class, as shown by Christensen [14]. For example, the classical Adams spectral sequence is constructed in the stable homotopy category with the injective class consisting of retracts of products $\prod_i \Sigma^{n_i} H\mathbb{F}_p$. On the other hand, higher Toda brackets can also be defined in an arbitrary triangulated category. This was done by Shipley in [40], based on Cohen's construction [15] for spaces and spectra, and was studied further by Sagave [36]. The goal of this paper is to describe precisely how the Adams d_r can be described as a particular subset of an $(r+1)$–fold Toda bracket which can be viewed as an r^{th} order cohomology operation, all in the context of a general triangulated category.

Triangulated categories arise throughout mathematics, so our work applies in various situations. As an example, we give calculations involving the Adams spectral sequence in the stable module category of a group algebra. Even in stable homotopy theory, there are a variety of Adams spectral sequences, such as the Adams–Novikov spectral sequence or the motivic Adams spectral sequence, and our results apply to all of them. Moreover, by working with minimal structure, our approach gains a certain elegance.

Organization and main results

In Section 2, we review the construction of the Adams spectral sequence in a triangulated category equipped with a projective class or an injective class. In Section 3, we review the construction of 3–fold Toda brackets in a triangulated category and some of their basic properties. Section 4 describes how the Adams d_2 is given by 3–fold Toda brackets. This section serves as a warm-up for Section 6.

In Section 5, we recall the construction of higher Toda brackets in a triangulated category via filtered objects. We provide a family of alternate constructions, and prove that they are all equivalent. The main result is Theorem 5.11, which says roughly the following.

Theorem *There is an inductive way to compute an n–fold Toda bracket $\langle f_n, \dots, f_1 \rangle \subseteq \mathcal{T}(\Sigma^{n-2} X_0, X_n)$, where the inductive step picks three consecutive maps and reduces the length by one. The $(n-2)!$ ways of doing this yield the same subset, up to an explicit sign.*

As a byproduct, we obtain Corollary 5.13, which would be tricky to prove directly from the filtered object definition.

Corollary *Toda brackets are self-dual up to suspension: $\langle f_n, \dots, f_1 \rangle \subseteq \mathcal{T}(\Sigma^{n-2} X_0, X_n)$ corresponds to the Toda bracket computed in the opposite category*

$$\langle f_1, \dots, f_n \rangle \subseteq \mathcal{T}^{\mathrm{op}}(\Sigma^{-(n-2)} X_n, X_0) = \mathcal{T}(X_0, \Sigma^{-(n-2)} X_n).$$

Section 6 establishes how the Adams d_r is given by $(r+1)$–fold Toda brackets. Our main results are Theorems 6.1 and 6.5, which say roughly the following.

Theorem *Let $[x] \in E_r^{s,t}$ be a class in the E_r term of the Adams spectral sequence. As subsets of $E_1^{s+r,t+r-1}$, we have*

$$\begin{aligned} d_r[x] &= \langle \Sigma^{r-1} d_1, \dots, \Sigma^2 d_1, \Sigma d_1, \Sigma p_{s+1}, \delta_s x \rangle \\ &= \langle \Sigma^{r-1} d_1 \overset{!}{;} \dots \overset{!}{;} \Sigma d_1 \overset{!}{;} d_1, x \rangle. \end{aligned}$$

Here, d_1, p_{s+1} and δ_s are maps appearing in the Adams resolution of Y, where each d_1 is a primary cohomology operation. The first expression for $d_r[x]$ is an $(r+1)$–fold Toda bracket. The second expression (with the superscripts !) denotes an appropriate subset of the bracket $\langle \Sigma^{r-1} d_1, \dots, \Sigma d_1, d_1, x \rangle$ with some choices dictated by the Adams resolution of Y. This description exhibits $d_r[x]$ as an r^{th} order cohomology operation applied to x.

In Section 7, we show that when certain sparseness assumptions are made, the subset $\langle \Sigma^{r-1} d_1 \overset{!}{;} \dots \overset{!}{;} \Sigma d_1 \overset{!}{;} d_1, x \rangle$ coincides with the full Toda bracket, and we give examples of this phenomenon. See Theorem 7.14, Proposition 7.15 and Example 7.17. The main application is to computing maps in the homotopy category of R–module spectra, for a ring spectrum R whose coefficient ring $\pi_* R$ is sufficiently sparse, such as ku. See Example 7.21.

In Appendix A, we compute examples of Toda brackets in stable module categories. In particular, Proposition A.1 provides an example where the inclusion $d_2[x] \subseteq \langle \Sigma d_1, d_1, x \rangle$ is proper. Appendix B provides for the record a short, simple proof of a theorem due to Heller, that 3–fold Toda brackets determine the triangulated structure. As a corollary, we note that the 3–fold Toda brackets indirectly determine the higher Toda brackets.

Related work

Detailed treatments of secondary operations can be found in [1, Section 3.6], where Adams used secondary cohomology operations to solve the Hopf invariant one problem; see Mosher and Tangora [32, Chapter 16] and Harper [19, Chapter 4].

There are various approaches to higher order cohomology operations and higher Toda brackets in the literature, many of which use some form of enrichment in spaces, chain complexes or groupoids; see for instance Spanier [42], Maunder [29], Kochman [24] and Klaus [23]. In this paper, we work solely with the triangulated structure, without enhancement, and provide no comparison to those other approaches.

In [6; 7], Baues and Jibladze express the Adams d_2 in terms of secondary cohomology operations, and this is generalized to higher differentials by Baues and Blanc in [5]. Their approach starts with an injective resolution as in diagram (2-3), and witnesses the equations $d_1 d_1 = 0$ by providing suitably coherent null-homotopies, described using mapping spaces. Using this coherence data, the authors express a representative of $d_r[x]$ as a specific element of the Toda bracket $\langle \Sigma^{r-1} d_1, \dots, \Sigma d_1, d_1, x \rangle$. While this approach makes use of an enrichment, we suspect that by translating the (higher dimensional) null-homotopies into lifts to fibers or extensions to cofibers, one could relate their expression for $d_r[x]$ to ours.

Acknowledgments We thank Robert Bruner, Dan Isaksen, Peter Jorgensen, Fernando Muro, Irakli Patchkoria, Steffen Sagave and Dylan Wilson for helpful conversations, as well as the referee for their useful comments. Frankland also thanks the Max-Planck-Institut für Mathematik for its hospitality. Frankland was partially funded by a grant of the DFG SPP 1786: Homotopy Theory and Algebraic Geometry.

2 The Adams spectral sequence

In this section, we recall the construction of the Adams spectral sequence in a triangulated category, along with some of its features. We follow [14, Section 4], or rather its dual. Some references for the classical Adams spectral sequence are [2, Section III.15], [26, Chapter 16] and [10]. Background material on triangulated categories can be found in [33, Chapter 1; 26, Appendix 2; 44, Chapter 10]. We assume that the suspension functor Σ is an equivalence, with chosen inverse Σ^{-1}. Moreover, we assume we have chosen natural isomorphisms $\Sigma\Sigma^{-1} \cong \mathrm{id}$ and $\Sigma^{-1}\Sigma \cong \mathrm{id}$ making Σ and Σ^{-1} into an adjoint equivalence. We silently use these isomorphisms when needed, eg when we say that a triangle of the form $\Sigma^{-1} Z \to X \to Y \to Z$ is distinguished.

Definition 2.1 [14, Proposition 2.6] A *projective class* in a triangulated category $\mathcal{T}$ is a pair $(\mathcal{P}, \mathcal{N})$, where $\mathcal{P}$ is a class of objects and $\mathcal{N}$ is a class of maps satisfying the following properties:

(1) A map $f\colon X \to Y$ is in $\mathcal{N}$ if and only if the induced map
$$f_*\colon \mathcal{T}(P, X) \to \mathcal{T}(P, Y)$$
is zero for all P in $\mathcal{P}$. In other words, $\mathcal{N}$ consists of the $\mathcal{P}$*–null* maps.

(2) An object P is in $\mathcal{P}$ if and only if the induced map
$$f_*\colon \mathcal{T}(P, X) \to \mathcal{T}(P, Y)$$
is zero for all f in $\mathcal{N}$.

(3) For every object X, there is a distinguished triangle $P \to X \xrightarrow{f} Y \to \Sigma P$, where P is in $\mathcal{P}$ and f is in $\mathcal{N}$.

In particular, the class $\mathcal{P}$ is closed under arbitrary coproducts and retracts. The objects in $\mathcal{P}$ are called *projective*.

Definition 2.2 A projective class is *stable* if it is closed under shifts, ie $P \in \mathcal{P}$ implies $\Sigma^n P \in \mathcal{P}$ for all $n \in \mathbb{Z}$.

We will assume for convenience that our projective class is stable. We suspect that many of the results can be adapted to unstable projective classes, with a careful treatment of shifts.

Definition 2.3 Let $\mathcal{P}$ be a projective class and $f\colon X \to Y$ be a map.

(1) f is $\mathcal{P}$*–epic* if the map
$$f_*\colon \mathcal{T}(P, X) \to \mathcal{T}(P, Y)$$
is surjective for all $P \in \mathcal{P}$. Equivalently, the map to the cofiber $Y \to C_f$ is $\mathcal{P}$–null.

(2) f is $\mathcal{P}$*–monic* if the map
$$f_*\colon \mathcal{T}(P, X) \to \mathcal{T}(P, Y)$$
is injective for all $P \in \mathcal{P}$. Equivalently, the map from the fiber $\Sigma^{-1} C_f \to X$ is $\mathcal{P}$–null.

Example 2.4 Let $\mathcal{T}$ be the stable homotopy category and $\mathcal{P}$ the projective class consisting of retracts of wedges of spheres $\bigvee_i S^{n_i}$. This is called the *ghost projective class*, studied for instance in [14, Section 7].

Now we dualize everything.

Definition 2.5 An *injective class* in a triangulated category $\mathcal{T}$ is a projective class in the opposite category $\mathcal{T}^{\mathrm{op}}$. Explicitly, it is a pair $(\mathcal{I}, \mathcal{N})$, where $\mathcal{I}$ is a class of objects and $\mathcal{N}$ is a class of maps satisfying the following properties:

(1) A map $f\colon X \to Y$ is in $\mathcal{N}$ if and only if the induced map
$$f^*\colon \mathcal{T}(Y, I) \to \mathcal{T}(X, I)$$
is zero for all I in $\mathcal{I}$.

(2) An object I is in $\mathcal{I}$ if and only if the induced map
$$f^*\colon \mathcal{T}(Y, I) \to \mathcal{T}(X, I)$$
is zero for all f in $\mathcal{N}$.

(3) For every object X, there is a distinguished triangle $\Sigma^{-1} I \to W \xrightarrow{f} X \to I$, where I is in $\mathcal{I}$ and f is in $\mathcal{N}$.

In particular, the class $\mathcal{I}$ is closed under arbitrary products and retracts. The objects in $\mathcal{I}$ are called *injective*. Just as for projective classes, we will assume for convenience that our injective class is stable.

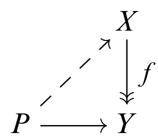
Example 2.6 Let $\mathcal{T}$ be the stable homotopy category. Take $\mathcal{N}$ to be the class of maps inducing zero on mod p cohomology and $\mathcal{I}$ to be the retracts of (arbitrary) products $\prod_i \Sigma^{n_i} H\mathbb{F}_p$ with $n_i \in \mathbb{Z}$. One can generalize this example to any cohomology theory (spectrum) E instead of $H\mathbb{F}_p$, letting $\mathcal{I}_E$ denote the injective class consisting of retracts of products $\prod_i \Sigma^{n_i} E$.

Definition 2.7 Let $\mathcal{I}$ be an injective class and $f\colon X \to Y$ be a map.

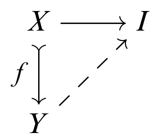
(1) f is *$\mathcal{I}$–monic* if the map
$$f^*\colon \mathcal{T}(Y, I) \to \mathcal{T}(X, I)$$
is surjective for all $I \in \mathcal{I}$. Equivalently, the map from the fiber $\Sigma^{-1} C_f \to X$ is $\mathcal{I}$–null.

(2) f is *$\mathcal{I}$–epic* if the map
$$f^*\colon \mathcal{T}(Y, I) \to \mathcal{T}(X, I)$$
is injective for all $I \in \mathcal{I}$. Equivalently, the map to the cofiber $Y \to C_f$ is $\mathcal{I}$–null.

Remark 2.8 The projectives and $\mathcal{P}$–epic maps determine each other via the lifting property:



Dually, the injectives and $\mathcal{I}$–monic maps determine each other via the extension property:



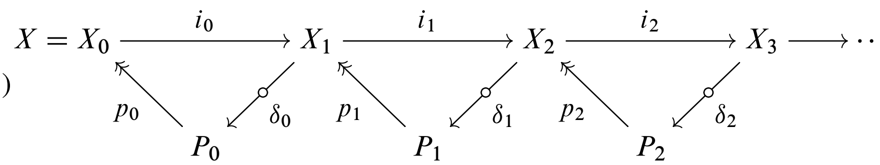
This is part of the equivalent definition of a projective (resp. injective) class described in [14, Proposition 2.4].

Convention 2.9 We will implicitly use the natural isomorphism

$$\mathcal{T}(A, B) \cong \mathcal{T}(\Sigma^k A, \Sigma^k B)$$

sending a map f to $\Sigma^k f$.

Definition 2.10 An *Adams resolution* of an object X in $\mathcal{T}$ with respect to a projective class $(\mathcal{P}, \mathcal{N})$ is a diagram

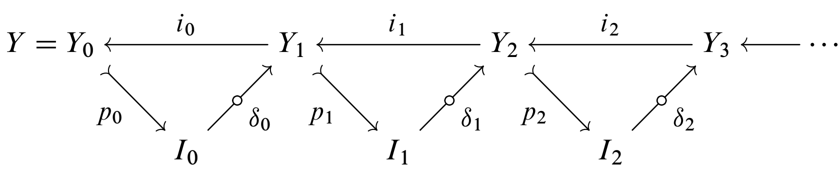
(2-1)


where every P_s is projective, every map i_s is in $\mathcal{N}$, and every triangle

$$P_s \xrightarrow{p_s} X_s \xrightarrow{i_s} X_{s+1} \xrightarrow{\delta_s} \Sigma P_s$$

is distinguished. Here the arrows $\delta_s\colon X_{s+1} \to P_s$ (drawn with a circle on the arrow) denote degree-shifting maps, namely, maps $\delta_s\colon X_{s+1} \to \Sigma P_s$.

Dually, an *Adams resolution* of an object Y in $\mathcal{T}$ with respect to an injective class $(\mathcal{I}, \mathcal{N})$ is a diagram

(2-2)


where every I_s is injective, every map i_s is in $\mathcal{N}$, and every triangle

$$\Sigma^{-1} I_s \xrightarrow{\Sigma^{-1}\delta_s} Y_{s+1} \xrightarrow{i_s} Y_s \xrightarrow{p_s} I_s$$

is distinguished.

From now on, fix a triangulated category $\mathcal{T}$ and a (stable) injective class $(\mathcal{I}, \mathcal{N})$ in $\mathcal{T}$.

Lemma 2.11 *Every object Y of $\mathcal{T}$ admits an Adams resolution.*

Given an object X and an Adams resolution of Y, applying $\mathcal{T}(X, -)$ yields an exact couple

$$\begin{array}{ccc} \bigoplus_{s,t} \mathcal{T}(\Sigma^{t-s} X, Y_s) & \xrightarrow{\quad i = \bigoplus (i_s)_* \quad} & \bigoplus_{s,t} \mathcal{T}(\Sigma^{t-s} X, Y_s) \\ & {\scriptstyle \delta = \bigoplus(\delta_s)_*} \nwarrow \qquad \swarrow {\scriptstyle p = \bigoplus(p_s)_*} & \\ & \bigoplus_{s,t} \mathcal{T}(\Sigma^{t-s} X, I_s) & \end{array}$$

and thus a spectral sequence with E_1 term

$$E_1^{s,t} = \mathcal{T}(\Sigma^{t-s} X, I_s) \cong \mathcal{T}(\Sigma^t X, \Sigma^s I_s)$$

and differentials

$$d_r \colon E_r^{s,t} \to E_r^{s+r,t+r-1}$$

given by $d_r = p \circ i^{-(r-1)} \circ \delta$, where i^{-1} means choosing an i–preimage. This is called the *Adams spectral sequence* with respect to the injective class $\mathcal{I}$ abutting to $\mathcal{T}(\Sigma^{t-s} X, Y)$.

Lemma 2.12 *The E_2 term is given by*

$$E_2^{s,t} = \operatorname{Ext}_{\mathcal{I}}^{s,t}(X, Y) := \operatorname{Ext}_{\mathcal{I}}^{s}(\Sigma^t X, Y),$$

where $\operatorname{Ext}_{\mathcal{I}}^{s}(X, Y)$ denotes the s^{th} derived functor of $\mathcal{T}(X, -)$ (relative to the injective class $\mathcal{I}$) applied to the object Y.

Proof The Adams resolution of Y yields an $\mathcal{I}$–injective resolution of Y

$$0 \longrightarrow Y \xrightarrow{p_0} I_0 \xrightarrow{(\Sigma p_1)\delta_0} \Sigma I_1 \xrightarrow{(\Sigma^2 p_2)(\Sigma \delta_1)} \Sigma^2 I_2 \longrightarrow \cdots . \tag{2-3}$$

□

Remark 2.13 We do not assume that the injective class $\mathcal{I}$ generates, ie that every nonzero object X admits a nonzero map $X \to I$ to an injective. Hence, we do not expect the Adams spectral sequence to be conditionally convergent in general; compare [14, Proposition 4.4].

Example 2.14 Let E be a commutative (homotopy) ring spectrum. A spectrum is called *E–injective* if it is a retract of $E \wedge W$ for some W [22, Definition 2.22]. A map of spectra $f\colon X \to Y$ is called *E–monic* if the map $E \wedge f\colon E \wedge X \to E \wedge Y$ is a split monomorphism. The E–injective objects and E–monic maps form an injective class in the stable homotopy category. The Adams spectral sequence associated to this injective class is the *Adams spectral sequence based on E–homology*, as described in [35, Definition 2.2.4], also called the *unmodified Adams spectral sequence* in [22, Section 2.2]. Further assumptions are needed in order to identify the E_2 term as Ext groups in E_*E–comodules.

Definition 2.15 The *$\mathcal{I}$–cohomology* of an object X is the family of abelian groups $H^I(X) := \mathcal{T}(X, I)$ indexed by the injective objects $I \in \mathcal{I}$.

A *primary operation* in $\mathcal{I}$–cohomology is a natural transformation $H^I(X) \to H^J(X)$ of functors $\mathcal{T}^{\mathrm{op}} \to \mathrm{Ab}$. Equivalently, by the (additive) Yoneda lemma, a primary operation is a map $I \to J$ in $\mathcal{T}$.

Example 2.16 The differential d_1 is given by primary operations. More precisely, let $x \in E_1^{s,t}$ be a map $x\colon \Sigma^{t-s}X \to I_s$. Then $d_1(x) \in E_1^{s+1,t}$ is the composite

$$\Sigma^{t-s}X \xrightarrow{x} I_s \xrightarrow{\delta_s} \Sigma Y_{s+1} \xrightarrow{\Sigma p_{s+1}} \Sigma I_{s+1}.$$

In other words, $d_1(x)$ is obtained by applying the primary operation

$$d_1 := (\Sigma p_{s+1})\delta_s\colon I_s \to \Sigma I_{s+1}$$

to x.

Proposition 2.17 *A primary operation $\theta\colon I \to J$ appears as $d_1\colon I_s \,{-\!\!\circ\!\!\to}\, I_{s+1}$ in some Adams resolution if and only if θ admits a factorization into an $\mathcal{I}$–epic followed by an $\mathcal{I}$–mono.*

Proof The condition is necessary by construction. In the factorization $d_1 = (\Sigma p_{s+1})\delta_s$, the map δ_s is $\mathcal{I}$–epic while p_{s+1} is $\mathcal{I}$–monic.

To prove sufficiency, assume given a factorization $\theta = iq\colon I \to W \to J$, where $q\colon I \twoheadrightarrow W$ is $\mathcal{I}$–epic and $i\colon W \hookrightarrow J$ is $\mathcal{I}$–monic. Taking the fiber of q twice yields the distinguished triangle

$$\Sigma^{-1}W \longrightarrow Y_0 \rightarrowtail I \overset{q}{\twoheadrightarrow} W,$$

which we relabel

$$Y_1 \xrightarrow{i_0} Y_0 \overset{p_0}{\rightarrowtail} I \overset{\delta_0}{\twoheadrightarrow} \Sigma Y_1.$$

Relabeling the given map $i\colon W \hookrightarrow J$ as $\Sigma p_1\colon \Sigma Y_1 \hookrightarrow \Sigma I_1$, we can continue the usual construction of an Adams resolution of Y_0 as illustrated in diagram (2-2), in which $\theta = iq$ appears as the composite $(\Sigma p_1)\delta_0$. Note that by the same argument, for any $s \geq 0$, θ appears as $d_1\colon I_s \to I_{s+1}$ in some (other) Adams resolution. □

Example 2.18 Not every primary operation appears as d_1 in an Adams resolution. For example, consider the stable homotopy category with the projective class $\mathcal{P}$ generated by the sphere spectrum $S = S^0$, that is, $\mathcal{P}$ consists of retracts of wedges of spheres. The $\mathcal{P}$–epis (resp. $\mathcal{P}$–monos) consist of the maps which are surjective (resp. injective) on homotopy groups. The primary operation $2\colon S \to S$ does *not* admit a factorization into an $\mathcal{I}$–epic followed by an $\mathcal{I}$–mono.

Indeed, assume that $2 = iq\colon S \twoheadrightarrow W \hookrightarrow S$ is such a factorization. We will show that this implies $\pi_2(S/2) = \mathbb{Z}/2 \oplus \mathbb{Z}/2$, contradicting the known fact $\pi_2(S/2) = \mathbb{Z}/4$. Here $S/2$ denotes the mod 2 Moore spectrum, sitting in the cofiber sequence $S \xrightarrow{2} S \to S/2$.

By the octahedral axiom applied to the factorization $2 = iq$, there is a diagram

$$\begin{array}{ccccccc}
S & \xrightarrow{q} & W & \longrightarrow & C_q & \xrightarrow{\delta'} & S^1 \\
\| & & \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle \alpha} & & \| \\
S & \xrightarrow{2} & S & \longrightarrow & S/2 & \xrightarrow{\delta} & S^1 \\
 & & \downarrow{\scriptstyle j} & & \downarrow{\scriptstyle \beta} & & \\
 & & C_i & = & C_i & &
\end{array}$$

with distinguished rows and columns. The long exact sequence in homotopy yields $\pi_n C_q = {}_2\pi_{n-1} S$, where the induced map $\pi_n(\delta')\colon \pi_n C_q \to \pi_n S^1$ corresponds to the inclusion ${}_2\pi_{n-1} S \hookrightarrow \pi_{n-1} S$. Likewise, we have $\pi_n C_i = (\pi_n S)/2$, where the induced map $\pi_n(j)\colon \pi_n S \to \pi_n C_i$ corresponds to the quotient map $\pi_n S \twoheadrightarrow (\pi_n S)/2$. The defining cofiber sequence $S \xrightarrow{2} S \to S/2$ yields the exact sequence

$$\pi_n S \xrightarrow{2} \pi_n S \xrightarrow{\pi_n} (S/2) \xrightarrow{\pi_n \delta} \pi_{n-1} S \xrightarrow{2} \pi_{n-1} S,$$

which in turn yields the short exact sequence

$$0 \longrightarrow (\pi_n S)/2 \longrightarrow \pi_n(S/2) \xrightarrow{\pi_n \delta} {}_2\pi_{n-1} S \longrightarrow 0.$$

The map $\pi_n(\alpha)\colon {}_2\pi_{n-1} S \to \pi_n(S/2)$ is a splitting of this sequence, because of the equality $\pi_n(\delta)\pi_n(\alpha) = \pi_n(\delta\alpha) = \pi_n(\delta')$. However, the short exact sequence does not split in the case $n = 2$, by the isomorphism $\pi_2(S/2) = \mathbb{Z}/4$. For references, see [38, Proposition II.6.48], [37, Proposition 4] and [27].

3 3–fold Toda brackets

In this section, we review different constructions of 3–fold Toda brackets and some of their properties.

Definition 3.1 Let $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3$ be a diagram in a triangulated category $\mathcal{T}$. We define subsets of $\mathcal{T}(\Sigma X_0, X_3)$ as follows:

- The *iterated cofiber Toda bracket* $\langle f_3, f_2, f_1\rangle_{\mathrm{cc}} \subseteq \mathcal{T}(\Sigma X_0, X_3)$ consists of all maps $\psi\colon \Sigma X_0 \to X_3$ that appear in a commutative diagram

(3-1)
$$\begin{CD} X_0 @>{f_1}>> X_1 @>>> C_{f_1} @>>> \Sigma X_0 \\ @| @| @VV{\varphi}V @VV{\psi}V \\ X_0 @>{f_1}>> X_1 @>{f_2}>> X_2 @>{f_3}>> X_3 \end{CD}$$

where the top row is distinguished.

- The *fiber-cofiber Toda bracket* $\langle f_3, f_2, f_1\rangle_{\mathrm{fc}} \subseteq \mathcal{T}(\Sigma X_0, X_3)$ consists of all composites $\beta \circ \Sigma\alpha\colon \Sigma X_0 \to X_3$, where α and β appear in a commutative diagram

(3-2)
$$\begin{CD} X_0 @>{f_1}>> X_1 \\ @V{\alpha}VV @| \\ \Sigma^{-1}C_{f_2} @>>> X_1 @>{f_2}>> X_2 @>>> C_{f_2} \\ @. @. @| @VV{\beta}V \\ @. @. X_2 @>{f_3}>> X_3 \end{CD}$$

where the middle row is distinguished.

- The *iterated fiber Toda bracket* $\langle f_3, f_2, f_1\rangle_{\mathrm{ff}} \subseteq \mathcal{T}(\Sigma X_0, X_3)$ consists of all maps $\Sigma\delta\colon \Sigma X_0 \to X_3$ where δ appears in a commutative diagram

(3-3)
$$\begin{CD} X_0 @>{f_1}>> X_1 @>{f_2}>> X_2 @>{f_3}>> X_3 \\ @V{\delta}VV @V{\gamma}VV @| @| \\ \Sigma^{-1}X_3 @>>> \Sigma^{-1}C_{f_3} @>>> X_2 @>{f_3}>> X_3 \end{CD}$$

where the bottom row is distinguished.

Remark 3.2 In the literature, there are variations of these definitions, which sometimes differ by a sign. With the notion of cofiber sequence implicitly used in [43], our definitions agree with Toda's. The Toda bracket also depends on the choice of triangulation. Given a triangulation, there is an associated negative triangulation whose distinguished triangles are those triangles whose negatives are distinguished in the original triangulation; see [3]. Negating a triangulation negates the 3–fold Toda brackets. Dan Isaksen has pointed out to us that in the stable homotopy category there are 3–fold Toda brackets which are not equal to their own negatives. For example, Toda showed in [43, Section VI.v and Theorems 7.4 and 14.1] that the Toda bracket $\langle 2\sigma, 8, \nu \rangle$ has no indeterminacy and contains an element ζ of order 8. We give another example in Example A.4.

The following proposition can be found in [36, Remark 4.5 and Figure 2] and was kindly pointed out by Fernando Muro. It is also proved in [31, Section 4.6]. We provide a different proof, more in the spirit of this article. In the case of spaces, it was originally proved by Toda [43, Proposition 1.7].

Proposition 3.3 *The iterated cofiber, fiber-cofiber and iterated fiber definitions of Toda brackets coincide. More precisely, for any diagram* $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3$ *in* $\mathcal{T}$, *we have the following equalities of subsets of* $\mathcal{T}(\Sigma X_0, X_3)$*:*

$$\langle f_3, f_2, f_1 \rangle_{\mathrm{cc}} = \langle f_3, f_2, f_1 \rangle_{\mathrm{fc}} = \langle f_3, f_2, f_1 \rangle_{\mathrm{ff}}.$$

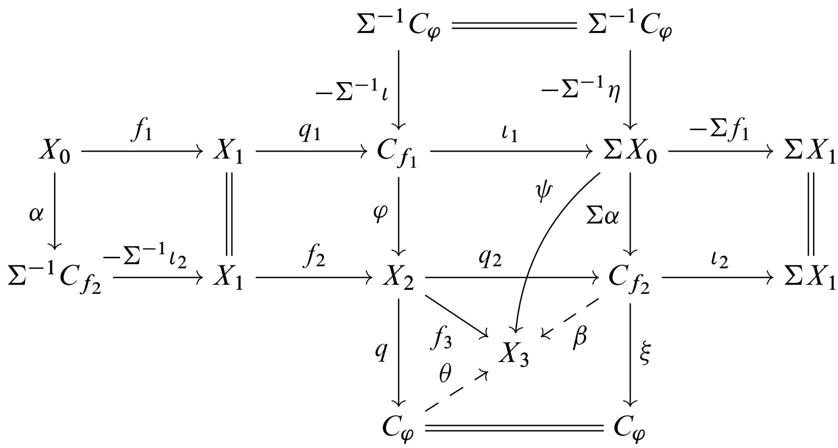
Proof We will prove the first equality; the second equality is dual.

($\supseteq$) Let $\beta(\Sigma\alpha) \in \langle f_3, f_2, f_1 \rangle_{\mathrm{fc}}$ be obtained from maps α and β as in diagram (3-2). Now consider the diagram with distinguished rows

$$\begin{array}{ccccccc}
X_0 & \xrightarrow{f_1} & X_1 & \longrightarrow & C_{f_1} & \longrightarrow & \Sigma X_0 \\
\downarrow{\scriptstyle \alpha} & & \| & & \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \Sigma\alpha} \\
\Sigma^{-1} C_{f_2} & \longrightarrow & X_1 & \xrightarrow{f_2} & X_2 & \longrightarrow & C_{f_2} \\
 & & & & \| & & \downarrow{\scriptstyle \beta} \\
 & & & & X_2 & \xrightarrow{f_3} & X_3
\end{array}$$

where there exists a filler $\varphi\colon C_{f_1} \to X_2$. The commutativity of the tall rectangle on the right exhibits the membership $\beta(\Sigma\alpha) \in \langle f_3, f_2, f_1 \rangle_{\mathrm{cc}}$.

($\subseteq$) Let $\psi \in \langle f_3, f_2, f_1 \rangle_{\mathrm{cc}}$ be as in diagram (3-1). The octahedral axiom comparing the cofibers of q_1, φ and $\varphi \circ q_1 = f_2$ yields a commutative diagram



where the rows and columns are distinguished. By exactness of the sequence

$$\mathcal{T}(C_{f_2}, X_3) \xrightarrow{(\Sigma\alpha)^*} \mathcal{T}(\Sigma X_0, X_3) \xrightarrow{(-\Sigma^{-1}\eta)^*} \mathcal{T}(\Sigma^{-1} C_\varphi, X_3)$$

there exists a map $\beta\colon C_{f_2} \to X_3$ satisfying $\psi = \beta(\Sigma\alpha)$ if and only if the restriction of ψ to the fiber $\Sigma^{-1} C_\varphi$ of $\Sigma\alpha$ is zero. That condition does hold: one readily checks the equality $\psi(-\Sigma^{-1}\eta) = 0$. The chosen map $\beta\colon C_{f_2} \to X_3$ might *not* satisfy the equation $\beta q_2 = f_3$, but we will correct it to another map β' which does. The error term $f_3 - \beta q_2$ is killed by restriction along φ, and therefore factors through the cofiber of φ, ie there exists a factorization

$$f_3 - \beta q_2 = \theta\iota$$

for some $\theta\colon C_\varphi \to X_3$. The corrected map $\beta' := \beta + \theta\xi\colon C_{f_2} \to X_3$ satisfies $\beta' q_2 = f_3$. Moreover, this corrected map β' still satisfies $\beta'(\Sigma\alpha) = \psi = \beta(\Sigma\alpha)$, since the correction term satisfies $\theta\xi(\Sigma\alpha) = 0$. □

Thanks to the proposition, we can write $\langle f_3, f_2, f_1 \rangle$ if we do not need to specify a particular definition of the Toda bracket.

We also recall this well-known fact, and leave the proof as an exercise.

Lemma 3.4 *For any diagram* $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3$ *in* $\mathcal{T}$*, the subset* $\langle f_3, f_2, f_1 \rangle$ *of* $\mathcal{T}(\Sigma X_0, X_3)$ *is a coset of the subgroup*

$$(f_3)_* \mathcal{T}(\Sigma X_0, X_2) + (\Sigma f_1)^* \mathcal{T}(\Sigma X_1, X_3). \qquad \square$$

The displayed subgroup is called the *indeterminacy*, and when it is trivial, we say that the Toda bracket *has no indeterminacy*.

Lemma 3.5 *Consider maps* $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3 \xrightarrow{f_4} X_4$. *The following inclusions of subsets of* $\mathcal{T}(\Sigma X_0, X_4)$ *hold:*

(a) $$f_4\langle f_3, f_2, f_1\rangle \subseteq \langle f_4 f_3, f_2, f_1\rangle,$$

(b) $$\langle f_4, f_3, f_2\rangle f_1 \subseteq \langle f_4, f_3, f_2 f_1\rangle,$$

(c) $$\langle f_4 f_3, f_2, f_1\rangle \subseteq \langle f_4, f_3 f_2, f_1\rangle,$$

(d) $$\langle f_4, f_3, f_2 f_1\rangle \subseteq \langle f_4, f_3 f_2, f_1\rangle.$$

Proof Inclusions (a)–(b) are straightforward.

For (c)–(d), using the iterated cofiber definition, the subset $\langle f_4 f_3, f_2, f_1\rangle_{\mathrm{cc}}$ consists of the maps $\psi\colon \Sigma X_0 \to X_4$ appearing in a commutative diagram

$$\begin{array}{ccccccccc} X_0 & \xrightarrow{f_1} & X_1 & \longrightarrow & C_{f_1} & \longrightarrow & & \longrightarrow & \Sigma X_0 \\ \| & & \| & & \downarrow{\scriptstyle\varphi} & & & & \downarrow{\scriptstyle\psi} \\ X_0 & \xrightarrow{f_1} & X_1 & \xrightarrow{f_2} & X_2 & \xrightarrow{f_3} & X_3 & \xrightarrow{f_4} & X_4 \end{array}$$

where the top row is distinguished. Given such a diagram, the diagram

$$\begin{array}{ccccccccc} X_0 & \xrightarrow{f_1} & X_1 & \longrightarrow & C_{f_1} & \longrightarrow & & \longrightarrow & \Sigma X_0 \\ \| & & \| & & & \searrow^{f_3\varphi} & & & \downarrow{\scriptstyle\psi} \\ X_0 & \xrightarrow{f_1} & X_1 & \xrightarrow{f_2} & X_2 & \xrightarrow{f_3} & X_3 & \xrightarrow{f_4} & X_4 \end{array}$$

exhibits the membership $\psi \in \langle f_4, f_3 f_2, f_1\rangle_{\mathrm{cc}}$. A similar argument can be used to prove the inclusion $\langle f_4, f_3, f_2 f_1\rangle_{\mathrm{ff}} \subseteq \langle f_4, f_3 f_2, f_1\rangle_{\mathrm{ff}}$. □

Example 3.6 The inclusion $\langle f_4 f_3, f_2, f_1\rangle \subseteq \langle f_4, f_3 f_2, f_1\rangle$ need not be an equality. For example, consider the maps $X \xrightarrow{0} Y \xrightarrow{1} Y \xrightarrow{0} Z \xrightarrow{1} Z$. The Toda brackets being compared are

$$\langle 1_Z 0, 1_Y, 0\rangle = \langle 0, 1_Y, 0\rangle = \{0\},$$
$$\langle 1_Z, 01_Y, 0\rangle = \langle 1_Z, 0, 0\rangle = \mathcal{T}(\Sigma X, Z).$$

Definition 3.7 In the setup of Definition 3.1, the *restricted Toda brackets* are the subsets of the Toda bracket

$$\langle f_3, f_2 \overset{\alpha}{,} f_1 \rangle_{\mathrm{fc}} \subseteq \langle f_3, f_2, f_1 \rangle_{\mathrm{fc}},$$
$$\langle f_3 \overset{\beta}{,} f_2, f_1 \rangle_{\mathrm{fc}} \subseteq \langle f_3, f_2, f_1 \rangle_{\mathrm{fc}}$$

consisting of all composites $\beta(\Sigma\alpha)\colon \Sigma X_0 \to X_3$, where α and β appear in a commutative diagram (3-2) where the middle row is distinguished, with the prescribed map $\alpha\colon X_0 \to \Sigma^{-1} C_{f_2}$ (resp. $\beta\colon C_{f_2} \to X_3$).

The lift to the fiber $\alpha\colon X_0 \to \Sigma^{-1} C_{f_2}$ is a witness of the equality $f_2 f_1 = 0$. Dually, the extension to the cofiber $\beta\colon C_{f_2} \to X_3$ is a witness of the equality $f_3 f_2 = 0$.

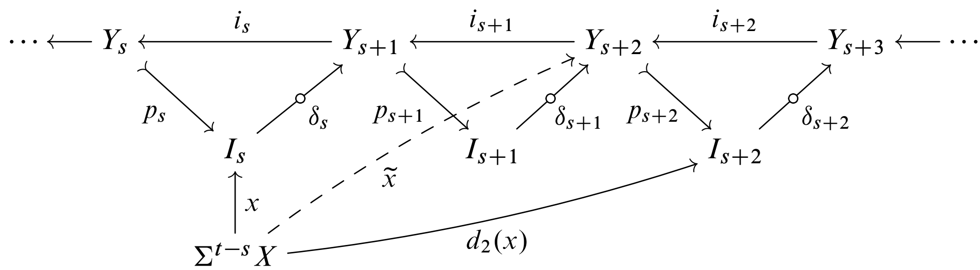
Remark 3.8 Let $X_1 \xrightarrow{f_2} X_2 \xrightarrow{q_2} C_{f_2} \xrightarrow{\iota_2} \Sigma X_1$ be a distinguished triangle. By definition, we have equalities of subsets

$$\langle f_3, f_2 \overset{\alpha}{,} f_1 \rangle_{\mathrm{fc}} = \langle f_3, f_2 \overset{1}{,} -\Sigma^{-1}\iota_2 \rangle_{\mathrm{fc}}(\Sigma\alpha),$$
$$\langle f_3 \overset{\beta}{,} f_2, f_1 \rangle_{\mathrm{fc}} = \beta \langle q_2 \overset{1}{,} f_2, f_1 \rangle_{\mathrm{fc}}.$$

4 Adams d_2 in terms of 3–fold Toda brackets

In this section, we show that the Adams differential d_r can be expressed in several ways using 3–fold Toda brackets. One of these expressions is as a secondary cohomology operation.

Given an injective class $\mathcal{I}$, an Adams resolution of an object Y as in diagram (2-2), and an object X, consider a class $[x] \in E_2^{s,t}$ represented by a cycle $x \in E_1^{s,t} = \mathcal{T}(\Sigma^{t-s} X, I_s)$. Recall that $d_2[x] \in E_2^{s+2,t+1}$ is obtained as illustrated in the diagram



Explicitly, since x satisfies $d_1(x) = (\Sigma p_{s+1})\delta_s x = 0$, we can choose a lift $\widetilde{x}\colon \Sigma^{t-s} X \to \Sigma Y_{s+2}$ of $\delta_s x$ to the fiber of Σp_{s+1}. Then the differential d_2 is given by

$$d_2[x] = [(\Sigma p_{s+2})\widetilde{x}].$$

From now on, we will unroll the distinguished triangles and keep track of the suspensions. Following Convention 2.9, we will use the identifications

$$E_1^{s+2,t+1} = \mathcal{T}(\Sigma^{t-s-1}X, I_{s+2}) \cong \mathcal{T}(\Sigma^{t-s}X, \Sigma I_{s+2}) \cong \mathcal{T}(\Sigma^{t-s+1}X, \Sigma^2 I_{s+2}).$$

Proposition 4.1 *Denote by $d_2[x] \subseteq E_1^{s+2,t+1}$ the subset of all representatives of the class $d_2[x] \in E_2^{s+2,t+1}$. Then the following equalities hold:*

(a)
$$\begin{aligned} d_2[x] &= \langle \Sigma d_1 \overset{\Sigma^2 p_{s+2}}{,} \Sigma p_{s+1}, \delta_s x \rangle_{\mathrm{fc}} \\ &= \langle \Sigma d_1, \Sigma p_{s+1}, \delta_s x \rangle, \end{aligned}$$

(b)
$$\begin{aligned} d_2[x] &= (\Sigma^2 p_{s+2}) \langle \Sigma \delta_{s+1} \overset{1}{,} \Sigma p_{s+1}, \delta_s x \rangle_{\mathrm{fc}} \\ &= (\Sigma^2 p_{s+2}) \langle \Sigma \delta_{s+1}, \Sigma p_{s+1}, \delta_s x \rangle, \end{aligned}$$

(c)
$$d_2[x] = \langle \Sigma d_1 \overset{\beta}{,} d_1, x \rangle_{\mathrm{fc}},$$

where β is the composite $C \xrightarrow{\widetilde{\beta}} \Sigma^2 Y_{s+2} \xrightarrow{\Sigma^2 p_{s+2}} \Sigma^2 I_{s+2}$ and $\widetilde{\beta}$ is obtained from the octahedral axiom applied to the factorization $d_1 = (\Sigma p_{s+1})\delta_s\colon I_s \to \Sigma Y_{s+1} \to \Sigma I_{s+1}$.

In (c), β is a witness to the fact that the composite $(\Sigma d_1) d_1$ of primary operations is zero, and so the restricted Toda bracket is a secondary operation.

Proof Note that t plays no role in the statement, so we will assume without loss of generality that $t = s$ holds.

(a) The first equality holds by definition of $d_2[x]$, namely choosing a lift of $\delta_s x$ to the fiber of Σp_{s+1}. The second equality follows from the fact that $\Sigma^2 p_{s+2}$ is the *unique* extension of $\Sigma d_1 = (\Sigma^2 p_{s+2})(\Sigma \delta_{s+1})$ to the cofiber of Σp_{s+1}. Indeed, $\Sigma \delta_{s+1}$ is $\mathcal{I}$–epic and ΣI_{s+2} is injective, so that the restriction map

$$(\Sigma \delta_{s+1})^*\colon \mathcal{T}(\Sigma^2 Y_{s+2}, \Sigma^2 I_{s+2}) \to \mathcal{T}(\Sigma I_{s+1}, \Sigma^2 I_{s+2})$$

is injective.

(b) The first equality holds by Remark 3.8. The second equality holds because $\Sigma \delta_{s+1}$ is $\mathcal{I}$–epic and ΣI_{s+2} is injective, as in part (a).

(c) The map $d_1\colon I_s \to \Sigma I_{s+1}$ is the composite $I_s \xrightarrow{\delta_s} \Sigma Y_{s+1} \xrightarrow{\Sigma p_{s+1}} \Sigma I_{s+1}$. The octahedral axiom applied to this factorization yields the dotted arrows in a commutative diagram

$$
\begin{array}{ccccccc}
I_s & \xrightarrow{\delta_s} & \Sigma Y_{s+1} & \xrightarrow{\Sigma i_s} & \Sigma Y_s & \xrightarrow{-\Sigma p_s} & \Sigma I_s \\
\| & & \downarrow{\scriptstyle \Sigma p_{s+1}} & & \downarrow{\scriptstyle \widetilde{\alpha}} & & \| \\
I_s & \xrightarrow{d_1} & \Sigma I_{s+1} & \xrightarrow{q} & C_{d_1} & \xrightarrow{\iota} & \Sigma I_s \\
 & & \downarrow{\scriptstyle \Sigma \delta_{s+1}} & & \downarrow{\scriptstyle \widetilde{\beta}} & & \\
 & & \Sigma^2 Y_{s+2} & = & \Sigma^2 Y_{s+2} & & \\
 & & \downarrow{\scriptstyle -\Sigma^2 i_{s+1}} & & \downarrow & & \\
 & & \Sigma^2 Y_{s+1} & \xrightarrow{\Sigma^2 i_s} & \Sigma^2 Y_s & &
\end{array}
$$

where the rows and columns are distinguished and the equation $(-\Sigma^2 i_{s+1})\widetilde{\beta} = (\Sigma\delta_s)\iota$ holds. The restricted bracket $\langle \Sigma d_1 \overset{\beta}{,} d_1, x\rangle_{\mathrm{fc}}$ consists of the maps $\Sigma X \to \Sigma^2 I_{s+2}$ appearing as downward composites in the commutative diagram

$$
\begin{array}{ccccccc}
 & & & & \Sigma X & \xrightarrow{-\Sigma x} & \Sigma I_s \\
 & & & & \downarrow{\scriptstyle \Sigma\alpha} & & \| \\
I_s & \xrightarrow{d_1} & \Sigma I_{s+1} & \xrightarrow{q} & C_{d_1} & \xrightarrow{\iota} & \Sigma I_s \\
 & & \| & \swarrow{\scriptstyle \widetilde{\beta}} & \downarrow{\scriptstyle \beta} & & \\
 & & \Sigma I_{s+1} \xrightarrow{\Sigma\delta_{s+1}} & \Sigma^2 Y_{s+2} & \xrightarrow{\Sigma^2 p_{s+2}} \Sigma^2 I_{s+2} & & \\
 & & \Sigma I_{s+1} & \xrightarrow{\Sigma d_1} & \Sigma^2 I_{s+2} & &
\end{array}
$$

($\supseteq$) Let $\beta(\Sigma\alpha) \in \langle d_1 \overset{\beta}{,} d_1, x\rangle_{\mathrm{fc}}$. By definition of β, we have $\beta(\Sigma\alpha) = (\Sigma^2 p_{s+2})\widetilde{\beta}(\Sigma\alpha)$. Then $\widetilde{\beta}(\Sigma\alpha)\colon \Sigma X \to \Sigma^2 Y_{s+2}$ is a valid choice of the lift $\widetilde{x}$ in the definition of $d_2[x]$:

$$
\begin{aligned}
(\Sigma^2 i_{s+1})\widetilde{\beta}(\Sigma\alpha) &= -(\Sigma\delta_s)\iota(\Sigma\alpha) \\
&= -(\Sigma\delta_s)(-\Sigma x) \\
&= \Sigma(\delta_s x).
\end{aligned}
$$

($\subseteq$) Given a representative $(\Sigma p_{s+2})\widetilde{x} \in d_2[x]$, we will show that $\Sigma\widetilde{x}\colon \Sigma X \to \Sigma^2 Y_{s+2}$ factors as $\Sigma X \xrightarrow{\Sigma\alpha} C_{d_1} \xrightarrow{\widetilde{\beta}} \Sigma^2 Y_{s+2}$ for some $\Sigma\alpha$, yielding a factorization of the desired form:

$$
\begin{aligned}
(\Sigma^2 p_{s+2})(\Sigma\widetilde{x}) &= (\Sigma^2 p_{s+2})\widetilde{\beta}(\Sigma\alpha) \\
&= \beta(\Sigma\alpha).
\end{aligned}
$$

By construction, the map $(\Sigma^2 i_s)(-\Sigma^2 i_{s+1})\colon \Sigma^2 Y_{s+2} \to \Sigma^2 Y_s$ is a cofiber of $\widetilde{\beta}$. The condition

$$(\Sigma^2 i_s)(\Sigma^2 i_{s+1})(\Sigma\widetilde{x}) = (\Sigma^2 i_s)\Sigma(\delta_s x) = 0$$

guarantees the existence of some lift $\Sigma\alpha\colon \Sigma X \to C_{d_1}$ of $\Sigma\widetilde{x}$. The chosen lift $\Sigma\alpha$ might *not* satisfy $\iota(\Sigma\alpha) = -\Sigma x$, but we will correct it to a lift $\Sigma\alpha'$ which does. The two sides of the equation become equal after applying $-\Sigma\delta_s$, ie $(-\Sigma\delta_s)(-\Sigma x) = (-\Sigma\delta_s)\iota(\Sigma\alpha)$ holds. Hence, the error term factors as

$$-\Sigma x - \iota\Sigma\alpha = (-\Sigma p_s)(\Sigma\theta)$$

for some $\Sigma\theta\colon \Sigma X \to \Sigma Y_s$, since $-\Sigma p_s$ is a fiber of $-\Sigma\delta_s$. The corrected map $\Sigma\alpha' := \Sigma\alpha + \widetilde{\alpha}(\Sigma\theta)\colon \Sigma X \to C_{d_1}$ satisfies $\iota(\Sigma\alpha') = -\Sigma x$ and still satisfies $\widetilde{\beta}(\Sigma\alpha') = \widetilde{\beta}(\Sigma\alpha) = \Sigma\widetilde{x}$, since the correction term $\widetilde{\alpha}(\Sigma\theta)$ satisfies $\widetilde{\beta}\widetilde{\alpha}(\Sigma\theta) = 0$. □

Proposition 4.2 *The following inclusions of subsets hold in* $E_1^{s+2,t+1}$*:*

$$d_2[x] \subseteq (\Sigma^2 p_{s+2})\langle \Sigma\delta_{s+1}, d_1, x\rangle \subseteq \langle \Sigma d_1, d_1, x\rangle.$$

Proof The first inclusion is

$$d_2[x] = (\Sigma^2 p_{s+2})\langle \Sigma\delta_{s+1}, \Sigma p_{s+1}, \delta_s x\rangle \subseteq (\Sigma^2 p_{s+2})\langle \Sigma\delta_{s+1}, (\Sigma p_{s+1})\delta_s, x\rangle,$$

whereas the second inclusion is

$$(\Sigma^2 p_{s+2})\langle \Sigma\delta_{s+1}, d_1, x\rangle \subseteq \langle (\Sigma^2 p_{s+2})(\Sigma\delta_{s+1}), d_1, x\rangle,$$

both using Lemma 3.5. □

Proposition 4.3 *The inclusion* $(\Sigma^2 p_{s+2})\langle \Sigma\delta_{s+1}, d_1, x\rangle \subseteq \langle \Sigma d_1, d_1, x\rangle$ *need **not** be an equality in general.*

It was pointed out to us by Robert Bruner that this can happen in principle. We give an explicit example in Proposition A.1.

5 Higher Toda brackets

We saw in Section 3 that there are several equivalent ways to define 3–fold Toda brackets. Following the approach given in [30], we show that the fiber-cofiber definition generalizes nicely to n–fold Toda brackets. There are $(n-2)!$ ways to make this generalization, and we prove that they are all the same up to a specified sign. We also show that this Toda bracket is self-dual.

Other sources that discuss higher Toda brackets in triangulated categories are [40, Appendix A], [18, Chapter IV, Section 2] and [36, Section 4], which all give definitions that follow Cohen's approach for spectra or spaces [15]. We show that our definition agrees with those of [40] and [36]. (We believe that it sometimes differs in sign from [15]. We have not compared carefully with [18].)

Definition 5.1 Let $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3$ be a diagram in a triangulated category $\mathcal{T}$. We define the *Toda family* of this sequence to be the collection $\mathrm{T}(f_3, f_2, f_1)$ consisting of all pairs $(\beta, \Sigma\alpha)$, where α and β appear in a commutative diagram

$$\begin{array}{ccccccc} X_0 & \xrightarrow{f_1} & X_1 & & & & \\ {\scriptstyle\alpha}\downarrow & & \| & & & & \\ \Sigma^{-1}C_{f_2} & \longrightarrow & X_1 & \xrightarrow{f_2} & X_2 & \longrightarrow & C_{f_2} \\ & & & & \| & & \downarrow{\scriptstyle\beta} \\ & & & & X_2 & \xrightarrow{f_3} & X_3 \end{array}$$

with distinguished middle row. Equivalently,

$$\begin{array}{ccccccc} & & & & \Sigma X_0 & \xrightarrow{-\Sigma f_1} & \Sigma X_1 \\ & & & & {\scriptstyle\Sigma\alpha}\downarrow & & \| \\ X_1 & \xrightarrow{f_2} & X_2 & \longrightarrow & C_{f_2} & \longrightarrow & \Sigma X_1 \\ & & \| & & \downarrow{\scriptstyle\beta} & & \\ & & X_2 & \xrightarrow{f_3} & X_3 & & \end{array}$$

where the middle row is again distinguished. (The negative of Σf_1 appears, since when a triangle is rotated, a sign is introduced.) Note that the maps in each pair form a composable sequence $\Sigma X_0 \xrightarrow{\Sigma\alpha} C_{f_2} \xrightarrow{\beta} X_3$, with varying intermediate object, and that the collection of composites $\beta \circ \Sigma\alpha$ is exactly the Toda bracket $\langle f_3, f_2, f_1 \rangle$, using the fiber-cofiber definition; see diagram (3-2). (Also note that the Toda family is generally a proper class, but this is only because the intermediate object can be varied up to isomorphism, and so we will ignore this.)

More generally, if S is a set of composable triples of maps, starting at X_0 and ending at X_3, we define $\mathrm{T}(S)$ to be the union of $\mathrm{T}(f_3, f_2, f_1)$ for each triple (f_3, f_2, f_1) in S.

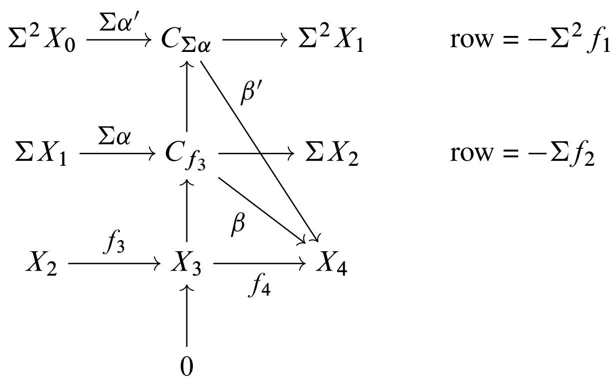
Definition 5.2 Let $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} \cdots \xrightarrow{f_n} X_n$ be a diagram in a triangulated category $\mathcal{T}$. We define the *Toda bracket* $\langle f_n, \dots, f_1 \rangle$ inductively as follows. If $n = 2$, it is the set consisting of just the composite $f_2 f_1$. If $n > 2$, it is the union of the sets $\langle \beta, \Sigma\alpha, \Sigma f_{n-3}, \dots, \Sigma f_1 \rangle$, where $(\beta, \Sigma\alpha)$ is in $\mathrm{T}(f_n, f_{n-1}, f_{n-2})$.

In fact, there are $(n-2)!$ such definitions, depending on a sequence of choices of which triple of consecutive maps to apply the Toda family construction to. In Theorem 5.11 we will enumerate these choices and show that they all agree up to sign.

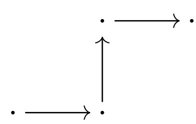
Example 5.3 Let us describe 4–fold Toda brackets in more detail. We have

$$\langle f_4, f_3, f_2, f_1 \rangle = \bigcup_{\beta,\alpha} \langle \beta, \Sigma\alpha, \Sigma f_1 \rangle = \bigcup_{\beta,\alpha} \bigcup_{\beta',\alpha'} \{\beta' \circ \Sigma\alpha'\}$$

with $(\beta, \Sigma\alpha) \in \mathrm{T}(f_4, f_3, f_2)$ and $(\beta', \Sigma\alpha') \in \mathrm{T}(\beta, \Sigma\alpha, \Sigma f_1)$. These maps fit into a commutative diagram



where the horizontal composites are specified as above, and each "snake"



is a distinguished triangle. The middle column is an example of a 3–*filtered object* as defined below.

Next, we will show that Definition 5.2 coincides with the definitions of higher Toda brackets in [40, Appendix A] and [36, Section 4], which we recall here.

Definition 5.4 Let $n \geq 1$ and consider a diagram in $\mathcal{T}$

$$Y_0 \xrightarrow{\lambda_1} Y_1 \xrightarrow{\lambda_2} Y_2 \longrightarrow \cdots \xrightarrow{\lambda_{n-1}} Y_{n-1}$$

consisting of $n-1$ composable maps. An *n–filtered object* Y based on $(\lambda_{n-1},\dots,\lambda_1)$ consists of a sequence of maps

$$0 = F_0Y \xrightarrow{i_0} F_1Y \xrightarrow{i_1} \cdots \xrightarrow{i_{n-1}} F_nY = Y$$

together with distinguished triangles

$$F_jY \xrightarrow{i_j} F_{j+1}Y \xrightarrow{q_{j+1}} \Sigma^j Y_{n-1-j} \xrightarrow{e_j} \Sigma F_j Y$$

for $0 \le j \le n-1$, such that for $1 \le j \le n-1$, the composite

$$\Sigma^j Y_{n-1-j} \xrightarrow{e_j} \Sigma F_j Y \xrightarrow{\Sigma q_j} \Sigma^j Y_{n-j}$$

is equal to $\Sigma^j \lambda_{n-j}$. In particular, the n–filtered object Y comes equipped with maps

$$\sigma'_Y \colon Y_{n-1} \cong F_1Y \to Y,$$
$$\sigma_Y \colon Y = F_nY \to \Sigma^{n-1} Y_0.$$

Definition 5.5 Let $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} \cdots \xrightarrow{f_n} X_n$ be a diagram in a triangulated category $\mathcal{T}$. The *Toda bracket* in the sense of Shipley and Sagave, $\langle f_n,\dots,f_1\rangle_{\mathrm{SS}} \subseteq \mathcal{T}(\Sigma^{n-2}X_0, X_n)$, is the set of all composites appearing in the middle row of a commutative diagram

$$\begin{array}{ccccc} & & X_{n-1} & & \\ & & \downarrow{\scriptstyle \sigma'_X} & \searrow{\scriptstyle f_n} & \\ \Sigma^{n-2}X_0 & \dashrightarrow & X & \dashrightarrow & X_n \\ & \searrow{\scriptstyle \Sigma^{n-2}f_1} & \downarrow{\scriptstyle \sigma_X} & & \\ & & \Sigma^{n-2}X_1 & & \end{array}$$

where X is an $(n-1)$–filtered object based on $(f_{n-1},\dots,f_3,f_2)$.

Example 5.6 For a 3–fold Toda bracket $\langle f_3,f_2,f_1\rangle_{\mathrm{SS}}$, a 2–filtered object X based on f_2 amounts to a cofiber of $-f_2$, more precisely, a distinguished triangle

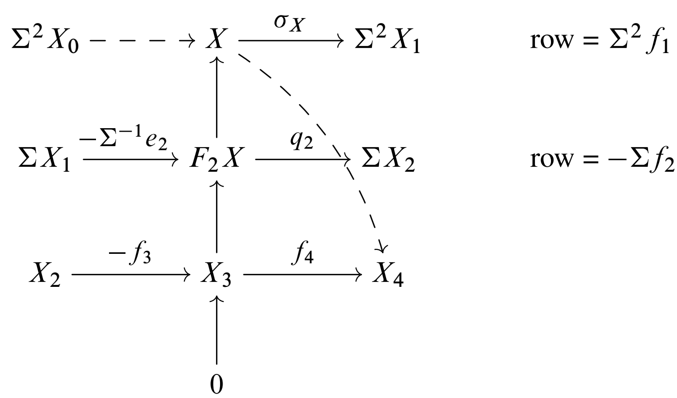
$$X_2 \xrightarrow{\sigma'_X} X \xrightarrow{\sigma_X} \Sigma X_1 \xrightarrow{\Sigma f_2} \Sigma X_2.$$

Using this, one readily checks the equality $\langle f_3,f_2,f_1\rangle_{\mathrm{SS}} = \langle f_3,f_2,f_1\rangle_{\mathrm{fc}}$, as noted in [36, Definition 4.5].

Example 5.7 For a 4–fold Toda bracket $\langle f_4, f_3, f_2, f_1\rangle_{\mathrm{SS}}$, a 3–filtered object X based on (f_3, f_2) consists of the data displayed in the diagram

$$\begin{array}{ccccccc}
 & & F_3X = X & \xrightarrow{q_3=\sigma_X} & \Sigma^2 X_1 & & \\
 & & \uparrow{\scriptstyle i_2} & & & & \\
\Sigma X_1 & \xrightarrow{-\Sigma^{-1}e_2} & F_2X & \xrightarrow{q_2} & \Sigma X_2 & & \text{row} = -\Sigma f_2 \\
 & & \uparrow{\scriptstyle i_1} & & & & \\
X_2 & \xrightarrow{-\Sigma^{-1}e_1} & F_1X & \xrightarrow[\cong]{q_1} & X_3 & & \text{row} = -f_3 \\
 & & \uparrow{\scriptstyle i_0} & & & & \\
 & & F_0X = 0 & & & &
\end{array}$$

where the two snakes are distinguished. The bracket consists of the maps $\Sigma^2 X_0 \to X_4$ appearing as composites of the dotted arrows in a commutative diagram



where the two snakes are distinguished. By negating the first and third map in each snake, this recovers the description in Example 5.3, thus proving the equality of subsets

$$\langle f_4, f_3, f_2, f_1\rangle_{\mathrm{SS}} = \langle f_4, f_3, f_2, f_1\rangle.$$

Proposition 5.8 *Definitions 5.2 and 5.5 agree. In other words, we have the equality*

$$\langle f_n, \dots, f_1\rangle_{\mathrm{SS}} = \langle f_n, \dots, f_1\rangle$$

of subsets of $\mathcal{T}(\Sigma^{n-2} X_0, X_n)$.

Proof This is a straightforward generalization of Example 5.7. □

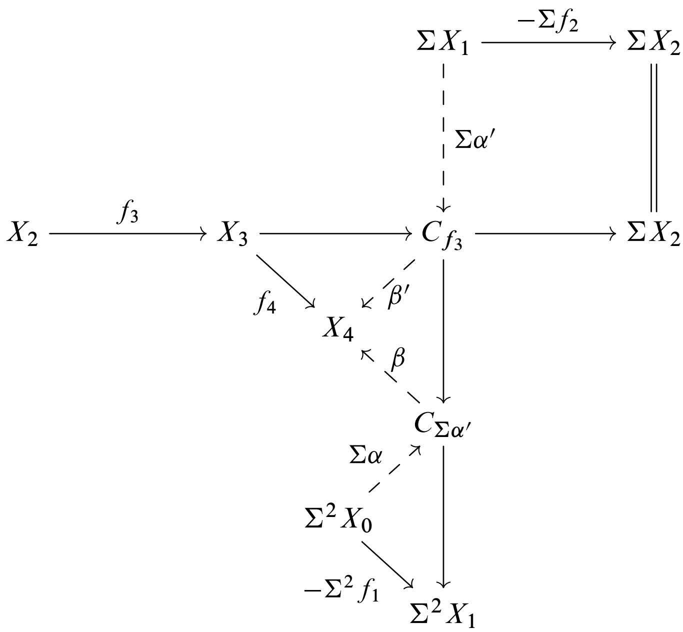
We define the *negative* of a Toda family $T(f_3, f_2, f_1)$ to consist of pairs $(\beta, -\Sigma\alpha)$ for $(\beta, \Sigma\alpha) \in T(f_3, f_2, f_1)$. (Since changing the sign of two maps in a triangle doesn't affect whether it is distinguished, it would be equivalent to put the minus sign with the β.)

Lemma 5.9 *Let* $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3 \xrightarrow{f_4} X_4$ *be a diagram in a triangulated category* $\mathcal{T}$. *Then the two sets of pairs* $T(T(f_4, f_3, f_2), \Sigma f_1)$ *and* $T(f_4, T(f_3, f_2, f_1))$ *are negatives of each other.*

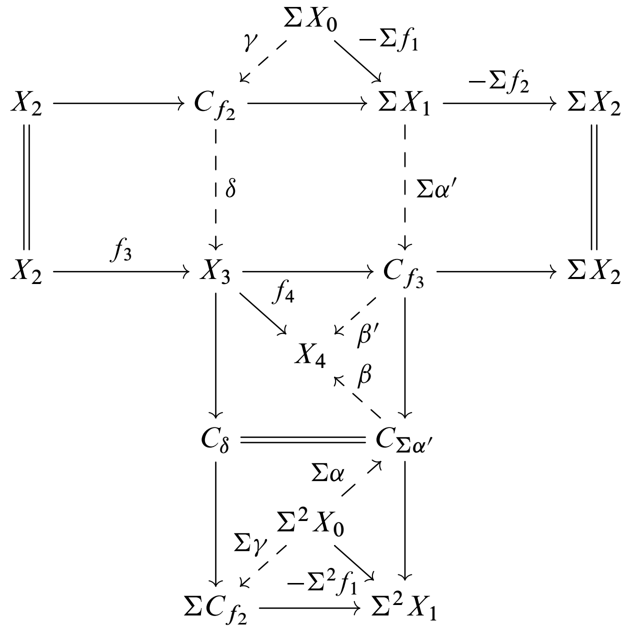
This is stronger than saying the two ways of computing the Toda bracket $\langle f_4, f_3, f_2, f_1\rangle$ are negatives, and the stronger statement will be used inductively to prove Theorem 5.11.

Proof We will show that the negative of $T(T(f_4, f_3, f_2), \Sigma f_1)$ is contained in the family $T(f_4, T(f_3, f_2, f_1))$. The reverse inclusion is proved dually.

Suppose $(\beta, \Sigma\alpha)$ is in $T(T(f_4, f_3, f_2), \Sigma f_1)$, that is, $(\beta, \Sigma\alpha)$ is in $T(\beta', \Sigma\alpha', \Sigma f_1)$ for some $(\beta', \Sigma\alpha')$ in $T(f_4, f_3, f_2)$. This means that we have the following commutative diagram, in which the long row and column are distinguished triangles:



Using the octahedral axiom, there exists a map $\delta\colon C_{f_2} \to X_3$ in the following diagram making the two squares commute, and such that the diagram can be extended as shown, with all rows and columns distinguished:



Define $\Sigma\gamma$ to be the composite $\Sigma^2 X_0 \to C_{\Sigma\alpha'} = C_\delta \to \Sigma C_{f_2}$, where the first map is $\Sigma\alpha$. Then the small triangles at the top and bottom of the last diagram commute as well. Therefore, (δ, γ) is in $\mathrm{T}(f_3, f_2, f_1)$. Moreover, this diagram shows that $(\beta, -\Sigma\alpha)$ is in $\mathrm{T}(f_4, \delta, \gamma)$, completing the argument. □

To concisely describe different ways of computing higher Toda brackets, we introduce the following notation. For $0 \le j \le n-3$, write $\mathrm{T}_j(f_n, f_{n-1}, \dots, f_1)$ for the set of tuples

$$\{(f_n, f_{n-1}, \dots, f_{n-j+1}, \beta, \Sigma\alpha, \Sigma f_{n-j-3}, \dots, \Sigma f_1)\},$$

where $(\beta, \Sigma\alpha)$ is in $\mathrm{T}(f_{n-j}, f_{n-j-1}, f_{n-j-2})$. (There are j maps to the left of the three used for the Toda family.) If S is a set of n–tuples of composable maps, we define $\mathrm{T}_j(S)$ to be the union of the sets $\mathrm{T}_j(f_n, f_{n-1}, \dots, f_1)$ for $(f_n, f_{n-1}, \dots, f_1)$ in S. With this notation, the standard Toda bracket $\langle f_n, \dots, f_1\rangle$ consists of the composites of all the pairs occurring in the iterated Toda family

$$\mathrm{T}(f_n, \dots, f_1) := \mathrm{T}_0(\mathrm{T}_0(\mathrm{T}_0(\cdots \mathrm{T}_0(f_n, \dots, f_1)\cdots))).$$

A general Toda bracket is of the form $\mathrm{T}_{j_1}(\mathrm{T}_{j_2}(\mathrm{T}_{j_3}(\cdots \mathrm{T}_{j_{n-2}}(f_n, \dots, f_1)\cdots)))$, where $j_1, j_2, \dots, j_{n-2}$ is a sequence of natural numbers with $0 \le j_i < i$ for each i. There are $(n-2)!$ such sequences.

Remark 5.10 There are six ways to compute the 5–fold Toda bracket $\langle f_5, f_4, f_3, f_2, f_1\rangle$

as the set of composites of the pairs of maps in one of the following sets:

$$\begin{aligned}
&\mathrm{T}_0(\mathrm{T}_0(\mathrm{T}_0(f_5, f_4, f_3, f_2, f_1))) = \mathrm{T}(\mathrm{T}(\mathrm{T}(f_5, f_4, f_3), \Sigma f_2), \Sigma^2 f_1),\\
&\mathrm{T}_0(\mathrm{T}_0(\mathrm{T}_1(f_5, f_4, f_3, f_2, f_1))) = \mathrm{T}(\mathrm{T}(f_5, \mathrm{T}(f_4, f_3, f_2)), \Sigma^2 f_1),\\
&\mathrm{T}_0(\mathrm{T}_1(\mathrm{T}_1(f_5, f_4, f_3, f_2, f_1))) = \mathrm{T}(f_5, \mathrm{T}(\mathrm{T}(f_4, f_3, f_2), \Sigma f_1)),\\
&\mathrm{T}_0(\mathrm{T}_1(\mathrm{T}_2(f_5, f_4, f_3, f_2, f_1))) = \mathrm{T}(f_5, \mathrm{T}(f_4, \mathrm{T}(f_3, f_2, f_1))),\\
&\mathrm{T}_0(\mathrm{T}_0(\mathrm{T}_2(f_5, f_4, f_3, f_2, f_1))),\\
&\mathrm{T}_0(\mathrm{T}_1(\mathrm{T}_0(f_5, f_4, f_3, f_2, f_1))).
\end{aligned}$$

The last two cannot be expressed directly just using T.

Now we can prove the main result of this section.

Theorem 5.11 *The Toda bracket computed using the sequence $j_1, j_2, \dots, j_{n-2}$ equals the standard Toda bracket up to the sign $(-1)^{\sum j_i}$.*

Proof Let $j_1, j_2, \dots, j_{n-2}$ be a sequence with $0 \le j_i < i$ for each i. Lemma 5.9 tells us that if we replace consecutive entries $k, k+1$ with k, k in any such sequence, the two Toda brackets agree up to a sign. To begin with, we ignore the signs. We will prove by induction on ℓ that the initial portion $j_1, \dots, j_\ell$ of such a sequence can be converted into any other sequence, using just the move allowed by Lemma 5.9 and its inverse, and without changing j_i for $i > \ell$. For $\ell = 1$, there is only one sequence 0. For $\ell = 2$, there are two sequences: $0, 0$ and $0, 1$, and Lemma 5.9 applies. For $\ell > 2$, suppose our goal is to produce the sequence $j'_1, \dots, j'_\ell$. We break the argument into three cases:

$\boldsymbol{j'_\ell = j_\ell}$ We can directly use the induction hypothesis to adjust the entries in the first $\ell - 1$ positions.

$\boldsymbol{j'_\ell > j_\ell}$ By induction, we can change the first $\ell - 1$ entries in the sequence j so that the entry in position $\ell - 1$ is j_ℓ, since $j_\ell < j'_\ell \le \ell - 1$. Then, using Lemma 5.9, we can change the entry in position ℓ to $j_\ell + 1$. Continuing in this way, we get j'_ℓ in position ℓ, and then we are in the first case.

$\boldsymbol{j'_\ell < j_\ell}$ Since the moves are reversible, this is equivalent to the second case.

To handle the sign, first note that signs propagate through the Toda family construction. More precisely, suppose S is a set of n–tuples of maps, and let S' be a set obtained by negating the k^{th} map in each n–tuple for some fixed k. Then $\mathrm{T}_j(S)$ has the same relationship to $\mathrm{T}_j(S')$, possibly for a different value of k.

As a result, applying the move of Lemma 5.9 changes the resulting Toda bracket by a sign. That move also changes the parity of $\sum_i j_i$. Since we get a plus sign when each j_i is zero, it follows that the difference in sign in general is $(-1)^{\sum_i j_i}$. □

An animation of this argument is available at [13]. It was pointed out by Dylan Wilson that the combinatorial part of the above proof is equivalent to the well-known fact that if a binary operation is associative on triples, then it is associative on n–tuples.

In order to compare our Toda brackets to the Toda brackets in the opposite category, we need one lemma.

Lemma 5.12 *If $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} X_3$ is a diagram in a triangulated category $\mathcal{T}$, then the Toda family $\mathrm{T}(\Sigma f_3, \Sigma f_2, \Sigma f_1)$ is the negative of the suspension of $\mathrm{T}(f_3, f_2, f_1)$. That is, it consists of $(\Sigma\beta, -\Sigma^2\alpha)$ for $(\beta, \Sigma\alpha)$ in $\mathrm{T}(f_3, f_2, f_1)$.*

Proof Given a distinguished triangle $\Sigma^{-1} C_{f_2} \xrightarrow{k} X_1 \xrightarrow{f_2} X_2 \xrightarrow{\iota} C_{f_2}$, a distinguished triangle involving Σf_2 is

$$C_{f_2} \xrightarrow{-\Sigma k} \Sigma X_1 \xrightarrow{\Sigma f_2} \Sigma X_2 \xrightarrow{\Sigma \iota} \Sigma C_{f_2}.$$

Because of the minus sign at the left, the maps that arise in the Toda family based on this triangle are $-\Sigma^2\alpha$ and $\Sigma\beta$, where $\Sigma\alpha$ and β arise in the Toda family based on the starting triangle. □

Given a triangulated category $\mathcal{T}$, the opposite category $\mathcal{T}^{\mathrm{op}}$ is triangulated in a natural way. The suspension in $\mathcal{T}^{\mathrm{op}}$ is Σ^{-1} and a triangle

$$Y_0 \xrightarrow{g_1} Y_1 \xrightarrow{g_2} Y_2 \xrightarrow{g_3} \Sigma^{-1} Y_0$$

in $\mathcal{T}^{\mathrm{op}}$ is distinguished if and only if the triangle

$$\Sigma\Sigma^{-1} Y_0 \xleftarrow{g_1'} Y_1 \xleftarrow{g_2} Y_2 \xleftarrow{g_3} \Sigma^{-1} Y_0$$

in $\mathcal{T}$ is distinguished, where g_1' is the composite of g_1 with the natural isomorphism $Y_0 \cong \Sigma\Sigma^{-1} Y_0$.

Corollary 5.13 *The Toda bracket is self-dual up to suspension. More precisely, let $X_0 \xrightarrow{f_1} X_1 \xrightarrow{f_2} X_2 \xrightarrow{f_3} \cdots \xrightarrow{f_n} X_n$ be a diagram in a triangulated category $\mathcal{T}$. Then the subset*

$$\langle f_1, \ldots, f_n \rangle^{\mathcal{T}^{\mathrm{op}}} \subseteq \mathcal{T}^{\mathrm{op}}(\Sigma^{-(n-2)} X_n, X_0) = \mathcal{T}(X_0, \Sigma^{-(n-2)} X_n)$$

defined by taking the Toda bracket in $\mathcal{T}^{\mathrm{op}}$ *is sent to the subset*

$$\langle f_n, \ldots, f_1 \rangle^{\mathcal{T}} \subseteq \mathcal{T}(\Sigma^{n-2} X_0, X_n)$$

defined by taking the Toda bracket in $\mathcal{T}$ *under the bijection* $\Sigma^{n-2}\colon \mathcal{T}(X_0, \Sigma^{-(n-2)} X_n) \to \mathcal{T}(\Sigma^{n-2} X_0, X_n)$.

Proof First we compare Toda families in $\mathcal{T}$ and $\mathcal{T}^{\mathrm{op}}$. It is easy to see that the Toda family $\mathrm{T}^{\mathcal{T}^{\mathrm{op}}}(f_1, f_2, f_3)$ computed in $\mathcal{T}^{\mathrm{op}}$ consists of the pairs $(\alpha, \Sigma^{-1}\beta)$ for $(\Sigma\alpha, \beta)$ in the Toda family $\mathrm{T}^{\mathcal{T}}(f_3, f_2, f_1)$ computed in $\mathcal{T}$. In short, one has to desuspend and transpose the pairs.

Using this, one can see that the iterated Toda family

$$\mathrm{T}^{\mathcal{T}^{\mathrm{op}}}(\mathrm{T}^{\mathcal{T}^{\mathrm{op}}} \cdots \mathrm{T}^{\mathcal{T}^{\mathrm{op}}}(f_1, f_2, f_3), \ldots, \Sigma^{-(n-3)} f_n)$$

is equal to the transpose of

$$\Sigma^{-1}\mathrm{T}^{\mathcal{T}}\big(\Sigma^{-(n-3)} f_n, \Sigma^{-1}\mathrm{T}^{\mathcal{T}}(\Sigma^{-(n-4)} f_{n-1}, \Sigma^{-1}\mathrm{T}^{\mathcal{T}} \cdots \Sigma^{-1}\mathrm{T}^{\mathcal{T}}(f_3, f_2, f_1) \cdots)\big).$$

By Lemma 5.12, the desuspensions pass through all of the Toda family constructions, introducing an overall sign of $(-1)^{1+2+3+\cdots+(n-3)}$, and producing

$$\Sigma^{-(n-2)}\mathrm{T}^{\mathcal{T}}(f_n, \mathrm{T}^{\mathcal{T}}(f_{n-1}, \mathrm{T}^{\mathcal{T}} \cdots \mathrm{T}^{\mathcal{T}}(f_3, f_2, f_1) \cdots)).$$

By Theorem 5.11, composing the pairs gives the usual Toda bracket up to the sign $(-1)^{0+1+2+\cdots+(n-3)}$. The two signs cancel, yielding the result. □

We do not know a direct proof of this corollary. To summarize, our insight is that by generalizing the corollary to all $(n-2)!$ methods of computing the Toda bracket, we were able to reduce the argument to the 4–fold case (Lemma 5.9) and some combinatorics.

Remark 5.14 As with the 3–fold Toda brackets (see Remark 3.2), the higher Toda brackets depend on the triangulation. If the triangulation is negated, the n–fold Toda brackets change by the sign $(-1)^n$.

6 Higher order operations determine d_r

In this section, we show that the higher Adams differentials can be expressed in terms of higher Toda brackets, in two ways. One of these expressions is as an r^{th} order cohomology operation.

Given an injective class $\mathcal{I}$, an Adams resolution of an object Y as in diagram (2-2) and an object X, consider a class $[x] \in E_r^{s,t}$ represented by an element $x \in E_1^{s,t} = \mathcal{T}(\Sigma^{t-s} X, I_s)$. The class $d_r[x]$ is the set of all $(\Sigma p_{s+r})\widetilde{x}$, where $\widetilde{x}$ runs over lifts of $\delta_s x$ through the $(r-1)$–fold composite $\Sigma(i_{s+1} \cdots i_{s+r-1})$ which appears across the top edge of the Adams resolution.

Our first result will be a generalization of Proposition 4.1(a), expressing d_r in terms of an $(r+1)$–fold Toda bracket.

Theorem 6.1 *As subsets of $E_1^{s+r,t+r-1}$, we have*

$$d_r[x] = \langle \Sigma^{r-1} d_1, \ldots, \Sigma^2 d_1, \Sigma d_1, \Sigma p_{s+1}, \delta_s x \rangle.$$

Proof We compute the Toda bracket, applying the Toda family construction starting from the right, which introduces a sign of $(-1)^{1+2+\cdots+(r-2)}$, by Theorem 5.11. We begin with the Toda family $\mathrm{T}(\Sigma d_1, \Sigma p_{s+1}, \delta_s x)$. There is a distinguished triangle

$$\Sigma Y_{s+2} \xrightarrow{\Sigma i_{s+1}} \Sigma Y_{s+1} \xrightarrow{\Sigma p_{s+1}} \Sigma I_{s+1} \xrightarrow{\Sigma \delta_{s+1}} \Sigma^2 Y_{s+2},$$

with no needed signs. The map Σd_1 factors through $\Sigma \delta_{s+1}$ as $\Sigma^2 p_{s+2}$, and this factorization is unique because $\Sigma \delta_{s+1}$ is $\mathcal{I}$–epic and $\Sigma^2 I_{s+2}$ is injective. The other maps in the Toda family are Σx_1, where x_1 is a lift of $\delta_s x$ through Σi_{s+1}. So

$$\mathrm{T}(\Sigma d_1, \Sigma p_{s+1}, \delta_s x) = \{(\Sigma^2 p_{s+2},\ \Sigma x_1) \mid x_1 \text{ a lift of } \delta_s x \text{ through } \Sigma i_{s+1}\}.$$

(The Toda family also includes $(\Sigma^2 p_{s+2}\, \phi,\ \phi^{-1}(\Sigma x_1))$, where ϕ is any isomorphism, but these contribute nothing additional to the later computations.) The composites of such pairs give $d_2[x]$, up to suspension, recovering Proposition 4.1(a).

Continuing, for each such pair we compute

$$\begin{aligned} \mathrm{T}(\Sigma^2 d_1, \Sigma^2 p_{s+2}, \Sigma x_1) &= -\Sigma \mathrm{T}(\Sigma d_1, \Sigma p_{s+2}, x_1) \\ &= -\Sigma\{(\Sigma^2 p_{s+3},\ \Sigma x_2) \mid x_2 \text{ a lift of } x_1 \text{ through } \Sigma i_{s+2}\}. \end{aligned}$$

The first equality is Lemma 5.12, and the second reuses the work done in the previous paragraph, with s increased by 1. Composing these pairs gives $-d_3[x]$. The sign which is needed to produce the standard Toda bracket is $(-1)^1$, and so the signs cancel.

At the next step, we compute

$$\begin{aligned} \mathrm{T}(\Sigma^3 d_1, \Sigma^3 p_{s+3}, -\Sigma^2 x_2) &= -\Sigma^2 \mathrm{T}(\Sigma d_1, \Sigma p_{s+3}, x_2) \\ &= -\Sigma^2\{(\Sigma^2 p_{s+4}, \Sigma x_3) \mid x_3 \text{ a lift of } x_2 \text{ through } \Sigma i_{s+3}\}. \end{aligned}$$

Again, the composites give $-d_4[x]$. Since it was a double suspension that passed through the Toda family, no additional sign was introduced. Similarly, the sign to convert to the standard Toda bracket is $(-1)^{1+2}$, and since 2 is even, no additional sign was introduced. Therefore, the signs still cancel.

The pattern continues. In total, there are $1+2+\cdots+(r-2)$ suspensions that pass through the Toda family, and the sign to convert to the standard Toda bracket is also based on that number, so the signs cancel. □

Remark 6.2 Theorem 6.1 can also be proved using the definition of Toda brackets based on r–filtered objects, as in Definitions 5.4 and 5.5. However, one must work in the opposite category $\mathcal{T}^{\mathrm{op}}$. In that category, there is a unique r–filtered object, up to isomorphism, based on the maps in the Toda bracket. One of the dashed arrows in the diagram from Definition 5.5 is also unique, and the other corresponds naturally to the choice of lift in the Adams differential.

In the remainder of this section, we describe the analog of Proposition 4.1(c). We begin by defining restricted higher Toda brackets, in terms of restricted Toda families.

Consider a Toda family $\mathrm{T}(gh_1, g_1h_0, g_0h)$, where the maps factor as shown, there are distinguished triangles

(6-1) $$Z_i \xrightarrow{g_i} J_i \xrightarrow{h_i} Z_{i+1} \xrightarrow{k_i} \Sigma Z_i$$

for $i=0,1$, and g and h are arbitrary maps $Z_2 \to A$ and $B \to Z_0$, respectively. This information determines an essentially unique element of the Toda family in the following way. The octahedral axiom applied to the factorization g_1h_0 yields the dotted arrows in a commutative diagram

$$\begin{array}{ccccccc}
J_0 & \xrightarrow{h_0} & Z_1 & \xrightarrow{k_0} & \Sigma Z_0 & \xrightarrow{-\Sigma g_0} & \Sigma J_0 \\
\| & & \downarrow{\scriptstyle g_1} & & \downarrow{\scriptstyle \alpha_2} & & \| \\
J_0 & \xrightarrow{g_1 h_0} & J_1 & \xrightarrow{q} & W_2 & \xrightarrow{\iota} & \Sigma J_0 \\
 & & \downarrow{\scriptstyle h_1} & & \downarrow{\scriptstyle \beta_2} & & \\
 & & Z_2 & = & Z_2 & & \\
 & & \downarrow{\scriptstyle k_1} & & \downarrow{\scriptstyle \gamma_2} & & \\
 & & \Sigma Z_1 & \xrightarrow{\Sigma k_0} & \Sigma^2 Z_0 & &
\end{array}$$

where the rows and columns are distinguished and $\gamma_2 := (\Sigma k_0)k_1$. It is easy to see that $-\Sigma(g_0 h)$ lifts through ι as $\alpha_2(\Sigma h)$, and that gh_1 extends over q as $g\beta_2$. We define the *restricted Toda family* to be the set $\mathrm{T}(gh_1 \overset{!}{,} g_1 h_0 \overset{!}{,} g_0 h)$ consisting of the pairs $(g\beta_2,\ \alpha_2(\Sigma h))$ that arise in this way. Since α_2 and β_2 come from a distinguished triangle involving a fixed map γ_2, such pairs are unique up to the usual ambiguity of replacing the pair with $(g\beta_2\phi,\ \phi^{-1}\alpha_2(\Sigma h))$, where ϕ is an isomorphism. Similarly, given any map $x\colon B \to J_0$, we define $\mathrm{T}(gh_1 \overset{!}{,} g_1 h_0, x)$ to be the set consisting of the pairs $(g\beta_2,\ \Sigma\alpha)$, where β_2 arises as above and $\Sigma\alpha$ is any lift of $-\Sigma x$ through ι.

Definition 6.3 Given distinguished triangles as in (6-1), for $i = 1, \dots, n-1$, and maps $g\colon Z_n \to A$ and $x\colon B \to J_1$, we define the *restricted Toda bracket*

$$\langle gh_{n-1} \overset{!}{,} g_{n-1}h_{n-2} \overset{!}{,} \dots \overset{!}{,} g_3 h_2 \overset{!}{,} g_2 h_1, x\rangle$$

inductively as follows. If $n = 2$, it is the set consisting of just the composite $gh_1 x$. If $n = 3$, it is the set of composites of the pairs in $\mathrm{T}(gh_2 \overset{!}{,} g_2 h_1, x)$. If $n > 3$, it is the union of the sets

$$\langle g\beta_2 \overset{!}{,} \alpha_2(\Sigma h_{n-3}) \overset{!}{,} \Sigma(g_{n-3}h_{n-4}) \overset{!}{,} \dots, \Sigma x\rangle,$$

where $(g\beta_2, \alpha_2(\Sigma h_{n-3}))$ is in $\mathrm{T}(gh_{n-1} \overset{!}{,} g_{n-1}h_{n-2} \overset{!}{,} g_{n-2}h_{n-3})$.

Remark 6.4 Despite the notation, we want to make it clear that these restricted Toda families and restricted Toda brackets depend on the choice of factorizations and on the distinguished triangles in (6-1). Moreover, the elements of the restricted Toda families are not simply pairs, but also include the factorizations of the maps in those pairs, and the distinguished triangle involving α_2 and β_2. This information is used in the $(n-1)$–fold restricted Toda bracket that is used to define the n–fold restricted Toda bracket.

Recall that the maps d_1 are defined to be $(\Sigma p_{s+1})\delta_s$, and that we have distinguished triangles

$$Y_s \xrightarrow{p_s} I_s \xrightarrow{\delta_s} \Sigma Y_{s+1} \xrightarrow{\Sigma i_s} \Sigma Y_s$$

for each s. The same holds for suspensions of d_1, with the last map changing sign each time it is suspended. Thus for $x\colon \Sigma^{t-s} X \to I_s$ in the E_1 term, the $(r+1)$–fold restricted Toda bracket $\langle \Sigma^{r-1} d_1 \overset{!}{,} \dots \overset{!}{,} \Sigma d_1 \overset{!}{,} d_1, x\rangle$ makes sense for each r, where we are implicitly using the defining factorizations and the triangles from the Adams resolution.

Theorem 6.5 *As subsets of $E_1^{s+r,t+r-1}$, we have*

$$d_r[x] = \langle \Sigma^{r-1} d_1 \overset{!}{,} \dots \overset{!}{,} \Sigma d_1 \overset{!}{,} d_1, x\rangle.$$

This is a generalization of Proposition 4.1(c). The data in the Adams resolution is the witness that the composites of the primary operations are zero in a sufficiently coherent way to permit an r^{th} order cohomology operation to be defined.

Proof The restricted Toda bracket $\langle \Sigma^{r-1} d_1 ¦ \ldots ¦ \Sigma d_1 ¦ d_1, x\rangle$ is defined recursively, working from the left. Each of the $r-2$ doubly restricted Toda families has essentially one element. The first one involves maps α_2, β_2 and γ_2 that form a distinguished triangle, and γ_2 is equal to $[(-1)^r \Sigma^r i_{s+r-2}][-(-1)^r \Sigma^r i_{s+r-1}]$. We will denote the corresponding maps in the following octahedra α_k, β_k and γ_k, where each γ_k equals $[(-1)^r \Sigma^r i_{s+r-k}]\gamma_{k-1}$, and so $\gamma_k = -(-1)^{rk} \Sigma^r (i_{s+r-k} \cdots i_{s+r-1})$. One is left to compute the singly restricted Toda family $\langle \Sigma^r p_{s+r} \beta_{r-1} ¦ \alpha_{r-1} \Sigma^{r-2} \delta_s, \Sigma^{r-2} x\rangle$, where α_{r-1} and β_{r-1} fit into a distinguished triangle

$$\Sigma^{r-1} Y_{s+1} \xrightarrow{\alpha_{r-1}} W_{r-1} \xrightarrow{\beta_{r-1}} \Sigma^r Y_{s+r} \xrightarrow{\gamma_{r-1}} \Sigma^r Y_{s+1},$$

and $\gamma_{r-1} = -\Sigma^r (i_{s+1} \cdots i_{s+r-1})$. Thus, to compute the last restricted Toda bracket, one uses the following diagram, obtained as usual from the octahedral axiom:

$$\begin{array}{ccccccc}
 & & & & & & \Sigma^{t-s+r-1} X \\
 & & & & & & \big\downarrow {\scriptstyle -\Sigma^{r-1} x} \\
\Sigma^{r-2} I_s & \xrightarrow{\Sigma^{r-2}\delta_s} & \Sigma^{r-1} Y_{s+1} & \xrightarrow{(-1)^r \Sigma^{r-1} i_s} & \Sigma^{r-1} Y_s & \xrightarrow{-\Sigma^{r-1} p_s} & \Sigma^{r-1} I_s \\
\| & & {\scriptstyle \alpha_{r-1}} \big\downarrow & & \big\downarrow {\scriptstyle \alpha_r} & & \| \\
\Sigma^{r-2} I_s & \longrightarrow & W_{r-1} & \xrightarrow{q_{r-1}} & W_r & \xrightarrow{\iota_{r-1}} & \Sigma^{r-1} I_s \\
 & & {\scriptstyle \beta_{r-1}} \big\downarrow & & \big\downarrow {\scriptstyle \beta_r} & & \\
\Sigma^r I_{s+r} & \xleftarrow{\Sigma^r p_{s+r}} & \Sigma^r Y_{s+r} & = & \Sigma^r Y_{s+r} & & \\
 & & {\scriptstyle \gamma_{r-1}} \big\downarrow & & \big\downarrow {\scriptstyle \gamma_r} & & \\
 & & \Sigma^r Y_{s+1} & \xrightarrow{(-1)^r \Sigma^r i_s} & \Sigma^r Y_s & &
\end{array}$$

Up to suspension, both $d_r[x]$ and the last restricted Toda bracket are computed by composing certain maps $\widetilde{x}\colon \Sigma^{t-s+r-2} X \to \Sigma^r Y_{s+r}$ with $\Sigma^r p_{s+r}$. For $d_r[x]$, the maps $\widetilde{x}$ must lift $\Sigma^{r-1}(\delta_s x)$ through $-\gamma_{r-1}$. For the last bracket, the maps $\widetilde{x}$ are of the form $\beta_r y$, where $y\colon \Sigma^{t-s+r-1} X \to W_r$ is a lift of $-\Sigma^{r-1} x$ through ι_{r-1}. As in the proof of Proposition 4.1(c), one can see that the possible choices of $\widetilde{x}$ coincide. $\square$

We next give a description of $d_r[x]$ using higher Toda brackets defined using filtered objects, as in Definitions 5.4 and 5.5. The computation of the restricted Toda bracket above produces a sequence

(6-2) $$0 = W_0 \xrightarrow{q_0} W_1 \xrightarrow{q_1} \cdots \xrightarrow{q_{r-1}} W_r,$$

where W_k is the fiber of the k–fold composite $\Sigma^r(i_{s+r-k} \cdots i_{s+r-1})$. (The map γ_k may differ in sign from this composite, but that doesn't affect the fiber.) For each k, we have a distinguished triangle

$$W_k \xrightarrow{q_k} W_{k+1} \xrightarrow{\iota_k} \Sigma^{r-1} I_{s+r-k-1} \xrightarrow{-(\Sigma\alpha_k)(\Sigma^{r-1}\delta_{s+r-k-1})} \Sigma W_k,$$

where we extend downwards to $k = 0$ by defining $W_1 = \Sigma^{r-1} I_{s+r-1}$ and using the nonobvious triangle

$$W_0 \xrightarrow{q_0=0} W_1 \xrightarrow{\iota_0=-1} \Sigma^{r-1} I_{s+r-1} \xrightarrow{0} \Sigma W_0.$$

One can check that

$$\begin{aligned}(\Sigma\iota_{k-1})(-\Sigma\alpha_k)(\Sigma^{r-1}\delta_{s+r-k-1}) &= (\Sigma^r p_{s+r-k})(\Sigma^{r-1}\delta_{s+r-k-1}) \\ &= \Sigma^{r-1} d_1 \\ &= \Sigma^k(\Sigma^{r-k-1} d_1),\end{aligned}$$

where $\Sigma^{r-k-1} d_1$ is the map appearing in the $(k+1)^{\text{st}}$ spot of the Toda bracket. In other words, the sequence (6-2) is an r–filtered object based on $(\Sigma^{r-2} d_1, \ldots, d_1)$.

The natural map $\sigma_W \colon W_r \to \Sigma^{r-1} I_s$ is ι_{r-1}, and the natural map $\sigma'_W \colon \Sigma^{r-1} I_{s+r-1} \cong W_1 \to W_r$ is the composite $q_{r-1} \cdots q_1 \iota_0 = -q_{r-1} \cdots q_1$. The Toda bracket computed using the filtered object W consists of all composites appearing in the middle row of this commutative diagram:

(6-3) $$\begin{array}{ccccc} & & \Sigma^{r-1} I_{s+r-1} & & \\ & & \downarrow{\scriptstyle \sigma'_W} & \searrow^{\Sigma^{r-1} d_1} & \\ \Sigma^{t-s+r-1} X & \overset{a}{\dashrightarrow} & W_r & \underset{b}{\dashrightarrow} & \Sigma^r I_{s+r} \\ & \searrow_{\Sigma^{r-1} x} & \downarrow{\scriptstyle \sigma_W} & & \\ & & \Sigma^{r-1} I_s & & \end{array}$$

We claim that there is a natural choice of extension b. Since

$$\Sigma^{r-1} d_1 = (\Sigma^r p_{s+r})(\Sigma^{r-1}\delta_{s+r-1}),$$

it suffices to extend $\Sigma^{r-1}\delta_{s+r-1}$ over σ'_W. Well, β_2 by definition is an extension of $\Sigma^{r-1}\delta_{s+r-1}$ over q_1, and each subsequent β_k gives a further extension. Because $\iota_0 = -1$, $-(\Sigma^r p_{s+r})\beta_r$ is a valid choice for b.

On the other hand, as described at the end of the previous proof, the lifts a of $\Sigma^{r-1}x$ through $\sigma_W = \iota_{r-1}$, when composed with $-(\Sigma^r p_{s+r})\beta_r$, give exactly the Toda bracket computed there.

In summary, we have the following theorem.

Theorem 6.6 *Given an Adams resolution of Y and $r \geq 2$, there is an associated r–filtered object W and a choice of a map b in diagram* (6-3)*, such that for any X and class $[x] \in E_r^{s,t}$, we have*

$$d_r[x] = \langle \Sigma^{r-1}d_1, \dots, \Sigma d_1, d_1, x\rangle,$$

where the Toda bracket is computed only using the r–filtered object W and the chosen extension b.

7 Sparse rings of operations

In this section, we focus on injective and projective classes which are generated by an object with a "sparse" endomorphism ring. In this context, we can give conditions under which the restricted Toda bracket appearing in Theorem 6.5 is equal to the unrestricted Toda bracket, producing a cleaner correspondence between Adams differentials and Toda brackets. We begin in Section 7.1 by giving the results in the case of an injective class, and then briefly summarize the dual results in Section 7.2. Section 7.3 gives examples.

Let us fix some notation and terminology, also discussed in [36; 34; 39, Section 2; 8].

Definition 7.1 Let N be a natural number. A graded abelian group R_* is *N–sparse* if R_* is concentrated in degrees which are multiples of N, ie $R_i = 0$ whenever $i \not\equiv 0 \pmod{N}$.

7.1 Injective case

Notation 7.2 Let E be an object of the triangulated category $\mathcal{T}$. Define the *E–cohomology* of an object X to be the graded abelian group E^*X given by $E^nX := \mathcal{T}(X, \Sigma^n E)$. Postcomposition makes E^*X into a left module over the graded endomorphism ring E^*E.

Assumption 7.3 For the remainder of Section 7.1, we assume the following:

(1) The triangulated category $\mathcal{T}$ has infinite products.
(2) The graded ring E^*E is N–sparse for some $N \geq 2$.

Let $\mathcal{I}_E$ denote the injective class generated by E, as in Example 2.6. Explicitly, $\mathcal{I}_E$ consists of retracts of (arbitrary) products $\prod_i \Sigma^{n_i} E$.

Lemma 7.4 *With this setup, we have the following:*

(1) *Let I be an injective object such that E^*I is N–sparse. Then I is a retract of a product $\prod_i \Sigma^{m_i N} E$.*
(2) *If, moreover, W is an object such that E^*W is N–sparse, then we have $\mathcal{T}(W, \Sigma^t I) = 0$ for $t \not\equiv 0 \pmod{N}$.*

Proof (1) I is a retract of a product $P = \prod_i \Sigma^{n_i} E$, with a map $\iota\colon I \hookrightarrow P$ and retraction $\pi\colon P \twoheadrightarrow I$. Consider the subproduct $P' = \prod_{N|n_i} \Sigma^{n_i} E$, with inclusion $\iota'\colon P' \hookrightarrow P$ (via the zero map into the missing factors) and projection $\pi'\colon P \twoheadrightarrow P'$. Then the equality

$$\iota'\pi'\iota = \iota\colon I \to P$$

holds, using the fact that E^*I is N–sparse. Therefore, we obtain $\pi\iota'\pi'\iota = \pi\iota = 1_I$, so that I is a retract of P'.

(2) By the first part, $\mathcal{T}(W, \Sigma^t I)$ is a retract of

$$\begin{aligned}\mathcal{T}(W, \Sigma^t \prod_i \Sigma^{m_i N} E) &= \mathcal{T}(W, \prod_i \Sigma^{m_i N+t} E)\\ &= \prod_i \mathcal{T}(W, \Sigma^{m_i N+t} E)\\ &= \prod_i E^{m_i N+t} W\\ &= 0,\end{aligned}$$

using the assumption that E^*W is N–sparse. □

Lemma 7.5 *Let $I_0 \xrightarrow{f_1} I_1 \xrightarrow{f_2} I_2 \to \cdots \xrightarrow{f_r} I_r$ be a diagram in $\mathcal{T}$, with $r \leq N+1$. Assume that each object I_j is injective and that each $E^*(I_j)$ is N–sparse. Then the iterated Toda family $\mathrm{T}(f_r, f_{r-1}, \ldots, f_1)$ is either empty or consists of a single composable pair $\Sigma^{r-2} I_0 \to C \to I_r$, up to automorphism of C.*

Proof In the case $r = 2$, there is nothing to prove, so we may assume $r \geq 3$. The iterated Toda family is obtained by $r-2$ iterations of the 3–fold Toda family construction. The first iteration computes the Toda family of the diagram

$$I_{r-3} \xrightarrow{f_{r-2}} I_{r-2} \xrightarrow{f_{r-1}} I_{r-1} \xrightarrow{f_r} I_r.$$

Choose a cofiber of f_{r-1}, ie a distinguished triangle $I_{r-2} \xrightarrow{f_{r-1}} I_{r-1} \to C_1 \to \Sigma I_{r-2}$. A lift of f_{r-2} to the fiber $\Sigma^{-1} C_1$, if it exists, is determined up to

$$\mathcal{T}(I_{r-3}, \Sigma^{-1} I_{r-1}) = \mathcal{T}(\Sigma I_{r-3}, I_{r-1}),$$

which is zero by Lemma 7.4(2). Likewise, an extension of f_r to the cofiber C_1, if it exists, is determined up to

$$\mathcal{T}(\Sigma I_{r-2}, I_r) = 0.$$

Hence, $\mathrm{T}(f_r, f_{r-1}, f_{r-2})$ is either empty or consists of a single pair $(\beta_1, \Sigma\alpha_1)$, up to automorphisms of C_1. It is easy to see that the object C_1 has the following property:

(7-1) If $E^* W = 0$ for $* \equiv 0, 1 \pmod{N}$, then $\mathcal{T}(W, C_1) = 0$.

For $r \geq 4$, the next iteration computes the Toda family of the diagram

$$\Sigma I_{r-4} \xrightarrow{\Sigma f_{r-3}} \Sigma I_{r-3} \xrightarrow{\Sigma\alpha_1} C_1 \xrightarrow{\beta_1} I_r.$$

The respective indeterminacies are

$$\mathcal{T}(\Sigma^2 I_{r-4}, C_1),$$

which is zero by property (7-1), and

$$\mathcal{T}(\Sigma^2 I_{r-3}, I_r),$$

which is zero by Lemma 7.4(2), since $N \geq 3$ in this case. Hence, $\mathrm{T}(\beta_1, \Sigma\alpha_1, \Sigma f_{r-3})$ is either empty or consists of a single pair $(\beta_2, \Sigma\alpha_2)$, up to automorphism of the cofiber C_2 of $\Sigma\alpha_1$. Repeating the argument inductively, the successive iterations compute the Toda family of a diagram

$$\Sigma^j I_{r-3-j} \xrightarrow{\Sigma^j f_{r-2-j}} \Sigma^j I_{r-2-j} \xrightarrow{\Sigma\alpha_j} C_j \xrightarrow{\beta_j} I_r$$

for $0 \leq j \leq r-3$, where C_j has the following property:

(7-2) If $E^* W = 0$ for $* \equiv 0, 1, \ldots, j \pmod{N}$, then $\mathcal{T}(W, C_j) = 0$.

The indeterminacies $\mathcal{T}(\Sigma^{j+1} I_{r-3-j}, C_j)$ and $\mathcal{T}(\Sigma^{j+1} I_{r-2-j}, I_r)$ again vanish. Hence, $\mathrm{T}(\beta_j, \Sigma\alpha_j, \Sigma^j f_{r-2-j})$ is either empty or consists of a single pair $(\beta_{j+1}, \Sigma\alpha_{j+1})$, up to automorphism of C_{j+1}. Note that the argument works until the last iteration $j = r-3$, by the assumption $r-2 < N$. □

We will need the following condition on an object Y.

Condition 7.6 Y admits an $\mathcal{I}_E$–Adams resolution $Y_\bullet$ (see (2-2)) such that for each injective I_j in the resolution, $E^*(\Sigma^j I_j)$ is N–sparse.

Remark 7.7 (1) Condition 7.6 implies that E^*Y is itself N–sparse, because of the surjection $E^* I_0 \twoheadrightarrow E^* Y$.

(2) The condition can be generalized to: there is an integer m such that for each j, $E^*(\Sigma^j I_j)$ is concentrated in degrees $* \equiv m \pmod N$. We take $m = 0$ for notational convenience.

(3) We will see in Propositions 7.9 and 7.10 situations in which Condition 7.6 holds.

Theorem 7.8 *Let X and Y be objects in $\mathcal{T}$ and consider the Adams spectral sequence abutting to $\mathcal{T}(X,Y)$ with respect to the injective class $\mathcal{I}_E$. Assume that Y satisfies Condition 7.6. Then for all $r \leq N$, the Adams differential is given, as a subset of $E_1^{s+r,t+r-1}$, by*

$$d_r[x] = \langle \Sigma^{r-1} d_1, \ldots, \Sigma d_1, d_1, x \rangle.$$

In other words, the restricted bracket appearing in Theorem 6.5 coincides with the full Toda bracket.

Proof We will show that

$$\langle \Sigma^{r-1} d_1 \overset{!}{,} \ldots \overset{!}{,} \Sigma d_1 \overset{!}{,} d_1, x \rangle = \langle \Sigma^{r-1} d_1, \ldots, \Sigma d_1, d_1, x \rangle.$$

Consider the diagram

$$\begin{array}{ccccccccc} I_s & \xrightarrow{d_1} & \Sigma I_{s+1} & \xrightarrow{\Sigma d_1} & \Sigma^2 I_{s+2} & \longrightarrow \cdots \longrightarrow & \Sigma^{r-1} I_{r-1} & \xrightarrow{\Sigma^{r-1} d_1} & \Sigma^r I_{s+r} \\ x \uparrow & & & & & & & & \\ X & & & & & & & & \end{array}$$

whose Toda bracket is being computed. The corresponding Toda family is

$$\mathrm{T}(\Sigma^{r-1} d_1, \ldots, \Sigma d_1, d_1, x) = \mathrm{T}\big(\mathrm{T}(\Sigma^{r-1} d_1, \ldots, \Sigma d_1, d_1), \Sigma^{r-2} x\big).$$

We know that

$$\mathrm{T}(\Sigma^{r-1} d_1 \overset{!}{,} \ldots \overset{!}{,} \Sigma d_1 \overset{!}{,} d_1) \subseteq \mathrm{T}(\Sigma^{r-1} d_1, \ldots, \Sigma d_1, d_1).$$

By Lemma 7.5, the Toda family on the right has at most one element, up to automorphism. But fully restricted Toda families are always nonempty, so the inclusion must be an equality. Write $\Sigma^{r-2} I_s \xrightarrow{f} C \xrightarrow{g} \Sigma^r I_{s+r}$ for an element of these families. It remains to show that the inclusion

$$\langle g \overset{!}{,} f, \Sigma^{r-2} x \rangle \subseteq \langle g, f, \Sigma^{r-2} x \rangle$$

is an equality, ie that the extension of g to the cofiber of f is unique. This follows from the equality $\mathcal{T}(\Sigma^{r-1} I_s, \Sigma^r I_{s+r}) = 0$, which uses the assumption on the injective objects I_j and that $r - 1 < N$. □

Next, we describe situations in which Theorem 7.8 applies.

Proposition 7.9 *Assume that every product of the form $\prod_i \Sigma^{m_i N} E$ has cohomology $E^*\left(\prod_i \Sigma^{m_i N} E\right)$ which is N–sparse. Then every object Y such that $E^* Y$ is N–sparse also satisfies* Condition 7.6.

Proof Let (y_i) be a set of nonzero generators of $E^* Y$ as an $E^* E$–module. Then the corresponding map $Y \to \prod_i \Sigma^{|y_i|} E$ is $\mathcal{I}_E$–monic into an injective object; we take this map as the first step $p_0\colon Y_0 \to I_0$, with cofiber ΣY_1. By our assumption on Y, each degree $|y_i|$ is a multiple of N, and thus $E^* I_0$ is N–sparse, by the assumption on E. The distinguished triangle $Y_1 \to Y_0 \xrightarrow{p_0} I_0 \to \Sigma Y_1$ induces a long exact sequence on E–cohomology which implies that the map $I_0 \to \Sigma Y_1$ is injective on E–cohomology. It follows that $E^*(\Sigma Y_1)$ is N–sparse as well. Repeating this process, we obtain an $\mathcal{I}_E$–Adams resolution of Y such that for every j, $E^*(\Sigma^j Y_j)$ and $E^*(\Sigma^j I_j)$ are N–sparse. □

The condition on E is discussed in Example 7.17.

Proposition 7.10 *Assume that the ring $E^* E$ is left coherent, and that $E^* Y$ is N–sparse and finitely presented as a left $E^* E$–module. Then Y satisfies Condition 7.6.*

Proof Since $E^* Y$ is finitely generated over $E^* E$, the map $p_0\colon Y \to I_0$ can be chosen so that $I_0 = \prod_i \Sigma^{m_i N} E \cong \bigoplus_i \Sigma^{m_i N} E$ is a finite product. It follows that $E^* I_0$ is N–sparse and finitely presented. We have that $E^{*-1} Y_1 = \ker(p_0^*\colon E^* I_0 \twoheadrightarrow E^* Y)$. This is N–sparse, since $E^* I_0$ is, and is finitely presented over $E^* E$, since both $E^* I_0$ and $E^* Y$ are, and $E^* E$ is coherent [9, Section I.2, Exercises 11–12]. Repeating this process, we obtain an $\mathcal{I}_E$–Adams resolution of Y such that for every j, $\Sigma^j I_j$ is a finite product of the form $\prod_i \Sigma^{m_i N} E$. □

7.2 Projective case

The main applications of Theorem 7.8 are to projective classes instead of injective classes. For future reference, we state here the dual statements of the previous subsection and adopt a notation inspired from stable homotopy theory.

Notation 7.11 Let R be an object of the triangulated category $\mathcal{T}$. Define the *homotopy* (with respect to R) of an object X as the graded abelian group $\pi_* X$ given by

$$\pi_n X := \mathcal{T}(\Sigma^n R, X).$$

Precomposition makes $\pi_* X$ a right module over the graded endomorphism ring $\pi_* R$.

Assumption 7.12 For the remainder of Section 7.2, we assume the following:

(1) The triangulated category $\mathcal{T}$ has infinite coproducts.

(2) The graded ring $\pi_* R$ is N–sparse for some $N \geq 2$.

Let $\mathcal{P}_R$ denote the stable projective class spanned by R, as in Example 2.4. Explicitly, $\mathcal{P}_R$ consists of retracts of (arbitrary) coproducts $\bigoplus_i \Sigma^{n_i} R$.

Condition 7.13 X admits a $\mathcal{P}_R$–Adams resolution $X_\bullet$ as in diagram (2-1) such that $\pi_*(\Sigma^{-j} P_j)$ is N–sparse for each projective P_j.

Theorem 7.14 *Let X and Y be objects in $\mathcal{T}$ and consider the Adams spectral sequence abutting to $\mathcal{T}(X, Y)$ with respect to the projective class $\mathcal{P}_R$. Assume that X satisfies Condition 7.13. Let $[y] \in E_r^{s,t}$ be a class represented by $y \in E_1^{s,t} = \mathcal{T}(\Sigma^{t-s} P_s, Y)$. Then for all $r \leq N$, the Adams differential is given, as a subset of $E_1^{s+r,t+r-1}$, by*

$$d_r[y] = \langle y, d_1, \Sigma^{-1} d_1, \ldots, \Sigma^{-(r-1)} d_1 \rangle.$$

Note that we used Corollary 5.13 to ensure that the equality holds as stated, not merely up to sign.

Proposition 7.15 *Assume that every coproduct of the form $\bigoplus_i \Sigma^{m_i N} R$ has homotopy $\pi_*\big(\bigoplus_i \Sigma^{m_i N} R\big)$ which is N–sparse. Then every object X such that $\pi_* X$ is N–sparse also satisfies* Condition 7.13.

Recall the following terminology.

Definition 7.16 An object X of $\mathcal{T}$ is *compact* if the functor $\mathcal{T}(X, -)$ preserves infinite coproducts.

Example 7.17 If the object R is compact in $\mathcal{T}$, then R satisfies the assumption of Proposition 7.15. This follows from the isomorphism

$$\pi_*\Big(\bigoplus_i \Sigma^{m_i N} R\Big) \cong \bigoplus_i \pi_*(\Sigma^{m_i N} R) = \bigoplus_i \Sigma^{m_i N} \pi_* R$$

and the assumption that $\pi_* R$ is N–sparse. The same argument works if R is a retract of a coproduct of compact objects.

Dually, if E is cocompact in $\mathcal{T}$, then E satisfies the assumption of Proposition 7.9. This holds more generally if E is a retract of a product of cocompact objects.

Remark 7.18 Some of the related literature deals with compactly generated triangulated categories. As noted in Remark 2.13, we do *not* assume that the object R is a generator, ie that the condition $\pi_* X = 0$ implies $X = 0$.

Proposition 7.19 *Assume that the ring $\pi_* R$ is right coherent, and that $\pi_* X$ is N–sparse and finitely presented as a right $\pi_* R$–module. Then X satisfies Condition 7.13.*

The following is a variant of [34, Lemma 2.2.2], where we do not assume that R is a generator. It identifies the E_2 term of the spectral sequence associated to the projective class $\mathcal{P}_R$. The proof is straightforward.

Proposition 7.20 *Assume that the object R is compact.*

(1) *Let P be in the projective class $\mathcal{P}_R$. Then the map of abelian groups*

$$\mathcal{T}(P, Y) \to \mathrm{Hom}_{\pi_* R}(\pi_* P, \pi_* Y)$$

is an isomorphism for every object Y.

(2) *There is an isomorphism*

$$\mathrm{Ext}^s_{\mathcal{P}_R}(X, Y) \cong \mathrm{Ext}^s_{\pi_* R}(\pi_* X, \pi_* Y)$$

which is natural in X and Y.

7.3 Examples

Theorem 7.14 applies to modules over certain ring spectra. We describe some examples, along the lines of [34, Examples 2.4.6–7].

Example 7.21 Let R be an A_∞ ring spectrum, and let $h\mathrm{Mod}_R$ denote the homotopy category of the stable model category of (right) R–modules [39, Example 2.3(ii); 17, Section III]. Then R itself, the free R–module of rank 1, is a compact generator for $h\mathrm{Mod}_R$. The R–homotopy of an R–module spectrum X is the usual homotopy of X, as suggested by the notation

$$h\mathrm{Mod}_R(\Sigma^n R, X) \cong h\mathrm{Mod}_S(S^n, X) = \pi_n X.$$

In particular, the graded endomorphism ring $\pi_* R$ is the usual coefficient ring of R.

The projective class $\mathcal{P}_R$ is the ghost projective class [14, Section 7.3], generalizing Example 2.4, where R was the sphere spectrum S. The Adams spectral sequence relative to $\mathcal{P}_R$ is the universal coefficient spectral sequence

$$\mathrm{Ext}^s_{\pi_* R}(\Sigma^t \pi_* X, \pi_* Y) \implies h\mathrm{Mod}_R(\Sigma^{t-s} X, Y)$$

as described in [17, Section IV.4] and [14, Corollary 7.12]. We used Proposition 7.20 to identify the E_2 term.

Some A_∞ ring spectra R with sparse homotopy $\pi_* R$ are discussed in [34, Sections 4.3, 5.3 and 6.4]. In view of Proposition 7.20, the Adams spectral sequence in $h\mathrm{Mod}_R$ collapses at the E_2 page if $\pi_* R$ has (right) global dimension less than 2.

The Johnson–Wilson spectrum $E(n)$ has coefficient ring

$$\pi_* E(n) = \mathbb{Z}_{(p)}[v_1, \dots, v_n, v_n^{-1}], \quad |v_i| = 2(p^i - 1),$$

which has global dimension n and is $2(p-1)$–sparse. Hence, Theorem 7.14 applies in this case to the differentials d_r with $r \le 2(p-1)$, while d_r is zero for $r > n$. Likewise, connective complex K–theory ku has coefficient ring

$$\pi_* ku = \mathbb{Z}[u], \quad |u| = 2,$$

which has global dimension 2 and is 2–sparse.

Example 7.22 Let R be a differential graded (*dg* for short) algebra over a commutative ring k, and consider the category of dg R–modules dgMod_R. The homology $H_* X$ of a dg R–module is a (graded) $H_* R$–module. The derived category $D(R)$ is defined as the localization of dgMod_R with respect to quasi-isomorphisms. The free dg R–module R is a compact generator of $D(R)$. The R–homotopy of an object X of $D(R)$ is its homology $\pi_* X = H_* X$. In particular, the graded endomorphism ring of R in $D(R)$ is the graded k–algebra $H_* R$.

The Adams spectral sequence relative to $\mathcal{P}_R$ is an Eilenberg–Moore spectral sequence

$$\mathrm{Ext}^s_{H_* R}(\Sigma^t H_* X, H_* Y) \implies D(R)(\Sigma^{t-s} X, Y)$$

from ordinary Ext to differential Ext, as described in [4, Section 8, 10]. See also [25, Section III.4; 21, Example 10.2(b); 16].

Remark 7.23 Example 7.22 can be viewed as a special case of Example 7.21. Letting HR denote the Eilenberg–MacLane spectrum associated to R, the categories Mod_{HR} and dgMod_R are Quillen equivalent, by [39, Example 2.4(i)] and [41, Corollary 2.15], yielding a triangulated equivalence $h\mathrm{Mod}_{HR} \cong D(R)$. The generator HR corresponds to the generator R via this equivalence.

Example 7.24 Let R be a ring, viewed as a dg algebra concentrated in degree 0. Then Example 7.22 yields the ordinary derived category $D(R)$. The graded endomorphism ring of R in $D(R)$ is H_*R, which is R concentrated in degree 0. This is N–sparse for any $N \geq 2$.

The Adams spectral sequence relative to $\mathcal{P}_R$ is the hyperderived functor spectral sequence

$$\mathrm{Ext}^s_{H_*R}(\Sigma^t H_*X, H_*Y) = \prod_{i\in\mathbb{Z}} \mathrm{Ext}^s_R(H_{i-t}X, H_iY) \Longrightarrow D(R)(\Sigma^{t-s}X, Y) = \mathbf{Ext}^{s-t}_R(X, Y)$$

from ordinary Ext to hyper Ext, as described in [44, Section 5.7 and 10.7].

Appendix A: Computations in the stable module category of a group

In this appendix, we give some computations in the stable module category of a group algebra kG, where k is a field and G is a finite group. These computations are used in Proposition 4.3.

Write R for the group algebra kG. We will work in the stable module category $\mathcal{T} := \mathrm{StMod}(R)$. This is the category whose objects are (left) R–modules, and whose morphisms from M to N consist of the R–module homomorphisms from M to N modulo those that factor through a projective module. An isomorphism in $\mathrm{StMod}(R)$ is called a *stable equivalence*, and two R–modules M and N are stably equivalent if and only if there are projectives P and Q such that $M \oplus P \cong N \oplus Q$. The category $\mathrm{StMod}(R)$ is triangulated. The suspension ΣM is defined by choosing an embedding of M into an injective module and taking the quotient, the desuspension ΩM is defined by choosing a surjection from a projective to M and taking the kernel, and these are inverse to each other because the projectives and injectives coincide. Given a short exact sequence

$$0 \to M_1 \to M_2 \to M_3 \to 0$$

and an embedding of M_1 into an injective module I, one can choose maps

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & 0 \\ & & \| & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M_1 & \longrightarrow & I & \longrightarrow & \Sigma M_1 & \longrightarrow & 0 \end{array} \tag{A-1}$$

making the diagram commute in Mod_R. The distinguished triangles are defined to be those triangles isomorphic in $\mathrm{StMod}(R)$ to one of the form

$$M_1 \to M_2 \to M_3 \to \Sigma M_1$$

constructed in this way.

Rather than working with respect to an injective class in $\mathcal{T}$, we will consider the *ghost projective class* $\mathcal{P}$, which is generated by the trivial module k. More precisely, $\mathcal{P}$ consists of the retracts of coproducts $\bigoplus_i \Sigma^{n_i} k$, and the associated ideal consists of the maps which induce the zero map in Tate cohomology. See [12, Section 4.2] for details.

Proposition A.1 *Let G be the cyclic group $C_4 = \langle g \mid g^4 = 1\rangle$, let $k = \mathbb{F}_2$, and write $R = kG$. There exists an R–module M, an Adams resolutions of M with respect to the ghost projective class, and a map $\kappa\colon M \to M$ such that the inclusion $\langle \kappa, d_1, \delta\rangle(\Sigma p) \subseteq \langle \kappa, d_1, d_1\rangle$ from Proposition 4.2 (dualized) is proper.*

Proof To produce our counterexample, we will consider the Adams spectral sequence abutting to $\mathrm{StMod}(M, \Omega^* M)$, where M is a two-dimensional module with basis vectors that are interchanged by g.

In order to make concrete computations, it will be helpful to observe that, as a k–algebra, R is the truncated polynomial algebra

$$R = k[g]/(g^4 - 1) = k[g]/(g-1)^4 = k[x]/x^4,$$

where we define $x := g - 1 \in R$. In this notation, the trivial module k is R/x and the module M is R/x^2.

We will need to compute their desuspensions, which are given, as R–modules, by

$$\begin{aligned} \Omega k &= \ker(R \twoheadrightarrow k) = k\{x, x^2, x^3\} \cong R/x^3, \\ \Omega^2 k &= \ker(R \twoheadrightarrow R/x^3) = k\{x^3\} \cong R/x = k, \\ \Omega M &= \ker(R \twoheadrightarrow R/x^2) = k\{x^2, x^3\} \cong R/x^2 = M, \end{aligned}$$

where curly brackets denote the k–vector space with the given generating set.

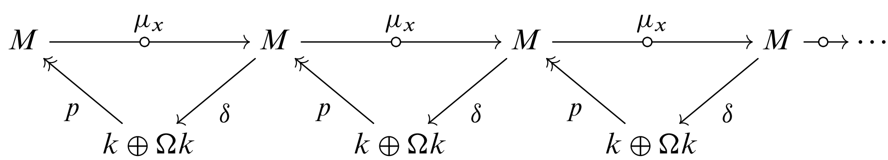
In order to produce a $\mathcal{P}$–epic map to M, we need to know the maps from suspensions of k to M. Since k is 2–periodic, the following calculations give us what we need:

$$\mathcal{T}(k,M) = \mathrm{Mod}_R(k,M)/\!\sim \,\cong \mathrm{Mod}_R(R/x, R/x^2)/\!\sim\, = k\{\mu_x\}/\!\sim\, = k\{\mu_x\},$$
$$\mathcal{T}(\Omega k,M) = \mathrm{Mod}_R(\Omega k,M)/\!\sim\, = \mathrm{Mod}_R(R/x^3, R/x^2)/\!\sim\, = k\{\mu_1,\mu_x\}/\!\sim\, = k\{\mu_1\},$$

where $f \sim g$ if $f-g$ factors through a projective, and $\mu_r\colon R/x^m \to R/x^n$ denotes the R–module map given by multiplication by $r \in R$ (when this is well-defined). Here, we used the fact that $\mu_x\colon R/x^3 \to R/x^2$ is stably null, since it factors as

$$R/x^3 \xrightarrow{\mu_x} R \xrightarrow{\mu_1} R/x^2.$$

Using this, we obtain a $\mathcal{P}$–epic map $p := \mu_x + \mu_1\colon k \oplus \Omega k \to M$. Since p is surjective, its fiber is its kernel. This kernel is generated by $(1,x)$ and is readily seen to be isomorphic to M. Under the identification of ΩM with M, the natural map $\Omega M \to M$ (using the dual of Equation (A-1)) is μ_x. Since we are working at the prime 2, fiber sequences and cofiber sequences agree, so we obtain the following Adams resolution of M:



where $\delta = \left[\begin{smallmatrix}\mu_1\\ \mu_x\end{smallmatrix}\right]$, and we have chosen to put the degree shifts on the horizontal arrows.

We will be considering the Adams spectral sequence formed by applying the functor $\mathcal{T}(-,M)$. The map $d_1 = \delta p\colon k \oplus \Omega k \to k \oplus \Omega k$ is $\left[\begin{smallmatrix}0 & \mu_1\\ \mu_{x^2} & \mu_x\end{smallmatrix}\right]$, which simplifies to $\left[\begin{smallmatrix}0 & \mu_1\\ \mu_{x^2} & 0\end{smallmatrix}\right]$, using the fact that $\mu_x\colon \Omega k \to \Omega k$ is stably null, because we have that it factors as $\Omega k \xrightarrow{\mu_x} R \xrightarrow{\mu_1} \Omega k$. The stable maps $k \oplus \Omega k \to M$ are of the form $[a\mu_x \;\; b\mu_1]$ for a and b in k, and all composites $[a\mu_x \;\; b\mu_1]$ are stably null. Using this twice for d_1's in different positions, one sees that if $\kappa\colon k \oplus \Omega k \to M$ is any map, then $d_2[\kappa]$ is defined and has no indeterminacy.

Now consider $\langle \kappa, d_1, \delta\rangle(\Sigma p)$. One part of the indeterminacy here consists of maps of the form $f\Sigma(\delta)\Sigma(p) = f\Sigma(d_1)$, for $f\colon \Sigma(k \oplus \Omega k) \to M$. As above, all such composites are zero. The other part of the indeterminacy consists of maps of the form $\kappa f \Sigma(p)$, for $f\colon \Sigma M \to k \oplus \Omega k$, and again, one can show that all such composites are zero. So $\langle \kappa, d_1, \delta\rangle(\Sigma p)$ has no indeterminacy and therefore equals $d_2[\kappa]$.

Finally, consider $\langle \kappa, d_1, d_1 \rangle$. The part of the indeterminacy involving d_1 is again zero. The other part consists of all composites κf, for $f\colon \Sigma(k \oplus \Omega k) \to k \oplus \Omega k$. Since there is an isomorphism $\Sigma(k \oplus \Omega k) \to k \oplus \Omega k$, this indeterminacy is nonzero if and only if κ is nonzero.

Since nonzero maps $\kappa\colon k \oplus \Omega k \to M$ exist, we conclude that the containment

$$\langle \kappa, d_1, \delta \rangle (\Sigma p) \subseteq \langle \kappa, d_1, d_1 \rangle$$

can be proper. □

Remark A.2 If in the proof above we take κ to be the map $[\mu_x \;\; 0]\colon k \oplus \Omega k \to M$, then using the same techniques one can show that

$$\begin{aligned} \langle \kappa, d_1, \delta \rangle &= \{1_M\}, \\ \langle \kappa, d_1, \delta \rangle (\Sigma p) = \{\Sigma p\} &= d_2[\kappa] = \{[\mu_1 \;\; \mu_x]\}, \\ \langle \kappa, d_1, d_1 \rangle &= \{[\mu_1 \;\; b\mu_x] \mid b \in \mathbb{F}_2\}, \end{aligned}$$

as subsets of $\mathcal{T}(\Omega k \oplus k, M) \cong \mathcal{T}(\Sigma(k \oplus \Omega k), M)$, where we identify M with ΩM and ΣM, as before.

Remark A.3 Theorem 7.14 does not apply to the example in Proposition A.1. Indeed, the graded endomorphism ring of k in $\mathrm{StMod}(kG)$ is the Tate cohomology ring $\widetilde{H}^n(G;k) = \mathrm{StMod}(kG)(\Omega^n k, k)$ [11, Section 6]. This ring is not sparse, as we have $\widetilde{H}^{-1}(G;k) \neq 0$.

Example A.4 The following example illustrates the fact that a Toda bracket need not be equal to its own negative, as noted in Remark 3.2.

Consider the ground field $k = \mathbb{F}_3$ and the group algebra $R = kC_3 \cong k[x]/x^3$, where we denote $x = g - 1 \in R$ for $g \in C_3$ a generator. Consider the R–modules $k = R/x$ and $M = R/x^2$. Let us compute the Toda bracket of the diagram

$$M \xrightarrow{\mu_1} k \xrightarrow{\mu_x} M \xrightarrow{\mu_1} k$$

in the triangulated category $\mathcal{T} = \mathrm{StMod}(R)$. We will use appropriate isomorphisms $\Sigma k \cong M$ and $\Sigma M \cong k$, and in particular compute the Toda bracket as a subset of $\mathcal{T}(k,k) \cong \mathcal{T}(\Sigma M, k)$. Via these isomorphisms, the suspension $\Sigma\mu_1\colon \Sigma M \to \Sigma k$

equals $\mu_x\colon k \to M$. Consider the commutative diagram in $\mathcal{T}$

$$\begin{array}{ccccccc}
 & & & & k & \xrightarrow{-\mu_x} & M \\
 & & & & {\scriptstyle \Sigma\alpha}\downarrow & & \| \\
k & \xrightarrow{\mu_x} & M & \xrightarrow{\mu_1} & k & \xrightarrow{\mu_x} & M \\
 & & \| & & \downarrow{\scriptstyle \beta} & & \\
 & & M & \xrightarrow{\mu_1} & k & &
\end{array}$$

where the middle row is distinguished. The only choices for the dotted arrows are $\Sigma\alpha = -1_k$ and $\beta = 1_k$, from which we conclude

$$\langle \mu_1, \mu_x, \mu_1 \rangle_{\mathrm{fc}} = \{-1_k\} \subset \mathcal{T}(k,k).$$

Appendix B: 3–fold Toda brackets determine the triangulated structure

Heller proved the following theorem in [20, Theorem 13.2]. We present an arguably simpler proof here. The argument was kindly provided by Fernando Muro.

Theorem B.1 *In a triangulated category $\mathcal{T}$, the diagram $X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} \Sigma X$ is a distinguished triangle if and only if the following two conditions hold:*

(1) *The sequence of abelian groups*

$$\mathcal{T}(A, \Sigma^{-1}Z) \xrightarrow{(\Sigma^{-1}h)_*} \mathcal{T}(A, X) \xrightarrow{f_*} \mathcal{T}(A, Y) \xrightarrow{g_*} \mathcal{T}(A, Z) \xrightarrow{h_*} \mathcal{T}(A, \Sigma X)$$

is exact for every object A of $\mathcal{T}$.

(2) *The Toda bracket $\langle h, g, f \rangle \subseteq \mathcal{T}(\Sigma X, \Sigma X)$ contains the identity map $1_{\Sigma X}$.*

Proof ($\Rightarrow$) A distinguished triangle satisfies the first condition. For the second condition, consider the following commutative diagram:

$$\begin{array}{ccccccc}
X & \xrightarrow{f} & Y & \xrightarrow{g} & Z & \xrightarrow{h} & \Sigma X \\
\| & & \| & & \downarrow{\scriptstyle 1_Z} & & \downarrow{\scriptstyle 1_{\Sigma X}} \\
X & \xrightarrow{f} & Y & \xrightarrow{g} & Z & \xrightarrow{h} & \Sigma X
\end{array}$$

Since the top row is distinguished, this diagram exhibits the membership $1_{\Sigma X} \in \langle h, g, f \rangle$.

($\Leftarrow$) Assume that $1_{\Sigma X} \in \langle h, g, f \rangle$ holds. By definition of the Toda bracket, there exists a map $\varphi\colon C_f \to Z$ making the diagram

$$\begin{array}{ccccccc} X & \xrightarrow{f} & Y & \xrightarrow{q} & C_f & \xrightarrow{\iota} & \Sigma X \\ \| & & \| & & \downarrow{\scriptstyle\varphi} & & \downarrow{\scriptstyle 1_{\Sigma X}} \\ X & \xrightarrow{f} & Y & \xrightarrow{g} & Z & \xrightarrow{h} & \Sigma X \end{array}$$

commute, where the top row is distinguished. To show that the bottom row is distinguished, it suffices to show that $\varphi\colon C_f \to Z$ is an isomorphism. By the Yoneda lemma, it suffices to show that $\varphi_*\colon \mathcal{T}(A, C_f) \to \mathcal{T}(A, Z)$ is an isomorphism for every object A of $\mathcal{T}$.

Consider the following diagram:

$$\begin{array}{ccccccccc} X & \xrightarrow{f} & Y & \xrightarrow{q} & C_f & \xrightarrow{\iota} & \Sigma X & \xrightarrow{-\Sigma f} & \Sigma Y \\ \| & & \| & & \downarrow{\scriptstyle\varphi} & & \downarrow{\scriptstyle 1_{\Sigma X}} & & \downarrow{\scriptstyle 1_{\Sigma Y}} \\ X & \xrightarrow{f} & Y & \xrightarrow{g} & Z & \xrightarrow{h} & \Sigma X & \xrightarrow{-\Sigma f} & \Sigma Y \end{array} \tag{B-1}$$

Applying $\mathcal{T}(A, -)$ yields the following diagram of abelian groups:

$$\begin{array}{ccccccccc} \mathcal{T}(A,X) & \xrightarrow{f_*} & \mathcal{T}(A,Y) & \xrightarrow{q_*} & \mathcal{T}(A,C_f) & \xrightarrow{\iota_*} & \mathcal{T}(A,\Sigma X) & \xrightarrow{(-\Sigma f)_*} & \mathcal{T}(A,\Sigma Y) \\ \| & & \| & & \downarrow{\scriptstyle\varphi_*} & & \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle 1} \\ \mathcal{T}(A,X) & \xrightarrow{f_*} & \mathcal{T}(A,Y) & \xrightarrow{g_*} & \mathcal{T}(A,Z) & \xrightarrow{h_*} & \mathcal{T}(A,\Sigma X) & \xrightarrow{(-\Sigma f)_*} & \mathcal{T}(A,\Sigma Y) \end{array}$$

The top row is exact, since the top row of (B-1) is a cofiber sequence, and the bottom row is exact, using the first condition. By the five lemma, φ_* is an isomorphism. $\square$

Remark B.2 Here are some remarks about the first condition.

(1) It implies $gf = g_* f_*(1_X) = 0$ and $hg = h_* g_*(1_Y) = 0$.

(2) It is equivalent to the exactness of the long sequence (infinite in both directions)

$$\cdots \to \mathcal{T}(A, \Sigma^n X) \xrightarrow{(\Sigma^n f)_*} \mathcal{T}(A, \Sigma^n Y) \xrightarrow{(\Sigma^n g)_*} \mathcal{T}(A, \Sigma^n Z) \xrightarrow{(\Sigma^n h)_*} \mathcal{T}(A, \Sigma^{n+1} X) \to \cdots$$

for every object A of $\mathcal{T}$.

(3) It is a weaker version of what is sometimes called a *pretriangle* [33, Section 1.1]. Indeed, the condition states that the sequence

$$H(\Sigma^{-1} Z) \xrightarrow{H(\Sigma^{-1} h)} H(X) \xrightarrow{H(f)} H(Y) \xrightarrow{H(g)} H(Z) \xrightarrow{H(h)} H(\Sigma X)$$

is exact for every decent homological functor $H\colon \mathcal{T} \to \mathrm{Ab}$ of the form $H = \mathcal{T}(A, -)$.

Corollary B.3 *Given the suspension functor $\Sigma\colon \mathcal{T} \to \mathcal{T}$, 3–fold Toda brackets in $\mathcal{T}$ determine the triangulated structure. In particular, 3–fold Toda brackets determine the higher Toda brackets, via the triangulation.*

Remark B.4 It is unclear to us if the higher Toda brackets can be expressed directly in terms of 3–fold brackets.

References

[1] **J F Adams**, *On the non-existence of elements of Hopf invariant one*, Ann. of Math. 72 (1960) 20–104 MR

[2] **J F Adams**, *Stable homotopy and generalised homology*, Univ. of Chicago Press (1974) MR

[3] **P Balmer**, *Triangulated categories with several triangulations* (2002) Available at `http://www.math.ucla.edu/~balmer/Pubfile/TriangulationS.pdf`

[4] **T Barthel**, **J P May**, **E Riehl**, *Six model structures for DG-modules over DGAs: model category theory in homological action*, New York J. Math. 20 (2014) 1077–1159 MR

[5] **H-J Baues**, **D Blanc**, *Higher order derived functors and the Adams spectral sequence*, J. Pure Appl. Algebra 219 (2015) 199–239 MR

[6] **H-J Baues**, **M Jibladze**, *Secondary derived functors and the Adams spectral sequence*, Topology 45 (2006) 295–324 MR

[7] **H-J Baues**, **M Jibladze**, *Dualization of the Hopf algebra of secondary cohomology operations and the Adams spectral sequence*, J. K-Theory 7 (2011) 203–347 MR

[8] **D Benson**, **H Krause**, **S Schwede**, *Realizability of modules over Tate cohomology*, Trans. Amer. Math. Soc. 356 (2004) 3621–3668 MR

[9] **N Bourbaki**, *Commutative algebra, Chapters 1–7*, Hermann, Paris (1972) MR

[10] **R R Bruner**, *An Adams spectral sequence primer* (2009) Available at `http://www.math.wayne.edu/~rrb/papers/adams.pdf`

[11] **J F Carlson**, *Modules and group algebras*, Birkhäuser, Basel (1996) MR

[12] **S K Chebolu**, **J D Christensen**, **J Mináč**, *Ghosts in modular representation theory*, Adv. Math. 217 (2008) 2782–2799 MR

[13] **J D Christensen** Python code for comparing sequences Available at `https://trinket.io/python/8932cfbf2f`

[14] **J D Christensen**, *Ideals in triangulated categories: phantoms, ghosts and skeleta*, Adv. Math. 136 (1998) 284–339 MR

[15] **J M Cohen**, *The decomposition of stable homotopy*, Ann. of Math. 87 (1968) 305–320 MR

[16] **S Eilenberg**, **J C Moore**, *Homology and fibrations, I: Coalgebras, cotensor product and its derived functors*, Comment. Math. Helv. 40 (1966) 199–236 MR

[17] **A D Elmendorf**, **I Kriz**, **M A Mandell**, **J P May**, *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs 47, Amer. Math. Soc., Providence, RI (1997) MR

[18] **S I Gelfand**, **Y I Manin**, *Methods of homological algebra*, 2nd edition, Springer (2003) MR

[19] **J R Harper**, *Secondary cohomology operations*, Graduate Studies in Mathematics 49, Amer. Math. Soc., Providence, RI (2002) MR

[20] **A Heller**, *Stable homotopy categories*, Bull. Amer. Math. Soc. 74 (1968) 28–63 MR

[21] **M Hovey**, **J H Palmieri**, **N P Strickland**, *Axiomatic stable homotopy theory*, Mem. Amer. Math. Soc. 610, Amer. Math. Soc., Providence, RI (1997) MR

[22] **M Hovey**, **N P Strickland**, *Morava K–theories and localisation*, Mem. Amer. Math. Soc. 666, Amer. Math. Soc., Providence, RI (1999) MR

[23] **S Klaus**, *Towers and pyramids, I*, Forum Math. 13 (2001) 663–683 MR

[24] **S O Kochman**, *Uniqueness of Massey products on the stable homotopy of spheres*, Canad. J. Math. 32 (1980) 576–589 MR

[25] **I Kříž**, **J P May**, *Operads, algebras, modules and motives*, Astérisque 233, Soc. Math. France, Paris (1995) MR

[26] **H R Margolis**, *Spectra and the Steenrod algebra*, North-Holland Mathematical Library 29, North-Holland, Amsterdam (1983) MR

[27] **A Mathew**, *Second homotopy group of the mod 2 Moore spectrum* Available at `http://mathoverflow.net/questions/100272/`

[28] **C R F Mauder**, *On the differentials in the Adams spectral sequence*, Proc. Cambridge Philos. Soc. 60 (1964) 409–420 MR

[29] **C R F Maunder**, *Cohomology operations of the N^{th} kind*, Proc. London Math. Soc. 13 (1963) 125–154 MR

[30] **J C McKeown**, *Toda bracket*, nLab (2012) version 1 Available at `https://ncatlab.org/nlab/show/Toda+bracket`

[31] **L Meier**, *United elliptic homology*, PhD thesis, Universität Bonn (2012) Available at `http://www.math.uni-bonn.de/~grk1150/DISS/dissertation-meier.pdf`

[32] **R E Mosher**, **M C Tangora**, *Cohomology operations and applications in homotopy theory*, Harper & Row, New York (1968) MR

[33] **A Neeman**, *Triangulated categories*, Annals of Mathematics Studies 148, Princeton Univ. Press (2001) MR

[34] **I Patchkoria**, *On the algebraic classification of module spectra*, Algebr. Geom. Topol. 12 (2012) 2329–2388 MR

[35] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Mathematics 121, Academic Press, Orlando, FL (1986) MR

[36] **S Sagave**, *Universal Toda brackets of ring spectra*, Trans. Amer. Math. Soc. 360 (2008) 2767–2808 MR

[37] **S Schwede**, *Algebraic versus topological triangulated categories*, from "Triangulated categories" (T Holm, P Jørgensen, R Rouquier, editors), London Math. Soc. Lecture Note Ser. 375, Cambridge Univ. Press (2010) 389–407 MR

[38] **S Schwede**, *Symmetric spectra*, preprint (2012) version 3.0 Available at `http://www.math.uni-bonn.de/people/schwede/SymSpec-v3.pdf`

[39] **S Schwede**, **B Shipley**, *Stable model categories are categories of modules*, Topology 42 (2003) 103–153 MR

[40] **B Shipley**, *An algebraic model for rational S^1–equivariant stable homotopy theory*, Q. J. Math. 53 (2002) 87–110 MR

[41] **B Shipley**, *$H\mathbb{Z}$–algebra spectra are differential graded algebras*, Amer. J. Math. 129 (2007) 351–379 MR

[42] **E Spanier**, *Higher order operations*, Trans. Amer. Math. Soc. 109 (1963) 509–539 MR

[43] **H Toda**, *Composition methods in homotopy groups of spheres*, Annals of Mathematics Studies 49, Princeton Univ. Press (1962) MR

[44] **C A Weibel**, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics 38, Cambridge Univ. Press (1994) MR

Department of Mathematics, University of Western Ontario
London, ON, Canada

Institut für Mathematik, Universität Osnabrück
Osnabrück, Germany

`jdc@uwo.ca`, `martin.frankland@uni-osnabrueck.de`

Received: 31 October 2015 Revised: 29 June 2016

Algebraic & Geometric Topology 17 (2017) 2737–2751

A generalized axis theorem for cube complexes

DANIEL J WOODHOUSE

We consider a finitely generated virtually abelian group G acting properly and without inversions on a CAT(0) cube complex X. We prove that G stabilizes a finite-dimensional CAT(0) subcomplex $Y \subseteq X$ that is isometrically embedded in the combinatorial metric. Moreover, we show that Y is a product of finitely many quasilines. The result represents a higher-dimensional generalization of Haglund's axis theorem.

20F65

1 Introduction

A CAT(0) *cube complex* X is a cell complex that satisfies two properties: it is a geodesic metric space satisfying the CAT(0) comparison triangle condition, and each n–cell is isometric to $[0,1]^n$. We will call this metric the CAT(0) *metric* d_X and refer to Bridson and Haefliger [2] for a comprehensive account. A *hyperplane* $\Lambda \subseteq X$ is the subset of points equidistant between two adjacent vertices. Despite the brevity of this definition, hyperplanes are better understood via their combinatorial definition, and the reader is urged to consult the literature; see Sageev [10], Haglund [6] and Wise [12] for the required background. There also exists an alternative metric on the 0–cubes of X, which we will refer to as the *combinatorial metric* d_X^c, sometimes referred to as the ℓ^1*–metric*. The combinatorial distance between two 0–cubes is the length of the shortest combinatorial path in X joining the 0–cubes. Equivalently, the combinatorial distance between two 0–cubes is the number of hyperplanes in X separating them. We will always assume that a group G acting on a CAT(0) cube complex preserves its cell structure and maps cubes isometrically to cubes. A group G acts without *inversions* if the stabilizer of a hyperplane also stabilizes each complementary component. The requirement that the action be without inversions is not a serious restriction as G acts without inversions on the cubical subdivision.

A connected CAT(0) cube complex X is a *quasiline* if it is quasiisometric to $\mathbb{R}$. The *rank* of a virtually abelian group commensurable to $\mathbb{Z}^n$ is n. The goal of this paper will be the following theorem:

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2737

Theorem 4.3 *Let G be virtually $\mathbb{Z}^n$. Suppose G acts properly and without inversions on a* CAT(0) *cube complex X. Then G stabilizes a finite-dimensional subcomplex $Y \subseteq X$ that is isometrically embedded in the combinatorial metric, and $Y \cong \prod_{i=1}^m C_i$, where each C_i is a cubical quasiline and $m \geq n$. Moreover,* $\mathrm{Stab}_G(\Lambda)$ *is a codimension-1 subgroup for each hyperplane Λ in Y.*

Note that Y will not in general be a convex subcomplex.

Corollary 1.1 *Let A be a finitely generated virtually abelian group acting properly on a* CAT(0) *cube complex X. Then A acts metrically properly on X.*

Corollary 1.2 *Let G be a finitely generated group acting properly on a* CAT(0) *cube complex X. Then virtually $\mathbb{Z}^n$ subgroups are undistorted in G.*

Let g be an isometry of X, and let $x \in X$. The *displacement of g at x*, denoted by $\tau_x(g)$, is the distance $\mathsf{d}_X(x, gx)$. The *translation length* of g, denoted by $\tau(g)$, is $\inf\{\tau_x(g) \mid x \in X\}$. Similarly, if x is a 0–cube of X, we can define the *combinatorial displacement of g at x*, denoted by $\tau_x^c(g)$, as $\mathsf{d}_X^c(x, gx)$ and the *combinatorial translation length*, denoted by $\tau^c(g)$, is $\inf\{\tau_x^c(g) \mid x \in X\}$. Note that τ and τ^c are conjugacy invariant. An isometry g of a CAT(0) space is *semisimple* if $\tau_x(g) = \tau(g)$ for some $x \in X$, and G acts *semisimply* on a CAT(0) space X if each $g \in G$ is semisimple.

If a virtually $\mathbb{Z}^n$ group G acts metrically properly by semisimple isometries on a CAT(0) space X, then the flat torus theorem of Bridson and Haefliger [2] provides a G–invariant, convex, flat $\mathbb{E}^n \subseteq X$. A group acting on a CAT(0) cube complex does not, in general, have to do so semisimply. See Algom-Kfir, Wajnryb and Witowicz [1] for examples of nonsemisimple isometries in Thompson's group F acting on an infinite-dimensional CAT(0) cube complex. Alternatively, in Gersten [5], a free-by-cyclic group G is shown not to permit a semisimple action on a CAT(0) space. Yet in Wise [13] it is shown that G does act freely on a CAT(0) cube complex. Thus Theorem 4.3 can be applied to such actions, whereas the classical flat torus theorem cannot.

A virtually abelian subgroup is *highest* if it is not virtually contained in a higher rank abelian subgroup. If G is a highest virtually abelian subgroup of a group acting properly and cocompactly on a CAT(0) cube complex X, then G cocompactly stabilizes a convex subcomplex Y which is a product of quasilines, as above; see Wise and Woodhouse [14]. However, this theorem fails without the highest hypothesis. Moreover, most actions do not arise in the above fashion.

Despite the fact that the flat torus theorem will not hold under the hypotheses of Theorem 4.3, we can deduce the following:

Corollary 4.4 *Let G be virtually $\mathbb{Z}^n$. Suppose G acts properly and without inversions on a* CAT(0) *cube complex X. Then G cocompactly stabilizes a subspace $F \subseteq X$ homeomorphic to $\mathbb{R}^n$ such that for each hyperplane $\Lambda \subseteq X$, the intersection $\Lambda \cap F$ is either empty or homeomorphic to $\mathbb{R}^{n-1}$.*

The initial motivation for Theorem 4.3 and Corollary 4.4 was to resolve the following question, posed by Wise. Although we have not found a combinatorial flat, Corollary 4.4 is perhaps better suited to applications (see Woodhouse [15]).

Problem 1.3 Let $\mathbb{Z}^2$ act freely on a CAT(0) cube complex Y. Does there exists a $\mathbb{Z}^2$–equivariant map $F \to Y$, where F is a square 2–complex homeomorphic to $\mathbb{R}^2$, and such that no two hyperplanes of F map to the same hyperplane in Y?

A *combinatorial geodesic axis for g* is a g–invariant, isometrically embedded in the combinatorial metric, subcomplex $\gamma \subseteq X$ with $\gamma \cong \mathbb{R}$. Note that γ realizes the minimal combinatorial translation length of g. Theorem 4.3 is a high-dimensional generalization of Haglund's combinatorial geodesic axis theorem. Haglund's proof involved an argument by contradiction, exploiting the geometry of hyperplanes. We reprove the result in Section 5 by using the dual cube complex construction of Sageev. The results are further support for Haglund's slogan "in CAT(0) cube complexes the combinatorial geometry is as nice as the CAT(0) geometry".

The following is an application of Theorem 4.3, and the argument is inspired by the solvable subgroup theorem of Bridson and Haefliger [2]. Note that since we do not require that the action of G on a CAT(0) cube complex be semisimple the following is not covered by the solvable subgroup theorem.

Corollary 1.4 *Let H be virtually $\mathbb{Z}^n$, and let $\phi\colon H \to H$ be an injection with $\phi \neq \phi^i$ for all $i > 1$. Then $G = \langle H, t \mid t^{-1}ht = \phi(h),\ h \in H\rangle$ cannot act properly on a* CAT(0) *cube complex.*

Proof Suppose that G acts properly on a CAT(0) cube complex X. After subdividing X we can assume that G acts without inversions. As H is finitely generated, there exists an a in the finite generating set such that $\phi^i(a) \neq a$ for all $i \in \mathbb{N}$, otherwise $\phi^i = \phi$ for some i, contradicting our hypothesis. Thus, $|\{\phi^i(a)\}| = \infty$. By Theorem 4.3 there is an H–equivariant isometrically embedded subcomplex $Y \subseteq X$ such that $Y \cong \prod_{i=1}^m C_i$, where each C_i is a cubical quasiline.

As Y is isometrically embedded in X in the combinatorial metric, the combinatorial translation length $\tau^c(\phi^i(a))$ is the same in Y as it is in X. The set $\{\tau^c(\phi^i(a))\}_{i \in \mathbb{N}}$

must be unbounded since the action of H on Y is proper and Y is locally finite. However, since τ^c is conjugacy invariant in G, we conclude that $\tau^c(\phi^i(a)) = \tau^c(\phi^j(a))$ for all $i, j \in \mathbb{N}$. Thus, we arrive at the contradiction that $\{\tau^c(\phi^i(a))\}_{i\in\mathbb{N}}$ is both bounded and unbounded. □

However, we have the following example of a solvable group which does act freely on a CAT(0) cube complex.

Example 1.5 Let $H = \langle a_1, a_2, \dots \mid [a_i, a_j] \text{ for } i \neq j \rangle$. Note that H is the fundamental group of the nonpositively curved cube complex Y obtained from a 0–cube v, and 1–cubes $e_1, e_2, e_3 \dots$ with n–cubes inserted for every cardinality-n collection of 1–cubes to create an n–torus. One should think of Y as an infinite cubical torus. The oriented loop e_i represents the element a_i.

Let $\phi\colon H \to H$ be the monomorphism such that $\phi(a_i) = a_{i+1}$. Let $G = H*_\phi = \langle t, a_1, a_2, \dots \mid [a_i, a_j] \text{ for } i \neq j,\ t^{-1}a_i t = a_{i+1} \rangle$ be the associated ascending HNN extension. Note that G is generated by a_1 and t. There is a graph of spaces X obtained by letting Y be the vertex space and $Y \times [0,1]$ be the edge space and identifying $(v,1)$ and $(v,0)$ with v, and the 1–cube $e_i \times \{1\}$ with e_i and $e_i \times \{0\}$ with e_{i+1}. Note that X is nonpositively curved, and therefore $G = \pi_1 X$ acts freely on the CAT(0) cube complex $\widetilde{X}$, the universal cover of X.

Acknowledgements I would like to thank Daniel T Wise, Mark F Hagen, Jack Button, Piotr Przytycki and Dan Guralnik.

2 Dual cube complexes

Let S be a set. A *wall* $\Lambda = \{\overleftarrow{\Lambda}, \overrightarrow{\Lambda}\}$ in S is a partition of S into two disjoint, nonempty subsets. The subsets $\overleftarrow{\Lambda}$ and $\overrightarrow{\Lambda}$ are the *halfspaces* of Λ. A wall Λ *separates* $x, y \in S$ if they belong to distinct halfspaces of Λ. Let $K \subseteq S$. A wall Λ *intersects* K if K nontrivially intersects both $\overleftarrow{\Lambda}$ and $\overrightarrow{\Lambda}$. Let $\mathcal{W}$ be a set of walls in S; then $(S, \mathcal{W})$ is a wallspace if for all $x, y \in S$, the number of walls separating x and y is finite. If Λ intersects K, then the *restriction of* Λ *to* K is the wall in K determined by $\Lambda|_K = \{\overleftarrow{\Lambda} \cap K, \overrightarrow{\Lambda} \cap K\}$.

In this paper, duplicate walls are not permitted in $\mathcal{W}$. Let $\mathcal{H}$ be the set of all halfspaces corresponding to the walls in $\mathcal{W}$.

Example 2.1 Let X be a CAT(0) cube complex, and let $\Lambda \subseteq X$ be a hyperplane in X. The complement $X - \Lambda$ has two components, therefore defining a wall in X such that $\overleftarrow{\Lambda}$ is an open halfspace not containing Λ and $\overrightarrow{\Lambda}$ is a closed halfspace containing Λ. Note that $\overleftarrow{\Lambda} \sqcup \overrightarrow{\Lambda} = X$. Let $L(\Lambda)$ and $R(\Lambda)$ denote the maximal subcomplexes contained in $\overleftarrow{\Lambda}$ and $\overrightarrow{\Lambda}$, respectively. Note that $L(\Lambda)$ and $R(\Lambda)$ are convex subcomplexes. Let $\mathcal{W}$ be the set of walls determined by the hyperplanes in X. Then $(X, \mathcal{W})$ is the wallspace associated to X. Note that we are using Λ to denote both the hyperplane and the wall corresponding to the hyperplane.

A function $c\colon \mathcal{W} \to \mathcal{H}$ is a 0–*cube* if $c[\Lambda] \in \{\overleftarrow{\Lambda}, \overrightarrow{\Lambda}\}$ and the following two conditions are satisfied:

(1) For all $\Lambda_1, \Lambda_2 \in \mathcal{W}$, the intersection $c[\Lambda_1] \cap c[\Lambda_2]$ is nonempty.

(2) For all $x \in S$, the set $\{\Lambda \in \mathcal{W} \mid x \notin c[\Lambda]\}$ is finite.

The *dual cube complex* $C(S, \mathcal{W})$ is the connected CAT(0) cube complex obtained by letting the union of all 0–cubes be the 0–skeleton. Two 0–cubes $c_1 \neq c_2$ are endpoints of a 1–cube if $c_1[\Lambda] = c_2[\Lambda]$ for all but precisely one $\Lambda \in \mathcal{W}$. An n–cube is then inserted wherever there is the 1–skeleton of an n–cube. The hyperplanes in $C(S, \mathcal{W})$ are identified naturally with the walls in $\mathcal{W}$. A proof of the fact that $C(S, \mathcal{W})$ is in fact a CAT(0) cube complex can be found in [9].

A point $x \in S$ determines a 0–cube c_x defined such that $x \in c_x[\Lambda]$ for all $\Lambda \in \mathcal{W}$. Condition (1) holds immediately since $x \in c_x[\Lambda]$ for all $\Lambda \in \mathcal{W}$. Condition (2) holds for c_x, since if $y \in S$ a wall Λ does not separate x and y, we can deduce that $y \in c_x[\Lambda]$, hence all but finitely many Λ satisfy $y \in c_x[\Lambda]$. Such 0–cubes are called the *canonical* 0–*cubes*.

Lemma 2.2 *Let X be a* CAT(0) *cube complex. Let $\mathcal{W}$ be a set of walls obtained from the hyperplanes in X. Let Z be a connected subcomplex of X, and let $\mathcal{W}_Z \subseteq \mathcal{W}$ be the subset of walls intersecting Z. Let $\mathcal{V}$ be the set of walls in $\mathcal{W}_Z$ restricted to Z. Then $(Z, \mathcal{V})$ is a wallspace and $C(Z, \mathcal{V})$ embeds in $C(X, \mathcal{W})$ isometrically in the combinatorial metric.*

Proof We first claim that the map $\mathcal{W}_Z \to \mathcal{V}$ is an injection. Suppose that $\Lambda_1, \Lambda_2 \in \mathcal{W}_Z$ are distinct walls. As Λ_1 and Λ_2 intersect Z, and since Z is connected, there are 1–cubes e_1 and e_2 in Z that are dual to the hyperplanes corresponding to Λ_1 and Λ_2. Therefore, both 0–cubes in e_1 belong in a single halfspace of $\Lambda_2|_Z$, so $\Lambda_1|_Z \neq \Lambda_2|_Z$.

We construct a map $\phi\colon C(Z, \mathcal{V}) \to C(X, \mathcal{W})$ on the 0–skeleton first. Let c be a 0–cube in $C(Z, \mathcal{V})$. We let $\phi(c) \in C(X, \mathcal{W})$ be the uniquely defined 0–cube such that $\phi(c)[\Lambda] \supseteq c[\Lambda|_Z]$ for $\Lambda|_Z \in \mathcal{V}$, and $\phi(c)[\Lambda] \supseteq Z$ for $\Lambda \in \mathcal{W} - \mathcal{W}_Z$. To verify that $\phi(c)$ is a 0–cube, first observe that $\phi(c)[\Lambda_1] \cap \phi(c)[\Lambda_2]$ is nonempty since $\Lambda_1|_Z \cap \Lambda_2|_Z \subseteq X$. Secondly, if $x \in X$ we need to show that $x \in \phi(c)[\Lambda]$

for all but finitely many $\Lambda \in \mathcal{W}$. Choose $z \in Z$; then $z \in c[\Lambda|_Z]$ for all $\Lambda|_Z \in \mathcal{V} - \{\Lambda_1|_Z, \dots, \Lambda_k|_Z\}$, hence $z \in \phi(c)[\Lambda]$ for all $\Lambda \in \mathcal{W}_Z - \{\Lambda_1, \dots, \Lambda_k\}$. Let $\{\Lambda_{k+1}, \dots, \Lambda_{k+\ell}\}$ be the set of walls in $\mathcal{W}$ separating x and z. Then $x \in \phi(c)[\Lambda]$ for all $\Lambda \in \mathcal{W} - \{\Lambda_1, \dots \Lambda_{k+\ell}\}$.

The 0–cubes are embedded since if $c_1 \neq c_2$, there exists $\Lambda|_Z \in \mathcal{V}$ such that $c_1[\Lambda|_Z] \neq c_2[\Lambda|_Z]$, hence $\phi(c_1)[\Lambda] \neq \phi(c_2)[\Lambda]$. If c_1 and c_2 are adjacent 0–cubes in $C(Z, \mathcal{V})$, then $c_1[\Lambda|_Z] = c_2[\Lambda|_Z]$ for all $\Lambda|_Z \in \mathcal{V}$, with the exception of precisely one wall $\widehat{\Lambda}|_Z$. Therefore, we can deduce that $\phi(c_1)[\Lambda] = \phi(c_2)[\Lambda]$ for all walls in $\mathcal{W}$, with the precise exception of $\widehat{\Lambda}$. Therefore, the 1–skeleton of $C(Z, \mathcal{V})$ embeds in $C(X, \mathcal{W})$, which is sufficient for ϕ to extend to an embedding of the entire cube complex.

Consider $C(Z, \mathcal{V})$ as a subcomplex of $C(X, \mathcal{W})$. The set of hyperplanes in $C(Z, \mathcal{V})$ embeds into the set of hyperplanes in $C(X, \mathcal{W})$. To see that $C(Z, \mathcal{V})$ is an isometrically embedded subcomplex, let z_1 and z_2 be 0–cubes in Z and γ be a geodesic combinatorial path in $C(Z, \mathcal{V})$ joining them. Each hyperplane dual to γ in $C(Z, \mathcal{V})$ intersects γ precisely once, and since the hyperplanes in $C(Z, \mathcal{V})$ inject to hyperplanes in $C(X, \mathcal{W})$, it is geodesic there as well. □

Given a wall Λ associated to a hyperplane in X we let $N(\Lambda)$ denote the *carrier* of Λ, by which we mean the union of all cubes intersected by Λ.

The following lemma describes what is called the *restriction quotient* in [3].

Lemma 2.3 *Let S be a set and let $\mathcal{W}$ be a set of walls of S. Let G be a group acting on $(S, \mathcal{W})$. Let $\mathcal{V} \subseteq \mathcal{W}$ be a G–invariant subset. Then there is a G–equivariant function ϕ: $C(S, \mathcal{W})^0 \to C(S, \mathcal{V})^0$. Moreover, $\phi^{-1}(z)$ is nonempty for all 0–cubes z in $C(S, \mathcal{V})$.*

Proof Let c be a 0–cube in $C(S, \mathcal{W})$. Let $\phi(c)[\Lambda] = c[\Lambda]$ for $\Lambda \in \mathcal{V}$. It is immediate that ϕ is G–equivariant.

To verify $\phi(c)[\Lambda]$ is a 0–cube in $C(S, \mathcal{V})$ first note that $\phi(c_1)[\Lambda_1] \cap \phi(c_2)[\Lambda_2] \neq \varnothing$ for all $\Lambda_1, \Lambda_2 \in \mathcal{V}$, since $c_1[\Lambda_1] \cap c_2[\Lambda_2] \neq \varnothing$ for all $\Lambda_1, \Lambda_2 \in \mathcal{W}$. Secondly, for all $x \in S$ observe that $x \in \phi(c)[\Lambda]$ for all but finitely many $\Lambda \in \mathcal{V}$. Indeed, this is true for all but finitely many $\Lambda \in \mathcal{W}$.

To see that $\phi^{-1}(z)$ is nonempty for all 0–cubes z in $C(S, \mathcal{V})$, we determine a 0–cube x in $C(S, \mathcal{W})$ such that $\phi(x) = z$. Fix $s \in S$. Let $x[\Lambda] = z[\Lambda]$ for $\Lambda \in \mathcal{V}$. Suppose that $\Lambda \in \mathcal{W} - \mathcal{V}$. If $\vec{\Lambda} \supseteq z[\Lambda']$ for some $\Lambda' \in \mathcal{V}$ let $x[\Lambda] = \vec{\Lambda}$. Similarly if $\overleftarrow{\Lambda} \supseteq z[\Lambda']$. Otherwise, if Λ intersects $z[\Lambda']$ for all $\Lambda' \in \mathcal{V}$ then let $s \in x[\Lambda]$.

To verify that x is a 0–cube, consider the following cases to show $x[\Lambda_1]\cap x[\Lambda_2]\neq\varnothing$ for $\Lambda_1,\Lambda_2\in\mathcal{W}$. If $\Lambda_1,\Lambda_2\in\mathcal{V}$ then $x[\Lambda_1]\cap x[\Lambda_2]=z[\Lambda_1]\cap z[\Lambda_2]\neq\varnothing$. Suppose that $\Lambda_1\in\mathcal{W}-\mathcal{V}$ and $x[\Lambda_1]\subseteq z[\Lambda_1']$ for some $\Lambda_1'\in\mathcal{V}$. If $\Lambda_2\in\mathcal{V}$, then $x[\Lambda_1]\cap x[\Lambda_2]\supseteq z[\Lambda_1']\cap z[\Lambda_2]\neq\varnothing$. If $\Lambda_2\in\mathcal{W}-\mathcal{V}$ and $x[\Lambda_2]\subseteq z[\Lambda_2']$ for some $\Lambda_2'\in\mathcal{V}$ then $x[\Lambda_1]\cap x[\Lambda_2]\subseteq z[\Lambda_1']\cap z[\Lambda_2']\neq\varnothing$. If Λ_2 intersects $z[\Lambda]$ for all $\Lambda\in\mathcal{V}$, then $x[\Lambda_1]\cap x[\Lambda_2]\supseteq z[\Lambda_1']\cap x[\Lambda_2]\neq\varnothing$. Finally if both $s\in x[\Lambda_1]$ and $s\in x[\Lambda_2]$, then their intersection will contain at least s.

Finally, we verify that for $s'\in S$ there are only finitely many $\Lambda\in\mathcal{W}$ such that $s'\notin x[\Lambda]$. Suppose, by way of contradiction, that there is an infinite subset of walls $\{\Lambda_1,\Lambda_2,\dots\}\subseteq\mathcal{W}$ such that $s'\notin x[\Lambda_i]$ for all $i\in\mathbb{N}$. We can assume, by excluding at most finitely many walls, that each $\Lambda_i\in\mathcal{W}-\mathcal{V}$. Similarly, by excluding finitely many walls, we can assume that Λ_i does not separate s and s'. Therefore, $s\notin x[\Lambda_i]$ for $i\in\mathbb{N}$. Therefore, by construction of x, there exist $\Lambda_i'\in\mathcal{V}$ such that $z[\Lambda_i']\subseteq x[\Lambda_i]$, which implies that $s'\notin z[\Lambda_i']$. There are infinitely many distinct Λ_i', as otherwise there is a $\Lambda'\in\mathcal{V}$ such that $z[\Lambda']\subseteq x[\Lambda_i]$ for infinitely many i, which would imply that infinitely many Λ_i separate s' from an element in the complement of $z[\Lambda']$. Therefore, infinitely many distinct walls $\Lambda_i'\in\mathcal{V}$ have $s'\notin z[\Lambda_i']$, contradicting that z is a 0–cube in $C(S,\mathcal{V})$. □

3 Minimal $\mathbb{Z}^n$–invariant convex subcomplexes

The following is Theorem 2 from [4]. As this paper is written in Russian, we give a proof in an appendix based on the work in [8] as well as stating the definition of codimension-1.

Theorem 3.1 (Gerasimov [4]) *Let G be a finitely generated group that acts on a* CAT(0) *cube complex X without a fixed point or inversions. Then there is a hyperplane in X that is stabilized by a codimension-*1 *subgroup of G.*

The goal of this section is to prove the following:

Lemma 3.2 *Let G be a finitely generated group acting without fixed point or inversions on a* CAT(0) *cube complex X. There exists a minimal, G–invariant, convex subcomplex $X_o\subseteq X$ such that X_o contains only finitely many hyperplane orbits, and every X_o hyperplane stabilizer is a codimension-*1 *subgroup of G.*

Proof Since G is finitely generated, by taking the convex hull of a G–orbit we obtain a G–invariant convex subcomplex $X_o\subseteq X$ containing finitely many G–orbits of hyperplanes. Assume that X_o is a minimal such subcomplex in terms of the number of hyperplane orbits.

Let $(X_o, \mathcal{W})$ be the wallspace obtained from the hyperplanes in X_o. Suppose that $\mathrm{Stab}_G(\Lambda)$ is not a codimension-1 subgroup of G for some $\Lambda \in \mathcal{W}$. Let $G\Lambda \subseteq \mathcal{W}$ be the G–orbit of Λ. By Lemma 2.3 there is an G–invariant map $\phi\colon X_o^0 \to C(X_o, G\Lambda)^0$. Since $\mathrm{Stab}_G(\Lambda)$ is not commensurable to a codimension-1 subgroup, Theorem 3.1 implies that there is a fixed 0–cube x in $C(X_o, G\Lambda)$. Lemma 2.3 then implies that $\phi^{-1}(x)$ is nonempty. Assuming that $\phi^{-1}(x) \subseteq \overleftarrow{\Lambda}$, then the intersection $\bigcap_{g\in G} gL(\Lambda)$ contains a proper, convex, G–invariant subcomplex of X_o, with one less hyperplane orbit. This contradicts the minimality of X_o. □

The following corollary follows since all codimension-1 subgroups of a rank n virtually abelian group are of rank $n-1$.

Corollary 3.3 *Let G be a rank n, virtually abelian group acting without fixed point or inversions on a* CAT(0) *cube complex X. Then there exists a minimal, G–invariant, convex subcomplex $X_o \subseteq X$ such that X_o contains only finitely many hyperplane orbits, and every hyperplane stabilizer is a rank $n-1$ subgroup of G.*

4 Proof of the main theorem

Definition 4.1 Regard $\mathbb{R}$ as a CAT(0) cube complex whose 0–skeleton is $\mathbb{Z}$. Let g be an isometry of X. A *geodesic combinatorial axis* for g is a g–invariant subcomplex homeomorphic to $\mathbb{R}$ that embeds isometrically in X.

Definition 4.2 Let (M, d) be a metric space. The subspaces $N_1, N_2 \subseteq M$ are *coarsely equivalent* if each lies in an r–neighborhood of the other for some $r > 0$.

Theorem 4.3 *Let G be virtually $\mathbb{Z}^n$. Suppose G acts properly and without inversions on a* CAT(0) *cube complex X. Then G stabilizes a finite-dimensional subcomplex $Y \subseteq X$ that is isometrically embedded in the combinatorial metric, and $Y \cong \prod_{i=1}^m C_i$, where each C_i is a cubical quasiline and $m \geq n$. Moreover,* $\mathrm{Stab}_G(\Lambda)$ *is a codimension-1 subgroup for each hyperplane Λ in Y.*

Proof By Corollary 3.3 there is a minimal, nonempty, convex subcomplex $X_o \subseteq X$ stabilized by G, containing finitely many hyperplane orbits, and $\mathrm{Stab}_G(\Lambda)$ is a rank $n-1$ subgroup of G for each hyperplane $\Lambda \subseteq X_o$.

Let $x \in X_o$ be a 0–cube. Let Υ be the Cayley graph of G with respect to a finite generating set S. Let $N = \max\{\mathrm{d}_X^c(x, gx) \mid g \in S\}$. Let $\phi\colon \Upsilon \to X_o$ be a G–equivariant map that maps the vertex corresponding to 1_G to x, and edges to geodesic combinatorial paths in X_o. Note that the image $\phi(e)$ of each edge e in Υ has length at most N and is intersected at most once by each hyperplane. Let $Q = \phi(\Upsilon)$. As G acts properly on X, and cocompactly on Υ, the graph Q is quasiisometric to G. Let $\mathcal{W}_Q$ be the set of hyperplanes intersecting Q, and let $(Q, \mathcal{W}_Q)$ be the associated wallspace. By Lemma 2.2 we know that $C(Q, \mathcal{W}_Q)$ is an isometrically embedded subcomplex of X_o. Fix a proper action of G on $\mathbb{R}^n$, and let $q\colon Q \to \mathbb{R}^n$ be a G–equivariant quasiisometry. Note that $\mathrm{Stab}_G(\Lambda)$ is a quasiisometrically embedded codimension-1 subgroup of G, for all $\Lambda \in \mathcal{W}_Q$. We claim that $q(\Lambda \cap Q)$ is coarsely equivalent to a codimension-1 affine subspace $H \subseteq \mathbb{R}^n$.

As G is virtually $\mathbb{Z}^n$ and $\mathrm{Stab}_G(\Lambda)$ is a codimension-1, there exists $g \in S$ such that $\langle g \rangle$ is not virtually contained in $\mathrm{Stab}_G(\Lambda)$. There are finitely many $\mathrm{Stab}_G(\Lambda)$–orbits of vertices in $\Upsilon/\langle g \rangle$, so let $A = \{a_0, \dots a_k\}$ be representatives in Υ such that Λ separates $\phi(a_i)$ and $\phi(ga_i)$. Let γ_i be the biinfinite geodesic in Υ containing $\langle g \rangle a_i$. Then $\Lambda \cap \phi(\gamma_i)$ is contained in the $N(N+1)$ neighborhood of $\phi(a_i)$ in $\phi(\gamma_i)$, since otherwise Λ would intersect a pair of 1–cubes in $\phi(\gamma_i)$ that lie in the same $\langle g \rangle$–orbit, implying that $\langle g \rangle$ is virtually contained in $\mathrm{Stab}_G(\Lambda)$. Thus, $\Lambda \cap h\phi(\gamma_i)$ is contained in the $N(N+1)$ neighborhood of $h\phi(a_i)$ in $h\phi(\gamma_i)$ for all $h \in \mathrm{Stab}_G(\Lambda)$.

Now suppose that Λ intersects Q outside of the $N(N+2)$ neighborhood of $\mathrm{Stab}_G(\Lambda)A$. Then Λ must intersect $\phi(e)$, where e is an edge connecting $h_1\gamma_i$ and $h_2\gamma_j$ for some $h_1, h_2 \in \mathrm{Stab}_G(\Lambda)$. Up to taking the inverse of g, we can assume that ge is further away from h_1a_i and h_2a_j than e. Then Λ must intersect $\phi(ge)$ since Λ is 2–sided, intersects $\phi(e)$ precisely once, and cannot intersect the intervals in $h_1\gamma_i$ and $h_2\gamma_j$ that lie between e and ge. Similarly, Λ intersects $\phi(g^n e)$ for all $n > 0$ implying that Λ intersects a pair of 1–cubes in the same $\langle g \rangle$–orbit, further implying that $\langle g \rangle$ is virtually contained in $\mathrm{Stab}_G(\Lambda)$ and contradicting the initial assumption on g. Thus, Λ cannot intersect Q outside of the $N(N+2)$ neighborhood of $\mathrm{Stab}_G(\Lambda)A$. Thus $q(\Lambda \cap Q)$ is coarsely equivalent to a codimension-1 affine subspace $H \subseteq \mathbb{R}^n$. Moreover, $q(\overleftarrow{\Lambda} \cap Q)$ and $q(\overrightarrow{\Lambda} \cap Q)$ are coarsely equivalent to the halfspaces of H.

Let $n > 0$. Since there are finitely many orbits of hyperplanes in X_o, there are only finitely many commensurability classes of stabilizers. Therefore, we may partition $\mathcal{W}_Q$ as the disjoint union $\bigsqcup_{i=1}^m \mathcal{W}_i$, where each $\mathcal{W}_i$ contains all walls with commensurable stabilizers. For each $\Lambda_i \in \mathcal{W}_i$ let $q(\Lambda_i \cap Q)$ be coarsely equivalent to a codimension-1 affine subspace $H_i \subseteq \mathbb{R}^n$, stabilized by $\mathrm{Stab}_G(\Lambda_i)$. If $i \neq j$ then H_i and H_j are nonparallel affine subspaces, and therefore Λ_i and Λ_j will intersect in Q. Therefore, every wall in $\mathcal{W}_i$ intersects every wall in $\mathcal{W}_j$ if $i \neq j$, and thus $C(Q, \mathcal{W}_Q) \cong \prod_{i=1}^m C(Q, \mathcal{W}_i)$.

Finally, we show that $C(Q, \mathcal{W}_i)$ is a quasiline for each $1 \leq i \leq m$. As G permutes the factors in $\prod_{i=1}^m C(Q, \mathcal{W}_i)$, there is a finite index subgroup $G' \leqslant G$ that preserves each factor. For each i, the stabilizers $\mathrm{Stab}_G(\Lambda)$ are commensurable for all $\Lambda \in \mathcal{W}_i$.

Therefore, there is a cyclic subgroup Z_i that is not virtually contained in any $\mathrm{Stab}_G(\Lambda)$ and thus acts freely on $C(Q, \mathcal{W}_i)$. As the stabilizers of $\Lambda \in \mathcal{W}_i$ are commensurable, all $q(\Lambda \cap Q)$ will be quasiequivalent to parallel codimension-1 affine subspaces of $\mathbb{R}^n$, which implies that only finitely many Z_i–translates of Λ can pairwise intersect. As there are finitely many Z_i–orbits of Λ in $\mathcal{W}_i$, there is an upper bound on the number of pairwise intersecting hyperplanes in $\mathcal{W}_i$. Thus, there are finitely many Z_i–orbits of maximal cubes in $C(Q, \mathcal{W}_i)$, which implies that $C(Q, \mathcal{W}_i)$ is CAT(0) cube complex quasiisometric to $\mathbb{R}$. □

We can now prove Corollary 4.4.

Corollary 4.4 *Let G be virtually $\mathbb{Z}^n$. Suppose G acts properly and without inversions on a* CAT(0) *cube complex X. Then G cocompactly stabilizes a subspace $F \subseteq X$ homeomorphic to $\mathbb{R}^n$ such that for each hyperplane $\Lambda \subseteq X$, the intersection $\Lambda \cap F$ is either empty or homeomorphic to $\mathbb{R}^{n-1}$.*

Proof By Theorem 4.3, there is a G–equivariant, isometrically embedded, subcomplex $Y \subseteq X$ such that $Y = \prod_{i=1}^m C_i$, where each C_i is a quasiline, and $\mathrm{Stab}_G(\Lambda)$ is a codimension-1 subgroup. Considering Y with the CAT(0) metric, note that Y is a complete CAT(0) metric space in its own right, and G acts semisimply on Y. By the flat torus theorem [2] there is an isometrically embedded flat $F \subseteq Y$. Note that $F \subseteq X$ is not isometrically embedded. As $\mathrm{Stab}_G(\Lambda)$ is a codimension-1 subgroup of G for each hyperplane Λ in X, the intersection $\Lambda \cap F = (\Lambda \cap Y) \cap F$ is either empty or, as $F \subseteq Y$ is isometrically embedded, the hyperplane intersection is an isometrically embedded copy of $\mathbb{R}^{n-1}$. □

5 Haglund's axis

The goal of this section is to reprove the following result of Haglund as a consequence of Corollary 4.4.

Theorem 5.1 (Haglund [6]) *Let G be a group acting on a* CAT(0) *cube complex without inversions. Every element $g \in G$ either fixes a 0–cube of G, or stabilizes a combinatorial geodesic axis.*

Proof As finite groups don't contain codimension-1 subgroups, Theorem 3.1 implies that if g is finite order then it fixes a 0–cube. Suppose that G does not fix a 0–cube, then $\langle g \rangle$ must act properly on X. By Corollary 4.4, there is a line $L \subset X$ stabilized by G, that intersects each hyperplane at most once at a single point in L. Let $\mathcal{W}_L$ be the set of hyperplanes intersecting L. Note that the intersection points of the walls in $\mathcal{W}_L$ with L is a locally finite subset.

Fix a basepoint $p \in L$ that doesn't belong to a hyperplane intersecting L, and let x be the canonical 0–cube corresponding to p. Let $\Lambda_1, \dots, \Lambda_k$ be the set of hyperplanes separating p and gp, and assume that $p \in \overleftarrow{\Lambda}_i$. Reindex the hyperplanes such that $\overleftarrow{\Lambda}_1 \cap L \subseteq \overleftarrow{\Lambda}_2 \cap L \subseteq \cdots \subseteq \overleftarrow{\Lambda}_k \cap L$. The ordering of the hyperplanes separating p and gp determines a combinatorial geodesic joining x and gx of length k, where the i^{th} edge is a 1–cube dual to Λ_i. This can be extended $\langle g \rangle$–equivariantly, to obtain a combinatorial geodesic axis L_c, since each hyperplanes intersects L_c at most once. □

Appendix: Codimension-1 subgroups

Definition A.1 Let G be a finitely generated group. Let Υ denote the Cayley graph of G with respect to some finite generating set. A subgroup $H \leqslant G$ is *codimension-*1 if K/Υ has more than one end.

Let $\oplus$ denote the operation of symmetric difference. A subset $A \subseteq G$ is *H–finite* if $A \subseteq HF$ where F is some finite subset of G. We will use the following equivalent formulation (see [11]) of codimension-1: A subgroup $H \leq G$ is a codimension-1 subgroup if there exists some $A \subseteq G$ such that:

(1) $A = HA$.

(2) A is *H–almost invariant*, that is to say that $A \oplus Ag$ is H–finite for any $g \in G$.

(3) A is *H–proper*, that is to say that neither A nor $G - A$ is H–finite.

We will reprove the following theorem from [4] using techniques from [8].

Theorem A.2 *Let G be a finitely generated group acting on a* CAT(0) *cube complex X without edge inversions or fixing a 0–cube. Then the stabilizer of some hyperplane in X is a codimension-*1 *subgroup of G.*

Proof Suppose that no hyperplane stabilizer is a codimension-1 subgroup of G. We will find a 0–cube fixed by G.

Let $\mathcal{H}$ denote the set of hyperplanes in X. We can assume that X has finitely many G–orbits of hyperplanes after possibly passing to the convex hull of a single 0–cube orbit in X. If x and y are 0–cubes in X, then let $\Delta(x, y) \subseteq \mathcal{H}$ denote the set of hyperplanes separating x and y. Note that

$$\mathsf{d}_X^c(x, y) = |\Delta(x, y)|.$$

Let $\Lambda_1, \dots, \Lambda_n$ be a minimal set of representatives of the orbits of hyperplanes. Let x_0 be some fixed choice of 0–cube in X. Let

$$H_i = \mathrm{Stab}_G(\Lambda_i) \quad \text{and} \quad A_i = \{g \in G \mid gx_0 \in \overleftarrow{\Lambda}_i\}.$$

We can verify that A_i satisfies the first two criteria in Definition A.1.

(1) It is immediate that $A_i = H_i A_i$, as G doesn't invert the hyperplanes in X.

(2) Let *xor* denote the exclusive or. For $f \in G$ we can deduce that $A_i \oplus A_i f$ is H_i–finite:

$$\begin{aligned} g \in A_i \oplus A_i f \quad &\iff \quad g x_0 \in \overleftarrow{\Lambda}_i \text{ xor } g f^{-1} x_0 \in \overleftarrow{\Lambda}_i \\ &\iff \quad x_0 \in g^{-1} \overleftarrow{\Lambda}_i \text{ xor } f^{-1} x_0 \in g^{-1} \overleftarrow{\Lambda}_i \\ &\iff \quad g \in G \text{ is such that } g^{-1}\Lambda_i \text{ separates } x_0 \text{ and } f^{-1}x_0. \end{aligned}$$

As $(X, \mathcal{H})$ is a wallspace, there are only finitely many $g \in G$ such that $g^{-1}\Lambda_i$ separates x_0 and $f^{-1}x_0$. If $g_1\Lambda_i, \dots, g_k\Lambda_i$ are the translates then

$$A_i \oplus A_i f = \{g_1, \dots, g_k\} H_i,$$

which implies almost H_i–invariance.

Therefore, A_i cannot be H_i–proper for any i, as we have assumed that none of the H_i are codimension-1. This means that either A_i or $G - A_i$ is H_i–finite. After possibly reversing the orientation of Λ_i we can assume that A_i is H_i–finite, so $A_i \subseteq H_i F_i$ where $F_i \subseteq G$ is finite.

Claim $\mathsf{d}_X(x_0, f x_0) < 2\max_i(|F_i|)$ *for all* $f \in G$.

Proof

$$\begin{aligned} g\Lambda_i \in \Delta(x_0, f x_0) \quad &\iff \quad x_0[g\Lambda_i] \neq f x_0[g\Lambda_i] \\ &\iff \quad x_0 \in g\overleftarrow{\Lambda}_i \text{ xor } f x_0 \in g\overleftarrow{\Lambda}_i \\ &\iff \quad g^{-1}x_0 \in \overleftarrow{\Lambda}_i \text{ xor } g^{-1} f x_0 \in \overleftarrow{\Lambda}_i \\ &\iff \quad g^{-1} \in A_i \text{ xor } g^{-1} \in A f^{-1} \\ &\iff \quad g^{-1} \in A_i \oplus A_i f^{-1}. \end{aligned}$$

As the final set is covered by $2|F_i|$ translates of H_i, we can deduce that there are at most $2|F_i|$ hyperplanes in $\Delta(x_0, f x_0)$. □

Thus, we can conclude that the G–orbit of x_0 is a bounded set. If G has a finite orbit in X, then the convex hull of the orbit is a compact, finite-dimensional, complete CAT(0) cube complex, and we can apply Corollary II.2.8(1) from [2] to find a fixed point p. If p is a 0–cube then we are done. Otherwise, p is in the interior of some n–cube that is fixed by G, and since G doesn't invert hyperplanes we can deduce that G fixes a 0–cube in that cube. If the G–orbits in X are infinite, then their convex hull may not be complete, so the above argument will not hold. Instead, we will follow the strategy of [8] and embed the cube complex into a Hilbert space.

Let $\mathcal{C}(\mathcal{H})$ denote the *connected cube*, a graph with vertices given by functions $c\colon \mathcal{H} \to \{0,1\}$ with finite support, and edges that join a pair of distinct vertices if and only if they differ on precisely one hyperplane.

Fix a 0–cube x_0. Then there is an embedding

$$\phi\colon X^1 \hookrightarrow \mathcal{C}(\mathcal{H})$$

that maps the 0–cube x to c_x, where

$$c_x(\Lambda) = \begin{cases} 1 & \text{if } x[\Lambda] \neq x_0[\Lambda], \\ 0 & \text{if } x[\Lambda] = x_0[\Lambda]. \end{cases}$$

A hyperplane $\Lambda \in \mathcal{H}$ *separates* two vertices c_1 and c_2 in $\mathcal{C}(\mathcal{H})$ if $c_1(\Lambda) \neq c_2(\Lambda)$. Note that Λ separates 0–cubes x, y in X if and only if it separates $\phi(x)$ and $\phi(y)$. Therefore, we can define $\Delta(c_1, c_2)$ for vertices in $\mathcal{C}(\mathcal{H})$ and conclude that if x, y are 0–cubes in X then $\Delta(x, y) = \Delta(\phi(x), \phi(y))$. This implies that ϕ is an isometric embedding in the combinatorial metric.

We will show that a bounded orbit in X implies there is a fixed 0–cube in $\mathcal{C}(\mathcal{H})$ and then argue that we can go one step further and find a fixed 0–cube in X.

Let $\ell^2(\mathcal{H})$ be the Hilbert space of square summable functions $s\colon \mathcal{H} \to \mathbb{R}$. There is an embedding $\rho\colon \mathcal{C}(\mathcal{H}) \to \ell^2(\mathcal{H})$ given by

$$\rho(c)(\Lambda) = c[\Lambda].$$

It is straightforward to verify that $\|\rho(c_1) - \rho(c_2)\|^2 = \mathrm{d}_{\mathcal{C}(\mathcal{H})}(c_1, c_2)$. There is a G–action on $\ell^2(\mathcal{H})$ such that if $s \in \ell^2(\mathcal{H})$, $\Lambda \in \mathcal{H}$ and $g \in G$, then

$$gs(\Lambda) = \begin{cases} s(g^{-1}\Lambda) & \text{if } c_{x_0}(g^{-1}\Lambda) = c_{x_0}(\Lambda), \\ 1 - s(g^{-1}\Lambda) & \text{if } c_{x_0}(g^{-1}\Lambda) \neq c_{x_0}(\Lambda). \end{cases}$$

It is again straightforward to verify that this action is by isometries, and that ρ is G–equivariant.

As Gx_0 is bounded, so is $G(\rho \circ \phi(x_0))$. It then follows that G has a fixed point in $\ell^2(\mathcal{H})$ (a proof is in [8], which also cites Lemma 3.8 in [7]). Let $s\colon \mathcal{H} \to \mathbb{R}$ be the fixed point. For all $g \in G$ we can deduce that $s(g\Lambda)$ is either $s(\Lambda)$ or $1 - s(\Lambda)$. Therefore s can only take two values on the hyperplanes in a single G–orbit. As s has to be square summable the two values have to be 0 and 1, and s can only take the value 1 on finitely many hyperplanes. Thus, s is the image of a point c in $\mathcal{C}(\mathcal{S})$.

Let $c \in \mathcal{C}(\mathcal{S})$ be a G–invariant vertex which minimizes the distance to the image of X^1 in $\mathcal{C}(\mathcal{S})$. Let Z be a G–orbit of 0–cubes in X such that $\phi(Z)$ realizes the minimal distance from c.

Let $\mathcal{V}$ be the set of hyperplanes that intersect $\{c\} \cup \mathcal{V}$. Every hyperplane in $\mathcal{V}$ must intersect Z, otherwise if $\mathcal{F} \subseteq \mathcal{V}$ is the finite, G–invariant subset of hyperplanes separating c from Z then we can define a 0–cube c' such that

$$c'(\Lambda) = \begin{cases} c(\Lambda) & \text{if } \Lambda \notin \mathcal{F}, \\ 1 - c(\Lambda') & \text{if } \Lambda \in \mathcal{F}, \end{cases}$$

and deduce that c' is G–invariant and is $|\mathcal{F}|$ closer to Z than c.

Let $z_0, z_1, z_2, \dots$ be an enumeration of 0–cubes in Z. Each hyperplane separating z_0 and z_1 must lie in either $\Delta(z_0, c)$ or $\Delta(z_1, c)$. As z_0 is minimal distance in X from c, the edges in X incident to z_0 must be dual to hyperplanes not in $\Delta(z_0, c)$, and instead belongs to $\Delta(z_1, c)$. Therefore, the hyperplane $\Lambda_0 \in \mathcal{V}$ dual to the first edge in a combinatorial geodesic joining z_0 to z_1 must lie in $\Delta(z_1, c)$. Similarly, there exists a hyperplane Λ_1 dual to the first edge of the combinatorial geodesic in X joining z_1 to z_2 that belongs to $\Delta(z_2, c)$ but not $\Delta(z_1, c)$. Note that Λ_1 cannot intersect Λ_0 in X, otherwise Λ_0 would be dual to an edge incident to z_1, which would imply that there exists a 0–cube in X adjacent to z_1 that is closer to c. Therefore Λ_0, Λ_1 separates z_0 from z_2 in X. Iterating this argument produces a sequence of disjoint hyperplanes $\Lambda_0, \Lambda_2, \Lambda_3, \dots$ such that $\Lambda_0, \dots, \Lambda_k$ separates z_0 from z_{k+1} in X. This contradicts the hypothesis that Z is a bounded set in X. □

References

[1] **Y Algom-Kfir**, **B Wajnryb**, **P Witowicz**, *A parabolic action on a proper,* CAT(0) *cube complex*, J. Group Theory 16 (2013) 965–984 MR

[2] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR

[3] **P-E Caprace**, **M Sageev**, *Rank rigidity for* CAT(0) *cube complexes*, Geom. Funct. Anal. 21 (2011) 851–891 MR

[4] **V N Gerasimov**, *Semi-splittings of groups and actions on cubings*, from "Algebra, geometry, analysis and mathematical physics", Izdat. Ross. Akad. Nauk Sib. Otd. Inst. Mat., Novosibirsk (1997) 91–109 MR In Russian

[5] **S M Gersten**, *The automorphism group of a free group is not a* CAT(0) *group*, Proc. Amer. Math. Soc. 121 (1994) 999–1002 MR

[6] **F Haglund**, *Isometries of* CAT(0) *cube complexes are semi-simple*, preprint (2007) arXiv

[7] **P de la Harpe**, **A Valette**, *La propriété* (T) *de Kazhdan pour les groupes localement compacts*, Astérisque 175, Soc. Math. France, Paris (1989) MR

[8] **G A Niblo**, **M A Roller**, *Groups acting on cubes and Kazhdan's property* (T), Proc. Amer. Math. Soc. 126 (1998) 693–699 MR

[9] **M Sageev**, *Ends of group pairs and non-positively curved cube complexes*, Proc. London Math. Soc. 71 (1995) 585–617 MR

[10] **M Sageev**, *Codimension-*1 *subgroups and splittings of groups*, J. Algebra 189 (1997) 377–389 MR

[11] **P Scott**, *Ends of pairs of groups*, J. Pure Appl. Algebra 11 (1977) 179–198 MR

[12] **D T Wise**, *From riches to raags: 3–manifolds, right-angled Artin groups, and cubical geometry*, CBMS Regional Conference Series in Mathematics 117, Amer. Math. Soc., Providence, RI (2012) MR

[13] **D T Wise**, *Cubular tubular groups*, Trans. Amer. Math. Soc. 366 (2014) 5503–5521 MR

[14] **D T Wise**, **D J Woodhouse**, *A cubical flat torus theorem and the bounded packing property*, Israel J. Math. 217 (2017) 263–281 MR

[15] **D J Woodhouse**, *Classifying virtually special tubular groups*, preprint (2016) arXiv To appear in Groups Geom. Dyn.

Mathematics Department, Technion – Israel Institute of Technology
Haifa, Israel

woodhouse.da@technion.ac.il

Received: 3 February 2016 Revised: 9 May 2017

Algebraic & Geometric Topology 17 (2017) 2753–2762

On growth of systole along congruence coverings of Hilbert modular varieties

PLINIO G P MURILLO

We study how the systole of principal congruence coverings of a Hilbert modular variety grows when the degree of the covering goes to infinity. We prove that, given a Hilbert modular variety M_k of real dimension $2n$ defined over a number field k, the sequence of principal congruence coverings M_I eventually satisfies

$$\mathrm{sys}_1(M_I) \geq \frac{4}{3\sqrt{n}} \log(\mathrm{vol}(M_I)) - c,$$

where c is a constant independent of M_I.

22E40, 11R80; 53C22

1 Introduction

The *systole* of a riemannian manifold is the least length of a noncontractible closed geodesic in M and it is denoted by $\mathrm{sys}_1(M)$. In 1994, P Buser and P Sarnak [2] constructed the first explicit examples of surfaces with systole growing logarithmically with the genus using a sequence of principal congruence coverings of an arithmetic compact Riemann surface. These sequences of surfaces $\{S_p\}$ satisfy the inequality

$$\mathrm{sys}_1(S_p) \geq \tfrac{4}{3} \log(\mathrm{genus}(S_p)) - c,$$

where c is a constant independent of p. This result was generalized in 2007 by M Katz, M Schaps and U Vishne [6] to principal congruence coverings of any compact arithmetic Riemann surface and arithmetic hyperbolic 3–manifolds. It is known that a sequence of principal congruence coverings of a compact arithmetic hyperbolic manifold asymptotically attains the logarithmic growth of the systole (see Gromov [4, 3.C.6]) but the examples above are the only cases where the explicit constant in the systole growth is known so far. In particular, it would be interesting to understand how the asymptotic constant depends on the dimension.

The purpose of this paper is to generalize the construction of Buser and Sarnak to Hilbert modular varieties which are noncompact riemannian manifolds of dimension $2n$. We will show that the sequence of principal congruence coverings $M_I \to M_k$ of a

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2753

Hilbert modular variety eventually satisfies

$$\operatorname{sys}_1(M_I) \geq \frac{4}{3\sqrt{n}} \log(\operatorname{vol}(M_I)) - c, \tag{1}$$

where c is a constant independent of I. We also prove that inequality (1) is asymptotically sharp. We refer to Theorem 4.2 and Theorem 4.3 for the precise statement of the results.

Since M_k is noncompact, it is a priori not clear if the systole of M_I is bounded above by a logarithmic function of its volume. In fact, an interesting more general question is to understand if the systole of a sequence of congruence coverings of a noncompact finite-volume arithmetic manifold of nonpositive curvature and which is not flat grows logarithmically in its volume. An affirmative answer seems very plausible but, to our knowledge, it has not been established in the literature. In this regard we will prove that the sequence of principal congruence coverings $M_I \to M_k$ of a Hilbert modular variety eventually satisfies

$$\operatorname{sys}_1(M_I) \leq \frac{4\sqrt{n}}{3} \log(\operatorname{vol}(M_I)) - d \tag{2}$$

for some constant d independent of M_I. These results give us the first examples of explicit constants for the growth of systole of a sequences of congruence coverings of arithmetic manifolds in dimensions greater than three.

We will begin in Section 2 recalling basic aspects of the action of $(\mathrm{PSL}_2(\mathbb{R}))^n$ on $(\mathbb{H}^2)^n$. We then define the congruence coverings M_I of a Hilbert modular variety M_k, and we prove inequality (2). In Section 3 we estimate the length of closed geodesics of M_I in terms on the norm of the ideal I. In Section 4 we relate the norm of the ideal I to $\operatorname{vol}(M_I)$, and we prove inequality (1) and the sharpness of the constant $4/(3\sqrt{n})$.

Acknowledgements I would like to thank Mikhail Belolipetsky for his advice, support and many valuable suggestions on the preliminary versions of this work. I want to thank Cayo Doria for helpful discussions which significantly improved the quality of this paper. I also thank the referee for valuable comments and corrections. This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES.

2 Preliminaries

2.1 The action of $(\mathrm{PSL}_2(\mathbb{R}))^n$ on $(\mathbb{H}^2)^n$

The group $\mathrm{PSL}_2(\mathbb{R})$ acts on the upper half plane model of the hyperbolic plane $\mathbb{H}^2$ by fractional linear transformations via

$$Bz = \frac{az+b}{cz+d} \quad \text{if } B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } z \in \mathbb{H}^2.$$

An element $B \in \mathrm{PSL}_2(\mathbb{R})$ is called *elliptic* if it has a fixed point in $\mathbb{H}^2$, *parabolic* if it has no fixed points in $\mathbb{H}^2$ and has only one fixed point in $\partial\mathbb{H}^2$, and *hyperbolic* if it has no fixed points in $\mathbb{H}^2$ and has two fixed points in $\partial\mathbb{H}^2$. An equivalent description is the following:

- B is *elliptic* if and only if $|\mathrm{tr}(B)| < 2$;
- B is *parabolic* if and only if $|\mathrm{tr}(B)| = 2$;
- B is *hyperbolic* if and only if $|\mathrm{tr}(B)| > 2$.

Here $\mathrm{tr}(B)$ denotes the trace of the matrix B.

Given a hyperbolic transformation B, the *translation length* of B, denoted by ℓ_B, is defined by

$$\ell_B = \inf\{d_{\mathbb{H}^2}(z, Bz) \mid z \in \mathbb{H}^2\}.$$

This infimum is attained at points on the unique geodesic $\overline{\alpha}_B$ in $\mathbb{H}^2$ joining the fixed points of B in $\partial\mathbb{H}^2$. The transformation B leaves $\overline{\alpha}_B$ invariant and acts on it as a translation. In particular, if a subgroup $\Lambda \subset \mathrm{PSL}_2(\mathbb{R})$ acts properly discontinuously and freely on $\mathbb{H}^2$, every hyperbolic element $B \in \Lambda$ determines a noncontractible closed geodesic α on the Riemann surface $\mathbb{H}^2/\Lambda$, whose length is equal to the translation length ℓ_B of B. Reciprocally, any closed geodesic α in $\mathbb{H}^2/\Lambda$ lifts to a geodesic $\overline{\alpha}_B$ in $\mathbb{H}^2$ fixed by a hyperbolic matrix $B \in \Lambda$.

On the other hand, since B is hyperbolic, B is conjugate to a matrix of the form

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix},$$

where $|\lambda| = e^{\ell_B/2}$. Hence $2\cosh(\ell_B/2) = |\mathrm{tr}(B)|$ and for any $z \in \mathbb{H}^2$ we have

$$d_{\mathbb{H}^2}(z, Bz) \geq 2\log(|\mathrm{tr}(B)| - 1) > 0. \tag{3}$$

We refer to [1, Chapter 7] for further details about the geometry of the isometries of the hyperbolic plane $\mathbb{H}^2$.

The action of $\mathrm{PSL}_2(\mathbb{R})$ on $\mathbb{H}^2$ extends to an action of the n–fold product $(\mathrm{PSL}_2(\mathbb{R}))^n$ on the n–fold product $(\mathbb{H}^2)^n$ in a natural way: if $z = (z_1, \dots, z_n) \in (\mathbb{H}^2)^n$ and $B = (B_1, \dots, B_n) \in (\mathrm{PSL}_2(\mathbb{R}))^n$, then

$$Bz := (B_1 z_1, \dots, B_n z_n),$$

where the action in every factor is the action by fractional linear transformations.

Let us recall the definition of a Hilbert modular variety (see [3]). Let k be a totally real number field of degree n, $\mathcal{O}_k$ the ring of integers of k and $\sigma_1, \dots, \sigma_n$ the n embeddings of k into the real numbers $\mathbb{R}$. The group $\mathrm{PSL}_2(\mathcal{O}_k)$ becomes an arithmetic noncocompact irreducible lattice of the semisimple Lie group $(\mathrm{PSL}_2(\mathbb{R}))^n$ via the map $\Delta(B) = (\sigma_1(B), \dots, \sigma_n(B))$, where $\sigma_i(B)$ denotes the matrix obtained by applying σ_i to the entries of B (see [7, Proposition 5.5.8]). Via this embedding, $\mathrm{PSL}_2(\mathcal{O}_k)$ acts on the n–fold product of hyperbolic planes $(\mathbb{H}^2)^n$ with finite covolume. The quotient $M_k = (\mathbb{H}^2)^n / \mathrm{PSL}_2(\mathcal{O}_k)$ is called a *Hilbert modular variety* and the group $\Gamma = \mathrm{PSL}_2(\mathcal{O}_k)$ is called a *Hilbert modular group*.

2.2 Congruence coverings of M_k

Let $I \subset \mathcal{O}_k$ be an ideal, the *principal congruence subgroup* $\Gamma(I) \subset \Gamma$ *at level* I is defined by

$$\Gamma(I) = \{A \in \mathrm{SL}_2(\mathcal{O}_k) \mid A \equiv \mathrm{Id} \mod I\}/\{1, -1\},$$

where Id denotes the identity 2×2 matrix. Since $\mathcal{O}_k/I$ is finite, $\Gamma(I)$ is a finite-index subgroup of Γ for any ideal I of $\mathcal{O}_k$. We associate to $\Gamma(I)$ a *congruence cover* $M_I = (\mathbb{H}^2)^n / \Gamma(I) \to M_k$. Note that Γ is an irreducible lattice in $(\mathrm{PSL}_2(\mathbb{R}))^n$ and so the varieties M_k and M_I do not split into products. We remark that M_k has quotient singularities, so the covering $M_I \to M_k$ should be interpreted in the orbifold sense. For large enough I the varieties M_I are manifolds by Selberg's lemma (see also Corollary 3.3).

This construction is a particular case of a more general situation: if G is a semisimple Lie group, a discrete subgroup $\Lambda \subset G$ is called arithmetic if there exists a number field K, a algebraic K–group H, and a surjective continuous homomorphism $\varphi\colon \mathrm{H}(K \otimes_{\mathbb{Q}} \mathbb{R}) \to G$ with compact kernel such that $\varphi(\mathrm{H}(\mathcal{O}_K))$ is commensurable to Λ, where $\mathrm{H}(\mathcal{O}_K)$ denotes the $\mathcal{O}_K$–points of H with respect to some fixed embedding of H into GL_m. For any ideal $I \subset \mathcal{O}_K$ the principal congruence subgroup of $\mathrm{H}(\mathcal{O}_K)$ at level I is defined by

$$\mathrm{H}(I) := \ker\big(\mathrm{H}(\mathcal{O}_K) \xrightarrow{\pi_I} \mathrm{H}(\mathcal{O}_K/I)\big),$$

where π_I is the reduction map modulo I. Any discrete subgroup of G containing some of these subgroups $\mathrm{H}(I)$ is called a *congruence subgroup of G*.

By Margulis' arithmeticity theorem (see [7, Chapter 5]), for $n \geq 2$ any irreducible lattice in $(\mathrm{PSL}_2(\mathbb{R}))^n$ is arithmetic. A conjecture of Serre, proved to be true in the nonuniform case, shows that any nonuniform lattice of $(\mathrm{PSL}_2(\mathbb{R}))^n$ is a congruence subgroup.

The coverings $M_I \to M_k$ are regular coverings because the subgroups $\Gamma(I)$ are normal subgroups of Γ. It is worth noting that in a sequence of nonregular congruence coverings of an arithmetic manifold the systole could grow slower than logarithmically with respect to the volume (see [5, Section 4.1]).

2.3 Upper bound for the systole growth of M_I

As was explained above, if Λ is any discrete group of isometries of $\mathbb{H}^2$ acting freely on $\mathbb{H}^2$, every hyperbolic element $\gamma \in \Lambda$ produces a noncontractible closed geodesic on $\mathbb{H}^2/\Lambda$. We can use this idea to see that the quotients M_I which we are interested in have closed geodesics, and subsequently we find a upper bound for $\mathrm{sys}_1(M_I)$.

We denote by $\mathrm{N}(I)$ the norm of an ideal $I \subset \mathcal{O}_k$, which is the cardinality of the quotient ring $\mathcal{O}_k/I$, and similarly $\mathrm{N}(r)$ denotes the field norm of an element r of the number field k.

Suppose $I \subset \mathcal{O}_k$ is an ideal with $\mathrm{N}(I) > 2$ and such that M_I is a riemannian manifold (see Corollary 3.3). The norm $\mathrm{N}(I)$ is a rational integer with $\mathrm{N}(I) \in I$, so if we take the matrix

$$B = \begin{pmatrix} 1-\mathrm{N}(I)^2 & \mathrm{N}(I) \\ -\mathrm{N}(I) & 1 \end{pmatrix},$$

then $B \in \Gamma(I)$ and $|\mathrm{tr}(\sigma_i(B))| > 2$ for any $i = 1, \dots, n$. This means that the matrices $\sigma_1(B) = \sigma_2(B) = \cdots = \sigma_n(B)$ are hyperbolic and if we take $\overline{\alpha}$ to be the only geodesic in $\mathbb{H}^2$ fixed by B, the curve $\overline{\beta} = \overline{\alpha} \times \cdots \times \overline{\alpha}$ is a geodesic in $(\mathbb{H}^2)^n$ that is fixed by $(\sigma_1(B), \dots, \sigma_n(B))$, and $\overline{\beta}$ projects to a noncontractible closed geodesic β in M_I. Note that this geodesic might not be the shortest one, so $\mathrm{sys}_1(M_I) \le \ell(\beta) = \sqrt{n}\,\ell_B$, where ℓ_B denotes the translation length of B along $\overline{\alpha}$.

We know that $2\cosh(\ell_B/2) = |\mathrm{tr}(B)| = \mathrm{N}(I)^2 - 2 < \mathrm{N}(I)^2$, and so

$$\mathrm{sys}_1(M_I) \le 4\sqrt{n}\log \mathrm{N}(I).$$

Now, as we will see in Section 4, there exists a constant C_k independent of I such that $[\Gamma : \Gamma(I)] \ge C_k\,\mathrm{N}(I)^3$ (Lemma 4.1), and then

$$\mathrm{sys}_1(M_I) \le \frac{4\sqrt{n}}{3}\log[\Gamma : \Gamma(I)] - \frac{4\sqrt{n}}{3}\log C_k. \tag{4}$$

This proves inequality (2) since $\mathrm{vol}(M_I) = [\Gamma : \Gamma(I)]\,\mathrm{vol}(M)$.

3 Distance estimate for congruence subgroups

In this section we will prove that the congruence subgroups $\Gamma(I)$ act freely on $(\mathbb{H}^2)^n$ when the norm of the ideal I is big enough and we will relate the length of closed geodesics in M_I to the norm of the ideal I. The first fact follows from Selberg's lemma [7, Section 4.8] but in our case the proof gives an explicit bound in terms of the norm of I. Some of the ideas are inspired by [6], where the authors studied the systole of compact arithmetic hyperbolic surfaces and 3–manifolds.

In this section, sometimes we will use the notation A or $(\sigma_1(A),\dots,\sigma_n(A))$ for the same element in Γ or its image in $(\mathrm{PSL}_2(\mathbb{R}))^n$ via the map Δ defined in Section 2.

For our purpose, it is convenient to express any element $A=\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)$ of Γ in the form

$$A=\begin{pmatrix} x_0+x_1 & x_2+x_3\\ x_2-x_3 & x_0-x_1\end{pmatrix},$$

where

$$x_0=\frac{a+d}{2},\quad x_1=\frac{a-d}{2},\quad x_2=\frac{b+c}{2},\quad x_3=\frac{b-c}{2}$$

are elements of the field K. We have $x_0^2-x_1^2-x_2^2+x_3^2=1$ and we write $y_0=x_0-1$. With this notation, if $I\subset\mathcal{O}_k$ is an ideal and $A\in\Gamma(I)$ then $2x_0-2\in I$ and $2x_i\in I$ for $i=1,2,3$. In terms of fractional ideals it means that y_0,x_1,x_2 and x_3 lie in $I/2$.

Lemma 3.1 *If $A\in\Gamma(I)$, then $y_0\in I^2/8$. In particular, if $y_0\neq 0$ then $|\mathrm{N}(y_0)|\geq \mathrm{N}(I)^2/8^n$.*

Proof We know that $A\in\Gamma(I)$ implies $x_0-1,x_1,x_2,x_3\in I/2$. Now, by replacing $x_0=1+y_0$ in the equation $x_0^2-x_1^2-x_2^2+x_3^2=1$ we obtain

$$2y_0=-y_0^2+x_1^2+x_2^2-x_3^2\in I^2/4.$$

Hence $y_0\in I^2/8$. □

Lemma 3.2 *If $A\in\Gamma(I)$ with $y_0\neq 0$ then $|\mathrm{tr}(\sigma_j(A))|\geq \mathrm{N}(I)^{2/n}/4-2$ for some $j\in\{1,\dots,n\}$.*

Proof By definition we have $\mathrm{N}(y_0)=\prod_{j=1}^n\sigma_j(y_0)$, so by Lemma 3.1, for some $j\in\{1,\dots,n\}$, we have $|\sigma_j(y_0)|\geq \mathrm{N}(I)^{2/n}/8$. Therefore

$$|\mathrm{tr}(\sigma_j(A))|=|2\sigma_j(x_0)|=|2\sigma_j(y_0)+2|\geq\frac{\mathrm{N}(I)^{2/n}}{4}-2. \qquad □$$

With this we can guarantee the riemannian structure for M_I:

Corollary 3.3 *For any ideal $I\subset\mathcal{O}_k$ with $\mathrm{N}(I)\geq 4^n$, the subgroup $\Gamma(I)$ acts freely on $(\mathbb{H}^2)^n$ and so $M_I=(\mathbb{H}^2)^n/\Gamma(I)$ admits a structure of a riemannian manifold with nonpositive sectional curvature.*

Proof The element $A=(\sigma_1(A),\dots,\sigma_n(A))\in\Gamma(I)$ has a fixed point on $(\mathbb{H}^2)^n$ if and only if $\sigma_i(A)$ has a fixed point in $\mathbb{H}^2$ for any $i=1,\dots,n$, but this happens if and only if $|\mathrm{tr}(\sigma_i(A))|<2$, which, by Lemma 3.2, is impossible if $\mathrm{N}(I)\geq 4^n$. □

Now observe that for $i=1,\dots,n$ and $A\in\Gamma$,

$$2|\sigma_i(y_0)|-2\leq|\mathrm{tr}(\sigma_i(A))|\leq 2+2|\sigma_i(y_0)|. \tag{5}$$

Proposition 3.4 *Let $I \subset \mathcal{O}_k$ be an ideal with $\mathrm{N}(I) \geq 40^{n/2}$ and $A \in \Gamma(I)$ with $y_0 \neq 0$. Then for any point $z = (z_1, \dots, z_n) \in (\mathbb{H}^2)^n$ we have*

$$d_{(\mathbb{H}^2)^n}(z, Az) \geq \frac{4}{\sqrt{n}} \log \mathrm{N}(I) - 2\sqrt{n} \log 40.$$

Proof By Lemma 3.2, $|\mathrm{tr}(\sigma_j(A))| \geq 8$ for some $j \in \{1, \dots, n\}$, hence we can subdivide our analysis into two different cases:

Case 1 $|\mathrm{tr}(\sigma_i(A))| \geq 8$ for any $i = 1, \dots, n$ In this case all of the matrices $\sigma_i(A)$ are hyperbolic and the right-hand side of (5) implies that $|\sigma_i(y_0)| \geq 3$ for $i = 1, \dots, n$.

Using (3), the left-hand side of (5), the fact that $|\sigma_i(y_0)| \geq 3$ for $i = 1, \dots, n$, the convexity of the function x^2 and Lemma 3.1 we obtain

$$\begin{aligned}
d_{(\mathbb{H}^2)^n}(z, Az) &= \sqrt{d^2_{\mathbb{H}^2}(z_1, \sigma_1(A)z_1) + \dots + d^2_{\mathbb{H}^2}(z_n, \sigma_n(A)z_n))} \\
&\geq 2\sqrt{\log^2(|\mathrm{tr}(\sigma_1(A))| - 1) + \dots + \log^2(|\mathrm{tr}(\sigma_n(A))| - 1)} \\
&\geq 2\sqrt{\log^2(2|\sigma_1(y_0)| - 3) + \dots + \log^2(2|\sigma_n(y_0)| - 3)} \\
&\geq 2\sqrt{\log^2|\sigma_1(y_0)| + \dots + \log^2|\sigma_n(y_0)|} \\
&\geq \frac{2}{\sqrt{n}}\big(\log|\sigma_1(y_0)| + \dots + \log|\sigma_n(y_0)|\big) \\
&= \frac{2}{\sqrt{n}} \log|\mathrm{N}(y_0)| \geq \frac{4}{\sqrt{n}} \log \mathrm{N}(I) - 2\sqrt{n} \log 8.
\end{aligned}$$

Case 2 There are exactly $k < n$ of the indices $1, \dots, n$ such that $|\mathrm{tr}(\sigma_j(A))| < 8$ Without loss of generality we assume that $|\mathrm{tr}(\sigma_j(A))| < 8$ for $j = 1, \dots, k$. By the left-hand side of (5), $|\sigma_j(y_0)| < 5$ for any such j and by Lemma 3.1 we have

$$\prod_{i=k+1}^{n} |\sigma_i(y_0)| = \frac{|\mathrm{N}(y_0)|}{\prod_{i=1}^{k} |\sigma_i(y_0)|} > \frac{1}{5^n \cdot 8^n} \mathrm{N}(I)^2.$$

Now, as $|\mathrm{tr}(\sigma_i(A))| \geq 8$ for $i = k+1, \dots, n$, for these indices $\sigma_i(A)$ is hyperbolic and $|\sigma_i(y_0)| \geq 3$ by the left-hand side of (5). By using (3) and the previous facts we obtain

$$\begin{aligned}
d_{(\mathbb{H}^2)^n}(z, Az) &= \sqrt{d^2_{\mathbb{H}^2}(z_1, \sigma_1(A)z_1) + \dots + d^2_{\mathbb{H}^2}(z_n, \sigma_n(A)z_n))} \\
&\geq \sqrt{d^2_{\mathbb{H}^2}(z_{k+1}, \sigma_{k+1}(A)z_{k+1}) + \dots + d^2_{\mathbb{H}^2}(z_n, \sigma_n(A)z_n))} \\
&\geq 2\sqrt{\log^2(|\mathrm{tr}(\sigma_{k+1}(A))| - 1) + \dots + \log^2(|\mathrm{tr}(\sigma_n(A))| - 1)} \\
&\geq 2\sqrt{\log^2(2|\sigma_{k+1}(y_0)| - 3) + \dots + \log^2(2|\sigma_n(y_0)| - 3)}
\end{aligned}$$

$$\begin{aligned}
&\geq 2\sqrt{\log^2|\sigma_{k+1}(y_0)| + \cdots + \log^2|\sigma_n(y_0)|} \\
&\geq \frac{2}{\sqrt{n-k}}\big(\log|\sigma_{k+1}(y_0)| + \cdots + \log|\sigma_n(y_0)|\big) \\
&= \frac{2}{\sqrt{n-k}} \log \prod_{i=k+1}^{n} |\sigma_i(y_0)| \geq \frac{4}{\sqrt{n}} \log \mathrm{N}(I) - 2\sqrt{n}\log 40.
\end{aligned}$$

In both cases we get

$$d_{(\mathbb{H}^2)^n}(z, Az) \geq \frac{4}{\sqrt{n}} \log(\mathrm{N}(I)) - 2\sqrt{n}\log(40). \qquad \square$$

Corollary 3.5 *For any ideal $I \subset \mathcal{O}_k$ with $\mathrm{N}(I) \geq 40^{n/2}$, the length of any noncontractible closed geodesic α in M_I satisfies*

$$\ell(\alpha) \geq \frac{4}{\sqrt{n}} \log \mathrm{N}(I) - 2\sqrt{n}\log 40.$$

Proof By Corollary 3.3, M_I is a riemannian manifold with the metric induced from $(\mathbb{H}^2)^n$. If we lift α to a geodesic $\widetilde{\alpha} = (\widetilde{\alpha}_1, \dots, \widetilde{\alpha}_n)$ in its universal cover $(\mathbb{H}^2)^n$ there is an element $A \in \Gamma(I)$ acting on $\widetilde{\alpha}$ as a translation and for any z in the graph of $\widetilde{\alpha}$ we have $\ell(\alpha) = d_{(\mathbb{H}^2)^n}(z, Az)$. Since α is noncontractible, $\widetilde{\alpha}$ is not a point, then for some $i \in \{1, \dots, n\}$ $\widetilde{\alpha}_i$ is a nontrivial geodesic in $\mathbb{H}^2$, and so $\sigma_i(A)$ acts on it as a translation. This implies that $\sigma_i(A)$ is hyperbolic and, in particular, $|\mathrm{tr}(A)| \neq 2$. Since $|\mathrm{tr}(A)| \neq 2$ implies $y_0 \neq 0$, the result now follows from Proposition 3.4. $\square$

4 Proof of the main results

To finish the proofs of the theorems we need to find uniform bounds for the quotient $[\Gamma : \Gamma(I)]/\mathrm{N}(I)^3$, for ideals $I \subset \mathcal{O}_k$ with norm sufficiently large.

Lemma 4.1 *For almost any ideal $I \subset \mathcal{O}_k$ we have*

(6)
$$\zeta_k(2)^{-1}\,\mathrm{N}(I)^3 \leq [\Gamma : \Gamma(I)] < \mathrm{N}(I)^3,$$

where ζ_k denotes the Dedekind zeta function of k.

Proof A well-known corollary of the strong approximation theorem (see Theorem 7.15 of [8]) implies that for almost all ideals $I \subset \mathcal{O}_k$ the reduction map

$$\mathrm{SL}_2(\mathcal{O}_k) \xrightarrow{\pi_I} \mathrm{SL}_2(\mathcal{O}_k/I)$$

is surjective. For those ideals the index $[\Gamma : \Gamma(I)]$ is equal to the cardinality of $\mathrm{SL}_2(\mathcal{O}_k/I)$, which is given by the formula

$$\mathrm{N}(I)^3 \prod_{\mathfrak{p}|I} \Big(1 - \frac{1}{\mathrm{N}(\mathfrak{p})^2}\Big).$$

From this the right-hand side of inequality (6) follows easily. On the other hand, the product formula for the Dedekind zeta function of k says that

$$\zeta_k(2) = \prod_{\mathfrak{p}\subset\mathcal{O}_k} \frac{1}{1-\mathrm{N}(\mathfrak{p})^{-2}} \geq \prod_{\mathfrak{p}|I} \frac{1}{1-\mathrm{N}(\mathfrak{p})^{-2}}.$$

This proves the second inequality. □

Theorem 4.2 *Let k be a totally real number field of degree n and $\mathcal{O}_k$ be the ring of integers of k. Any sequence of ideals in $\mathcal{O}_k$ with $\mathrm{N}(I) \to \infty$ eventually satisfies*

$$\mathrm{sys}_1(M_I) \geq \frac{4}{3\sqrt{n}} \log(\mathrm{vol}(M_I)) - c,$$

where $\Gamma(I)$ is the principal congruence subgroup of $\Gamma = \mathrm{PSL}_2(\mathcal{O}_k)$ at level I, $M_I = (\mathbb{H}^2)^n/\Gamma(I)$ and c is a constant independent of I.

Proof For any ideal I with $\mathrm{N}(I) \geq 40^{n/2}$, Corollary 3.3 implies that M_I is a riemannian manifold with the metric induced by the product metric on $(\mathbb{H}^2)^n$. Now, by Corollary 3.5 and Lemma 4.1, we conclude that

$$\mathrm{sys}_1(M_I) \geq \frac{4}{3\sqrt{n}} \log[\Gamma : \Gamma(I)] - 2\sqrt{n}\log 40$$

when $\mathrm{N}(I) \to \infty$. □

To finish, we prove that among congruence coverings of Hilbert modular varieties the constant $4/(3\sqrt{n})$ in the growth of the systole in general cannot be improved to any $\gamma > 4/(3\sqrt{n})$.

Theorem 4.3 *Let k be a totally real number field of degree n and $\mathcal{O}_k$ be the ring of integers of k. Then there exists a sequence of ideals in $\mathcal{O}_k$ with $\mathrm{N}(I) \to \infty$ such that*

$$\mathrm{sys}_1(M_I) \leq \frac{4}{3\sqrt{n}} \log(\mathrm{vol}(M_I)) + c_1,$$

where $\Gamma(I)$ is the principal congruence subgroup of $\Gamma = \mathrm{PSL}_2(\mathcal{O}_k)$ at level I, $M_I = (\mathbb{H}^2)^n/\Gamma(I)$ and c_1 is a constant independent of M_I.

Proof Let p be a rational integer and consider the ideal $I_p = p\mathcal{O}_k$ in $\mathcal{O}_k$. Since $\mathrm{N}(I_p) = p^n$, by following the same argument as in Section 2.3 with the matrix

$$B = \begin{pmatrix} 1-p^2 & p \\ -p & 1 \end{pmatrix},$$

we obtain that $\mathrm{sys}_1(M_{I_p}) \le 4\sqrt{n}\log(p)$ when p is large enough. Therefore, Lemma 4.1 implies that

$$\mathrm{sys}_1(M_{I_p}) \le \frac{4}{3\sqrt{n}}\log[\Gamma : \Gamma(I_p)] + \frac{4}{3\sqrt{n}}\log\zeta_k(2)$$

when $p \to \infty$, and then we obtain the result with

$$c_1 = \frac{4}{3\sqrt{n}}\log\frac{\zeta_k(2)}{\mathrm{vol}(M_k)},$$

where $M_k = (\mathbb{H}^2)^n/\Gamma$. □

References

[1] **A F Beardon**, *The geometry of discrete groups*, Graduate Texts in Mathematics 91, Springer (1995) MR Corrected reprint

[2] **P Buser**, **P Sarnak**, *On the period matrix of a Riemann surface of large genus*, Invent. Math. 117 (1994) 27–56 MR

[3] **E Freitag**, *Hilbert modular forms*, Springer (1990) MR

[4] **M Gromov**, *Systoles and intersystolic inequalities*, from "Actes de la Table Ronde de Géométrie Différentielle" (A L Besse, editor), Sémin. Congr. 1, Soc. Math. France, Paris (1996) 291–362 MR

[5] **L Guth**, **A Lubotzky**, *Quantum error correcting codes and 4–dimensional arithmetic hyperbolic manifolds*, J. Math. Phys. 55 (2014) art. id. 082202, 13 pages MR

[6] **M G Katz**, **M Schaps**, **U Vishne**, *Logarithmic growth of systole of arithmetic Riemann surfaces along congruence subgroups*, J. Differential Geom. 76 (2007) 399–422 MR

[7] **D W Morris**, *Introduction to arithmetic groups*, Deductive Press (2015) MR

[8] **V Platonov**, **A Rapinchuk**, *Algebraic groups and number theory*, Pure and Applied Mathematics 139, Academic Press, Boston (1994) MR

Instituto de Matemática Pura e Aplicada
Rio de Janeiro, Brazil

plinio@impa.br

http://www.impa.br/~plinio

Received: 9 June 2016 Revised: 6 April 2017

msp

Algebraic & Geometric Topology 17 (2017) 2763–2806

Stable Postnikov data of Picard 2–categories

NICK GURSKI
NILES JOHNSON
ANGÉLICA M OSORNO
MARC STEPHAN

Picard 2–categories are symmetric monoidal 2–categories with invertible 0–, 1– and 2–cells. The classifying space of a Picard 2–category $\mathcal{D}$ is an infinite loop space, the zeroth space of the K–theory spectrum $K\mathcal{D}$. This spectrum has stable homotopy groups concentrated in levels 0, 1 and 2. We describe part of the Postnikov data of $K\mathcal{D}$ in terms of categorical structure. We use this to show that there is no strict skeletal Picard 2–category whose K–theory realizes the 2–truncation of the sphere spectrum. As part of the proof, we construct a categorical suspension, producing a Picard 2–category ΣC from a Picard 1–category C, and show that it commutes with K–theory, in that $K\Sigma C$ is stably equivalent to ΣKC.

55S45; 18C20, 18D05, 19D23, 55P42

1 Introduction

This paper is part of a larger effort to refine and expand the theory of algebraic models for homotopical data, especially that of *stable* homotopy theory. Such modeling has been of interest since May [46] and Segal [53] gave K–theory functors which build connective spectra from symmetric monoidal categories. Moreover, Thomason [57] proved that symmetric monoidal categories have a homotopy theory which is equivalent to that of all connective spectra.

Our current work is concerned with constructing models for stable homotopy 2–types using symmetric monoidal 2–categories. Preliminary foundations for this appear, for example, in Gurski and Osorno [31], Gurski, Johnson and Osorno [29], Johnson and Osorno [33] and Schommer-Pries [52]. In forthcoming work [30], we prove that all stable homotopy 2–types are modeled by a special kind of symmetric monoidal 2–categories, which we describe below and call *strict Picard 2–categories*.

Research leading to the methods in [30] has shown that the most difficult aspect of this problem is replacing a symmetric monoidal 2–category modeling an arbitrary

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2763

connective spectrum (see Gurski, Johnson and Osorno [29]) by a strict Picard 2–category with the same stable homotopy 2–type. This paper can then be interpreted as setting a minimum level of complexity for such a categorical model of stable homotopy 2–types. Furthermore, we intend to construct the Postnikov tower for a stable homotopy 2–type entirely within a categorical context, and the results here give some guidance as to the assumptions we can make on those Postnikov towers.

This paper has three essential goals. First, we explicitly describe part of the Postnikov tower for strict Picard 2–categories. Second, and of independent interest, we show that the K–theory functor commutes with suspension up to stable equivalence. This allows us to bootstrap previous results on Picard 1–categories to give algebraic formulas for the two nontrivial Postnikov layers of a Picard 2–category. Third, we combine these to show that, while strict Picard 2–categories are expected to model all stable homotopy 2–types, strict and *skeletal* Picard 2–categories cannot. We prove that there is no strict and skeletal Picard 2–category modeling the truncation of the sphere spectrum.

1.1 Background and motivation

Homotopical invariants, and therefore homotopy *types*, often have a natural interpretation as categorical structures. The fundamental groupoid is a complete invariant for homotopy 1–types, while pointed connected homotopy 2–types are characterized by their associated crossed module or $\mathcal{C}at^1$–group structure; see Brown and Spencer [11], Conduché [16], Loday [43], MacLane and Whitehead [44] and Whitehead [58]. Such characterizations provide the low-dimensional cases of Grothendieck's *homotopy hypothesis* [23].

Homotopy hypothesis *There is an equivalence of homotopy theories between $\mathcal{G}pd^n$, weak n–groupoids equipped with categorical equivalences, and $\mathcal{T}op^n$, homotopy n–types equipped with weak homotopy equivalences.*

Restricting attention to stable phenomena, we replace homotopy n–types with stable homotopy n–types: spectra X such that $\pi_i X = 0$ unless $0 \leq i \leq n$. On the categorical side, we take a cue from May [46] and Thomason [57] and replace n–groupoids with a grouplike, symmetric monoidal version that we call Picard n–categories. The stable version of the homotopy hypothesis is then the following.

Stable homotopy hypothesis *There is an equivalence of homotopy theories between $\mathcal{P}ic^n$, Picard n–categories equipped with categorical equivalences, and $\mathcal{S}p_0^n$, stable homotopy n–types equipped with stable equivalences.*

For $n = 0$, $\mathcal{P}ic^0$ is the category of abelian groups $\mathcal{A}b$ with weak equivalences given by group isomorphisms. It is equivalent to the homotopy theory of Eilenberg–Mac Lane spectra. For $n = 1$, a proof of the stable homotopy hypothesis appears in Johnson and Osorno [33], and a proof for $n = 2$ will appear in [30]. The advantage of being able to work with categorical weak equivalences is that the maps in the homotopy category between two stable 2–types modeled by strict Picard 2–categories are realized by symmetric monoidal pseudofunctors between the two strict Picard 2–categories, instead of having to use general zigzags. In fact, as will appear in [30], the set of homotopy classes between two strict Picard 2–categories $\mathcal{D}$ and $\mathcal{D}'$ is the quotient of the set of symmetric monoidal pseudofunctors $\mathcal{D} \to \mathcal{D}'$ by the equivalence relation $F \sim G$ if there exists a pseudonatural transformation $F \Rightarrow G$.

More than a proof of the stable homotopy hypothesis, we seek a complete dictionary translating between stable homotopical invariants and the algebra of Picard n–categories. The search for such a dictionary motivated three questions that lie at the heart of this paper. First, how can we express invariants of stable homotopy types in algebraic terms? Second, how can we construct stable homotopy types of interest, such as Postnikov truncations of the sphere spectrum, from a collection of invariants? Third, can we make simplifying assumptions, such as strict inverses, about Picard n–categories without losing homotopical information?

The results in this paper provide key steps toward answering these questions. In particular, we characterize the three stable homotopy groups of a strict Picard 2–category in terms of equivalence classes of objects, isomorphism classes of 1–cells and 2–cells, respectively, and deduce that a map of Picard 2–categories is a stable equivalence if and only if it is a categorical equivalence (Proposition 3.3). This fact is used in [30] to prove the stable homotopy hypothesis for $n = 2$.

1.2 Postnikov invariants and strict skeletalization

It has long been folklore that the symmetry in a Picard 1–category should model the bottom k–invariant, k_0. Along with a proof of the stable homotopy hypothesis in dimension 1, this folklore result was established in Johnson and Osorno [33]. This shows that a Picard 1–category is characterized by exactly three pieces of data: an abelian group π_0 of isomorphism classes of objects, an abelian group π_1 of automorphisms of the unit object, and a group homomorphism $k_0\colon \pi_0 \otimes \mathbb{Z}/2 \to \pi_1$ (ie a stable quadratic map from π_0 to π_1) corresponding to the symmetry. Such a characterization is implied by the following result.

Theorem 1.1 [33, Theorem 2.2] *Every Picard category is equivalent to one which is both strict and skeletal.*

We call this phenomenon *strict skeletalization*. This theorem is quite surprising given that it is false without the symmetry. Indeed, Baez and Lauda [3] give a good account of the failure of strict skeletalization for 2–groups (the nonsymmetric version of Picard 1–categories), and how it leads to a cohomological classification for 2–groups. Johnson and Osorno [33] show, in effect, that the relevant obstructions are *unstable* phenomena which become trivial upon stabilization.

When we turn to the question of building models for specific homotopy types, the strict and skeletal ones are the simplest: given a stable 1–type X, a strict and skeletal model will have objects equal to the elements of $\pi_0 X$ and automorphisms of every object equal to the elements of $\pi_1 X$, with no morphisms between distinct objects. All that then remains is to define the correct symmetry isomorphisms, and these are determined entirely by the map k_0.

As an example, a strict and skeletal model for the 1–truncation of the sphere spectrum has objects the integers, each hom-set of automorphisms the integers mod 2, and k_0 given by the identity map on $\mathbb{Z}/2$ corresponding to the fact that the generating object 1 has a nontrivial symmetry with itself. One might be tempted to build a strict and skeletal model for the 2–type of the sphere spectrum (the authors here certainly were, and such an idea also appears in Bartlett [4, Example 5.2]). But here we prove that this is not possible for the sphere spectrum, and in fact a large class of stable 2–types.

Theorem 1.2 (Theorem 3.14) *Let $\mathcal{D}$ be a strict skeletal Picard 2–category with k_0 surjective. Then the 0–connected cover of $K\mathcal{D}$ splits as a product of Eilenberg–Mac Lane spectra. In particular, there is no strict and skeletal model of the 2–truncation of the sphere spectrum.*

Our proof of this theorem identifies both the bottom k–invariant k_0 and the first Postnikov layer $k_1 i_1$ (see Section 3) of $K\mathcal{D}$ explicitly, using the symmetric monoidal structure for any strict Picard 2–category $\mathcal{D}$. In addition, we provide a categorical model of the 1–truncation of $K\mathcal{D}$ in Proposition 3.6. This provides data which is necessary, although not sufficient, for a classification of stable 2–types akin to the cohomological classification in Baez and Lauda [3]. Remaining data, to be studied in future work, must describe the connection of π_2 with π_0. For instance, stable 2–types X with trivial π_1 are determined by a map $H(\pi_0 X) \to \Sigma^3 H(\pi_2 X)$ in the stable homotopy category. For general X, the third cohomology group of the 1–truncation of X with coefficients in $\pi_2 X$ has to be calculated. In the spectral sequence associated to the stable Postnikov tower of X (see Greenlees and May [22, Appendix B]), the connection between π_0 and π_2 becomes apparent in the form of a d_3 differential.

In addition to clarifying the relationship between Postnikov invariants and the property of being skeletal, Theorem 1.2 suggests a direction for future work developing a 2–categorical structure that adequately captures the homotopy theory of stable 2–types. Such structure ought to be more specific than that of strict Picard 2–categories but more general than strict, skeletal Picard 2–categories. Interpretations of this structure which are conceptual (in terms of other categorical structures) and computational (in terms of homotopical or homological invariants, say) will shed light on both the categorical and topological theory.

1.3 Categorical suspension

In order to give a formula for the first Postnikov layer, we must show that K–theory functors are compatible with suspension. More precisely, given a strict monoidal category C, one can construct a one-object 2–category ΣC, where the category of morphisms is given by C, with composition defined using the monoidal structure. Further, if C is a permutative category then ΣC is naturally a symmetric monoidal 2–category, with the monoidal structure also defined using the structure of C. Unstably, it is known that this process produces a categorical delooping: if C is a strict monoidal category with invertible objects, the classifying space $B(\Sigma C)$ is a delooping of BC; see Carrasco, Cegarra and Garzón [12] and Jardine [32]. We prove the stable analogue.

Theorem 1.3 (Theorem 3.11) *For any permutative category C, the spectra $K(\Sigma C)$ and $\Sigma(KC)$ are stably equivalent.*

Here $K(-)$ denotes both the K–theory spectrum associated to a symmetric monoidal category — see May [46] and Segal [53] — and the K–theory spectrum associated to a symmetric monoidal 2–category; see Gurski and Osorno [31] and Gurski, Johnson and Osorno [29].

This theorem serves at least three purposes beyond being a necessary calculation tool. A first step in the proof is Corollary 2.35, which shows that the categories of permutative categories and of one-object permutative Gray monoids are equivalent; this is a strong version of one case of the Baez–Dolan stabilization hypothesis [2], stronger than the usual proofs in low dimensions; see Cheng and Gurski [13; 14; 15]. The second purpose of this theorem is to justify, from a homotopical perspective, the definition of permutative Gray monoid, the construction of the K–theory spectrum, and the categorical suspension functor. The suspension functor of spectra and the K–theory spectrum of a permutative category are both central features of stable homotopy theory, so any generalization of the latter should respect the former. A final purpose of this theorem will appear in future work, namely in the categorical construction of stable Postnikov towers. Suspension spectra necessarily appear in these towers, and Theorem 3.11 and Corollary 2.35 together allow us to replicate these features of a Postnikov tower entirely within the world of symmetric monoidal 2–categories.

1.4 Relation to supersymmetry and supercohomology

The theory of Picard 2–categories informs recent work in mathematical physics related to higher supergeometry — see Kapranov [36] — and invertible topological field theories; see Freed [19]. Kapranov [36] links the $\mathbb{Z}$–graded Koszul sign rule appearing in supergeometry to the 1–truncation of the sphere spectrum. He describes how higher supersymmetry is governed by higher truncations of the sphere spectrum, which one expects to be modeled by the free Picard n–category on a single object. Likewise, Freed [19] describes examples using the Picard bicategory of complex invertible super algebras related to twisted K–theory; see Freed, Hopkins and Teleman [20].

The failure of strict skeletalization for a categorical model of the 2–truncation of the sphere spectrum shows that already for $n = 2$ capturing the full higher supersymmetry in algebraic terms is more complicated than one might expect.

Furthermore, it would be interesting to relate examples appearing in physics literature about topological phases of matter — see Gu and Wen [24] and Bhardwaj, Gaiotto and Kapustin [6] — to cohomology with coefficients in Picard n–categories. The *supercohomology* in Gu and Wen [24] is assembled from two different classical cohomology groups of a classifying space BG with a nontrivial symmetry. One expects that this supercohomology can be expressed as the cohomology of BG with coefficients in a Picard 1–category, and similarly, for the extension of this supercohomology in [6] as cohomology with coefficients in a Picard 2–category.

Outline

In Section 2 we sketch the basic theory of Picard categories and Picard 2–categories. This includes some background to fix notation and some recent results of Gurski, Johnson and Osorno [29] about symmetric monoidal 2–categories. In Section 3 we develop algebraic models for some of the Postnikov data of the spectrum associated to a Picard 2–category, giving formulas for the two nontrivial layers in terms of the symmetric monoidal structure. This section closes with applications showing that strict skeletal Picard 2–categories cannot model all stable 2–types. Section 4 establishes formal strictification results for 2–categorical diagrams using 2–monad theory. We use those results in Section 5 to prove that the K–theory functor commutes with suspension.

Acknowledgements The authors gratefully acknowledge travel funding and hospitality from The Ohio State University Mathematics Research Institute. Gurski was supported by EPSRC EP/K007343/1, and Stephan by SNSF grants 158932 and 151967. This material is partially based on work supported by the National Science Foundation under Grant No 0932078 000 while Osorno and Stephan were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2014 semester. This work was partially supported by a grant from the Simons Foundation (#359449, Osorno).

The authors thank Jack Morava for encouraging them to explore the relations to supersymmetry. They also thank the referee, whose comments helped improve the exposition.

2 Picard categories and Picard 2–categories

This section introduces the primary categorical structures of interest which we call Picard 2–categories, as well as the particularly relevant variant of strict skeletal Picard 2–categories. Note that we use the term 2–category in its standard sense [38], and in particular all composition laws are strictly associative and unital.

Notation 2.1 We let $\mathcal{C}at$ denote the category of categories and functors, and let $2\mathcal{C}at$ denote the category of 2–categories and 2–functors. Note that these are both 1–categories.

Notation 2.2 We let $\mathcal{C}at_2$ denote the 2–category of categories, functors, and natural transformations. This can be thought of as the 2–category of categories enriched in $\mathcal{S}et$. Similarly, we let $2\mathcal{C}at_2$ denote the 2–category of 2–categories, 2–functors and 2–natural transformations; the 2–category of categories enriched in $\mathcal{C}at$.

2.1 Picard categories

We will begin by introducing all of the 1–categorical notions before going on to discuss their 2–categorical analogues. First we recall the notion of a permutative category (ie symmetric strict monoidal category); the particular form of this definition allows an easy generalization to structures on 2–categories.

Definition 2.3 A *permutative category* C consists of a strict monoidal category $(C, \oplus, e)$ together with a natural isomorphism

$$\begin{array}{ccc} C \times C & \xrightarrow{\ \tau\ } & C \times C \\ & {\searrow_{\oplus}} \ \overset{\beta}{\Nearrow} \ {\swarrow_{\oplus}} & \\ & C & \end{array}$$

where $\tau\colon C\times C\to C\times C$ is the symmetry isomorphism in $\mathcal{C}at$, such that the following axioms hold for all objects x, y and z of C:

- $\beta_{y,x}\beta_{x,y}=\mathrm{id}_{x\oplus y}$.
- $\beta_{e,x}=\mathrm{id}_x=\beta_{x,e}$.
- $\beta_{x,y\oplus z}=(y\oplus\beta_{x,z})\circ(\beta_{x,y}\oplus z)$.

Remark 2.4 We will sometimes say that a symmetric monoidal structure on a category is strict if its underlying monoidal structure is. Note that this does not imply that the symmetry is the identity, even though the other coherence isomorphisms are. Thus a permutative category is nothing more than a strict symmetric monoidal category.

Notation 2.5 Let $\mathcal{P}erm\mathcal{C}at$ denote the category of permutative categories and symmetric, strict monoidal functors between them.

Next we require a notion of invertibility for the objects in a symmetric monoidal category.

Definition 2.6 Let $(C,\oplus,e)$ be a monoidal category. An object x is *invertible* if there exists an object y together with isomorphisms $x\oplus y\cong e$ and $y\oplus x\cong e$.

Definition 2.7 A *Picard category* is a symmetric monoidal category in which all of the objects and morphisms are invertible.

The terminology comes from the following example.

Example 2.8 Let R be a commutative ring, and consider the symmetric monoidal category of R–modules. We have the subcategory $\mathcal{P}ic R$ of invertible R–modules and isomorphisms between them. The set of isomorphism classes of objects of $\mathcal{P}ic R$ is the classical Picard group of R.

Remark 2.9 If we drop the symmetric structure in Definition 2.7 above, we get the notion of what is both called a categorical group [34] or a 2–group [3]. These are equivalent to crossed modules [58; 43], and hence are a model for pointed connected homotopy 2–types (ie spaces X for which $\pi_i(X)=0$ unless $i=1,2$).

One should consider Picard categories as a categorified version of abelian groups. Just as abelian groups model the homotopy theory of spectra with trivial homotopy groups aside from π_0, Picard categories do the same for spectra with trivial homotopy groups aside from π_0 and π_1.

Theorem 2.10 [33, Theorem 1.5] *There is an equivalence of homotopy theories between the category of Picard categories, $\mathcal{P}ic^1$, equipped with categorical equivalences, and the category of stable 1–types, $\mathcal{S}p_0^1$, equipped with stable equivalences.*

Forthcoming work [30] proves the 2–dimensional analogue of Theorem 2.10. This requires a theory of Picard 2–categories, which began in [29] and motivated the work of the current paper. We now turn to such theory.

2.2 Picard 2–categories

To give the correct 2–categorical version of Picard categories, we must first describe the analogue of a mere strict monoidal category: such a structure is called a Gray monoid. It is most succinctly defined using the Gray tensor product of 2–categories, written $\mathcal{A}\otimes\mathcal{B}$ for a pair of 2–categories $\mathcal{A}$ and $\mathcal{B}$. We will not give the full definition of $\otimes$ here (see [29; 27; 8; 9]) but instead give the reader the basic idea. The objects of $\mathcal{A}\otimes\mathcal{B}$ are tensors $a\otimes b$ for $a\in\mathcal{A}$ and $b\in\mathcal{B}$, but the 1–cells are *not* tensors of 1–cells as one would find in the cartesian product. Instead they are generated under composition by 1–cells $f\otimes 1$ and $1\otimes g$ for $f\colon a\to a'$ a 1–cell in $\mathcal{A}$ and $g\colon b\to b'$ a 1–cell in $\mathcal{B}$. These different kinds of generating 1–cells do not commute with each other strictly, but instead up to specified isomorphism 2–cells

$$\Sigma_{f,g}\colon (f\otimes 1)\circ(1\otimes g)\cong(1\otimes g)\circ(f\otimes 1),$$

which obey appropriate naturality and bilinearity axioms. We call these Σ the *Gray structure* 2*–cells*. The 2–cells of $\mathcal{A}\otimes\mathcal{B}$ are defined similarly, generated by $\alpha\otimes 1$, $1\otimes\beta$ and the $\Sigma_{f,g}$. The function $(\mathcal{A},\mathcal{B})\mapsto\mathcal{A}\otimes\mathcal{B}$ is the object part of a functor of categories

$$2\mathcal{C}at\times 2\mathcal{C}at\to 2\mathcal{C}at,$$

which is the tensor product for a symmetric monoidal structure on $2\mathcal{C}at$ with unit the terminal 2–category.

Definition 2.11 A *Gray monoid* is a monoid object $(\mathcal{D},\oplus,e)$ in the monoidal category $(2\mathcal{C}at,\otimes)$.

Remark 2.12 By the coherence theorem for monoidal bicategories [21; 27], every monoidal bicategory is equivalent (in the appropriate sense) to a Gray monoid. There is a stricter notion, namely that of a monoid object in $(2\mathcal{C}at,\times)$, but a general monoidal bicategory will not be equivalent to one of these.

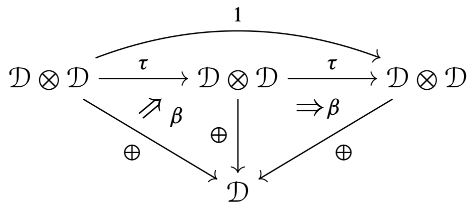
We now turn to the symmetry.

Definition 2.13 A *permutative Gray monoid* $\mathcal{D}$ consists of a Gray monoid $(\mathcal{D},\oplus,e)$ together with a 2–natural isomorphism

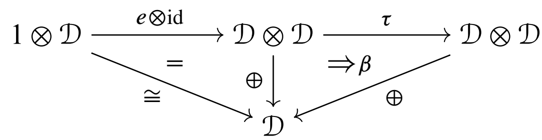
$$\begin{array}{ccc} \mathcal{D}\otimes\mathcal{D} & \xrightarrow{\tau} & \mathcal{D}\otimes\mathcal{D} \\ & {\scriptstyle\oplus}\searrow \;\; \Nearrow\beta \;\; \swarrow{\scriptstyle\oplus} & \\ & \mathcal{D} & \end{array}$$

where $\tau\colon \mathcal{D} \otimes \mathcal{D} \to \mathcal{D} \otimes \mathcal{D}$ is the symmetry isomorphism in $2\mathcal{C}at$ for the Gray tensor product, such that the following axioms hold:

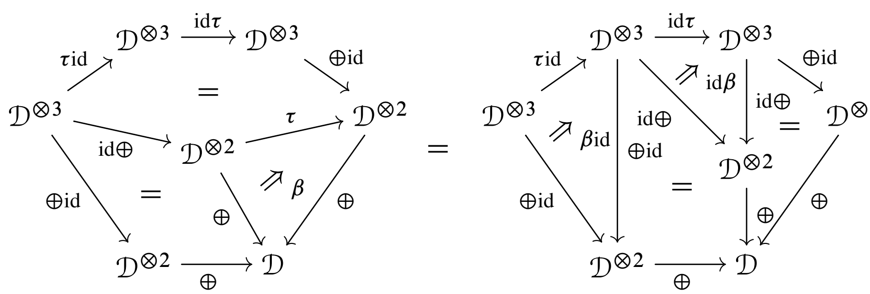
- The following pasting diagram is equal to the identity 2–natural transformation for the 2–functor $\oplus$:



- The following pasting diagram is equal to the identity 2–natural transformation for the canonical isomorphism $1 \otimes \mathcal{D} \cong \mathcal{D}$:



- The following equality of pasting diagrams holds, where we have abbreviated the tensor product to concatenation when labeling 1– or 2–cells:



Remark 2.14 A symmetric monoidal 2–category is a symmetric monoidal bicategory (see [29] for a sketch or [49] for full details) in which the underlying bicategory is a 2–category. Every symmetric monoidal bicategory is equivalent as such to a symmetric monoidal 2–category by strictifying the underlying bicategory and transporting the structure as in [26]. A deeper result is that every symmetric monoidal bicategory is equivalent as such to a permutative Gray monoid; this is explained fully in [29], making use of [52].

Notation 2.15 For convenience and readability, we use the following notational conventions for cells in a Gray monoid $\mathcal{D}$:

- For objects, we may use concatenation instead of explicitly indicating the monoidal product.

- For an object b and a 1–cell $f\colon a \to a'$, we denote by fb the 1–cell in $\mathcal{D}$ which is the image under $\oplus$ of $f \otimes 1\colon a \otimes b \to a' \otimes b$ in $\mathcal{D} \otimes \mathcal{D}$. We use similar notation for multiplication on the other side, and for 2–cells.
- We let $\Sigma_{f,g}$ also denote the image in $\mathcal{D}$ of the Gray structure 2–cells under $\oplus$,

$$\Sigma_{f,g}\colon (fb') \circ (ag) \cong (a'g) \circ (fb).$$

Notation 2.16 Let $\mathcal{P}erm\mathcal{G}ray\mathcal{M}on$ denote the category of permutative Gray monoids and strict symmetric monoidal 2–functors between them.

We are actually interested in permutative Gray monoids which model stable homotopy 2–types, and we therefore restrict to those in which all the cells are invertible. We begin by defining invertibility in a Gray monoid, then the notion of a Picard 2–category, and finish with that of a strict skeletal Picard 2–category.

Definition 2.17 Let $(\mathcal{D}, \oplus, e)$ be a Gray monoid.

(1) A 2–cell of $\mathcal{D}$ is invertible if it has an inverse in the usual sense.

(2) A 1–cell $f\colon x \to y$ is invertible if there exists a 1–cell $g\colon y \to x$ together with invertible 2–cells $g \circ f \cong \mathrm{id}_x$ and $f \circ g \cong \mathrm{id}_y$. In other words, f is invertible if it is an internal equivalence (denoted with the $\simeq$ symbol) in $\mathcal{D}$.

(3) An object x of $\mathcal{D}$ is invertible if there exists another object y together with invertible 1–cells $x \oplus y \simeq e$ and $y \oplus x \simeq e$.

Remark 2.18 The above definition actually used none of the special structure of a Gray monoid that is not also present in a more general monoidal bicategory.

Definition 2.19 A *Picard 2–category* is a symmetric monoidal 2–category (see Remark 2.14) in which all of the objects, 1–cells, and 2–cells are invertible. A *strict Picard 2–category* is a permutative Gray monoid which is a Picard 2–category.

Remark 2.20 The definition of a strict Picard 2–category does not require that cells be invertible in the strict sense, ie having inverses on the nose rather than up to mediating higher cells. It only requires that the underlying symmetric monoidal structure is strict in the sense of being a permutative Gray monoid.

Definition 2.21 A 2–category $\mathcal{A}$ is *skeletal* if the following condition holds: whenever there exists an invertible 1–cell $f\colon x \simeq y$, then $x = y$.

Remark 2.22 This definition might more accurately be named *skeletal on objects*, as one could impose a further condition of being skeletal on 1–cells as well. We have no need of this further condition, and so we work with this less restrictive notion of a skeletal 2–category. It is also important to remember that, in the definition above, the invertible 1–cell f need not be the identity 1–cell. The slogan is that "every equivalence is an autoequivalence": an object is allowed to have many nonidentity autoequivalences, and there can be 1–cells between different objects as long as they are not equivalences.

Definition 2.23 A *strict skeletal Picard 2–category* is a strict Picard 2–category whose underlying 2–category is skeletal.

2.3 Two adjunctions

Our goal in this subsection is to present two different adjunctions between strict Picard categories and strict Picard 2–categories. While we focus on the categorical algebra here, later we will give each adjunction a homotopical interpretation. The unit of the first adjunction will categorically model Postnikov 1–truncation (Proposition 3.6), universally making π_2 zero, while the counit of the second will categorically model the 0–connected cover (Proposition 3.10).

Recall that for any category C, we have its set of path components, denoted by $\pi_0 C$; these are given by the path components of the nerve of C, or equivalently by quotienting the set of objects by the equivalence relation generated by $x \sim y$ if there exists an arrow $x \to y$. This is the object part of a functor $\pi_0\colon \mathcal{C}at \to \mathcal{S}et$, and it is easy to verify that this functor preserves finite products. It is also left adjoint to the functor $d\colon \mathcal{S}et \to \mathcal{C}at$ which sends a set S to the discrete category with the same set of objects. Being a right adjoint, d preserves all products. The counit $\pi_0 \circ d \Rightarrow \mathrm{id}$ is the identity, and the unit $\mathrm{id} \Rightarrow d \circ \pi_0$ is the quotient functor $C \to d\pi_0 C$ sending every object to its path component and every morphism to the identity. Since d and π_0 preserve products, by applying them to hom-objects they induce change of enrichment functors d_* and $(\pi_0)_*$, respectively. We obtain the following result.

Lemma 2.24 *The adjunction $\pi_0 \dashv d$ lifts to a 2–adjunction*

$$2\mathcal{C}at_2 \underset{d_*}{\overset{(\pi_0)_*}{\rightleftarrows}} \mathcal{C}at_2, \quad \perp$$

Notation 2.25 We will write the functor $(\pi_0)_*$ as $\mathcal{D} \mapsto \mathcal{D}_1$ or $(-)_1$ to lighten the notation. This anticipates the homotopical interpretation in Proposition 3.6. Furthermore, we will write d_* as d; it will be clear from context which functor we are using.

Lemma 2.26 *The functor $(-)_1$ is strong symmetric monoidal $(2\mathcal{C}at, \otimes) \to (\mathcal{C}at, \times)$. The functor d is lax symmetric monoidal $(\mathcal{C}at, \times) \to (2\mathcal{C}at, \otimes)$.*

Proof The second statement follows from the first by doctrinal adjunction [37]. For the first, one begins by checking that

$$\mathcal{D}_1 \times \mathcal{E}_1 \cong (\mathcal{D} \otimes \mathcal{E})_1;$$

this is a simple calculation using the definition of $\otimes$ that we leave to the reader. If we let I denote the terminal 2–category, the unit for $\otimes$, then I_1 is the terminal category, so $(-)_1$ preserves units up to (unique) isomorphism. It is then easy to check that these isomorphisms interact with the associativity, unit, and symmetry isomorphisms to give a strong symmetric monoidal functor. □

Remark 2.27 It is useful to point out that if A and B are categories, then the comparison 2–functor

$$\chi_{A,B}\colon dA \otimes dB \to d(A \times B)$$

is the 2–functor which quotients all the 2–cells $\Sigma_{f,g}$ to be the identity. In view of the adjunction in Lemma 2.24, the 2–functor $\chi_{A,B}$ can be identified with the component of the unit at $dA \otimes dB$.

Our first adjunction between Picard 1– and 2–categories is contained in the following result.

Proposition 2.28 *The functors $\mathcal{D} \mapsto \mathcal{D}_1$ and d induce adjunctions between*

- *the categories $\mathcal{P}erm\mathcal{G}ray\mathcal{M}on$ and $\mathcal{P}erm\mathcal{C}at$, and*
- *the category of strict Picard 2–categories and the category of strict Picard categories.*

The counits of these adjunctions are both identities.

Proof It is immediate from Lemma 2.26 and the definitions that applying $\mathcal{D} \mapsto \mathcal{D}_1$ to a permutative Gray monoid gives a permutative category, and that the resulting permutative category is a strict Picard category if $\mathcal{D}$ is a strict Picard 2–category; this constructs both left adjoints. To construct the right adjoints, let $(C, \oplus, e)$ be a permutative category. We must equip dC with a permutative Gray monoid structure. The tensor product is given by

$$dC \otimes dC \xrightarrow{\chi_{C,C}} d(C \times C) \xrightarrow{d\oplus} dC$$

using Lemma 2.26 or the explicit description in Remark 2.27. The 2–natural isomorphism β^{dC} is $d(\beta^C) * \chi_{C,C}$, using the fact that $d(\tau^\times) \circ \chi = \chi \circ \tau^\otimes$ by the second part of Lemma 2.26. The permutative Gray monoid axioms for dC then reduce to the permutative category axioms for C and the lax symmetric monoidal functor axioms for d. Once again, dC is a strict Picard 2–category if C is a strict Picard category. The statement about counits follows from the corresponding statement about the counit for the adjunction $\pi_0 \dashv d$, and the unit is a strict symmetric monoidal 2–functor by inspection. The triangle identities then follow from those for $\pi_0 \dashv d$, concluding the construction of both adjunctions. □

Remark 2.29 The proof above is simple, but not entirely formal: while symmetric monoidal categories are the symmetric pseudomonoids in the symmetric monoidal 2–category $\mathcal{C}at$, permutative Gray monoids do not admit such a description due to the poor interaction between the Gray tensor product and 2–natural transformations.

We now move on to our second adjunction between permutative categories and permutative Gray monoids which restricts to one between strict Picard categories and strict Picard 2–categories. This adjunction models loop and suspension functors, and appears informally in work of Baez and Dolan [1] on stabilization phenomena in higher categories.

Lemma 2.30 *Let $(C, \oplus, e)$ be a permutative category with symmetry σ. Then the 2–category ΣC with one object $*$, hom-category $\Sigma C(*, *) = C$, and horizontal composition given by $\oplus$ admits the structure of a permutative Gray monoid $(\Sigma C, \widetilde{\oplus})$. The assignment $(C, \oplus) \mapsto (\Sigma C, \widetilde{\oplus})$ is the function on objects of a functor*

$$\Sigma\colon \mathcal{P}erm\mathcal{C}at \to \mathcal{P}erm\mathcal{G}ray\mathcal{M}on.$$

Proof Since C is a strict monoidal category, ΣC is a strict 2–category when horizontal composition is given by $\oplus$. We can define a 2–functor $\widetilde{\oplus}\colon \Sigma C \otimes \Sigma C \to \Sigma C$ as the unique function on 0–cells, by sending any cell of the form $a \otimes 1$ to a, any cell of the form $1 \otimes b$ to b, and $\Sigma_{a,b}$ to the symmetry $\sigma_{a,b}\colon a \oplus b \cong b \oplus a$. With the unique object as the unit, it is simple to check that this 2–functor makes ΣC into a Gray monoid. All that remains is to define β and check the three axioms. Since there is only one object and it is the unit, the second axiom shows that the unique component of β must be the identity 1–cell. Then naturality on 1–cells is immediate, and the only two-dimensional naturality that is not obvious is for the cells $\Sigma_{a,b}$. This axiom becomes the equation

$$\beta \oplus \Sigma_{a,b} = \Sigma_{b,a}^{-1} \oplus \beta,$$

which is merely the claim that $\sigma_{a,b}$ is a symmetry rather than a braid. It is then obvious that this assignment defines a functor as stated. □

Example 2.31 The permutative Gray monoid constructed in [52, Example 2.30] is a suspension ΣC for the following permutative category C:

- The objects of C are the elements of $\mathbb{Z}/2$ with the monoidal structure given by addition.
- Each endomorphism monoid of C is $\mathbb{Z}/2$ and there are no morphisms between distinct objects.
- The symmetry of the nonunit object with itself is the nontrivial morphism.

Remark 2.32 It is natural to expect that the permutative Gray monoid ΣC in the previous example models the 0–connected cover of the 2–type of the sphere spectrum, and indeed this will follow from Theorem 3.11. One might also hope that a skeletal model for the sphere spectrum can be constructed as a "many-object" version of ΣC together with an appropriate symmetry. However, Theorem 3.14 will prove that this is not possible.

Lemma 2.33 *Let $(\mathcal{D}, \oplus, e)$ be a permutative Gray monoid. Then the category $\mathcal{D}(e,e)$ is a permutative category, with tensor product given by composition. The assignment $\mathcal{D} \mapsto \mathcal{D}(e,e)$ is the function on objects of a functor*

$$\Omega\colon \mathcal{P}\mathit{ermGrayMon} \to \mathcal{P}\mathit{ermCat}.$$

Proof For a Gray monoid $\mathcal{D}$, the hom-category $\mathcal{D}(e,e)$ is a braided, strict monoidal category [21; 14] in which the tensor product is given by composition and the braid $f \circ g \cong g \circ f$ is the morphism $\Sigma_{f,g}$ in $\mathcal{D}(e,e)$; we note that $fe = f$ and $eg = g$ since all the 1–cells involved are endomorphisms of the unit object, and the unit object in a Gray monoid is a strict two-sided unit. The component $\beta_{e,e}$ is necessarily the identity, and the calculations in the proof of Lemma 2.30 show that $\Sigma_{f,g} = \Sigma_{g,f}^{-1}$, so we have a permutative structure on $\mathcal{D}(e,e)$. □

Proposition 2.34 *The functor $\Sigma\colon \mathcal{P}\mathit{ermCat} \to \mathcal{P}\mathit{ermGrayMon}$ is left adjoint to the functor $\Omega\colon \mathcal{P}\mathit{ermGrayMon} \to \mathcal{P}\mathit{ermCat}$.*

Proof It is easy to check that the composite $\Omega\Sigma$ is the identity functor on $\mathcal{P}\mathit{ermCat}$, and we take this equality to be the unit of the adjunction. The counit would be a functor $\Sigma(\mathcal{D}(e,e)) \to \mathcal{D}$ which we must define to send the single object of $\Sigma(\mathcal{D}(e,e))$ to the unit object e of $\mathcal{D}$ and then to be the obvious inclusion on the single hom-category. This is clearly a 2–functor, and the arguments in the proofs of the previous two lemmas show that this is a strict map of permutative Gray monoids.

The counit is then obviously the identity on the only hom-category when $\mathcal{D}$ has a single object, and this statement is in fact the commutativity of one of the triangle identities for the adjunction. It is simple to check that Ω applied to the counit is the identity as well since the counit is the identity functor when restricted to the hom-category of the unit objects, and this is the other triangle identity, completing the verification of the adjunction. □

Since the unit $1 \Rightarrow \Omega\Sigma$ is the identity, and the counit is an isomorphism on permutative Gray monoids with one object, we have the following corollary.

Corollary 2.35 *The adjunction $\Sigma \dashv \Omega$ in Proposition 2.34 restricts to the categories of strict Picard categories and strict Picard 2–categories. Moreover, this adjunction gives equivalences between*

- *the category of permutative categories and the category of one-object permutative Gray monoids, and*
- *the category of strict Picard categories and the category of one-object strict Picard 2–categories.*

Proof The first statement follows from the definitions, since both Σ and Ω send strict Picard objects in one category to strict Picard objects in the other. The other two statements are obvious from the proof above. □

3 Stable homotopy theory of Picard 2–categories

In this section we describe how to use the algebra of Picard 2–categories to express homotopical features of their corresponding connective spectra categorically. We begin with a brief review of stable Postnikov towers, mainly for the purpose of fixing notation. Subsequently, we identify algebraic models for this homotopical data in terms of the categorical structure present in a Picard 2–category.

For an abelian group π, the Eilenberg–Mac Lane spectrum of π is denoted by $H\pi$. Its n^{th} suspension is denoted by $\Sigma^n H\pi$, and has zeroth space given by the Eilenberg–Mac Lane space $K(\pi, n)$. With this notation, the stable Postnikov tower of a connective spectrum X is given as follows:

$$\begin{array}{ccccc} & & \vdots & & \\ & & \downarrow & & \\ \Sigma^2 H(\pi_2 X) & \xrightarrow{i_2} & X_2 & \xrightarrow{k_2} & \Sigma^4 H(\pi_3 X) \\ & & \downarrow & & \\ \Sigma^1 H(\pi_1 X) & \xrightarrow{i_1} & X_1 & \xrightarrow{k_1} & \Sigma^3 H(\pi_2 X) \\ & & \downarrow & & \\ & & X_0 & \xrightarrow{k_0} & \Sigma^2 H(\pi_1 X) \end{array}$$

Since X is connective, it follows that $X_0 = H(\pi_0 X)$ and k_0 is therefore a stable map from $H(\pi_0 X)$ to $\Sigma^2 H(\pi_1 X)$. When X is the K–theory spectrum of a strict Picard 2–category, we will model k_0 and $k_1 i_1$ algebraically via stable quadratic maps. A *stable quadratic map* is a homomorphism from an abelian group A to the 2–torsion of an abelian group B. The abelian group of stable homotopy classes $[HA, \Sigma^2 HB]$ is naturally isomorphic to the abelian group of stable quadratic maps $A \to B$ by [17, Equation (27.1)]. Moreover [18, Theorem 20.1] implies that, under this identification, $k_0\colon H(\pi_0 X) \to \Sigma^2 H(\pi_1 X)$ corresponds to the stable quadratic map $\pi_0 X \to \pi_1 X$ given by precomposition with the Hopf map $\eta\colon \Sigma\mathbb{S} \to \mathbb{S}$, where $\mathbb{S}$ denotes the sphere spectrum.

The stable Postnikov tower can be constructed naturally in X, so that if

$$X' \to X$$

is a map of spectra, we have the following commuting naturality diagram of stable Postnikov layers:

$$\begin{array}{ccccc}
\Sigma^n H(\pi_n X') & \xrightarrow{i'_n} & X'_n & \xrightarrow{k'_n} & \Sigma^{n+2} H(\pi_{n+1} X') \\
\downarrow & & \downarrow & & \downarrow \\
\Sigma^n H(\pi_n X) & \xrightarrow{i_n} & X_n & \xrightarrow{k_n} & \Sigma^{n+2} H(\pi_{n+1} X)
\end{array} \tag{3-1}$$

Picard 2–categories model stable 2–types via K–theory. The K–theory functors for symmetric monoidal n–categories, constructed in [53; 57; 45] for $n = 1$ and [29] for $n = 2$, give faithful embeddings of Picard n–categories into stable homotopy. For the purposes of this section we can take K–theory largely as a black box; in Section 5 we give the necessary definitions and properties.

3.1 Modeling stable Postnikov data

For a Picard category $(C, \oplus, e)$, the two possibly nontrivial stable homotopy groups of its K–theory spectrum $K(C)$ are given by

$$\pi_0 K(C) \cong \operatorname{ob} C/\{x \sim y \text{ if there exists a 1–cell } f\colon x \to y\},$$
$$\pi_1 K(C) \cong C(e, e).$$

The stable homotopy groups of the K–theory spectrum of a strict Picard 2–category can be calculated similarly. We denote the classifying space of a 2–category $\mathcal{D}$ by $B\mathcal{D}$ [12].

Lemma 3.2 *Let $\mathcal{D}$ be a strict Picard 2–category. The classifying space $B\mathcal{D}$ is equivalent to $\Omega^\infty K(\mathcal{D})$. The stable homotopy groups $\pi_i K(\mathcal{D})$ are zero except when*

$0 \le i \le 2$, *in which case they are given by the formulas below:*

$$\pi_0 K(\mathcal{D}) \cong \mathrm{ob}\,\mathcal{D}/\{x \sim y \text{ if there exists a 1–cell } f\colon x \to y\},$$
$$\pi_1 K(\mathcal{D}) \cong \mathrm{ob}\,\mathcal{D}(e,e)/\{f \sim g \text{ if there exists a 2–cell } \alpha\colon f \Rightarrow g\},$$
$$\pi_2 K(\mathcal{D}) \cong \mathcal{D}(e,e)(\mathrm{id}_e, \mathrm{id}_e).$$

Proof First, note that $\mathcal{D}$ has underlying 2–category a bigroupoid, and the above are the unstable homotopy groups of the pointed space $(B\mathcal{D}, e)$ by [12, Remark 4.4]. Since the objects of $\mathcal{D}$ are invertible, the space $B\mathcal{D}$ is group-complete, and hence it is the zeroth space of the Ω–spectrum $K(\mathcal{D})$. Thus the stable homotopy groups of $K(\mathcal{D})$ agree with the unstable ones for $B\mathcal{D}$. □

Proposition 3.3 *A map of strict Picard 2–categories induces a stable equivalence of K–theory spectra if and only if it is an equivalence of Picard 2–categories.*

Proof Note that the existence of inverses in a Picard 2–category implies that for any object x we have an equivalence of categories $\mathcal{D}(e,e) \simeq \mathcal{D}(x,x)$ induced by translation by x. Similarly, for any 1–morphism $f\colon e \to e$ there is an isomorphism of sets $\mathcal{D}(e,e)(\mathrm{id}_e, \mathrm{id}_e) \cong \mathcal{D}(e,e)(f,f)$ induced by translation by f.

A map $F\colon \mathcal{D} \to \mathcal{D}'$ of strict Picard 2–categories is a categorical equivalence if and only if it is an equivalence of underlying 2–categories, that is, if it is biessentially surjective and a local equivalence (see [26, Section 5] and [52, Theorem 2.25]). By Lemma 3.2 and the observation above, this happens exactly when f induces an isomorphism on the stable homotopy groups of the corresponding K–theory spectra. □

We will use the adjunctions from Section 2.3 to reduce the calculation of the stable quadratic maps corresponding to k_0 and $k_1 i_1$ of $K(\mathcal{D})$ to two instances of the calculation of k_0 in the 1–dimensional case.

Lemma 3.4 [33] *Let C be a strict Picard category with unit e and symmetry β. Then the bottom stable Postnikov invariant $k_0\colon H\pi_0 K(C) \to \Sigma^2 H\pi_1 K(C)$ is modeled by the stable quadratic map $k_0\colon \pi_0 K(C) \to \pi_1 K(C)$,*

$$[x] \mapsto (e \xrightarrow{\cong} xxx^*x^* \xrightarrow{\beta_{x,x}x^*x^*} xxx^*x^* \xrightarrow{\cong} e),$$

where x is an object in C and x^* denotes an inverse of x.

Remark 3.5 The middle term of the composite $k_0(x)$ was studied in [54; 34] and is called the *signature* of x.

Proof of Lemma 3.4 Note that $k_0\colon \pi_0 K(C) \to \pi_1 K(C)$ is a well-defined function (does not depend on the choices of x, x^* and $xx^* \cong e$). Indeed, given isomorphisms $x \cong y$, $xx^* \cong e$ and $yy^* \cong e$, there is a unique isomorphism $j\colon x^* \cong y^*$ such that

$$\begin{array}{ccc} yx^* & \longrightarrow & xx^* \\ {\scriptstyle yj}\downarrow & & \downarrow \\ yy^* & \longrightarrow & e \end{array}$$

commutes.

Moreover, it is clear that k_0 is compatible with equivalences of Picard categories. By [33, Theorem 2.2], we can thus replace C by a strict skeletal Picard category. In [33, Section 3], a natural action $S \times C \to C$ is defined, where S is a strict skeletal model for the 1–truncation of the sphere spectrum. It follows from the definition of the action that

$$\pi_1(BS) \times \pi_1(BC, x) \to \pi_1(BC, e)$$

sends (η, id_x) to $\beta_{x,x} x^* x^*$, where η denotes the generator of $\pi_1(BS) \cong \mathbb{Z}/2$. Finally, it follows from [33, Proposition 3.4] that the action $S \times C \to C$ models the truncation of the action of the sphere spectrum on KC, thus the image under the action of (η, id_x) agrees with the image of $[x]$ under the stable quadratic map associated to the bottom stable Postnikov invariant. □

Proposition 3.6 *Let $\mathcal{D}$ be a strict Picard 2–category and let $\mathcal{D} \to d(\mathcal{D}_1)$ be the unit of the adjunction in Proposition 2.28. Then*

$$K(\mathcal{D}) \to K(d(\mathcal{D}_1))$$

is the 1–truncation of $K(\mathcal{D})$.

Proof Using the formulas in Lemma 3.2, it is clear that $\mathcal{D} \to d(\mathcal{D}_1)$ induces an isomorphism on π_0 and π_1, and that $\pi_2 K(d(\mathcal{D}_1)) = 0$. Moreover, both K–theory spectra have $\pi_i = 0$ for $i > 2$, so $\mathcal{D}_1$ models the 1–truncation of $\mathcal{D}$. □

Lemma 3.7 *For any permutative category C, the K–theory spectrum of C is stably equivalent to the K–theory spectrum of the corresponding permutative Gray monoid, dC.*

Proof This follows directly from the formulas in [29], and in particular Remark 6.32. □

For any connective spectrum X, the bottom stable Postnikov invariant of X and its 1–truncation X_1 agree. Thus combining Lemma 3.4, Proposition 3.6 and Lemma 3.7 yields the following result.

Corollary 3.8 *Let $\mathcal{D}$ be a strict Picard 2–category with unit e and symmetry β. Then the bottom stable Postnikov invariant $k_0\colon H\pi_0 K(\mathcal{D}) \to \Sigma^2 H\pi_1 K(\mathcal{D})$ is modeled by the stable quadratic map $k_0\colon \pi_0 K(\mathcal{D}) \to \pi_1 K(\mathcal{D})$,*

$$[x] \mapsto [e \xrightarrow{\simeq} xxx^*x^* \xrightarrow{\beta_{x,x}x^*x^*} xxx^*x^* \xrightarrow{\simeq} e],$$

where x is an object in $\mathcal{D}$ and x^* denotes an inverse of x.

Remark 3.9 It can be checked directly that the function $k_0\colon \mathrm{ob}(\mathcal{D}) \to \pi_1(K\mathcal{D})$ is well-defined using the essential uniqueness of the inverse: given another object $\bar{x}$ together with an equivalence $e \simeq x\bar{x}$, there is an equivalence $x^* \simeq \bar{x}$ and an isomorphism 2–cell in the obvious triangle, which is unique up to unique isomorphism. This follows from the techniques in [26], and many of the details are explained there in Section 6.

In order to identify the composite $k_1 i_1$ categorically, we analyze the relationship between Postnikov layers and categorical suspension.

Proposition 3.10 *Let $\mathcal{D}$ be a strict Picard 2–category and let $\Sigma\Omega\mathcal{D} \to \mathcal{D}$ be the counit of the adjunction in Proposition 2.34. Then*

$$K(\Sigma\Omega\mathcal{D}) \to K(\mathcal{D})$$

is a 0–connected cover of $K(\mathcal{D})$.

Proof It is clear from the formulas in Lemma 3.2 that $\Sigma\Omega\mathcal{D} \to \mathcal{D}$ induces an isomorphism on π_1 and π_2, and moreover, the corresponding K–theory spectra have $\pi_i = 0$ for $i > 2$. Since $\Sigma\Omega\mathcal{D}$ has only one object, we have $\pi_0 K(\Sigma\Omega\mathcal{D}) = 0$, so $\Sigma\Omega\mathcal{D}$ models the 0–connected cover of $\mathcal{D}$. □

In addition to the elementary algebra and homotopy theory of Picard 2–categories discussed above, we require the following result.

Theorem 3.11 *Let C be a permutative category. Then $\Sigma K(C)$ and $K(\Sigma C)$ are stably equivalent.*

The proof of Theorem 3.11 requires a nontrivial application of 2–monad theory. We develop the relevant 2–monadic techniques in Section 4 and give the proof in Section 5. These two sections are independent of the preceding sections.

Lemma 3.12 *Let $(\mathcal{D}, \oplus, e)$ be a strict Picard 2–category. Then the composite*

$$k_1 i_1\colon \Sigma H\pi_1 K(\mathcal{D}) \to \Sigma^3 H\pi_2 K(\mathcal{D})$$

is modeled by the stable quadratic map $\pi_1 K(\mathcal{D}) \to \pi_2 K(\mathcal{D})$,

$$[f] \mapsto (\mathrm{id}_e \xrightarrow{\cong} f \circ f \circ f^* \circ f^* \xrightarrow{\Sigma_{f,f} f^* \circ f^*} f \circ f \circ f^* \circ f^* \xrightarrow{\cong} \mathrm{id}_e),$$

where $f\colon e \to e$ is a 1–cell in $\mathcal{D}$ and f^* denotes an inverse of f.

Proof We use superscripts to distinguish Postnikov data of different spectra. The composite $k_1^{\mathcal{D}} i_1^{\mathcal{D}}$ in the first Postnikov layer of the spectrum $K(\mathcal{D})$ identifies with the composite $k_1^{\Sigma\Omega\mathcal{D}} i_1^{\Sigma\Omega\mathcal{D}}$ since $K(\Sigma\Omega\mathcal{D})$ is the 0–connected cover of $K(\mathcal{D})$ by Proposition 3.10 and the Postnikov tower can be constructed naturally; see (3-1).

Since $K(\Sigma\Omega\mathcal{D}) \simeq \Sigma K(\Omega\mathcal{D})$ by Theorem 3.11 and $K(\Omega\mathcal{D})$ is connective, it follows that

$$k_1^{\Sigma\Omega\mathcal{D}} i_1^{\Sigma\Omega\mathcal{D}} = \Sigma(k_0^{\Omega\mathcal{D}} i_0^{\Omega\mathcal{D}}) = \Sigma(k_0^{\Omega\mathcal{D}})$$

in the stable homotopy category.

Finally, we deduce from Lemma 3.4 that the map $\Sigma(k_0^{\Omega\mathcal{D}})$ is represented by the desired group homomorphism. □

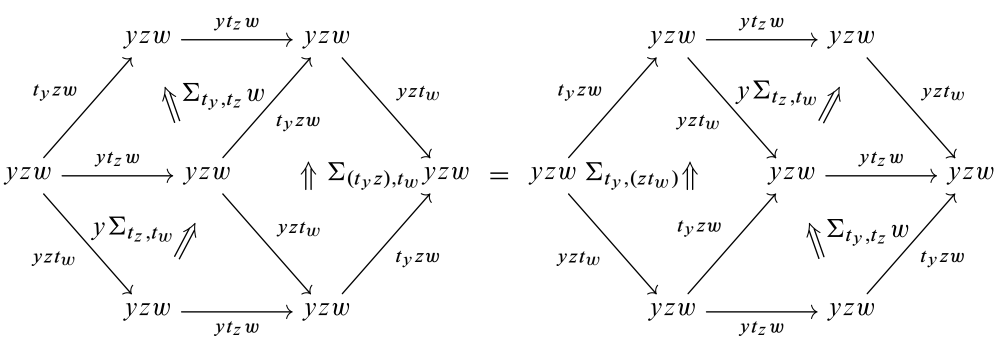
3.2 Application to strict skeletal Picard 2–categories

Now we make an observation about the structure 2–cells $\Sigma_{f,g}$ in a strict Picard 2–category. This algebra will be a key input for our main application, Theorem 3.14.

Lemma 3.13 *Let $(\mathcal{D}, \oplus, e)$ be a strict Picard 2–category. Let $g\colon e \to e$ be any 1–cell and let $s = \beta_{x,x} x^* x^*$ be a representative of the signature of some object x with inverse x^*. Then $\Sigma_{s,g}$ and $\Sigma_{g,s}$ are identity 2–cells in $\mathcal{D}$.*

Proof By naturality of the symmetry and interchange, $\Sigma_{\beta_{y,z},h}$ and $\Sigma_{h,\beta_{y,z}}$ are identity 2–cells for any 1–cell h [29, Proposition 3.41]. The result for $\Sigma_{g,s}$ follows by noting that $\Sigma_{g,f} w = \Sigma_{g,fw}$ for any 1–cells f and g and object w by the associativity axiom for a Gray monoid. Hence $\Sigma_{g,s} = \Sigma_{g,\beta_{x,x}x^*x^*} = \Sigma_{g,\beta_{x,x}} x^* x^*$, which is the identity 2–cell.

For the other equality, we note the final axiom of [27, Proposition 3.3] reduces to the following equality of pasting diagrams for objects y, z and w with endomorphisms t_y, t_z and t_w, respectively:



Thus the result for $\Sigma_{s,g}$ follows by taking $(y,z,w)=(xx,x^*x^*,e)$, $t_y=\beta_{x,x}$, $t_z=\mathrm{id}$ and $t_w=g$. □

We are now ready to give our main application regarding stable Postnikov data of strict skeletal Picard 2–categories.

Theorem 3.14 *Let $\mathcal{D}$ be a strict skeletal Picard 2–category and assume that*

$$k_0\colon \pi_0 K(\mathcal{D})\to\pi_1 K(\mathcal{D})$$

is surjective. Then $k_1 i_1$ is trivial.

Proof We prove that the stable quadratic map $\pi_1 K(\mathcal{D})\to\pi_2 K(\mathcal{D})$ from Lemma 3.12 that models the composite $k_1 i_1$ is trivial. Since k_0 is surjective by assumption, it suffices to consider $k_1 i_1(f)$ for f of the form

$$e\xrightarrow{w} xxx^*x^*\xrightarrow{\beta_{x,x}x^*x^*} xxx^*x^*\xrightarrow{w^*} e \tag{3-15}$$

for some object x with inverse x^*. Here w denotes the composite

$$e\xrightarrow{u} xx^*\xrightarrow{xux^*} xxx^*x^*$$

for a chosen equivalence $u\colon e\simeq xx^*$ and w^* denotes the corresponding reverse composite for a chosen $u^*\colon xx^*\simeq e$ inverse to u. Note that the isomorphism class of f is independent of the choices of the inverse object x^* and the equivalences u and u^* (see Remark 3.9). Since $\mathcal{D}$ is skeletal, it must be that $xx^*=e$. Therefore we can choose the equivalence $u\colon e\simeq xx^*$ to be id_e and then choose u^* to be id_e as well. With these choices, the composite f is actually equal to $\beta_{x,x}x^*x^*$. By Lemma 3.13, the Gray structure 2–cell $\Sigma_{f,f}$ is the identity 2–cell $\mathrm{id}_{f\circ f}$. This implies that $k_1 i_1(f)=\mathrm{id}_{\mathrm{id}_e}$. □

Remark 3.16 The result of Theorem 3.14 may be viewed as the computation of a differential in the spectral sequence arising from mapping into the stable Postnikov tower of $K\mathcal{D}$. This spectral sequence appears, for example, in [35] and is a cocellular construction of the Atiyah–Hirzebruch spectral sequence (see [22, Appendix B]).

Our most important application concerns the sphere spectrum.

Corollary 3.17 *Let $\mathcal{D}$ be a strict skeletal Picard 2–category. Then $\mathcal{D}$ cannot be a model for the 2–truncation of the sphere spectrum.*

Proof The nontrivial element in π_1 of the sphere spectrum is given by $k_0(1)$, so k_0 is surjective and therefore Theorem 3.14 applies. But $k_1 i_1$ is Sq^2, which is the nontrivial element of $H^2(\mathbb{Z}/2;\mathbb{Z}/2)$ [50, pages 117–118]. □

Remark 3.18 To understand the meaning of this result, recall that one can specify a unique Picard category by choosing two abelian groups for π_0 and π_1 together with a stable quadratic map k_0 for the symmetry. This is the content of Theorem 1.1. However, one does not specify a Picard 2–category by simply choosing three abelian groups and two group homomorphisms. This is tantamount to specifying a stable 2–type by choosing the bottom Postnikov invariant k_0 and the composite $k_1 i_1$. Theorem 3.14 shows that such data do not always assemble to form a strict Picard 2–category. For example, the construction of [4, 5.2] does not satisfy the axioms of a permutative Gray monoid.

4 Strictification via 2–monads

In this section we develop the 2–monadic tools used in the proof of Theorem 3.11. In Section 4.1 we recall some basic definitions as well as abstract coherence theory from the perspective of 2–monads. Our focus is on various strictification results for algebras and pseudoalgebras over 2–monads, and how strictification can often be expressed as a 2–adjunction with good properties. In Section 4.2 we apply this to construct a strictification of pseudodiagrams as a left 2–adjoint. The material in this section is largely standard 2–category theory, but we did not know a single reference which collected it all in one place.

The formalism of this section aids the proof of Theorem 3.11 in two ways. First, it allows us to produce strict diagrams of 2–categories by working with diagrams which are weaker (eg whose arrows take values in pseudofunctors) but more straightforward to define. This occurs in Section 5.1. Second, it allows us to construct strict equivalences of strict diagrams by working instead with pseudonatural equivalences between them. This occurs in Section 5.2.

4.1 Review of 2–monad theory

We recall relevant aspects of 2–monad theory and fix notation. These include maps of monads and abstract coherence theory [38; 51; 7; 39]. Let $\mathcal{A}$ be a 2–category and $(T\colon \mathcal{A} \to \mathcal{A}, \eta, \mu)$ be a 2–monad on $\mathcal{A}$. We then have the following 2–categories of algebras and morphisms with varying levels of strictness:

(1) T-$\mathbf{Alg}_s$ is the 2–category of strict T–algebras, strict morphisms, and algebra 2–cells. Its underlying category is just the usual category of algebras for the underlying monad of T on the underlying category of $\mathcal{A}$.

(2) T-$\mathbf{Alg}$ is the 2–category of strict T–algebras, pseudo-T–morphisms, and algebra 2–cells.

(3) $\mathbf{Ps}\text{-}T\text{-}\mathbf{Alg}$ is the 2–category of pseudo-T–algebras, pseudo-T–morphisms, and algebra 2–cells.

We have inclusions and forgetful functors as below:

$$\begin{array}{ccccc} T\text{-}\mathbf{Alg}_s & \xrightarrow{\ i\ } & T\text{-}\mathbf{Alg} & \longrightarrow & \mathbf{Ps}\text{-}T\text{-}\mathbf{Alg} \\ & \searrow^{U} & \downarrow{\scriptstyle U} & \swarrow_{U} & \\ & & \mathcal{A} & & \end{array}$$

A map of 2–monads is precisely the data necessary to provide a 2–functor between 2–categories of strict algebras.

Definition 4.1 Let S be a 2–monad on $\mathcal{A}$ and T a 2–monad on $\mathcal{B}$. A *strict map of 2–monads* $S \to T$ consists of a 2–functor $F\colon \mathcal{A} \to \mathcal{B}$ and a 2–natural transformation $\lambda\colon TF \Rightarrow FS$ satisfying two compatibility axioms [5]:

$$\lambda \circ \mu F = F\mu \circ \lambda S \circ T\lambda,$$
$$\lambda \circ \eta F = F\eta.$$

Proposition 4.2 *If $F\colon S \to T$ is a strict map of 2–monads, then F lifts to the indicated 2–functors in the following diagram:*

$$\begin{array}{ccc} S\text{-}\mathcal{Alg}_s & \xrightarrow{\ F\ } & T\text{-}\mathcal{Alg}_s \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle i} \\ S\text{-}\mathcal{Alg} & \xrightarrow{\ F\ } & T\text{-}\mathcal{Alg} \\ {\scriptstyle U}\downarrow & & \downarrow{\scriptstyle U} \\ \mathcal{A} & \xrightarrow[F]{} & \mathcal{B} \end{array}$$

Abstract coherence theory provides left 2–adjoints to $T\text{-}\mathbf{Alg}_s \hookrightarrow T\text{-}\mathbf{Alg}$ and the composite $T\text{-}\mathbf{Alg}_s \hookrightarrow \mathbf{Ps}\text{-}T\text{-}\mathbf{Alg}$. Lack discusses possible hypotheses in [39, Section 3], so we give the following theorem in outline form.

Theorem 4.3 [39, Section 3] *Under some assumptions on $\mathcal{A}$ and T, the inclusions*

$$i\colon T\text{-}\mathbf{Alg}_s \hookrightarrow T\text{-}\mathbf{Alg}, \quad j\colon T\text{-}\mathbf{Alg}_s \hookrightarrow \mathbf{Ps}\text{-}T\text{-}\mathbf{Alg}$$

have left 2–adjoints generically denoted by Q. Under even further assumptions, the units $1 \Rightarrow iQ$ and $1 \Rightarrow jQ$ and the counits $Qi \Rightarrow 1$ and $Qj \Rightarrow 1$ of these 2–adjunctions have components which are internal equivalences in T-$\mathbf{Alg}$ for $Q \dashv i$ and $\mathbf{Ps}$-T-$\mathbf{Alg}$ for $Q \dashv j$, respectively.

Remark 4.4 The proofs in [39] only concern the units, but the statement about counits follows immediately from the 2-out-of-3 property for equivalences and one of the triangle identities. We note that the components of the counits are actually always 1–cells in T-$\mathbf{Alg}_s$, so saying they are equivalences in T-$\mathbf{Alg}$ or $\mathbf{Ps}$-T-$\mathbf{Alg}$ requires implicitly applying i or j, respectively.

Notation 4.5 We will always denote inclusions of the form $T\text{-}\mathbf{Alg}_s \hookrightarrow T\text{-}\mathbf{Alg}$ by i, and inclusions of the form $T\text{-}\mathbf{Alg}_s \hookrightarrow \mathbf{Ps}\text{-}T\text{-}\mathbf{Alg}$ by j. If we need to distinguish between the left adjoints for i and j, we will denote them by Q_i and Q_j, respectively.

4.2 Two applications of 2–monads

We are interested in two applications of Theorem 4.3: one which gives 2–categories as the strict algebras (Proposition 4.12), and one which gives 2–functors with fixed domain and codomain as the strict algebras (Proposition 4.16). Combining these in Theorem 4.19, we obtain the main strictification result used in our analysis of K–theory and suspension in Section 5.

We begin with the 2–monad for 2–categories and refer the interested reader to [41; 42] for further details.

Definition 4.6 (1) A *category-enriched graph* or $\mathcal{C}at$*–graph* $(S, S(x,y))$ consists of a set of objects S and a category $S(x,y)$ for each pair of objects $x, y \in S$.

(2) A *map of* $\mathcal{C}at$*–graphs* $(F, F_{x,y})\colon (S, S(x,y)) \to (T, T(w,z))$ consists of a function $F\colon S \to T$ and a functor $F_{x,y}\colon S(x,y) \to T(Fx, Fy)$ for each pair of objects $x, y \in S$.

(3) A $\mathcal{C}at$*–graph* 2*–cell* $\alpha\colon (F, F_{x,y}) \Rightarrow (G, G_{x,y})$ only exists when $F = G$ as functions $S \to T$, and then consists of a natural transformation $\alpha_{x,y}\colon F_{x,y} \Rightarrow G_{x,y}$ for each pair of objects $x, y \in S$.

Notation 4.7 $\mathcal{C}at$–graphs, their maps, and 2–cells form a 2–category, $\mathcal{C}at\text{-}\mathcal{G}rph$, with the obvious composition and unit structures.

Definition 4.8 Let $\mathcal{A}$ and $\mathcal{B}$ be 2–categories, and $F, G\colon \mathcal{A} \to \mathcal{B}$ be a pair of 2–functors between them. An *icon* $\alpha\colon F \Rightarrow G$ exists only when $Fa = Ga$ for all objects $a \in \mathcal{A}$, and then consists of natural transformations

$$\alpha_{a,b}\colon F_{a,b} \Rightarrow G_{a,b}\colon \mathcal{A}(a,b) \to \mathcal{B}(Fa, Fb)$$

for all pairs of objects a and b such that the following diagrams commute:

$$
\begin{array}{ccc}
\mathrm{id}_{Fa} & \xrightarrow{=} & F\mathrm{id}_a \\
& \searrow^{=} & \downarrow{\scriptstyle \alpha_{\mathrm{id}}} \\
& & G\mathrm{id}_a
\end{array}
\qquad
\begin{array}{ccc}
Ff \circ Fg & \xrightarrow{=} & F(f \circ g) \\
{\scriptstyle \alpha_f * \alpha_g}\downarrow & & \downarrow{\scriptstyle \alpha_{f \circ g}} \\
Gf \circ Gg & \xrightarrow[=]{} & G(f \circ g)
\end{array}
$$

(Note that we suppress the 0–cell source and target subscripts for components of the transformations $\alpha_{a,b}$ and instead only list the 1–cell for which a given 2–cell is the component.)

Remark 4.9 We can define icons between pseudofunctors or lax functors with only minor modifications, replacing some equalities above with the appropriate coherence cell; see [42; 41].

Notation 4.10 2–categories, 2–functors, and icons form a 2–category, which we denote by $2\mathcal{C}at_{2,\mathrm{i}}$. 2–categories, pseudofunctors, and icons form a 2–category, which we denote by $2\mathcal{C}at_{\mathrm{p,i}}$. Bicategories, pseudofunctors, and icons also form a 2–category, which we denote by $\mathcal{B}icat_{\mathrm{p,i}}$.

Recall that a 2–functor $U\colon \mathcal{A} \to \mathcal{K}$ is 2–monadic if it has a left 2–adjoint F and $\mathcal{A}$ is 2–equivalent to the 2–category of algebras (UF)-$\mathbf{Alg}_s$ via the canonical comparison map.

Proposition 4.11 [42; 41] *The 2–functor $2\mathcal{C}at_{2,\mathrm{i}} \to \mathcal{C}at\text{-}\mathcal{G}rph$ is 2–monadic, and the left 2–adjoint is given by the $\mathcal{C}at$–enriched version of the free category functor.*

The following is our first application of Theorem 4.3.

Proposition 4.12 *The two inclusions*

$$i\colon 2\mathcal{C}at_{2,\mathrm{i}} \hookrightarrow 2\mathcal{C}at_{\mathrm{p,i}}, \quad j\colon 2\mathcal{C}at_{2,\mathrm{i}} \hookrightarrow \mathcal{B}icat_{p,i}$$

have left 2–adjoints, and the components of the units and counits of both adjunctions are internal equivalences in $2\mathcal{C}at_{\mathrm{p,i}}$ for $Q_i \dashv i$ and $\mathcal{B}icat_{p,i}$ for $Q_j \dashv j$, respectively.

Proof The induced monad T on $\mathcal{C}at\text{-}\mathcal{G}rph$ satisfies a version of the hypotheses for Theorem 4.3 (for example, it is a finitary monad) so we get left 2–adjoints to both inclusions

$$i\colon T\text{-}\mathbf{Alg}_s \to T\text{-}\mathbf{Alg}, \quad j\colon T\text{-}\mathbf{Alg}_s \to \mathbf{Ps}\text{-}T\text{-}\mathbf{Alg}.$$

Now T-$\mathbf{Alg}$ can be identified with $2\mathcal{C}at_{\mathrm{p,i}}$, and one can check that $\mathbf{Ps}$-T-$\mathbf{Alg}$ can be identified with $\mathcal{B}icat_{\mathrm{p,i}}$, and using these the two left 2–adjoints above are both given by the standard functorial strictification functor, often denoted by st (see [34] for the version with only a single object, ie monoidal categories). The objects of $\mathrm{st}(X)$ are the same as X, while the 1–cells are formal strings of composable 1–cells (including the empty string at each object). Internal equivalences in either T-**Alg** or **Ps**-T-**Alg** for the 2–monad T are bijective-on-objects biequivalences, and it is easy to check that the unit is such; see [42; 28] for further details. □

Remark 4.13 We should note that $\mathit{2Cat}_{2,\mathrm{i}}$ is complete and cocomplete as a 2–category, since it is the 2–category of algebras for a finitary 2–monad on a complete and cocomplete 2–category. This will be necessary for later constructions. On the other hand, $\mathit{2Cat}_{\mathrm{p,i}}$ is not cocomplete as a 2–category, but is as a bicategory: coequalizers of pseudofunctors rarely exist in the strict, 2–categorical sense, but all bicategorical colimits do exist.

Our second application of Theorem 4.3 deals with functor 2–categories. Here we fix a small 2–category $\mathcal{A}$ and a complete and cocomplete 2–category $\mathcal{K}$.

Notation 4.14 Let $[\mathcal{A},\mathcal{K}]$ denote the 2–category of 2–functors, 2–natural transformations, and modifications from $\mathcal{A}$ to $\mathcal{K}$. Let $\mathcal{B}\mathit{icat}(\mathcal{A},\mathcal{K})$ denote the 2–category of pseudofunctors, pseudonatural transformations, and modifications from $\mathcal{A}$ to $\mathcal{K}$. Let $\mathcal{G}\mathit{ray}(\mathcal{A},\mathcal{K})$ denote the 2–category of 2–functors, pseudonatural transformations, and modifications from $\mathcal{A}$ to $\mathcal{K}$. This is the internal hom-object corresponding to the Gray tensor product on $\mathit{2Cat}$ [21].

Remark 4.15 $\mathcal{B}\mathit{icat}(\mathcal{A},\mathcal{K})$ inherits its compositional and unit structures from the target 2–category $\mathcal{K}$ and is therefore a 2–category rather than a bicategory even though all of its cells are of the weaker, bicategorical variety.

Let $\mathrm{ob}\,\mathcal{A}$ denote the discrete 2–category with the same set of objects as $\mathcal{A}$. We have an inclusion $\mathrm{ob}\,\mathcal{A}\hookrightarrow\mathcal{A}$, which induces a 2–functor $U\colon [\mathcal{A},\mathcal{K}]\to[\mathrm{ob}\,\mathcal{A},\mathcal{K}]$.

Proposition 4.16 *The forgetful 2–functor $U\colon [\mathcal{A},\mathcal{K}]\to[\mathrm{ob}\,\mathcal{A},\mathcal{K}]$ is 2–monadic, and the left 2–adjoint is given by enriched left Kan extension. The induced 2–monad preserves all colimits, and so the inclusions*

$$i\colon [\mathcal{A},\mathcal{K}]\hookrightarrow \mathcal{G}\mathit{ray}(\mathcal{A},\mathcal{K}),\qquad j\colon [\mathcal{A},\mathcal{K}]\hookrightarrow \mathcal{B}\mathit{icat}(\mathcal{A},\mathcal{K})$$

have left 2–adjoints. The units and counits of these adjunctions have components which are internal equivalences in $\mathcal{G}\mathit{ray}(\mathcal{A},\mathcal{K})$ for $Q_i\dashv i$ and $\mathcal{B}\mathit{icat}(\mathcal{A},\mathcal{K})$ for $Q_j\dashv j$, respectively.

Proof That U is 2–monadic follows because it has a left 2–adjoint given by enriched left Kan extension and is furthermore conservative. Thus $[\mathcal{A}, \mathcal{K}]$ is 2–equivalent to the 2–category of strict algebras for $U \circ \mathrm{Lan}$. The 2–functor U also has a right adjoint given by right Kan extension since $\mathcal{K}$ is complete, so $U \circ \mathrm{Lan}$ preserves all colimits as it is a composite of two left 2–adjoints. The 2–category $[\mathrm{ob}\,\mathcal{A}, \mathcal{K}]$ is cocomplete since $\mathcal{K}$ is, hence $T = U \circ \mathrm{Lan}$ satisfies the strongest version of the hypotheses for Theorem 4.3. One can check that T-**Alg** is 2–equivalent to $\mathcal{G}ray(\mathcal{A}, \mathcal{K})$ and **Ps**-T-**Alg** is 2–equivalent to $\mathcal{B}icat(\mathcal{A}, \mathcal{K})$ [40]. This proves that the inclusions i and j in the statement have left 2–adjoints. The version of Theorem 4.3 which applies in this case proves, moreover, that the components of the units are internal equivalences in $\mathcal{G}ray(\mathcal{A}, \mathcal{K})$ and $\mathcal{B}icat(\mathcal{A}, \mathcal{K})$, respectively, and hence the claim about counits follows (see Remark 4.4). □

We require one further lemma before stating the main result of this section.

Lemma 4.17 *For a fixed 2–category $\mathcal{A}$, $\mathcal{B}icat(\mathcal{A}, -)$ is an endo-2–functor of the 2–category of 2–categories, 2–functors, and 2–natural transformations.*

Proof For any 2–category $\mathcal{B}$, we know that $\mathcal{B}icat(\mathcal{A}, \mathcal{B})$ is a 2–category. Furthermore, if $F\colon \mathcal{B} \to \mathcal{C}$ is a 2–functor, it is straightforward to check that $F_*\colon \mathcal{B}icat(\mathcal{A}, \mathcal{B}) \to \mathcal{B}icat(\mathcal{A}, \mathcal{C})$ is also a 2–functor. The only interesting detail to check is on the level of 2–cells, where we must show that if $\sigma\colon F \Rightarrow G$ is 2–natural, then so is σ_*. The component of σ_* at $H\colon \mathcal{A} \to \mathcal{B}$ is the pseudonatural transformation $\sigma H\colon FH \Rightarrow GH$ with $(\sigma H)_a = \sigma_{Ha}$ and similarly for pseudonaturality isomorphisms. We must verify that σ_* is 2–natural in H. Thus, for any $\alpha\colon H \Rightarrow K$, we must check that $G\alpha \circ \sigma H = \sigma K \circ F\alpha$ as pseudonatural transformations and then similarly for modifications. At an object a, we have components

$$(G\alpha \circ \sigma H)_a = G(\alpha_a) \circ \sigma_{Ha} = \sigma_{Ka} \circ F(\alpha_a) = (\sigma K \circ F\alpha)_a$$

by the 2–naturality of σ in Ha. A short and simple pasting diagram argument that we leave to the reader also shows that the pseudonaturality isomorphisms for $G\alpha \circ \sigma H$ and $\sigma K \circ F\alpha$ are the same, once again relying on the 2–naturality of σ in its argument. This completes the 1–dimensional part of 2–naturality, and the 2–dimensional part is a direct consequence of the 2–naturality of σ when written out on components. □

Remark 4.18 While the argument above is simple, it is not entirely formal. The "dual" version for $\mathcal{B}icat(-, \mathcal{A})$ does not hold due to an asymmetry in the definition of the pseudonaturality isomorphisms for a horizontal composite of pseudonatural transformations.

We are now ready to prove the main result of this section, namely that we can replace pseudofunctors $\mathcal{A} \to 2\mathcal{C}at_{\mathrm{p,i}}$ with equivalent 2–functors $\mathcal{A} \to 2\mathcal{C}at_{2,\mathrm{i}}$.

Theorem 4.19 *The inclusion*

$$J\colon [\mathcal{A}, 2\mathcal{C}at_{2,\mathrm{i}}] \hookrightarrow \mathcal{B}icat(\mathcal{A}, 2\mathcal{C}at_{\mathrm{p,i}})$$

has a left 2–adjoint Q. The unit and counit of this adjunction have components which are internal equivalences in $\mathcal{B}icat(\mathcal{A}, 2\mathcal{C}at_{\mathrm{p,i}})$.

Proof We will combine Propositions 4.12 and 4.16. The inclusion J factors into the two inclusions

$$[\mathcal{A}, 2\mathcal{C}at_{2,\mathrm{i}}] \overset{j}{\hookrightarrow} \mathcal{B}icat(\mathcal{A}, 2\mathcal{C}at_{2,\mathrm{i}}) \overset{i_*}{\hookrightarrow} \mathcal{B}icat(\mathcal{A}, 2\mathcal{C}at_{\mathrm{p,i}}).$$

Since $2\mathcal{C}at_{2,\mathrm{i}}$ is cocomplete, j has a left 2–adjoint Q_j by Proposition 4.16. The inclusion i has a left 2–adjoint Q_i by Proposition 4.12, so i_* has a left 2–adjoint $(Q_i)_*$ by Lemma 4.17. Both of these 2–adjunctions have units whose components are equivalences, so the composite $Q = Q_j(Q_i)_*$ does as well, from which the claim about counits follows. □

5 Categorical suspension models stable suspension

The purpose of this section is to prove Theorem 3.11, which states that K–theory commutes with suspension, in the appropriate sense. More precisely, we show that for any permutative category C, the K–theory spectrum of the one-object permutative Gray monoid ΣC is stably equivalent to the suspension of the K–theory spectrum of C.

This entails a comparison between constructions of K–theory for categories and 2–categories. Both constructions use the theory of Γ–spaces developed by Segal [53]. We recall this theory in Section 5.1. Our interest in Γ–spaces arises from the fact that they model the homotopy theory of connective spectra, as developed by Bousfield and Friedlander [10] in the simplicial setting. Thus, in what follows, we will work with Γ–simplicial sets to prove Theorem 3.11.

We model the spectra $K(\Sigma C)$ and ΣKC with Γ–simplicial sets which are constructed from certain Γ–objects in simplicial categories. These Γ–objects in simplicial categories are two different strictifications of the same pseudofunctor $\mathscr{F} \to \mathcal{B}icat(\Delta^{\mathrm{op}}, \mathcal{C}at_2)$, where $\mathscr{F}$ is the category of finite pointed sets and pointed maps. The first of these strictifications is provided in Definition 5.8 by applying the suspension of Γ–simplicial sets (Definition 5.5) to a strictification of the pseudofunctor $n \mapsto C^n$ (Construction 5.7), giving a model for ΣKC. The second is provided in Definition 5.16 and gives a model for $K(\Sigma C)$.

In Section 5.2 we use the formalism of Section 4 to compare the two strictifications via a zigzag of levelwise equivalences. The key step in this comparison is constructed in Theorem 5.21 by strictification of a pseudonatural equivalence.

5.1 Constructions of K–theory spectra and suspension

Let $\mathscr{F}$ denote the following skeletal model for the category of finite pointed sets and pointed maps. An object of $\mathscr{F}$ is determined by an integer $m \geq 0$, which represents the pointed set $\underline{m}_+ = \{0, 1, \dots, m\}$, where 0 is the basepoint. This category is isomorphic to the opposite of the category Γ defined by Segal [53].

Definition 5.1 Let $\mathcal{C}$ be a category with a terminal object $*$. A Γ*–object* in $\mathcal{C}$ is a functor $X\colon \mathscr{F} \to \mathcal{C}$ such that $X(\underline{0}_+) = *$.

We give the above definition in full generality, but are only interested in the cases when $\mathcal{C}$ is one of $\mathcal{C}at$, $2\mathcal{C}at$, the category of simplicial sets $s\mathcal{S}et$ or of topological spaces $\mathcal{T}op$. In each of these cases, we have finite products and a notion of weak equivalence. In $\mathcal{T}op$ and $s\mathcal{S}et$ this is the classical notion of weak homotopy equivalence, and in both $\mathcal{C}at$ and $2\mathcal{C}at$ we define a functor or 2–functor to be a weak equivalence if it induces a weak homotopy equivalence in $s\mathcal{S}et$ after applying the nerve [25; 12].

Definition 5.2 Let X be a Γ–object in $\mathcal{C}$. We say X is *special* if the Segal maps

$$X(\underline{n}_+) \to X(\underline{1}_+)^n$$

are weak equivalences.

The main result of [53] is that, given a Γ–space X, one can produce a connective spectrum $\widetilde{X}$. Moreover, if X is special then $\widetilde{X}$ is an almost Ω–spectrum such that $\Omega^\infty \widetilde{X}$ is a group completion of $X(\underline{1}_+)$. We recall how to express suspension of spectra in terms of Γ–simplicial sets using the standard "inclusion" $\Delta^{\mathrm{op}} \to \mathscr{F}$, as specified in [48, Lemma 3.5] and the following smash product. Let $\wedge\colon \mathscr{F} \times \mathscr{F} \to \mathscr{F}$ be the functor that sends $(\underline{n}_+, \underline{p}_+)$ to $(\underline{np})_+ = \underline{n}_+ \vee \cdots \vee \underline{n}_+$. Our reverse lexicographic convention differs from the smash product in [48, Construction 3.4], which considers $(\underline{np})_+$ as $\underline{p}_+ \vee \cdots \vee \underline{p}_+$.

Notation 5.3 Let

$$\Phi\colon \mathcal{B}icat(\mathcal{A} \times \mathcal{B}, \mathcal{C}) \to \mathcal{B}icat(\mathcal{A}, \mathcal{B}icat(\mathcal{B}, \mathcal{C}))$$

denote the biequivalence of functor bicategories given in [55], sending a pseudofunctor $F\colon \mathcal{A} \times \mathcal{B} \to \mathcal{C}$ to the pseudofunctor

$$\Phi(F)(a)(b) = F(a, b).$$

We also let Φ denote the isomorphism of functor 2–categories

$$[\mathcal{A} \times \mathcal{B}, \mathcal{C}] \xrightarrow{\cong} [\mathcal{A}, [\mathcal{B}, \mathcal{C}]].$$

In order to justify using the same notation Φ for both of these, we note that both versions (reading vertical arrows upwards or downwards) of the square

$$\begin{array}{ccc} [\mathcal{A} \times \mathcal{B}, \mathcal{C}] & \longrightarrow & \mathcal{B}icat(\mathcal{A} \times \mathcal{B}, \mathcal{C}) \\ \cong \updownarrow & & \updownarrow \simeq \\ [\mathcal{A}, [\mathcal{B}, \mathcal{C}]] & \longrightarrow & \mathcal{B}icat(\mathcal{A}, \mathcal{B}icat(\mathcal{B}, \mathcal{C})) \end{array} \tag{5-4}$$

commute, with the downward direction being given by Φ on the vertical arrows.

Definition 5.5 Let $X\colon \mathscr{F} \to sSet$ be a special Γ–simplicial set and let $X \circ \wedge$ denote the composite

$$\mathscr{F} \times \Delta^{\mathrm{op}} \xrightarrow{\wedge} \mathscr{F} \xrightarrow{X} sSet.$$

Let $d\colon [\Delta^{\mathrm{op}}, sSet] \to sSet$ denote the diagonal functor. We define the *suspension*, ΣX, as the special Γ–simplicial set $d \circ \Phi(X \circ \wedge)$.

Proposition 5.6 [53; 10] *Let X be a special Γ–simplicial set and $\widetilde{X}$ its associated spectrum. Then the spectrum associated to ΣX is stably equivalent to $\Sigma \widetilde{X}$.*

Given a permutative category C, there are several equivalent ways of constructing a special Γ–category. The following was first constructed by Thomason [56, Definition 4.1.2].

Construction 5.7 Let $(C, \oplus, e)$ be a permutative category. We can construct a pseudofunctor

$$C^{(-)}\colon \mathscr{F} \to Cat_2$$

which sends $\underline{m}_+$ to C^m. Given a morphism $\phi\colon \underline{m}_+ \to \underline{n}_+$, the corresponding functor $\phi_*\colon C^m \to C^n$ is defined uniquely by the requirement that the square below commutes for each projection $\pi_j\colon C^n \to C$:

$$\begin{array}{ccc} C^m & \longrightarrow & C^{\phi^{-1}(j)} \\ \phi_* \downarrow & & \downarrow \oplus \\ C^n & \xrightarrow[\pi_j]{} & C \end{array}$$

The top horizontal map is the projection onto the coordinates which appear in $\phi^{-1}(j)$. The $\oplus$ appearing on the right vertical map is the iterated application of the tensor product $\oplus$, with the convention that if $\phi^{-1}(j)$ is empty, then the map is the constant functor on the unit e. This assignment is not strictly functorial, but the permutative structure provides natural isomorphisms

$$\psi_* \circ \phi_* \cong (\psi \circ \phi)_*,$$

which are uniquely determined by the symmetry. These isomorphisms assemble to make $C^{(-)}$ a pseudofunctor.

Definition 5.8 The *K–theory* of C is the functor

$$KC = N \circ Q_j(C^{(-)})\colon \mathscr{F} \to sSet,$$

where N is the usual nerve functor $Cat \to sSet$ and Q_j is the left 2–adjoint from Proposition 4.16 when $\mathcal{K} = Cat_2$.

Remark 5.9 Although the pseudofunctor $C^{(-)}$ satisfies the property that it maps $\underline{0}_+$ to $*$, its strictification $Q_j(C^{(-)})$ does not. Thus $Q_j(C^{(-)})$ is a functor $\mathscr{F} \to Cat$, but it is not a Γ–category as in Definition 5.1. Since $Q_j(C^{(-)})$ is levelwise equivalent to $C^{(-)}$, and in particular $Q_j(C^{(-)})(\underline{0}_+)$ is contractible, we can replace $N \circ Q_j(C^{(-)})$ by a levelwise equivalent Γ–simplicial set. This replacement is made implicitly here, and throughout the remainder of the paper.

Lemma 5.10 *Consider the composite*

$$[\mathscr{F} \times \Delta^{\mathrm{op}}, Cat] \xrightarrow{\Phi} [\mathscr{F}, [\Delta^{\mathrm{op}}, Cat]] \xrightarrow{N_* \circ -} [\mathscr{F}, [\Delta^{\mathrm{op}}, sSet]] \xrightarrow{d \circ -} [\mathscr{F}, sSet].$$

If F is a levelwise weak equivalence of diagrams $\mathscr{F} \times \Delta^{\mathrm{op}} \to Cat$, then $dN_*\Phi(F)$ is a levelwise weak equivalence of diagrams $\mathscr{F} \to sSet$.

Proof This follows from [10, Theorem B.2], which states that if $f\colon X \to Y$ is a map of bisimplicial sets such that $X_{n,\bullet} \to Y_{n,\bullet}$ is a weak equivalence of simplicial sets for all $n \geq 0$, then $d(f)\colon d(X) \to d(Y)$ is a weak equivalence. □

To relate the Γ–simplicial set ΣKC to the K–theory of the permutative Gray monoid ΣC, we provide a new construction of a special Γ–2–category $\underline{K}(\Sigma C)$ and show it is levelwise weakly equivalent to the K–theory defined in [29].

Notation 5.11 Let $2Cat_{\mathrm{p,p,m}}$ denote the tricategory whose objects are 2–categories and whose higher cells are pseudofunctors, pseudonatural transformations, and modifications [27].

Lemma 5.12 *Let $(\mathcal{D}, \oplus, e)$ be a permutative Gray monoid. Then there is a pseudofunctor of tricategories $\mathcal{D}^{(-)}\colon \mathscr{F} \to 2Cat_{\mathrm{p,p,m}}$ with value at $\underline{m}_+$ given by $\mathcal{D}^m$. If $\mathcal{D}$ has a single object, then this becomes a pseudofunctor of 2–categories $\mathcal{D}^{(-)}\colon \mathscr{F} \to 2Cat_{\mathrm{p,i}}$.*

Proof The first claim is a special case of [31, Theorem 2.5]. For the second claim, by Corollary 2.35, it suffices to work with ΣD for a permutative category D. Recall from Construction 5.7 that we have the pseudofunctor

$$D^{(-)}\colon \mathscr{F} \to \mathit{Cat}_2.$$

The permutative structure on D in fact makes each D^m a strict monoidal category with pointwise tensor product and unit, and each functor $\phi_*\colon D^m \to D^n$ for $\phi\colon \underline{m}_+ \to \underline{n}_+$ a strong monoidal functor. One can verify that the isomorphisms $\psi_* \circ \phi_* \cong (\psi \circ \phi)_*$ are themselves monoidal, so we get a pseudofunctor

$$D^{(-)}\colon \mathscr{F} \to \mathit{StMonCat}_p$$

from $\mathscr{F}$ to the 2–category $\mathit{StMonCat}_p$ of strict monoidal categories, strong monoidal functors, and monoidal natural transformations. Note that $(\Sigma D)^m \cong \Sigma(D^m)$, so we define

$$(\Sigma D)^{(-)} = \Sigma \circ D^{(-)},$$

where Σ is now the 2–functor $\mathit{StMonCat}_p \to \mathit{2Cat}_{\mathrm{p,i}}$ which views each strict monoidal category as the hom-category of a 2–category with a single object. This composite is the desired pseudofunctor. □

Definition 5.13 [42] Let $\mathcal{A}$ be a 2–category. The *nerve* of $\mathcal{A}$ is the simplicial category $N\mathcal{A}\colon \Delta^{\mathrm{op}} \to \mathit{Cat}$ defined by

$$N\mathcal{A}_n = \mathit{2Cat}_{2,\mathrm{i}}([n], \mathcal{A}),$$

where $[n]$ is the standard category $0 \to 1 \to \cdots \to n$ treated as a discrete 2–category. This is the function on objects of a 2–functor from $\mathit{2Cat}_{2,\mathrm{i}}$ to $[\Delta^{\mathrm{op}}, \mathit{Cat}_2]$.

Remark 5.14 This is called the 2*–nerve* by Lack and Paoli. It is related but not equal to the general bicategorical nerve of [25; 12]. Detailed comparisons are given in [12].

Unpacking this definition, $N\mathcal{A}_0 = \mathrm{ob}\,\mathcal{A}$ as a discrete category. When $n \geq 1$,

$$N\mathcal{A}_n = \coprod_{a_0, \dots, a_n \in \mathrm{ob}\,\mathcal{A}} \mathcal{A}(a_{n-1}, a_n) \times \cdots \times \mathcal{A}(a_0, a_1).$$

Using this same formula, we define the nerve on $\mathit{Cat}\text{-}\mathit{Grph}$, which fits in the following commuting diagram:

$$\begin{array}{ccc} \mathit{2Cat}_{2,\mathrm{i}} & \xrightarrow{N} & [\Delta^{\mathrm{op}}, \mathit{Cat}_2] \\ \downarrow & & \downarrow \\ \mathit{Cat}\text{-}\mathit{Grph} & \xrightarrow[N]{} & [\mathrm{ob}\,\Delta^{\mathrm{op}}, \mathit{Cat}_2] \end{array}$$

Let S be the 2–monad on $\mathcal{C}at\text{-}\mathcal{G}rph$ whose algebra 2–category is $2\mathcal{C}at_{2,\mathrm{i}}$ (Proposition 4.11). Let T be the 2–monad on $[\mathrm{ob}\,\Delta^{\mathrm{op}}, \mathcal{C}at_2]$ whose algebra 2–category is $[\Delta^{\mathrm{op}}, \mathcal{C}at_2]$ (Proposition 4.16). We now apply Proposition 4.2 to show that the nerve extends to $2\mathcal{C}at_{\mathrm{p,i}}$.

Lemma 5.15 *The nerve N is a strict map of 2–monads $S \to T$ and therefore provides the middle map in the commutative diagram below:*

$$\begin{array}{ccc} 2\mathcal{C}at_{2,\mathrm{i}} & \xrightarrow{N} & [\Delta^{\mathrm{op}}, \mathcal{C}at_2] \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle i} \\ 2\mathcal{C}at_{\mathrm{p,i}} & \xrightarrow{N} & \mathcal{G}ray(\Delta^{\mathrm{op}}, \mathcal{C}at_2) \\ {\scriptstyle U}\downarrow & & \downarrow{\scriptstyle U} \\ \mathcal{C}at\text{-}\mathcal{G}rph & \xrightarrow[N]{} & [\mathrm{ob}\,\Delta^{\mathrm{op}}, \mathcal{C}at_2] \end{array}$$

We now define the Γ–objects we will use to understand K–theory of a suspension.

Definition 5.16 Let C be a permutative category with ΣC its suspension permutative Gray monoid. Let $Q = Q_j(Q_i)_*$ denote the left 2–adjoint of the inclusion $J\colon [\mathscr{F}, 2\mathcal{C}at_{2,\mathrm{i}}] \hookrightarrow \mathcal{B}icat(\mathscr{F}, 2\mathcal{C}at_{\mathrm{p,i}})$ constructed in Theorem 4.19.

(1) Define $\underline{K}(\Sigma C)$ to be $Q((\Sigma C)^{(-)})$. This is a functor $\mathscr{F} \to 2\mathcal{C}at$.

(2) The composite $N \circ \underline{K}(\Sigma C)$ is a functor $\mathscr{F} \to [\Delta^{\mathrm{op}}, \mathcal{C}at]$. Define $K_{\mathrm{adj}}(\Sigma C)$ to be $\Phi^{-1}(N \circ \underline{K}(\Sigma C))$.

The composite
$$2\mathcal{C}at \xrightarrow{N} [\Delta^{\mathrm{op}}, \mathcal{C}at] \xrightarrow{N_*} [\Delta^{\mathrm{op}}, s\mathcal{S}et] \xrightarrow{d} s\mathcal{S}et$$
is one of the versions of the nerve for 2–categories in [12]. Postcomposing $\underline{K}(\Sigma C)$ with this functor (and, as noted in Remark 5.9, implicitly replacing with a reduced diagram) yields a Γ–simplicial set which is a model of the K–theory of ΣC. We make this rigorous in the following lemma, which relates the definition of K–theory here with that introduced in [29], here denoted by $\widetilde{K}$.

For a permutative Gray monoid $\mathcal{D}$, $\widetilde{K}(\mathcal{D})$ is a special Γ–2–category such that an object at level n is an object in $\mathcal{D}$, together with an explicit way of decomposing it as a sum of n objects. This allows for strict functoriality with respect to $\mathscr{F}$. This construction generalizes the construction of [47; 45] for permutative categories.

Lemma 5.17 *Let $(C, \oplus, e)$ be a permutative category. There is a levelwise weak equivalence between the Γ–2–categories $\underline{K}(\Sigma C)$ and $\widetilde{K}(\Sigma C)$, hence a stable equivalence between the spectra these represent.*

Proof We shall prove that there is a levelwise weak equivalence $\underline{K}(\Sigma C) \to \widetilde{K}(\Sigma C)$ of Γ–2–categories. Since both of these are special, it suffices to construct such a map and check that it is a weak equivalence when evaluated at $\underline{1}_+$. The functor Q is a left adjoint, so strict maps $Z\colon \underline{K}(\Sigma C) = Q((\Sigma C)^{(-)}) \to \widetilde{K}(\Sigma C)$ are in bijection with pseudonatural transformations

$$\check{Z}\colon (\Sigma C)^{(-)} \to \widetilde{K}(\Sigma C)$$

in $\mathcal{B}icat(\mathscr{F}, 2\mathcal{C}at_{\mathrm{p,i}})$. This bijection is induced by composition with a universal pseudonatural transformation $\eta\colon (\Sigma C)^{(-)} \to Q((\Sigma C)^{(-)})$, so we have the commutative triangle shown below:

$$\begin{array}{ccc} (\Sigma C)^{(-)} & \xrightarrow{\eta} & Q((\Sigma C)^{(-)}) \\ & \underset{\check{Z}}{\searrow} & \downarrow{\scriptstyle Z} \\ & & \widetilde{K}(\Sigma C) \end{array}$$

We know that η is a levelwise weak equivalence by Theorem 4.19, so the component of Z at $\underline{1}_+$ is a weak equivalence if and only if the same holds for $\check{Z}$.

We will construct the pseudonatural transformation $\check{Z}$. In order to do so, we briefly review the data which define the cells of $\widetilde{K}(\Sigma C)(\underline{n}_+)$; we omit the axioms these data must satisfy and refer the reader to [29]. Because ΣC has a single object, an object of $\widetilde{K}(\Sigma C)(\underline{n}_+)$ consists of objects $c_{s,t}$ of the permutative category C for s and t disjoint subsets of $\underline{n} = \{1, \dots, n\}$. We denote such an object by $\{c_{s,t}\}$ or, when more detail is useful, a function

$$\{s, t \mapsto c_{s,t}\}.$$

A 1–cell $\{c_{s,t}\} \to \{d_{s,t}\}$ consists of objects x_s of C for $s \subset \underline{n}$ together with isomorphisms

$$\gamma_{s,t}\colon x_t \oplus x_s \oplus c_{s,t} \cong d_{s,t} \oplus x_{s \cup t}.$$

We denote this by $\{x_s, \gamma_{s,t}\}$ or, in functional notation,

$$\{s \mapsto x_s;\ s, t \mapsto \gamma_{s,t}\}.$$

A 2–cell $\{x_s, \gamma_{s,t}\} \Rightarrow \{y_s, \delta_{s,t}\}$ consists of morphisms $\alpha_s\colon x_s \to y_s$ in C. We denote this by $\{\alpha_s\}$ or with a corresponding functional notation.

Now $(\Sigma C)^{\underline{n}_+}$ is $(\Sigma C)^n \cong \Sigma(C^n)$ by definition. We define $\check{Z}$ on cells as follows.

- The unique 0–cell of $\Sigma(C^n)$ maps to the object of $\widetilde{K}(\Sigma C)(\underline{n}_+)$ with $c_{s,t} = e$ for all s and t.

- A 1–cell $(x_1, \dots, x_n)$ maps to the 1–cell
$$\Bigl\{s \mapsto \bigoplus_{i \in s} x_i;\ s,t \mapsto \lambda_{s,t}\Bigr\},$$
where $\lambda_{s,t}$ denotes the unique interleaving symmetry isomorphism
$$\Bigl(\bigoplus_{i\in s} x_i\Bigr) \oplus \Bigl(\bigoplus_{j\in t} x_j\Bigr) \cong \bigoplus_{k \in s\cup t} x_k.$$
- A 2–cell $(f_1, \dots, f_n)$ maps to the 2–cell
$$\Bigl\{s \mapsto \bigoplus_{i\in s} f_i\Bigr\}.$$

Using the permutative structure of C, it is straightforward to verify that the formulas above satisfy the axioms of [29, Section 6.1] and therefore define valid cells. Clearly, $\check{Z}$ sends the identity 1–cell of $\Sigma(C^n)$, namely $(e, \dots, e)$, to the identity 1–cell in $\widetilde{K}(\Sigma C)(\underline{n}_+)$. Now composition of 1–cells in $\Sigma(C^n)$ is given by the monoidal structure, so
$$(x_1, \dots, x_n) \circ (y_1, \dots, y_n) = (x_1 \oplus y_1, \dots, x_n \oplus y_n).$$
We have a similar formula for composition in $\widetilde{K}(\Sigma C)(\underline{n}_+)$, with the object part of $\{x_s, \gamma_{s,t}\} \circ \{y_s, \delta_{s,t}\}$ being given on s by $x_s \oplus y_s$. From these formulas, we see that $\check{Z}$ does not strictly preserve 1–cell composition since
$$\check{Z}(x_1, \dots, x_n) \circ \check{Z}(y_1, \dots, y_n) = \Bigl\{s \mapsto \Bigl(\bigoplus_{i\in s} x_i\Bigr) \oplus \Bigl(\bigoplus_{i\in s} y_i\Bigr);\ s,t \mapsto \mu_{s,t}\Bigr\},$$
where μ denotes the unique interleaving symmetry isomorphism. On the other hand,
$$\check{Z}(x_1 \oplus y_1, \dots, x_n \oplus y_n) = \Bigl\{s \mapsto \bigoplus_{i\in s}(x_i \oplus y_i);\ s,t \mapsto \lambda_{s,t}\Bigr\}.$$
These are isomorphic by a unique symmetry, and that data equips
$$\check{Z}(\underline{n}_+)\colon (\Sigma C)^n \to \widetilde{K}(\Sigma C)(\underline{n}_+)$$
with the structure of a normal (ie strictly unit-preserving) pseudofunctor.

Now let $\phi\colon \underline{m}_+ \to \underline{n}_+$ in $\mathscr{F}$. We must construct an invertible icon in the square below:
$$\begin{array}{ccc} (\Sigma C)^m & \xrightarrow{\check{Z}} & \widetilde{K}(\Sigma C)(\underline{m}_+) \\ \downarrow{\scriptstyle \phi_*} & & \downarrow{\scriptstyle \phi_*} \\ (\Sigma C)^n & \xrightarrow[\check{Z}]{} & \widetilde{K}(\Sigma C)(\underline{n}_+) \end{array}$$

We begin by noting that this diagram obviously commutes on the unique object, so there can exist an icon (see Definition 4.8) between the two composite pseudofunctors. The top and right composite sends a 1–cell $(x_1, \dots, x_n)$ to the 1–cell

$$\Big\{u \mapsto \bigoplus_{i \in \phi^{-1}(u)} x_i;\ u, v \mapsto \lambda_{\phi^{-1}(u), \phi^{-1}(v)}\Big\}.$$

The left and bottom composite then sends $(x_1, \dots, x_n)$ to the 1–cell with

$$\Big\{u \mapsto \bigoplus_{i \in u}\Big(\bigoplus_{j \in \phi^{-1}(i)} x_j\Big);\ u, v \mapsto \kappa_{u,v}\Big\},$$

where $\kappa_{u,v}$ interleaves the blocks $\big(\bigoplus_{j \in \phi^{-1}(i)} x_j\big)$.

There is an invertible 2–cell between these 1–cells, which is given by the symmetry isomorphism

$$\bigoplus_{i \in \phi^{-1}(u)} x_i \cong \bigoplus_{i \in u}\Big(\bigoplus_{j \in \phi^{-1}(i)} x_j\Big).$$

Coherence for symmetric monoidal categories, together with the naturality of symmetries, implies that the icon axioms hold. Further, the same coherence shows that these invertible icons are themselves the naturality isomorphisms which constitute a pseudonatural transformation between pseudofunctors $\mathscr{F} \to \mathcal{2Cat}_{\mathrm{p,i}}$.

Our final task is to verify that $\check{Z}(\underline{1}_+)$ is a weak equivalence. It is a simple calculation to check that in fact $\check{Z}(\underline{1}_+)$ induces an isomorphism of 2–categories $\widetilde{K}(\Sigma C)(\underline{1}_+) \cong \Sigma C$. □

Remark 5.18 One can check that the equivalence constructed in Lemma 5.17 is pseudonatural in the variable C.

5.2 Proof of Theorem 3.11

Given a permutative category C, we can construct two pseudofunctors from $\mathscr{F}$ to $\mathcal{Bicat}(\Delta^{\mathrm{op}}, \mathcal{Cat}_2)$. One is the composite

$$\mathscr{F} \xrightarrow{(\Sigma C)^{(-)}} \mathcal{2Cat}_{\mathrm{p,i}} \xrightarrow{N} \mathcal{Gray}(\Delta^{\mathrm{op}}, \mathcal{Cat}_2) \hookrightarrow \mathcal{Bicat}(\Delta^{\mathrm{op}}, \mathcal{Cat}_2),$$

where N denotes the nerve functor of Lemma 5.15. The other is given by $\Phi(C^{(-)} \circ \wedge)$, where

$$\Phi\colon \mathcal{Bicat}(\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{Cat}_2) \to \mathcal{Bicat}(\mathscr{F}, \mathcal{Bicat}(\Delta^{\mathrm{op}}, \mathcal{Cat}_2))$$

is the 2–functor from Notation 5.3 and $C^{(-)} \circ \wedge$ is the composite

$$\mathscr{F} \times \Delta^{\mathrm{op}} \xrightarrow{\wedge} \mathscr{F} \xrightarrow{C^{(-)}} \mathcal{C}at_2.$$

Proposition 5.19 *With notation as above,* $\Phi(C^{(-)} \circ \wedge) = N \circ (\Sigma C)^{(-)}$.

Proof This result follows from a direct comparison of $\Phi(C^{(-)} \circ \wedge)$ with $N \circ (\Sigma C)^{(-)}$. Both pseudofunctors send the object $\underline{m}_+$ in $\mathscr{F}$ to the 2–functor $\Delta^{\mathrm{op}} \to \mathcal{C}at_2$ given by

$$[p] \mapsto C^{m \cdot p} = (C^m)^p, \quad ([p] \xrightarrow{\alpha} [q]) \mapsto (C^{m \cdot p} \xrightarrow{(m \wedge \alpha)_*} C^{m \cdot q}).$$

For $\Phi(C^{(-)} \circ \wedge)$ this is immediate. For $N \circ (\Sigma C)^{(-)}$ this follows because ΣC has only one object and the horizontal composition of cells is given by the monoidal product in C.

Both pseudofunctors send a morphism $\phi\colon \underline{m}_+ \to \underline{n}_+$ in $\mathscr{F}$ to the pseudonatural transformation whose component at $[p] \in \Delta^{\mathrm{op}}$ is given by

$$C^{m \cdot p} \xrightarrow{(\phi \wedge p)_*} C^{n \cdot p}.$$

For $\Phi(C^{(-)} \circ \wedge)$, it is immediate that the pseudonaturality constraint has components given by

$$(\underline{n}_+ \wedge \alpha)_* \circ (\phi \wedge [p])_* \cong (\phi \wedge \alpha)_* \cong (\phi \wedge [q])_* \circ (\underline{m}_+ \wedge \alpha)_* \tag{5-20}$$

at $\alpha\colon [p] \to [q]$. These isomorphisms are the pseudofunctoriality constraints of $C^{(-)}$ and are instances of the symmetry in C (see Construction 5.7). A straightforward check shows that the pseudofunctoriality constraint of $N \circ (\Sigma C)^{(-)}$ is given by the same instances of the symmetry of C.

For a composable pair $\phi\colon \underline{m}_+ \to \underline{n}_+$ and $\psi\colon \underline{n}_+ \to \underline{k}_+$, the symmetry of C provides

$$(\psi \wedge [p])_* \circ (\phi \wedge [p])_* \cong ((\psi \circ \phi) \wedge [p])_*$$

and these are the components of the pseudofunctoriality of $\Phi(C^{(-)} \circ \wedge)$. The same computation holds for $N \circ (\Sigma C)^{(-)}$. □

We are now ready for the main theorem of this section, from which the proof of Theorem 3.11 follows. Let Q_j be as in Definition 5.8, the left 2–adjoint to the inclusion functor

$$j\colon [\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2] \hookrightarrow \mathcal{B}icat(\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2).$$

Theorem 5.21 *For any permutative category* C*, there is a zigzag of levelwise equivalences between* $Q_j(C^{(-)}) \circ \wedge$ *and* $K_{\mathrm{adj}}(\Sigma C)$.

Proof The components of the unit and counit of the 2–adjunction $Q_j \dashv j$ are internal equivalences in $\mathcal{B}icat(\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2)$ by Proposition 4.16. Assume that

$$\alpha\colon\ j(Q_j(C^{(-)}) \circ \wedge) \xrightarrow{\simeq} j(K_{\mathrm{adj}}(\Sigma C))$$

is a pseudonatural equivalence in $\mathcal{B}icat(\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2)$. Since a pseudonatural equivalence is an internal equivalence in $\mathcal{B}icat(\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2)$, we can apply Q_j and get an internal equivalence in $[\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2]$. This gives a zigzag

$$Q_j(C^{(-)}) \circ \wedge \xleftarrow{\varepsilon} Q_j j(Q_j(C^{(-)}) \circ \wedge) \xrightarrow{Q_j(\alpha)} Q_j j(K_{\mathrm{adj}}(\Sigma C)) \xrightarrow{\varepsilon} K_{\mathrm{adj}}(\Sigma C)$$

in $[\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2]$ in which the first and third arrows are levelwise equivalences as they are internal equivalences in $\mathcal{B}icat(\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2)$, and the second arrow is a levelwise equivalence as it is an internal equivalence (ie 2–equivalence) in $[\mathscr{F} \times \Delta^{\mathrm{op}}, \mathcal{C}at_2]$. It only remains to construct an equivalence α as above.

In order to construct the pseudonatural equivalence α, first recall from Definition 5.16(2) that

$$K_{\mathrm{adj}}(\Sigma C) = \Phi^{-1}\big(N \circ Q((\Sigma C)^{(-)})\big),$$

where Φ denotes the adjunction of Notation 5.3 and Q denotes the left adjoint constructed in Theorem 4.19. We define α as the composite

$$\begin{aligned} j(Q_j(C^{(-)}) \circ \wedge) &\xrightarrow{=} jQ_j(C^{(-)}) \circ \wedge \\ &\xrightarrow{\simeq} C^{(-)} \circ \wedge \\ &\xrightarrow{\simeq} \Phi^{-1}(N \circ (\Sigma C)^{(-)}) \\ &\xrightarrow{\simeq} \Phi^{-1}\big(N \circ JQ((\Sigma C)^{(-)})\big) \\ &\xrightarrow{=} j\Phi^{-1}\big(N \circ Q((\Sigma C)^{(-)})\big) \\ &\xrightarrow{=} jK_{\mathrm{adj}}(\Sigma C), \end{aligned}$$

which we now explain. The equality giving the first arrow is a simple calculation. The equivalence giving the second arrow is a pseudoinverse of the unit for $Q_j \dashv j$, whiskered by $\wedge$ and hence still an equivalence. The equivalence giving the third arrow is the adjoint of the equality in Proposition 5.19. The equivalence giving the fourth arrow is derived from the unit of $Q \dashv J$ which is itself an equivalence, so whiskering with N and applying Φ^{-1} still yields an equivalence. The equality giving the fifth arrow follows from the commutativity of (5-4), and the equality giving the final arrow is Definition 5.16(2). □

Remark 5.22 The zigzag in Theorem 3.11 is natural up to homotopy. More precisely, this zigzag consists of three maps, two of which are counits for the 2–adjunction $Q_j \dashv j$.

It is easy to see that $C \mapsto C^{(-)}$ sends symmetric, strong monoidal functors between permutative categories to pseudonatural transformations between their corresponding pseudofunctors $\mathscr{F} \to \mathcal{C}at_2$, so a symmetric, strong monoidal functor $F\colon C \to D$ will yield a 2–natural transformation

$$Q_j(C^{(-)}) \circ \wedge \to Q_j(D^{(-)}) \circ \wedge.$$

The counit ε is strictly natural with respect to such, so the first map in our zigzag is strictly natural in symmetric, strong monoidal functors. A similar argument holds for K_{adj}, so the third map in our zigzag is also strictly natural in symmetric, strong monoidal functors. The second map is what is called $Q_j(\alpha)$ in the proof above. It is more involved, but a careful check reveals that each of the maps of which it is a composite is pseudonatural in symmetric, strong monoidal functors, and so the same will be true after applying Q_j. Thus our zigzag is actually pseudonatural in the variable C, which in particular implies that it is natural up to homotopy when viewed as a zigzag of spectra.

Proof of Theorem 3.11 On one hand, the suspension of Γ–simplicial sets given in Definition 5.5 models the stable suspension by Proposition 5.6. Recalling [56; 48], the Γ–simplicial set $KC = N \circ Q_j(C^{(-)})$ from Definition 5.8 models the K–theory spectrum of C. Its suspension as a Γ–simplicial set, $\Sigma K(C)$, is given by composing the diagonal d with $\Phi(K(C) \circ \wedge)$. By naturality of Φ in its target 2–category, this is given by $dN_*\Phi(Q_j(C^{(-)}) \circ \wedge)$. By Lemma 5.10, a levelwise weak equivalence of functors $X, Y\colon \mathscr{F} \times \Delta^{\mathrm{op}} \to \mathcal{C}at_2$ induces a levelwise weak equivalence between $dN_*\Phi(X)$ and $dN_*\Phi(Y)$. Therefore it suffices to examine $Q_j(C^{(-)}) \circ \wedge$. On the other hand, in Definition 5.16 we have the Γ–2–category $\underline{K}(\Sigma C) = Q((\Sigma C)^{(-)})$ and the related adjoint $K_{\mathrm{adj}}(\Sigma C) = \Phi^{-1}(N \circ \underline{K}(\Sigma C))$. Lemma 5.17 shows that $dN_*\Phi(K_{\mathrm{adj}}(\Sigma C))$ models the K–theory spectrum of ΣC. Finally, the result follows by Theorem 5.21, which shows that there is a zigzag of levelwise equivalences between $Q_j(C^{(-)}) \circ \wedge$ and $K_{\mathrm{adj}}(\Sigma C)$. □

References

[1] **J C Baez**, **J Dolan**, *Higher-dimensional algebra and topological quantum field theory*, J. Math. Phys. 36 (1995) 6073–6105 MR

[2] **J C Baez**, **J Dolan**, *Higher-dimensional algebra, III: n–categories and the algebra of opetopes*, Adv. Math. 135 (1998) 145–206 MR

[3] **J C Baez**, **A D Lauda**, *Higher-dimensional algebra, V: 2–groups*, Theory Appl. Categ. 12 (2004) 423–491 MR

[4] **B Bartlett**, *Quasistrict symmetric monoidal 2–categories via wire diagrams*, preprint (2014) arXiv

[5] **J Beck**, *Distributive laws*, from "Seminar on triples and categorical homology theory" (B Eckmann, editor), Springer (1969) 119–140 MR

[6] **L Bhardwaj**, **D Gaiotto**, **A Kapustin**, *State sum constructions of spin-TFTs and string net constructions of fermionic phases of matter*, J. High Energy Phys. (2017) art. id. 096 MR

[7] **R Blackwell**, **G M Kelly**, **A J Power**, *Two-dimensional monad theory*, J. Pure Appl. Algebra 59 (1989) 1–41 MR

[8] **J Bourke**, **N Gurski**, *A cocategorical obstruction to tensor products of Gray-categories*, Theory Appl. Categ. 30 (2015) 387–409 MR

[9] **J Bourke**, **N Gurski**, *The Gray tensor product via factorisation*, Appl. Categ. Structures 25 (2017) 603–624 MR

[10] **A K Bousfield**, **E M Friedlander**, *Homotopy theory of* Γ*–spaces, spectra, and bisimplicial sets*, from "Geometric applications of homotopy theory, II" (M G Barratt, M E Mahowald, editors), Lecture Notes in Math. 658, Springer (1978) 80–130 MR

[11] **R Brown**, **C B Spencer**, G*–groupoids, crossed modules and the fundamental groupoid of a topological group*, Indag. Math. 38 (1976) 296–302 MR

[12] **P Carrasco**, **A M Cegarra**, **A R Garzón**, *Nerves and classifying spaces for bicategories*, Algebr. Geom. Topol. 10 (2010) 219–274 MR

[13] **E Cheng**, **N Gurski**, *The periodic table of* n*–categories for low dimensions, I: Degenerate categories and degenerate bicategories*, from "Categories in algebra, geometry and mathematical physics" (A Davydov, M Batanin, M Johnson, S Lack, A Neeman, editors), Contemp. Math. 431, Amer. Math. Soc., Providence, RI (2007) 143–164 MR

[14] **E Cheng**, **N Gurski**, *The periodic table of* n*–categories, II: Degenerate tricategories*, Cah. Topol. Géom. Différ. Catég. 52 (2011) 82–125 MR

[15] **E Cheng**, **N Gurski**, *Iterated icons*, Theory Appl. Categ. 29 (2014) 929–977 MR

[16] **D Conduché**, *Modules croisés généralisés de longueur* 2, J. Pure Appl. Algebra 34 (1984) 155–178 MR

[17] **S Eilenberg**, **S MacLane**, *On the groups* $H(\Pi, n)$*, II: Methods of computation*, Ann. of Math. 60 (1954) 49–139 MR

[18] **S Eilenberg**, **S MacLane**, *On the groups* $H(\Pi, n)$*, III: Operations and obstructions*, Ann. of Math. 60 (1954) 513–557 MR

[19] **D S Freed**, *Anomalies and invertible field theories*, from "String-Math 2013" (R Donagi, M R Douglas, L Kamenova, M Roček, editors), Proc. Sympos. Pure Math. 88, Amer. Math. Soc., Providence, RI (2014) 25–45 MR

[20] **D S Freed**, **M J Hopkins**, **C Teleman**, *Loop groups and twisted K–theory, I*, J. Topol. 4 (2011) 737–798 MR

[21] **R Gordon**, **A J Power**, **R Street**, *Coherence for tricategories*, Mem. Amer. Math. Soc. 558, Amer. Math. Soc., Providence, RI (1995) MR

[22] **J P C Greenlees**, **J P May**, *Generalized Tate cohomology*, Mem. Amer. Math. Soc. 543, Amer. Math. Soc., Providence, RI (1995) MR

[23] **A Grothendieck**, *Pursuing stacks*, unpublished manuscript (1983)

[24] **Z-C Gu**, **X-G Wen**, *Symmetry-protected topological orders for interacting fermions: Fermionic topological nonlinear σ models and a special group supercohomology theory*, Phys. Rev. B 90 (2014) art. id. 115141

[25] **N Gurski**, *Nerves of bicategories as stratified simplicial sets*, J. Pure Appl. Algebra 213 (2009) 927–946 MR

[26] **N Gurski**, *Biequivalences in tricategories*, Theory Appl. Categ. 26 (2012) 349–384 MR

[27] **N Gurski**, *Coherence in three-dimensional category theory*, Cambridge Tracts in Mathematics 201, Cambridge Univ. Press (2013) MR

[28] **N Gurski**, *The monoidal structure of strictification*, Theory Appl. Categ. 28 (2013) 1–23 MR

[29] **N Gurski**, **N Johnson**, **A M Osorno**, *K–theory for 2–categories*, preprint (2015) arXiv

[30] **N Gurski**, **N Johnson**, **A M Osorno**, *The 2–dimensional stable homotopy hypothesis*, preprint (2017)

[31] **N Gurski**, **A M Osorno**, *Infinite loop spaces, and coherence for symmetric monoidal bicategories*, Adv. Math. 246 (2013) 1–32 MR

[32] **J F Jardine**, *Supercoherence*, J. Pure Appl. Algebra 75 (1991) 103–194 MR

[33] **N Johnson**, **A M Osorno**, *Modeling stable one-types*, Theory Appl. Categ. 26 (2012) 520–537 MR

[34] **A Joyal**, **R Street**, *Braided tensor categories*, Adv. Math. 102 (1993) 20–78 MR

[35] **D W Kahn**, *The spectral sequence of a Postnikov system*, Comment. Math. Helv. 40 (1966) 169–198 MR

[36] **M Kapranov**, *Supergeometry in mathematics and physics*, preprint (2015) arXiv

[37] **G M Kelly**, *Doctrinal adjunction*, from "Category seminar" (G M Kelly, editor), Lecture Notes in Math. 420, Springer (1974) 257–280 MR

[38] **G M Kelly**, **R Street**, *Review of the elements of 2–categories*, from "Category seminar" (G M Kelly, editor), Lecture Notes in Math. 420, Springer (1974) 75–103 MR

[39] **S Lack**, *Codescent objects and coherence*, J. Pure Appl. Algebra 175 (2002) 223–241 MR

[40] **S Lack**, *A 2–categories companion*, from "Towards higher categories" (J C Baez, J P May, editors), IMA Vol. Math. Appl. 152, Springer (2010) 105–191 MR

[41] **S Lack**, *Icons*, Appl. Categ. Structures 18 (2010) 289–307 MR

[42] **S Lack**, **S Paoli**, *2–nerves for bicategories*, K–Theory 38 (2008) 153–175 MR

[43] **J-L Loday**, *Spaces with finitely many nontrivial homotopy groups*, J. Pure Appl. Algebra 24 (1982) 179–202 MR

[44] **S MacLane**, **J H C Whitehead**, *On the 3–type of a complex*, Proc. Nat. Acad. Sci. U. S. A. 36 (1950) 41–48 MR

[45] **M A Mandell**, *An inverse K–theory functor*, Doc. Math. 15 (2010) 765–791 MR

[46] **J P May**, *E_∞ spaces, group completions, and permutative categories*, from "New developments in topology" (G Segal, editor), London Math. Soc. Lecture Note Ser. 11, Cambridge Univ. Press (1974) 61–93 MR

[47] **J P May**, *The spectra associated to permutative categories*, Topology 17 (1978) 225–228 MR

[48] **J P May**, **R Thomason**, *The uniqueness of infinite loop space machines*, Topology 17 (1978) 205–224 MR

[49] **P McCrudden**, *Balanced coalgebroids*, Theory Appl. Categ. 7 (2000) 71–147 MR

[50] **R E Mosher**, **M C Tangora**, *Cohomology operations and applications in homotopy theory*, Harper & Row, New York (1968) MR

[51] **A J Power**, *A general coherence result*, J. Pure Appl. Algebra 57 (1989) 165–173 MR

[52] **C J Schommer-Pries**, *The classification of two-dimensional extended topological field theories*, preprint (2014) arXiv

[53] **G Segal**, *Categories and cohomology theories*, Topology 13 (1974) 293–312 MR

[54] **H X Sính**, Gr*–catégories*, PhD thesis, Université Paris VII (1975)

[55] **R Street**, *Fibrations in bicategories*, Cahiers Topologie Géom. Différentielle 21 (1980) 111–160 MR

[56] **R W Thomason**, *Homotopy colimits in the category of small categories*, Math. Proc. Cambridge Philos. Soc. 85 (1979) 91–109 MR

[57] **R W Thomason**, *Symmetric monoidal categories model all connective spectra*, Theory Appl. Categ. 1 (1995) 78–118 MR

[58] **J H C Whitehead**, *Combinatorial homotopy, II*, Bull. Amer. Math. Soc. 55 (1949) 453–496 MR

Department of Mathematics, Applied Mathematics and Statistics
Case Western Reserve University
Cleveland, OH, United States

Department of Mathematics, The Ohio State University Newark
Newark, OH, United States

Department of Mathematics, Reed College
Portland, OR, United States

Department of Mathematics, University of British Columbia
Vancouver, BC, Canada

nick.gurski@case.edu, niles@math.osu.edu, aosorno@reed.edu, mstephan@math.ubc.ca

http://casfaculty.case.edu/nick-gurski/, http://nilesjohnson.net, http://people.reed.edu/~aosorno/, http://www.math.ubc.ca/~mstephan/

Received: 8 July 2016 Revised: 1 March 2017

Algebraic & Geometric Topology 17 (2017) 2807–2823

Links with finite n–quandles

JIM HOSTE
PATRICK D SHANAHAN

Associated to every oriented link L in the 3–sphere is its fundamental quandle and, for each $n > 1$, there is a certain quotient of the fundamental quandle called the n–quandle of the link. We prove a conjecture of Przytycki which asserts that the n–quandle of an oriented link L in the 3–sphere is finite if and only if the fundamental group of the n–fold cyclic branched cover of the 3–sphere, branched over L, is finite. We do this by extending into the setting of n–quandles, Joyce's result that the fundamental quandle of a knot is isomorphic to a quandle whose elements are the cosets of the peripheral subgroup of the knot group. In addition to proving the conjecture, this relationship allows us to use the well-known Todd–Coxeter process to both enumerate the elements and find a multiplication table of a finite n–quandle of a link. We conclude the paper by using Dunbar's classification of spherical 3–orbifolds to determine all links in the 3–sphere with a finite n–quandle for some n.

57M25; 57M27

1 Introduction

While the algebraic study of racks and quandles dates back to the early 1900s, Fenn and Rourke in [5] credit Conway and Wraith with introducing the concepts in 1959 as an algebraic approach to study knots and links in 3–manifolds. In the late 1900s, several mathematicians began studying similar concepts under names such as kei, distributive groupoids, crystals, and automorphic sets. In 1982, Joyce [10] published a groundbreaking work which included introducing the term quandle, giving both topological and algebraic descriptions of the fundamental quandle of a link, and proving that the fundamental quandle of a knot is a complete invariant up to reversed mirror image. Much of Joyce's work was independently discovered by Matveev [11]. In this article, we consider a quotient of the fundamental quandle of a link called the fundamental n–quandle, defined for any natural number n. Whereas the quandle of a link is usually infinite and somewhat intractable, there are many examples of knots and links for which the n–quandle is finite for some n. In his PhD thesis, Winker [13] developed a method to produce the analog of the Cayley diagram for a quandle. In addition, Winker established a relationship between the n–quandle of the link L and the fundamental

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2807

group of $\widetilde{M}_n(L)$, the n–fold cyclic branched cover of the 3–sphere, branched over L. When combined with previous work of Joyce, this implied that if the n–quandle of a link L is finite, then so is $\pi_1(\widetilde{M}_n(L))$. Przytycki (private communication, 2013) then conjectured that this condition is both necessary and sufficient, which we prove to be true in this paper. Our proof involves first generalizing a key result of Joyce: the cosets of the peripheral subgroup of a knot group can be given a quandle structure making it isomorphic to the fundamental quandle of the knot. We extend this result to the n–quandle of a knot, showing that it can also be viewed as the set of cosets of the peripheral subgroup in a certain quotient of the knot group. This result allows Winker's diagramming method to be replaced by the well known Todd–Coxeter method of coset enumeration.

We assume the reader is familiar with the theory of racks and quandles, but include basic definitions for completeness. The reader is referred to Fenn and Rourke [5], Joyce [10; 9], Matveev [11], and Winker [13] for more information. A *quandle* is a set Q together with two binary operations $\rhd$ and $\rhd^{-1}$ which satisfies the following three axioms:

(Q1) $x \rhd x = x$ for all $x \in Q$.

(Q2) $(x \rhd y) \rhd^{-1} y = x = (x \rhd^{-1} y) \rhd y$ for all $x, y \in Q$.

(Q3) $(x \rhd y) \rhd z = (x \rhd z) \rhd (y \rhd z)$ for all $x, y, z \in Q$.

A *rack* is more general, requiring only (Q2) and (Q3). It is important to note that, in general, the quandle operations are not associative. In fact, using axioms (Q2) and (Q3) it is easy to show that

$$x \rhd (y \rhd z) = ((x \rhd^{-1} z) \rhd y) \rhd z. \tag{1}$$

This property allows one to write any expression involving $\rhd$ and $\rhd^{-1}$ in a unique left-associated form (see Winker [13]). Henceforth, expressions without parenthesis are assumed to be left-associated.

Given a quandle Q, each element $q \in Q$ defines a map $S_q\colon Q \to Q$ by $S_q(p) = p \rhd q$. It follows from axiom (Q2) that S_q is a bijection and $S_q^{-1}(p) = p \rhd^{-1} q$. From axiom (Q3), it follows that S_q is a quandle homomorphism. The automorphism S_q is called the *point symmetry at* q and the set of all point symmetries generate the *inner automorphism group* $\mathrm{Inn}(Q)$. A quandle Q is *algebraically connected* if $\mathrm{Inn}(Q)$ acts transitively on Q. An *algebraic component* of Q is a maximal algebraically connected subset of Q.

In [9], Joyce defines two functors from the category of groups to the category of quandles. These functors and their adjoints will be of importance in this paper. The first, denoted

Conj, takes a group G to a quandle $Q = \mathrm{Conj}(G)$ defined as the set G with operations given by conjugation. Specifically, $x \rhd y = y^{-1}xy$ and $x \rhd^{-1} y = yxy^{-1}$. Its adjoint, denoted Adconj takes the quandle Q to the group $\mathrm{Adconj}(Q)$ generated by the elements of Q and defined by the group presentation

$$\mathrm{Adconj}(Q) = \langle \overline{q} \text{ for all } q \text{ in } Q \mid \overline{p \rhd q} = \overline{q}^{-1}\,\overline{p}\,\overline{q} \text{ for all } p \text{ and } q \text{ in } Q \rangle.$$

A quandle Q is called an *n–quandle* if each point symmetry S_q has order dividing n. It is convenient to write $x \rhd^k y$ for $S_y^k(x)$, the k^{th} power of S_y evaluated at x. Thus Q is an n–quandle if for all x and y in Q, we have $x \rhd^n y = x$. A second functor from groups to n–quandles is defined for each natural number n and is denoted Q_n. Given a group G, the n–quandle $\mathrm{Q}_n(G)$ is the set

$$\mathrm{Q}_n(G) = \{x \in G \mid x^n = 1\}$$

again with the operations given by conjugation. The adjoint of this functor is AdQ_n. If Q is any n–quandle, the group $\mathrm{AdQ}_n(Q)$ is defined by the presentation

$$\mathrm{AdQ}_n(Q) = \langle \overline{q} \text{ for all } q \text{ in } Q \mid \overline{q}^n = 1, \overline{p \rhd q} = \overline{q}^{-1}\,\overline{p}\,\overline{q} \text{ for all } p \text{ and } q \text{ in } Q \rangle.$$

Quandles may be presented in terms of generators and relators in much the same way as groups. See [5] for a rigorous development of this topic. If the quandle Q is given by the finite presentation

$$Q = \langle q_1, q_2, \dots, q_i \mid r_1, r_2, \dots, r_j \rangle,$$

then Winker proves in [13] that $\mathrm{Adconj}(Q)$ and $\mathrm{AdQ}_n(Q)$ can be finitely presented as

$$\mathrm{Adconj}(Q) = \langle \overline{q}_1, \overline{q}_2, \dots, \overline{q}_i \mid \overline{r}_1, \overline{r}_2, \dots, \overline{r}_j \rangle \tag{2}$$

and

$$\mathrm{AdQ}_n(Q) = \langle \overline{q}_1, \overline{q}_2, \dots, \overline{q}_i \mid \overline{q}_1^n = 1, \overline{q}_2^n = 1, \dots, \overline{q}_i^n = 1, \overline{r}_1, \overline{r}_2, \dots, \overline{r}_j \rangle. \tag{3}$$

Here, each quandle relation r_i is an equation between two quandle elements each expressed using the generators, the operations $\rhd$ and $\rhd^{-1}$, and parenthesis to indicate the order of operations. The associated group relation $\overline{r}_i$ must now be formed in a corresponding way using conjugation. For example, if r is the quandle relation $x = y \rhd (z \rhd^{-1} w)$, then $\overline{r}$ is the relation $\overline{x} = \overline{w}\,\overline{z}^{-1}\,\overline{w}^{-1}\,\overline{y}\,\overline{w}\,\overline{z}\,\overline{w}^{-1}$.

Associated to every oriented knot or link L in the 3–sphere $\mathbb{S}^3$ is its *fundamental quandle* $Q(L)$ which is defined by means of a presentation derived from a regular diagram D of L with a arcs and c crossings. First assign quandle generators $x_1, x_2, \dots, x_a$ to each arc of D. Next, introduce a relation r_ℓ at each crossing of D as shown in

Figure 1. It is easy to check that the three axioms, (Q1), (Q2), and (Q3), are exactly what is needed to prove that $Q(L)$ is preserved by Reidemeister moves and hence is an invariant of the link. Passing from this presentation

$$Q(L) = \langle x_1, \dots, x_a \mid r_1, \dots, r_c \rangle$$

to a presentation for $\mathrm{Adconj}(Q(L))$ by using Winker's formula (2), we obtain the well-known Wirtinger presentation of $\pi_1(\mathbb{S}^3 - L)$. Thus for any link L, $\pi_1(\mathbb{S}^3 - L) \cong \mathrm{Adconj}(Q(L))$.

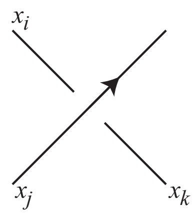


Figure 1: The relation $x_i = x_k \rhd x_j$ is associated to a crossing with arcs labeled as shown.

Joyce proves in [10] that $Q(L)$ is a complete invariant of knots up to reverse mirror image. A less sensitive, but presumably more tractable, invariant is the *fundamental n–quandle* $Q_n(L)$ which can be defined for each natural number n. If

$$Q(L) = \langle x_1, \dots, x_a \mid r_1, \dots, r_c \rangle$$

is the presentation of the fundamental quandle of L given by a diagram D and n is a fixed natural number, then the fundamental n–quandle of L is defined to be the quandle with presentation

$$Q_n(L) = \langle x_1, \dots, x_a \mid r_1, \dots, r_c, s_1, \dots, s_k \rangle$$

where the relations s_ℓ are of the form $x_i \rhd^n x_j = x_i$ for all distinct pairs of generators x_i and x_j. As before, it is easy to check that $Q_n(L)$ is an invariant of L and moreover that it is an n–quandle. Passing from this presentation of $Q_n(L)$ to a presentation for $\mathrm{AdQ}_n(Q_n(L))$ by using Winker's formula (3), we see that $\mathrm{AdQ}_n(Q_n(L))$ is a quotient of $\mathrm{Adconj}(Q(L))$. In particular, we may present $\mathrm{AdQ}_n(Q_n(L))$ by starting with the Wirtinger presentation of $\pi_1(\mathbb{S}^3 - L)$ and then adjoining the relations $x^n = 1$ for each Wirtinger generator x. While the fundamental quandle of a nontrivial knot is always infinite, the associated n–quandle is sometimes finite. Determining when this occurs is the focus of this paper.

If L is a link of more than one component, then both $Q(L)$ and $Q_n(L)$ are algebraically disconnected with one algebraic component $Q^i(L)$ and $Q^i_n(L)$, respectively, corresponding to each component K_i of L.

If K is a knot, let P be the peripheral subgroup of $G=\pi_1(\mathbb{S}^3-K)$ generated by the meridian μ and longitude λ of K. In [10], Joyce defines a quandle structure on the set of right cosets $P\backslash G$ by declaring $Pg \rhd^{\pm 1} Ph = Pgh^{-1}\mu^{\pm 1}h$. He denotes this quandle as $(P\backslash G;\mu)$ and then proves that it is isomorphic to $Q(K)$. This is the key step in Joyce's proof that the quandle is a complete knot invariant up to reverse mirror image. It also implies that the order of $Q(K)$ is the index of P in G and hence that $Q(K)$ is infinite when K is nontrivial. The key result of this paper is the following theorem which extends Joyce's result to the case of $Q_n(L)$.

Theorem 1.1 *If $L=\{K_1,K_2,\dots,K_s\}$ is a link in $\mathbb{S}^3$ and P_i is the subgroup of $\mathrm{AdQ}_n(Q_n(L))$ generated by the meridian μ_i and longitude λ_i of K_i, then the quandle $(P_i\backslash \mathrm{AdQ}_n(Q_n(L));\mu_i)$ is isomorphic to the algebraic component $Q_n^i(L)$ of $Q_n(L)$.*

Section 2 is devoted to proving Theorem 1.1. In Section 3 we use this result, as well as a theorem of Joyce, to prove the conjecture of Przytycki stated in the Abstract. Theorem 1.1 implies that the Todd–Coxeter process for coset enumeration can be used to describe $Q_n^i(L)$ provided it is finite. In Section 4 we describe this in greater detail and give examples. In the last section, we enumerate all links that have finite n–quandles for some n. In a separate set of papers, we plan to describe the n–quandles of these links, thereby providing a tabulation of all finite quandles that appear as the n–quandle of a link. The first of these papers is [7], where we describe the 2–quandle of every Montesinos link of the form $M(p_1/2,p_2/2,p/q;e)$. The authors extend their thanks to Daryl Cooper and Francis Bonahon for their assistance with Section 5. The authors also thank the referee for helpful comments.

2 Relating $Q_n(L)$ to cosets in $\mathrm{AdQ}_n(Q_n(L))$

To prove Theorem 1.1 we make use of topological descriptions of both the fundamental quandle $Q(L)$ and the n–quandle $Q_n(L)$. We begin by recalling Fenn and Rourke's formulation of $Q(L)$ given in [5] and then extend it to $Q_n(L)$. (Their formulation is actually for the rack associated to a framed link.) Let $X=\mathbb{S}^3-\mathring{N}(L)$ be the exterior of L and choose a basepoint b in X. Define $T(L)$ to be the set of all homotopy classes of paths $\alpha\colon [0,1]\to X$ such that $\alpha(0)=b$ and $\alpha(1)\in\partial X$. Moreover, we require that any homotopy be through a sequence of paths each of which starts at b and ends at ∂X. Define the two binary operations, $\rhd$ and $\rhd^{-1}$, on $T(L)$ by

$$\alpha \rhd^{\pm 1} \beta = \beta m^{\mp 1}\beta^{-1}\alpha \tag{4}$$

where m is a meridian of L. Namely, m is a loop in $\partial N(L)$ that begins and ends at $\beta(1)$, is essential in $\partial N(L)$, is nullhomotopic in $N(L)$, and has linking number $+1$

with L. Thus the arc $\alpha \rhd \beta$ is formed by starting at the basepoint b, going along β to $\partial N(L)$, traveling around m^{-1}, following β^{-1} back to the base point, and finally following α to its endpoint in $\partial N(L)$. See Figure 2. Note that the algebraic component $T^i(L)$ corresponding to the i^{th} component K_i of L consists of those paths ending at $\partial N(K_i)$. The equivalence of $Q(L)$ and $T(L)$ is proven in [5]. A similar description using "nooses" is given in [10]. In order to give a topological description of $Q_n(L)$ we introduce the following definition.

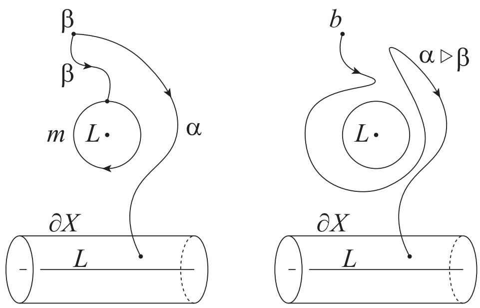


Figure 2: The topological definition of $\alpha \rhd \beta$

Definition 2.1 Suppose α is a path in X with $\alpha(0) = b$ and $\alpha(1) \in \{b\} \cup \partial X$. Suppose further that there exists t_0 with $0 \le t_0 \le 1$ such that $\alpha(t_0) \in \partial N(L)$. Let $\sigma_1(t) = \alpha(t t_0)$ and $\sigma_2(t) = \alpha((1-t)t_0 + t)$. We say that the path $\sigma_1 m^{\pm n} \sigma_2$ is obtained from α by a *$\pm n$–meridian move*. Two paths are called *n–meridionally equivalent* if they are related by a sequence of $\pm n$–meridian moves and homotopies.

We now define the n–quandle $T_n(L)$ as the set of n–meridional equivalence classes of paths with the quandle operations defined by (4). Again, paths that end at $\partial N(K_i)$ give the algebraic component $T_n^i(L)$ of $T_n(L)$.

Theorem 2.2 *The n–quandles $Q_n(L)$ and $T_n(L)$ are quandle-isomorphic.*

Proof In [5], the topological and algebraic-presentation definitions of the rack of a framed link are proven to be quandle isomorphic by constructing homomorphisms $f\colon T \to Q$ and $g\colon Q \to T$ and then showing that both $f \circ g$ and $g \circ f$ are the identity. The same maps can be used to show that $T_n(L)$ and $Q_n(L)$ are isomorphic. Rather than repeating and extending Fenn and Rourke's proof here, we simply enumerate the differences from which the interested reader can easily fill in the details of the proof.

- In [5] homotopies in T allow the endpoint of a path to move around on the chosen longitude of L given by the framing, while we allow homotopies in T_n to move the endpoint around in $\partial N(L)$. For our maps to be well-defined, this requires the idempotency axiom (Q1) which is not present in a rack.

- In T_n we allow n–meridional moves that are not present in T. In order for our maps to be well-defined this requires the addition of the corresponding relations $q_i \rhd^n q_j = q_i$ to Q_n. □

We are now prepared to prove Theorem 1.1.

Theorem 1.1 *If $L = \{K_1, K_2, \dots, K_s\}$ is a link in $\mathbb{S}^3$ and P_i is the subgroup of $\mathrm{AdQ}_n(Q_n(L))$ generated by the meridian μ_i and longitude λ_i of K_i, then the quandle $(P_i \backslash \mathrm{AdQ}_n(Q_n(L)); \mu_i)$ is isomorphic to the algebraic component $Q_n^i(L)$ of $Q_n(L)$.*

Proof Suppose that $L = \{K_1, K_2, \dots, K_s\}$. Without loss of generality, we shall prove the theorem for the first component K_1. We begin by fixing some element $\nu \in Q_n(L)$ which we think of as a path from the basepoint b in X to $\partial N(K_1)$. We now define a map $\tau\colon \mathrm{AdQ}_n(Q_n(L)) \to Q_n(L)$ by $\tau(\alpha) = \alpha^{-1}\nu$.

Claim 1 *The map τ is onto $Q_n^1(L)$.*

Proof Let σ be a path representing any element of $Q_n^1(L)$. Move σ by a homotopy until $\sigma(1) = \nu(1)$ and let α be the loop $\alpha = \nu\sigma^{-1}$. Now $\tau(\alpha) = \alpha^{-1}\nu = \sigma\nu^{-1}\nu = \sigma$.

Let P_ν be the subgroup of $\mathrm{AdQ}_n(Q_n(L))$ generated by the meridian $\mu_1 = \nu m \nu^{-1}$ and longitude $\lambda_1 = \nu \ell \nu^{-1}$ of K_1.

Claim 2 $$\tau^{-1}(\nu) = P_\nu.$$

Proof Notice first that $\tau^{-1}(\nu)$ is a subgroup of $\mathrm{AdQ}_n(Q_n(L))$. For suppose that $\alpha, \beta \in \tau^{-1}(\nu)$. Now $\tau(\alpha\beta^{-1}) = \beta\alpha^{-1}\nu = \beta\nu = \nu$ because $\alpha^{-1}\nu = \nu$ and $\beta^{-1}\nu = \nu$ implies $\nu = \beta\nu$. Thus to show that $P_\nu \subset \tau^{-1}(\nu)$ we need only show that $\mu, \lambda \in \tau^{-1}(\nu)$. But $\tau(\lambda) = \lambda^{-1}\nu = (\nu\ell\nu^{-1})^{-1}\nu = \nu\ell^{-1}\nu^{-1}\nu = \nu\ell^{-1} = \nu$ because $\ell \subset \partial X$. Similarly, $\mu \in \tau^{-1}(\nu)$.

Now suppose that $\alpha \in \tau^{-1}(\nu)$. This means that $\alpha^{-1}\nu$ can be taken to ν by a sequence of n–meridian moves separated by homotopies. We illustrate the situation in Figure 3. The first homotopy begins at $\alpha^{-1}\nu$ and ends at the path $\sigma_1\rho_1$ where $\sigma_1(1) = \rho_1(0)$ is a point in ∂X. We then do an n–meridian move, replacing $\sigma_1\rho_1$ with the path $\sigma_1 m^{\pm n}\rho_1$. This path is then homotopic to the path $\sigma_2\rho_2$ and so on until finally the last homotopy ends at ν. For simplicity, the Figure illustrates the case of three homotopies separated by two n–meridian moves. Notice that the "right edge" of the i^{th} homotopy defines a path in $\partial N(K_1)$ which we call β_i. These homotopies can be reparametrized so that the polygonal paths indicated in each homotopy depict the new level sets. The first homotopy can now be thought of as one between the loop α and the loop $\nu\beta_1\rho_1^{-1}\sigma_1^{-1}$. We then perform an n–meridian move to this loop and continue through

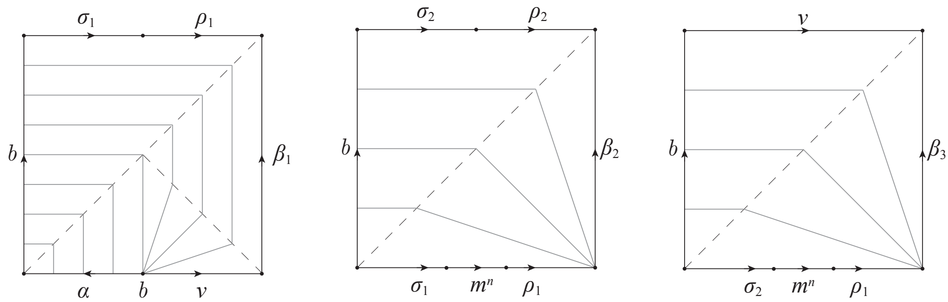


Figure 3: Homotopies separated by n–meridian moves

the second homotopy, ending at the loop $\nu\beta_1\beta_2\rho_2^{-1}\sigma_2^{-1}$. Eventually we arrive at the loop $\nu\beta_1\beta_2\dots\beta_k\nu^{-1}$, an element of P_ν. Thus α represents an element of P_ν and hence $\tau^{-1}(\nu)\subset P_\nu$.

Claim 3 *Let ϕ_1 be the automorphism of* $\mathrm{AdQ}_n(Q_n(L))$ *given by conjugation by μ_1. Then ϕ_1 fixes every element of P_ν.*

Proof Suppose that $\nu\beta\nu^{-1}\in P_\nu$. Now

$$\begin{aligned}\phi_1(\nu\beta\nu^{-1}) &= \mu_1^{-1}\nu\beta\nu^{-1}\mu_1\\ &= (\nu m\nu^{-1})^{-1}\nu\beta\nu^{-1}(\nu m\nu^{-1})\\ &= \nu m^{-1}\beta m\nu^{-1}\\ &= \nu m^{-1}m\beta\nu^{-1}\\ &= \nu\beta\nu^{-1}\end{aligned}$$

because loops in $\partial N(K_1)$ commute.

We can now turn the set of right cosets $P_\nu\backslash\mathrm{AdQ}_n(Q_n(L))$ into a quandle, which we denote as $(P_\nu\backslash\mathrm{AdQ}_n(Q_n(L));\mu_1)$ by defining

$$\begin{aligned}P_\nu\alpha \rhd^{\pm1} P_\nu\beta &= P_\nu\phi_1^{\pm1}(\alpha\beta^{-1})\beta\\ &= P_\nu\mu_1^{\mp1}\alpha\beta^{-1}\mu_1^{\pm1}\beta\\ &= P_\nu\alpha\beta^{-1}\mu_1^{\pm1}\beta\end{aligned}\tag{5}$$

because $\mu_1\in P_\nu$.

Claim 4 *The quandle operations defined in* (5) *are well-defined.*

Proof Suppose that $P_\nu\alpha = P_\nu a$ and $P_\nu\beta = P_\nu b$. Then

$$\begin{aligned}\alpha\beta^{-1}\mu_1^{\pm 1}\beta(ab^{-1}\mu_1^{\pm 1}b)^{-1} &= \alpha\beta^{-1}\mu_1^{\pm 1}\beta b^{-1}\mu_1^{\mp 1}ba^{-1}\\ &= \alpha\beta^{-1}\beta b^{-1}ba^{-1}\\ &= \alpha a^{-1} \in P_\nu\end{aligned}$$

because conjugation by $\mu_1^{\pm 1}$ fixes βb^{-1}, an element of P_ν. Hence $P_\nu\alpha \rhd^{\pm 1} P_\nu\beta = P_\nu a \rhd^{\pm 1} P_\nu b$.

Claim 5 *The map τ determines a quandle isomorphism from $(P_\nu\backslash \mathrm{AdQ}_n(Q_n(L)); \mu_1)$ to $Q_n^1(L)$.*

Proof Define $\overline{\tau}$: $(P_\nu\backslash \mathrm{AdQ}_n(Q_n(L)); \mu) \to Q_n^1(L)$ as $\overline{\tau}(P_\nu\alpha) = \tau(\alpha)$. Because $\tau^{-1}(\nu) = P_\nu$, it follows easily that $\overline{\tau}$ is both well-defined and injective. Because τ is onto $Q^1(L)$, we also have that $\overline{\tau}$ is onto $Q_n^1(L)$. Thus $\overline{\tau}$ is a bijection. However, $\overline{\tau}$ is also a quandle homomorphism because

$$\begin{aligned}\overline{\tau}(P_\nu\alpha \rhd P_\nu\beta) &= \overline{\tau}(P_\nu\alpha\beta^{-1}\mu_1^{-1}\beta)\\ &= \tau(\alpha\beta^{-1}\mu_1^{-1}\beta)\\ &= \beta^{-1}\mu_1^{-1}\beta\alpha^{-1}\nu\\ &= (\beta^{-1}\nu)m^{-1}(\beta^{-1}\nu)^{-1}(\alpha^{-1}\nu)\\ &= \tau(\alpha) \rhd \tau(\beta)\\ &= \overline{\tau}(P_\nu\alpha) \rhd \overline{\tau}(P_\nu\beta).\end{aligned}$$

This completes the proof of Theorem 1.1. □

3 Przytycki's conjecture

In this section we prove the conjecture of Przytycki stated in the abstract.

Theorem 3.1 *Let L be an oriented link in $\mathbb{S}^3$ and let $\widetilde{M}_n(L)$ be the n–fold cyclic branched cover of $\mathbb{S}^3$, branched over L. Then $Q_n(L)$ is finite, if and only if $\pi_1(\widetilde{M}_n(L))$ is finite.*

Before giving the proof of Theorem 3.1, we point out the relationship between $\pi_1(\widetilde{M}_n(L))$ and a certain subgroup of $\mathrm{AdQ}_n(Q_n(L))$. The reader is referred to [13] for more details. If $M_n(L)$ is the n–fold cyclic cover of $\mathbb{S}^3 - L$, then $\pi_1(M_n(L))$ is isomorphic to the subgroup E^0 of $\pi_1(\mathbb{S}^3 - L) \cong \mathrm{Adconj}(Q(L))$ consisting of those loops in $\mathbb{S}^3 - L$ that lift to loops in the cover. Equivalently, E^0 consists of loops having total linking number zero with L, that is, those loops α such that the sum of the linking numbers of α with each component of L is zero. The subgroup E^0 can also be described as those elements of $\pi_1(\mathbb{S}^3 - L)$ which, when written as words in the Wirtinger generators, have total exponent sum equal to zero. This concept is well-defined, and defines a subgroup, because each of the relators in the Wirtinger presentation has total exponent sum equal to zero. This last description extends to the quotient group $\mathrm{AdQ}_n(Q_n(L))$. Let E_n^0 be the subgroup of $\mathrm{AdQ}_n(Q_n(L))$ consisting of all elements with total exponent sum equal to zero modulo n. In order to obtain the fundamental group of the cyclic branched cover we must algebraically kill the n^{th} power of each Wirtinger generator in E^0, hence,

$$\pi_1(\widetilde{M}_n(L)) \cong E_n^0. \tag{6}$$

Notice further, that the index of E_n^0 in $\mathrm{AdQ}_n(Q_n(L))$ is n.

One direction of Theorem 3.1 follows from work that appears in the PhD thesis of Joyce [9]. For completeness, and because this result does not appear in Joyce's paper [10], we reproduce his proof here (with some modification).

Theorem 3.2 (Joyce) *If Q_n is any finite n–quandle, then $|\mathrm{AdQ}_n(Q_n)| \le n^{|Q_n|}$ and hence $\mathrm{AdQ}_n(Q_n)$ is finite.*

Proof Suppose that Q_n is a finite n–quandle with elements $\{q_1, q_2, \dots, q_k\}$. Now $\mathrm{AdQ}_n(Q_n)$ is generated by the ordered set of elements $\overline{q}_1, \overline{q}_2, \dots, \overline{q}_k$ so that every element in $\mathrm{AdQ}_n(Q_n)$ is a word in these generators and their inverses.

Claim 1 *If $w = \overline{q}_{i_1}^{\epsilon_1} \overline{q}_{i_2}^{\epsilon_2} \cdots \overline{q}_{i_m}^{\epsilon_m}$, where each exponent is ± 1, then we may rewrite w as $w = \overline{q}_{j_1}^{\eta_1} \overline{q}_{j_2}^{\eta_2} \cdots \overline{q}_{j_m}^{\eta_m}$, where each exponent is ± 1, $j_1 = \min(j_1, j_2, \dots, j_m)$ and $j_1 \le \min(i_1, i_2, \dots, i_m)$.*

Proof Suppose $\overline{q}_{i_k}^{\epsilon_k}$ is the first occurrence of the generator with smallest index and that $k > 1$. Now $q_{i_{k-1}} \rhd^{\epsilon_k} q_{i_k} = q_t$ for some t and so $\overline{q}_{i_{k-1}}^{\epsilon_{k-1}} \overline{q}_{i_k}^{\epsilon_k} = \overline{q}_{i_k}^{\epsilon_k} \overline{q}_t^{\epsilon_{k-1}}$. If we replace $\overline{q}_{i_{k-1}}^{\epsilon_{k-1}} \overline{q}_{i_k}^{\epsilon_k}$ with $\overline{q}_{i_k}^{\epsilon_k} \overline{q}_t^{\epsilon_{k-1}}$ in w, then either the first occurrence of the generator with smallest index has moved one place closer to the beginning of w, or a new generator of smaller index was introduced if $t < i_k$. Hence, after a finite number of steps of this kind, the first generator of w will have the smallest index and it will be no greater than any of the indices in the original word.

Claim 2 *If $w = \overline{q}_{i_1}^{\epsilon_1} \overline{q}_{i_2}^{\epsilon_2} \cdots \overline{q}_{i_m}^{\epsilon_m}$, where each exponent is ± 1, then we may rewrite w as $w = \overline{q}_{j_1}^{\eta_1} \overline{q}_{j_2}^{\eta_2} \cdots \overline{q}_{j_m}^{\eta_m}$, where each exponent is ± 1 and $j_1 \le j_2 \le \cdots \le j_m$.*

Proof We proceed by induction on m. The case with $m = 2$ is a direct consequence of Claim 1. Assume now that the result is true for words of length m and suppose that $w = \overline{q}_{i_1}^{\epsilon_1} \overline{q}_{i_2}^{\epsilon_2} \cdots \overline{q}_{i_{m+1}}^{\epsilon_{m+1}}$. Applying the inductive hypothesis to the last m generators of w, we may assume that $i_2 \le i_3 \le \cdots \le i_{m+1}$. If $i_1 \le i_2$, we are done. If not, apply Claim 1 to w, which will strictly decrease the index of the first generator in w, and then again apply the inductive hypothesis to the last m generators. This cannot continue forever because the index of the first generator in w cannot decrease below 1.

We may now write any word in $\mathrm{AdQ}_n(Q_n)$ as $\overline{q}_1^{r_1}\overline{q}_2^{r_2}\cdots\overline{q}_k^{r_k}$ and, using the fact that $\overline{q}_i^n = 1$, we may assume that $0 \le r_i < n$ for each i. There are at most $n^k = n^{|Q_n|}$ words of this kind. □

Proof of Theorem 3.1 Suppose L is an oriented link and $Q_n(L)$ is finite. By Theorem 3.2, it follows that $\mathrm{AdQ}_n(Q_n(L))$ is finite. Hence the subgroup E_n^0 of $\mathrm{AdQ}_n(Q_n(L))$ is finite and so $\pi_1(\widetilde{M}_n(L))$ is finite by (6).

Now suppose that $\pi_1(\widetilde{M}_n(L))$ is finite. Because E_n^0 has finite index in $\mathrm{AdQ}_n(Q_n(L))$, it follows that $\mathrm{AdQ}_n(Q_n(L))$ is finite. Hence, for each component K_i of L, the set of cosets $P_i \backslash \mathrm{AdQ}_n(Q_n(L))$ is finite and therefore, by Theorem 1.1, each algebraic component $Q_n^i(L)$ of $Q_n(L)$ is finite. □

4 Examples

From the proof of Theorem 3.1, all information about the knot invariant $Q_n(L)$ is encoded by the cosets of the subgroups P_i in the group $\mathrm{AdQ}_n(Q_n(L))$. For example, if $Q_n(L)$ is finite, then

$$|Q_n(L)| = \sum_{i=1}^{s}[\mathrm{AdQ}_n(Q_n(L)) : P_i].$$

Algorithmically computing the index of P_i in the group $\mathrm{AdQ}_n(Q_n(L))$ from a presentation of the group is a well-known problem in computational group theory. The first process to accomplish this task was introduced by Todd and Coxeter in 1936 [3] and is now a fundamental method in computational group theory. In addition to determining the index (if it is finite), the Todd–Coxeter process also provides a Cayley diagram that represents the action of right-multiplication on the cosets. In this section we will apply the Todd–Coxeter process to several examples and determine the quandle multiplication table from the Cayley diagram of the cosets. More detailed treatments of the Todd–Coxeter process can be found in [6] and [8].

Consider the right-hand trefoil knot K and fix $n = 3$. From the Wirtinger presentation we obtain the presentation

$$\mathrm{AdQ}_3(Q_3(K)) = \langle x, y \mid x^3 = 1, y^3 = 1, x^{-1}y^{-1}xyxy^{-1} = 1\rangle.$$

A meridian for K is $\mu = x$ and a (nonpreferred) longitude is $\lambda = yxxy$. The Todd–Coxeter process produces a *coset table* whose rows are numbered by indices $\alpha \in \{1, 2, \dots, \kappa\}$ that represent cosets of P. The columns are labeled by the generators and their inverses and encode the action of $\mathrm{AdQ}_3(Q_3(K))$ on the cosets by right-multiplication. An additional column will be added to give a representative $\phi(\alpha) \in \mathrm{AdQ}_3(Q_3(K))$ of coset α.

We initialize the coset table by letting 1 represent the trivial coset P, thus $\phi(1) = e$ is a representative of this coset (we use e here for the identity element of $\mathrm{AdQ}_3(Q_3(K))$ to avoid confusion). Since $\mu = x \in P$, we have $Px = P$, this information is encoded in a *helper table* where P is represented by index 1 and is encoded in the coset table as a relation $1x = 1$. Of course, it follows from this that $1x^{-1} = 1$ as well, so there are two defined entries in row 1 of the coset table:

	x	y	x^{-1}	y^{-1}	ϕ
1	1		1		e

	x	
1		1

Since $\lambda = yxxy \in P$ we also produce a helper table to encode $1yxxy = 1$. Additional entries in the table are required to represent the cosets $1y$, $1yx$, and $1yxx$. These entries are defined by adding indices 2, 3, and 4, respectively, and adding additional information to the coset table for these indices coming from the helper table. For example, 2 is defined to be the coset $1y$ and, thus, $1y = 2$ and $2y^{-1} = 1$ are encoded in the coset table. At this point a *deduction* also occurs. Since $1yxxy = 1$, we see in the helper table that $4y = 1$:

	x	y	x^{-1}	y^{-1}	ϕ
1	1	2	1	4	e
2	3			1	y
3	4		2		yx
4		1	3		yx^2

	y	x	x	y
1	2	3	4	1

This completes the initial set up of the coset table and is referred to as *scanning* the generators of P. The Todd–Coxeter process next proceeds to scan the relations of $\mathrm{AdQ}_3(Q_3(K))$ for all indices. This encodes the fact that if α is any coset and $w = e \in \mathrm{AdQ}_3(Q_3(K))$, then $\alpha w = \alpha$ in the coset table since $P\phi(\alpha)w = P\phi(\alpha)$ in $\mathrm{AdQ}_3(Q_3(K))$. We scan the three relations $x^3 = e$, $y^3 = e$, and $x^{-1}y^{-1}xyxy^{-1} = e$, in this order, for each index, defining new indices and obtaining new deductions along the way.

Scanning x^3 for $\alpha = 1$ gives no new information. Scanning y^3 gives no new definitions but does produce the deduction $2y = 4$ and scanning $x^{-1}y^{-1}xyxy^{-1}$ defines the indices 5 and 6 as shown in the coset tables below:

	x	y	x^{-1}	y^{-1}	ϕ
1	1	2	1	4	e
2	3	4		1	y
3	4		2		yx
4		1	3	2	yx^2

y	y	y	
1	2	4	1

	x	y	x^{-1}	y^{-1}	ϕ
1	1	2	1	4	e
2	3	4	6	1	y
3	4		2		yx
4	5	1	3	2	yx^2
5		6	4		yx^3
6	2			5	yx^3y

x^{-1}	y^{-1}	x	y	x	y^{-1}	
1	1	4	5	6	2	1

At this point we see that the representative for coset 5 is $\phi(5) = yx^3$. Since $x^3 = e$ in the group $\phi(5) = yx^3 = y = \phi(2)$ and so the cosets 5 and 2 are the same. This information is determined by a *coincidence* which occurs when scanning x^3 for $\alpha = 2$. Filling in the entries of the helper table from left to right, $2x = 3$, $3x = 4$, $4x = 5$. However we require $2xxx = 2$ thus we see that $5 = 2$. In the coset table we process this coincidence by replacing all values of 5 with 2, merging the data from row 5 into row 2, and then deleting row 5. In merging the data from 5 to 2 we see a new coincidence, namely $6 = 4$ and so we repeat the coincidence procedure for $6 = 4$ before moving on to the next scan:

	x	y	x^{-1}	y^{-1}	ϕ
1	1	2	1	4	e
2	3	4	~~6~~4	1	y
3	4		2		yx
4	~~5~~2	1	3	2	yx^2
~~5~~		~~6~~	~~4~~		yx^3
~~6~~	~~2~~			~~5~~~~2~~	yx^3y

	x	x	x
2	3	4	5 = 2

Scanning $x^{-1}y^{-1}xyxy^{-1}$ for $\alpha = 2$ completes the table. The process terminates after the table is complete and all relations have been scanned for all indices. In our example, no additional coincidences occur and the completed table is shown in Table 1.

It is important to note that the operation encoded by the coset table is that of right-multiplication. It is not the operations of $\triangleright^{\pm 1}$ in the quandle $P \backslash \mathrm{AdQ}_3(Q_3(K))$. The

	x	y	x^{-1}	y^{-1}	ϕ
1	1	2	1	4	e
2	3	4	4	1	y
3	4	3	2	3	yx
4	2	1	3	2	yx^2

Table 1: Completed coset table for $P\backslash \mathrm{AdQ}_3(Q_3(K))$, where K is the trefoil knot

$\triangleright$	P	Py	Pyx	Pyx^2
P	P	Pyx^2	Py	Pyx
Py	Pyx	Py	Pyx^2	P
Pyx	Pyx^2	P	Pyx	Py
Pyx^2	Py	Pyx	P	Pyx^2

Table 2: The multiplication table for $Q_3(K)$, where K is the trefoil knot

n	$\lvert P\rvert$	$\lvert Q_n(K)\rvert$	$\lvert \mathrm{AdQ}_n(Q_n(K))\rvert$	$\lvert\pi_1(\widetilde{M}_n(K))\rvert$
2	2	3	6	3
3	6	4	24	8
4	16	6	96	24
5	50	12	600	120

Table 3: The order of $\mathrm{AdQ}_n(Q_n(K))$ and index of P for the right-handed trefoil

multiplication table for the quandle can be easily worked out, however, from the coset table and the definition of the operations $Pg \triangleright^{\pm 1} Ph = Pgh^{-1}x^{\pm 1}h$ since $\mu = x$. From the completed coset table, the quandle $Q_3(K)$ has four elements P, Py, Pyx, and Pyx^2. So, for example, $Py \triangleright Pyx = Pyx^{-1}y^{-1}xyx$. This coset is represented by $1yx^{-1}y^{-1}xyx = 4$ in the coset table. Therefore, $Py \triangleright Pyx = Pyx^2$. The full multiplication table for $Q_3(K)$ is given in Table 2.

Applying the Todd–Coxeter method in the case of the trefoil for $n = 2, 3, 4, 5$, enumerating the cosets of both the trivial subgroup as well as $P = \langle \mu, \lambda \rangle$, we obtain the data in Table 3. These calculations agree with the well known fact that $\pi_1(\widetilde{M}_n(K))$ for the trefoil with $n = 2, 3, 4$, or 5 is, respectively, the cyclic group of order 3, the quaternion group of order 8, the binary tetrahedral group of order 24, and the binary icosahedral group of order 120. See [12].

As another example, consider the $(2, 2, 3)$–pretzel link L and fix $n = 2$. Starting with the standard pretzel diagram with Wirtinger generators x, y, z, we obtain the following presentation of $\mathrm{AdQ}_2(Q_2(L))$:

$$\langle x, y, z \mid x^{-1}z^{-1}xzxy^{-1}x^{-1}y = 1,\ y^{-1}x^{-1}yxyzyz^{-1}y^{-1}z^{-1} = 1,\\ y^{-1}x^{-1}yxyzyz^{-1}y^{-1}x^{-1}y^{-1}xyx^{-1}z^{-1}z = 1,\ x^2 = 1,\ y^2 = 1,\ z^2 = 1\rangle.$$

The link L has two components and subgroups generated by a meridian and longitude of each component are $P_1 = \langle x, x^{-1}zxy^{-1}\rangle$ and $P_2 = \langle y, y^{-1}x^{-1}yzyz^{-1}xzyzy^{-1}\rangle$. Applying the Todd–Coxeter process for each of these subgroups gives

$$|Q_2(L)| = [\mathrm{AdQ}_2(Q_2(L)) : P_1] + [\mathrm{AdQ}_2(Q_2(L)) : P_2] = 8 + 24 = 32.$$

These calculations agree with Theorem 1.1 of [7] where it is shown using Winker's diagramming method [13] that if L is the Montesinos link of the form $(1/2, 1/2, p/q; e)$, then $|Q_2(L)| = 2(q+1)|(e-1)q-p|$. For the $(2,2,3)$–pretzel link L we have $p = 1$, $q = 3$, and $e = 0$.

5 Links with finite *n*–quandles

The set of links which have a finite n–quandle for some n can be derived from Thurston's geometrization theorem. To see this, let L be a link and $n > 1$ an integer such that $Q_n(L)$ is finite. By Theorem 3.1, we have that $\pi_1(\widetilde{M}_n(L))$ is finite. Define $\mathcal{O}(L,n)$ to be the 3–orbifold with underlying space $\mathbb{S}^3$ and singular locus L where each component of L is labeled n. (Both [1] and [2] are excellent references for orbifolds.) We now have a manifold covering of the orbifold, $p\colon \widetilde{M}_n(L) \to \mathcal{O}(L,n)$, and the covering map p induces a homomorphism $p_*\colon \pi_1(\widetilde{M}_n(L)) \to \pi_1^{\mathrm{orb}}(\mathcal{O}(L,n))$ for which the index of $p_*(\pi_1(\widetilde{M}_n(L)))$ in $\pi_1^{\mathrm{orb}}(\mathcal{O}(L,n))$ is the branch index n. Since $\pi_1(\widetilde{M}_n(L))$ is finite, it follows that $\pi_1^{\mathrm{orb}}(\mathcal{O}(L,n))$ is finite. In addition, the universal orbifold cover of $\mathcal{O}(L,n)$ is a simply connected manifold (equal to the universal cover of $\widetilde{M}_n(L)$) and, since $\pi_1^{\mathrm{orb}}(\mathcal{O}(L,n))$ is finite, the universal cover is also compact. Now Thurston's geometrization theorem asserts that the only compact, simply connected 3–manifold is $\mathbb{S}^3$. Therefore, $\mathcal{O}(L,n)$ is a spherical 3–orbifold. In [4], Dunbar classifies all geometric, nonhyperbolic 3–orbifolds. The following, obtained from Dunbar, is the complete list of all spherical 3–orbifolds with underlying space $\mathbb{S}^3$ and singular locus L with each component labeled n. Therefore, it also represents the list of all links in $\mathbb{S}^3$ with finite $Q_n(L)$ for some n.

In Table 4, we list the links as they appear in [4]. A box labeled k denotes k left-handed half twists between the two strands and a box labeled m/n denotes the m/n rational tangle with $-n/2 \le m \le n/2$ and $m \ne 0$. See [4] for a detailed explanation.

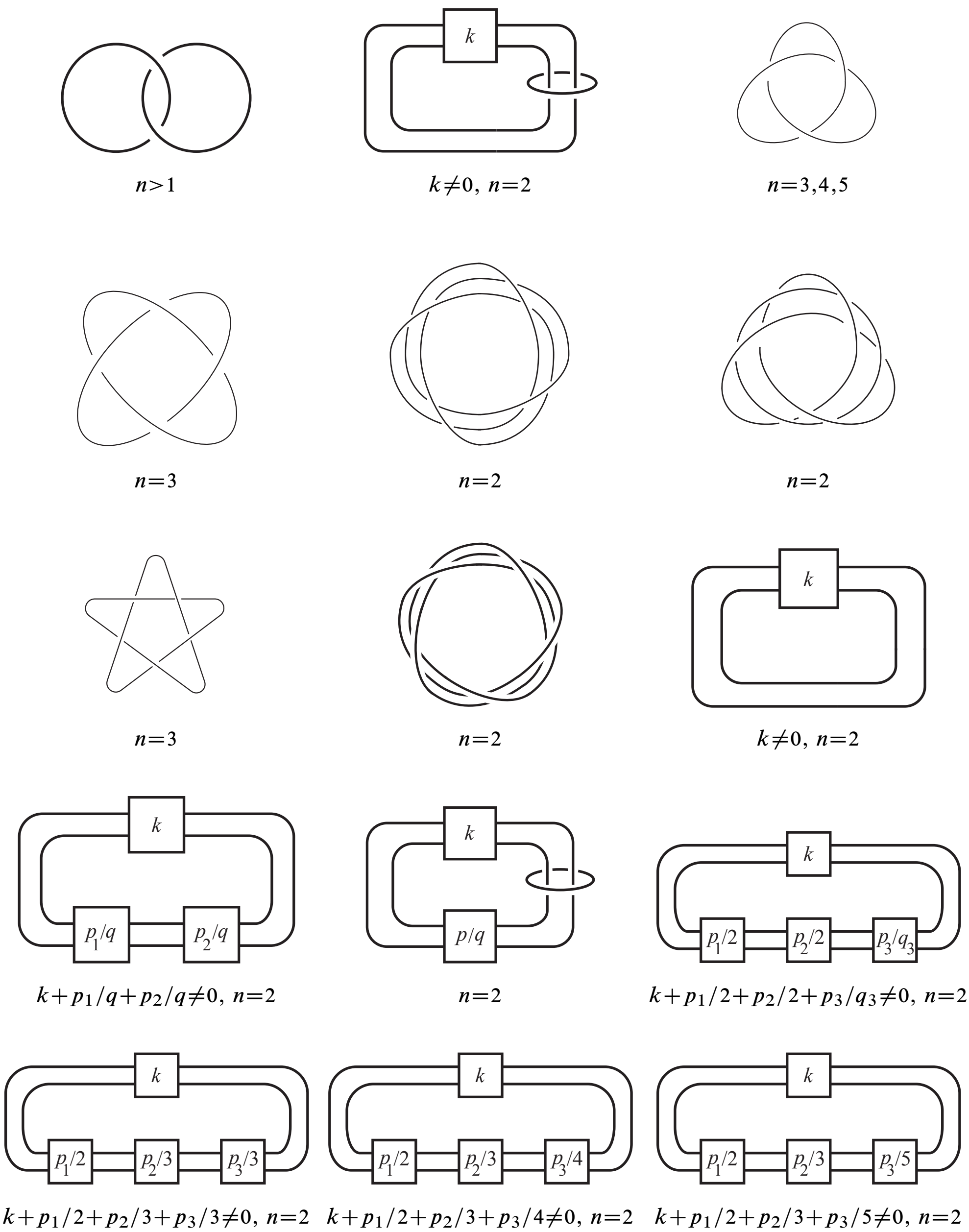


Table 4: Links $L \subset \mathbb{S}^3$ with finite $Q_n(L)$

References

[1] **M Boileau**, **B Leeb**, **J Porti**, *Geometrization of* 3*–dimensional orbifolds*, Ann. of Math. 162 (2005) 195–290 MR

[2] **D Cooper**, **C D Hodgson**, **S P Kerckhoff**, *Three-dimensional orbifolds and cone-manifolds*, MSJ Memoirs 5, Math. Soc. of Japan, Tokyo (2000) MR

[3] **H S M Coxeter**, **J A Todd**, *A practical method for enumerating cosets of a finite abstract group*, Proc. Edinburgh Math. Soc. Ser. II 5 (1936) 26–34

[4] **W D Dunbar**, *Geometric orbifolds*, Rev. Mat. Univ. Complut. Madrid 1 (1988) 67–99 MR

[5] **R Fenn**, **C Rourke**, *Racks and links in codimension two*, J. Knot Theory Ramifications 1 (1992) 343–406 MR

[6] **D F Holt**, **B Eick**, **E A O'Brien**, *Handbook of computational group theory*, Chapman & Hall, Boca Raton, FL (2005) MR

[7] **J Hoste**, **P D Shanahan**, *Involutory quandles of* $(2,2,r)$*–Montesinos links*, J. Knot Theory Ramifications 26 (2017) art. id. 1741003, 19 pages MR

[8] **D L Johnson**, *Presentations of groups*, 2nd edition, London Math. Soc. Student Texts 15, Cambridge Univ. Press (1997) MR

[9] **D E Joyce**, *An algebraic approach to symmetry with applications to knot theory*, PhD thesis, University of Pennsylvania (1979) MR

[10] **D Joyce**, *A classifying invariant of knots, the knot quandle*, J. Pure Appl. Algebra 23 (1982) 37–65 MR

[11] **S V Matveev**, *Distributive groupoids in knot theory*, Mat. Sb. 119(161) (1982) 78–88, 160 MR Translated in Math. USSR-Sb 47 (1984) 73–83

[12] **D Rolfsen**, *Knots and links*, Math. Lect. Ser. 7, Publish or Perish, Houston (1990) MR Corrected reprint

[13] **S K Winker**, *Quandles, knot invariants, and the* n*–fold branched cover*, PhD thesis, Univ. of Illinois at Chicago (1984) MR Available at `https://search.proquest.com/docview/303377949`

Department of Mathematics, Pitzer College
Claremont, CA, United States

Department of Mathematics, Loyola Marymount University
Los Angeles, CA, United States

`jim_hoste@pitzer.edu`, `pshanahan@lmu.edu`

Received: 14 July 2016 Revised: 10 March 2017

Algebraic & Geometric Topology 17 (2017) 2825–2840

Vanishing of L^2–Betti numbers and failure of acylindrical hyperbolicity of matrix groups over rings

Feng Ji
Shengkui Ye

Let R be an infinite commutative ring with identity and $n \geq 2$ an integer. We prove that for each integer $i = 0, 1, \dots, n-2$, the L^2–Betti number $b_i^{(2)}(G)$ vanishes when G is the general linear group $\mathrm{GL}_n(R)$, the special linear group $\mathrm{SL}_n(R)$ or the group $E_n(R)$ generated by elementary matrices. When R is an infinite principal ideal domain, similar results are obtained when G is the symplectic group $\mathrm{Sp}_{2n}(R)$, the elementary symplectic group $\mathrm{ESp}_{2n}(R)$, the split orthogonal group $O(n,n)(R)$ or the elementary orthogonal group $\mathrm{EO}(n,n)(R)$. Furthermore, we prove that G is not acylindrically hyperbolic if $n \geq 4$. We also prove similar results for a class of noncommutative rings. The proofs are based on a notion of n–rigid rings.

20F65

1 Introduction

In this article, we study the s–normality of subgroups of matrix groups over rings together with two applications. Firstly, the low-dimensional L^2–Betti numbers of matrix groups are proved to be zero. Secondly, the matrix groups are proved to be not acylindrically hyperbolic in the sense of Dahmani, Guirardel and Osin [6] and Osin [17]. Let us briefly review the relevant background.

Let G be a discrete group. Denote by

$$l^2(G) = \left\{ f\colon G \to \mathbb{C} \;\middle|\; \sum_{g \in G} \|f(g)\|^2 < +\infty \right\}$$

the Hilbert space with inner product $\langle f_1, f_2 \rangle = \sum_{x \in G} f_1(x)\overline{f_2(x)}$. Let $B(l^2(G))$ be the set of all bounded linear operators on the Hilbert space $l^2(G)$. By definition, the group von Neumann algebra $\mathcal{N}G$ is the completion of the complex group ring $\mathbb{C}[G]$ in $B(l^2(G))$ with respect to the weak operator topology. There is a continuous, additive von Neumann dimension that assigns to every right $\mathcal{N}G$–module M a value $\dim_{\mathcal{N}G}(M) \in [0, \infty]$; see Definition 6.20 of Lück [14]. For a group G, let EG be the universal covering space of its classifying space BG. Denote by $C_*^{\mathrm{sing}}(EG)$ the

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2825

singular chain complex of EG with the induced $\mathbb{Z}G$–structure. The L^2–homology is the singular homology $H_i^G(EG;\mathcal{N}G)$ with coefficients in $\mathcal{N}G$, ie the homology of the $\mathcal{N}G$–chain complex $\mathcal{N}G \otimes_{\mathbb{Z}G} C_*^{\text{sing}}(EG)$. The i^{th} L^2–Betti number of G is defined by

$$b_i^{(2)}(G) := \dim_{\mathcal{N}G}(H_i^G(EG;\mathcal{N}G)) \in [0,\infty].$$

The L^2–homology and L^2–Betti numbers are important invariants of spaces and groups. They have many applications to geometry and K–theory. For more details, see [14].

It has been proved that the L^2–Betti numbers are (almost) zero for several classes of groups including amenable groups, Thompson's group (see [14, Theorem 7.20]), the Baumslag–Solitar group (see Bader, Furman and Sauer [1] and Dicks and Linnell [7]), the mapping class group of a closed surface with genus $g \geq 2$ except $b_{3g-3}^{(2)}$ (see Kida [12], Corollary D.15) and so on; for more information, see [14, Chapter 7]. Let R be an associative ring with identity and $n \geq 2$ an integer. The general linear group $\mathrm{GL}_n(R)$ is the group of all $n \times n$ invertible matrices with entries in R. For an element $r \in R$ and any integers i, j such that $1 \leq i \neq j \leq n$, denote by $e_{ij}(r)$ the elementary $n \times n$ matrix with ones in the diagonal positions, r in the $(i,j)^{\text{th}}$ position and zeros elsewhere. The group $E_n(R)$ is generated by all such $e_{ij}(r)$, ie

$$E_n(R) = \langle e_{ij}(r) \mid 1 \leq i \neq j \leq n,\ r \in R \rangle.$$

When R is commutative, we define the special linear group $\mathrm{SL}_n(R)$ as the subgroup of $\mathrm{GL}_n(R)$ consisting of matrices with determinant 1. For example, in the case $R = \mathbb{Z}$, the integers, we have that $\mathrm{SL}_n(\mathbb{Z}) = E_n(R)$. The groups $\mathrm{GL}_n(R)$ and $E_n(R)$ are important in algebraic K–theory.

In this article, we prove the vanishing of lower L^2–Betti numbers for matrix groups over a large class of rings, including all infinite commutative rings. For this, we introduce the notion of n–rigid rings; for details, see Definition 3.1. Examples of n–rigid (for any $n \geq 1$) rings contain the following (see Section 3):

- infinite integral domains;
- $\mathbb{Z}$–torsion-free infinite noetherian rings (may be noncommutative);
- infinite commutative noetherian rings (moreover, any infinite commutative ring is 2–rigid);
- finite-dimensional algebras over n–rigid rings.

We prove the following results.

Theorem 1.1 *Suppose $n \geq 2$. Let R be an infinite $(n-1)$–rigid ring and $E_n(R)$ the group generated by elementary matrices. For each $i \in \{0, \dots, n-2\}$, the L^2–Betti number $b_i^{(2)}(E_n(R))$ vanishes.*

Since $b_1^{(2)}(E_2(\mathbb{Z})) \neq 0$, this result does not hold for $i = n-1$ in general.

Corollary 1.2 *Let R be any infinite commutative ring and $n \geq 2$. For each $i \in \{0, \dots, n-2\}$, the following L^2–Betti numbers vanish:*

$$b_i^{(2)}(\mathrm{GL}_n(R)) = b_i^{(2)}(\mathrm{SL}_n(R)) = b_i^{(2)}(E_n(R)) = 0.$$

Let $\mathrm{SL}_n(R)$ be a lattice in a semisimple Lie group, eg when $R = \mathbb{Z}$ or a subring of algebraic integers. It follows from results of Borel, which rely on global analysis on the associated symmetric space, that the L^2–Betti numbers of $\mathrm{SL}_n(R)$ vanish except possibly in the middle dimension of the symmetric space; see Borel [5] and Olbrich [16]. In particular, all the L^2–Betti numbers of $\mathrm{SL}_n(\mathbb{Z})$ $(n \geq 3)$ are zero; see Eckmann [8, Example 2.5]. For any infinite integral domain R and any $i \in \{0, \dots, n-2\}$, Bader, Furman and Sauer [1] prove that the L^2–Betti number $b_i^{(2)}(\mathrm{SL}_n(R))$ vanishes. Ershov and Jaikin-Zapirain [9] prove that the noncommutative universal lattice $E_n(\mathbb{Z}\langle x_1, \dots, x_k\rangle)$ (and therefore $E_n(R)$ for any finitely generated associative ring R) has Kazhdan's property (T) for $n \geq 3$. This implies that for any finitely generated associative ring R, the first L^2–Betti number of $E_n(R)$ vanishes; see Bekka and Valette [4].

We consider more matrix groups as follows. Let R be a commutative ring with identity. The symplectic group and the split orthogonal group are defined as

$$\mathrm{Sp}_{2n}(R) = \{A \in \mathrm{GL}_{2n}(R) \mid A^T \varphi_n A = \varphi_n\}, \quad O(n,n)(R) = \{A \in \mathrm{GL}_{2n}(R) \mid A^T \psi_n A = \psi_n\},$$

where A^T is the transpose of A and

$$\varphi_n = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \quad \psi_n = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}.$$

For symplectic and orthogonal groups, we obtain the following.

Theorem 1.3 *Let R be an infinite principal ideal domain (PID), $\mathrm{Sp}_{2n}(R)$ the symplectic group with its elementary subgroup $\mathrm{ESp}_{2n}(R)$, and $O(n,n)(R)$ the orthogonal group with its elementary subgroup $\mathrm{EO}(n,n)(R)$. We have the following.*

(i) *For each $i = 0, \dots, n-2$ $(n \geq 2)$, the following L^2–Betti numbers vanish:*

$$b_i^{(2)}(\mathrm{Sp}_{2n}(R)) = b_i^{(2)}(\mathrm{ESp}_{2n}(R)) = 0.$$

(ii) *For each $i = 0, \dots, n-2$ $(n \geq 2)$, the following L^2–Betti numbers vanish:*

$$b_i^{(2)}(O(n,n)(R)) = b_i^{(2)}(\mathrm{EO}(n,n)(R)) = 0.$$

The proofs of Theorem 1.1 and Theorem 1.3 are based on a study of the notion of weak normality of particular subgroups in matrix groups, introduced in [1] and by Peterson and Thom [18]. We present another application of the weak normality of subgroups in matrix groups as follows.

Acylindrically hyperbolic groups are defined by Dahmani, Guirardel and Osin [6] and Osin [17]. Let G be a group. An isometric G–action on a metric space S is said to be acylindrical if for every $\varepsilon > 0$, there exist $R, N > 0$ such that for every two points $x, y \in S$ with $d(x, y) \geq R$, there are at most N elements $g \in G$ which satisfy $d(x, gx) \leq \varepsilon$ and $d(y, gy) \leq \varepsilon$. A G–action by isometries on a hyperbolic geodesic space S is said to be elementary if the limit set of G on the Gromov boundary ∂S contains at most two points. A group G is called *acylindrically hyperbolic* if G admits a nonelementary acylindrical action by isometries on a (Gromov-δ) hyperbolic geodesic space. The class of acylindrically hyperbolic groups includes nonelementary hyperbolic and relatively hyperbolic groups, mapping class groups of closed surface Σ_g of genus $g \geq 1$, outer automorphism groups $\mathrm{Out}(F_n)$ $(n \geq 2)$ of free groups, directly indecomposable right angled Artin groups, 1–relator groups with at least three generators, most 3–manifold groups, and many other examples.

Although there are many analogies among matrix groups, mapping class groups and outer automorphism groups of free groups, we prove that they are different on acylindrical hyperbolicity, as follows.

Theorem 1.4 *Suppose that n is an integer.*

(i) *Let R be a 2–rigid (eg commutative) ring. The group $E_n(R)$ $(n \geq 3)$ is not acylindrically hyperbolic.*

(ii) *Let R be a commutative ring. The group G is not acylindrically hyperbolic if $G = \mathrm{GL}_n(R)$ $(n \geq 3)$ the general linear group, $\mathrm{SL}_n(R)$ $(n \geq 3)$ the special linear group, $\mathrm{Sp}_{2n}(R)$ $(n \geq 2)$ the symplectic group, $\mathrm{ESp}_{2n}(R)$ $(n \geq 2)$ the elementary symplectic group, $O(n,n)(R)$ $(n \geq 4)$ the orthogonal group, or $\mathrm{EO}(n,n)(R)$ $(n \geq 4)$ the elementary orthogonal group.*

When R is commutative, the failure of acylindrical hyperbolicity of the elementary groups $E_n(R)$, $\mathrm{ESp}_{2n}(R)$ and $\mathrm{EO}(n,n)(R)$ is already known to Mimura [15] by studying property TT for weakly mixing representations. But our approach is different and Theorem 1.4 is more general, even for elementary subgroups. Explicitly, for noncommutative rings, we have the following.

Corollary 1.5 *Let R be a noncommutative $\mathbb{Z}$–torsion-free infinite noetherian ring, an integral group ring over a polycyclic-by-finite group or a finite-dimensional algebra over either. For each nonnegative integer $i \leq n-2$, we have*

$$b_i^{(2)}(E_n(R)) = 0.$$

Furthermore, the group $E_n(R)$ $(n \geq 3)$ is not acylindrically hyperbolic.

2 s–normality

Recall from [1] that the n–step s–normality is defined as follows.

Definition 2.1 Let $n \geq 1$ be an integer. A subgroup H of a group G is called n–step s–normal if for any $(n+1)$–tuple $\omega = (g_0, g_1, \dots, g_n) \in G^{n+1}$, the intersection

$$H^\omega := \bigcap_{i=0}^{n} g_i H g_i^{-1}$$

is infinite. A 1–step s–normal group is simply called s–normal.

The following result is proved by Bader, Furman and Sauer; see [1, Theorem 1.3].

Lemma 2.2 *Let H be a subgroup of G. Assume that*

$$b_i^{(2)}(H^\omega) = 0$$

for all integers $i, k \geq 0$ with $i + k \leq n$ and every $\omega \in G^{k+1}$. In particular, H is an n–step s–normal subgroup of G. Then for every $i \in \{0, \dots, n\}$,

$$b_i^{(2)}(G) = 0.$$

The following result is important for our later arguments; see [14, Theorem 7.2, (1-2), page 294].

Lemma 2.3 *Let n be any nonnegative integer. Then:*

(i) *For any infinite amenable group G, the L^2–Betti numbers $b_n^{(2)}(G)$ vanish.*

(ii) *Let H be a normal subgroup of a group G with vanishing $b_i^{(2)}(H)$ for each $i \in \{0, 1, \dots, n\}$. Then for each $i \in \{0, 1, \dots, n\}$, we have $b_i^{(2)}(G) = 0$.*

We will also need the following fact; see [17, Corollaries 1.5, 7.3].

Lemma 2.4 *The class of acylindrically hyperbolic groups is closed under taking s–normal subgroups. Furthermore, acylindrically hyperbolic groups have finite center.*

3 Rigidity of rings

We introduce the notion of n–rigidity of rings. For a ring R, all R–modules are right modules and homomorphisms are right R–module homomorphisms.

Definition 3.1 For a positive integer n, an infinite ring R is called *n–rigid* if every R–homomorphism $R^n \to R^{n-1}$ of free modules has an infinite kernel.

A related concept is the strong rank condition: a ring R satisfies the strong rank condition if there is no injection $R^n \to R^{n-1}$ for any n; see Lam [13, page 12]. Clearly, n–rigidity for any n implies the strong rank condition for a ring. Fixing the standard basis of both R^n and R^{n-1}, the kernel of an R–homomorphism $\phi\colon R^n \to R^{n-1}$ corresponds to a system S of $n-1$ linear equations with n unknowns over R:

$$S: \quad \sum_{1\le i\le n} a_{ij}x_i = 0, \quad 1 \le j \le n-1,$$

with $a_{ij} \in R$, $1 \le i \le n$, $1 \le j \le n-1$. Therefore, the strong rank condition asserts that the system S has nontrivial solutions over R, while the n–rigidity property requires that S has infinitely many solutions.

Many rings are n–rigid. For example, infinite integral rings are n–rigid for any n by considering the dimensions over quotient fields. Moreover, let A be a ring satisfying the strong rank condition (eg a noetherian ring, see Theorem 3.15 of [13]). Suppose that A is a torsion-free $\mathbb{Z}$–module, where $\mathbb{Z}$ acts on A via $\mathbb{Z}\cdot 1_A$. Since the kernel $A^n \to A^{n-1}$ is a nontrivial $\mathbb{Z}$–module, the ring A is n–rigid for any n.

We present several basic facts on n–rigid rings as follows.

Lemma 3.2 *n–rigidity implies $(n-1)$–rigidity.*

Proof For any R–homomorphism $f\colon R^{n-1} \to R^{n-2}$, we could add a copy of R as direct summand to get a map $f \oplus \mathrm{id}\colon R^{n-1} \oplus R \to R^{n-2} \oplus R$. The two maps have the same kernel. □

Lemma 3.3 *Let R be an n–rigid ring for any $n \ge 1$. Suppose that an associative ring A is a finite-dimensional R–algebra (ie A is a free R–module of finite rank with compatible multiplications in A and R). Then A is n–rigid for any $n \ge 1$.*

Proof Let $f\colon A^n \to A^{n-1}$ be an A–homomorphism. If we view A as a finite-dimensional R–module, we see that f is also an R–homomorphism. Embed the target A^{n-1} into $R^{n\cdot\mathrm{rank}_R(A)-1}$. The kernel $\ker f$ is infinite by the assumption that R is $n\cdot\mathrm{rank}_R(A)$–rigid. □

Proposition 3.4 *Let R be an n–rigid ring and let $u_1, u_2, \dots, u_{n-1} \in R^m$ $(m \ge n)$ be arbitrary $n-1$ elements. Then the set*

$$\{\phi \in \mathrm{Hom}_R(R^m, R) \mid \phi(u_i) = 0,\ i = 1, 2, \dots, n-1\}$$

is infinite.

Proof When $m = n$, we define an R–homomorphism

$$\mathrm{Hom}_R(R^n, R) \to R^{n-1}, \quad f \mapsto (f(u_1), f(u_2), \dots, f(u_{n-1})).$$

Since $\mathrm{Hom}_R(R^n, R)$ is isomorphic to R^n, such an R–homomorphism has an infinite kernel. When $m > n$, we may project R^m to its last n–components and apply a similar proof. □

Lemma 3.5 *An infinite commutative ring R is 2–rigid.*

Proof Let $f\colon R^2 \to R$ be any R–homomorphism, and let

$$I = \langle xR + yR \mid (x, y) \in \ker f \rangle \trianglelefteq R.$$

Suppose that $\ker f$ is finite. When $(x, y) \in \ker f$, the set xR and yR are also finite. Thus I is finite. Let $a = f((1,0))$ and $b = f((0,1))$. Note that $(-b, a) \in \ker f$. For any $(x, y) \in R^2$, we have $ax + by \in I$. Since the set of right cosets R/I is infinite, we may choose (x, x) and (y, y) with x, y from distinct cosets such that

$$ax + bx = ay + by.$$

However, $(x - y, x - y) \in \ker f$, and thus $x - y \in I$. This is a contradiction. □

To state our result in the most general form, we introduce the following notion.

Definition 3.6 A ring R is called *size-balanced* if any finite right ideal of R generates a finite two-sided ideal of R.

It is immediate that any commutative ring is size-balanced.

Proposition 3.7 *A size-balanced infinite noetherian ring is n–rigid for any n.*

Proof Let $f\colon R^n \to R^{n-1}$ be any R–homomorphism. Let $A = (a_{ij})_{(n-1)\times n}$ be the matrix representation of f with respect to the standard basis, and let

$$I' = \langle x_1 R + x_2 R + \cdots + x_n R \mid (x_1, x_2, \dots, x_n) \in \ker f \rangle \trianglelefteq R.$$

First we notice that I' is nontrivial by the strong rank condition of noetherian rings; see Theorem 3.15 of [13]. Suppose that $\ker f$ is finite. For any

$$(x_1, x_2, \dots, x_n) \in \ker f$$

and $r \in R$, each $(x_1 r, x_2 r, \dots, x_n r) \in \ker f$. As $\ker f$ is finite, each right ideal $x_i R$ is finite, and hence so is I'. Let I be the two-sided ideal generated by the finite right ideal I'. It is finite as R is assumed to be size-balanced. Therefore, the quotient ring R/I is infinite and noetherian.

Let $\bar{f}\colon R/I \to R/I$ be the R/I–homomorphism induced by the matrix $\bar{A} = (\bar{a}_{ij})$, where $\bar{a}_{ij}$ is the image of a_{ij}. If $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n) \in \ker \bar{f}$ and x_i is any preimage of $\bar{x}_i$, we have

$$A(x_1, x_2, \dots, x_n)^T \in I^{n-1}.$$

As I is finite, so is I^{n-1}. If $\ker \bar{f}$ is infinite, there are two distinct elements in $\ker \bar{f}$ with preimage $(x_1, x_2, \dots, x_n)$ and $(y_1, y_2, \dots, y_n)$ in R^n such that

$$A(x_1, x_2, \dots, x_n)^T = A(y_1, y_2, \dots, y_n)^T \in I^{n-1}.$$

However, this implies that

$$(x_1, x_2, \dots, x_n) - (y_1, y_2, \dots, y_n) \in \ker f.$$

We have a contradiction as $\underline{(x_1, x_2, \dots, x_n)}$ and $(y_1, y_2, \dots, y_n)$ are distinct in $(R/I)^n$. Therefore, $\ker \bar{f}$ is finite. Moreover, $\ker \bar{f}$ is nontrivial by the strong rank condition of noetherian rings. Let I_1' be the preimage of the right ideal generated by components of elements in $\ker \bar{f}$ in R, which is a finite right ideal by a similar argument as above. It generates a finite two-sided ideal I_1 of R, and it properly contains I.

Repeating the argument, we get an infinite ascending sequence

$$I \subsetneqq I_1 \subsetneqq I_2 \subsetneqq \cdots$$

of finite ideals of R. This is a contradiction to the assumption that R is noetherian. □

Corollary 3.8 *Any commutative ring R containing an infinite noetherian subring is n–rigid for each n.*

Proof Let R_0 be an infinite noetherian subring of R. Let

$$S: \quad \sum_{1 \le i \le n} a_{ij} x_i = 0, \quad 1 \le j \le m$$

be a system of linear equations with $a_{ij} \in R$. Form the infinite commutative subring $R' = R_0[a_{ij}, 1 \le i \le n, 1 \le j \le m]$ of R. By the Hilbert basis theorem, R' is infinite noetherian. Proposition 3.7 asserts that the system S has infinitely many solutions in R', and hence in R. □

Example 3.9 Let G be a polycyclic-by-finite group and $R = \mathbb{Z}[G]$ be its integral group ring. It is known that R is infinite noetherian [11]. Moreover, R is size-balanced by the trivial reason that there are no nontrivial finite right ideals. According to Proposition 3.7, the ring R is n–rigid for any n.

Example 3.10 Let F be a nonabelian free group and $\mathbb{Z}[F]$ the group ring. Since $\mathbb{Z}[F]$ does not satisfy the strong rank condition [13, Exercise 29, page 21], the ring $\mathbb{Z}[F]$ is not n–rigid for any $n \geq 2$.

4 Proofs

Let

$$Q = \left\{ \begin{pmatrix} 1 & x \\ 0 & A \end{pmatrix} \,\middle|\, x \in R^{n-1},\ A \in \mathrm{GL}_{n-1}(R),\ \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix} \in E_n \right\}.$$

It is straightforward that Q contains the normal subgroup

$$S = \left\{ \begin{pmatrix} 1 & x \\ 0 & I_{n-1} \end{pmatrix} \,\middle|\, x \in R^{n-1} \right\},$$

an abelian group. Therefore, all the L^2–Betti numbers of S and Q are zero when the ring R is infinite.

Lemma 4.1 *Let $k < n$ $(n \geq 3)$ be two positive integers. Suppose that R is an infinite k–rigid ring. Then the subgroup Q is $(k-1)$–step s–normal in $E_n(R)$. In particular, Q is s–normal if R is infinite 2–rigid.*

Proof Without loss of generality, we assume $k = n-1$. Let $g_1, g_2, \ldots, g_{n-2}$ be any $n-2$ elements in $E_n(R)$. We will show that the intersection $Q \cap \bigcap_{i=1}^{n-2} g_i Q g_i^{-1}$ is infinite, which implies the $(k-1)$–step s–normality of H. Let $\{e_i\}_{i=1}^n$ be the standard basis of R^n. Denote by $U = R^{n-1}$ the R–submodule spanned by $\{e_i\}_{i=2}^n$ and $p\colon R^n \to U$ the natural projection.

For each g_i $(i = 1, 2, \ldots, n-2)$, suppose that

$$g_i e_1 = x_i e_1 + u_i$$

for $x_i \in R$ and $u_i \in U$. Let

$$\Phi = \{\phi \in \mathrm{Hom}_R(U, R) \mid \phi(u_i) = 0,\ i = 1, 2, \ldots, n-2\}.$$

For any $\phi \in \Phi$, define $T_\phi\colon R^n \to R^n$ by $T_\phi(v) = v + \phi \circ p(v) e_1$. It is obvious that

$$g_i^{-1} T_\phi g_i(e_1) = e_1$$

for each $i = 1, 2, \ldots, n-2$. Note that Q is the stabilizer of e_1. This shows that for each $\phi \in \Phi$, the transformation T_ϕ lies in $Q \cap \bigcap_{i=1}^{n-2} g_i Q g_i^{-1}$. Denote by T the subgroup

$$T = \{T_\phi \mid \phi \in \Phi\}. \tag{1}$$

By Proposition 3.4, Φ is infinite, and thus T is infinite. The proof is finished. □

Lemma 4.2 *The subgroup T from* (1) *is normal in $Q \cap \bigcap_{i=1}^{n-2} g_i Q g_i^{-1}$.*

Proof For any ϕ, write $e_\phi = (\phi(e_2), \dots, \phi(e_n))$. With respect to the standard basis, the representation matrix of the transformation T_ϕ is $\begin{pmatrix} 1 & e_\phi \\ 0 & I_{n-1} \end{pmatrix}$. For any $\begin{pmatrix} 1 & x \\ 0 & A \end{pmatrix} \in Q \cap \bigcap_{i=1}^{n-2} g_i Q g_i^{-1}$, the conjugate of the representation matrix has the following form:

$$\begin{pmatrix} 1 & x \\ 0 & A \end{pmatrix}^{-1} \begin{pmatrix} 1 & e_\phi \\ 0 & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & A \end{pmatrix} = \begin{pmatrix} 1 & e_\phi A \\ 0 & I_{n-1} \end{pmatrix}.$$

Define $\psi\colon U = R^{n-1} \to R$ by $\psi(x) = e_\phi A x$. For each $i = 1, \dots, n-2$, we have that $\begin{pmatrix} 1 & x \\ 0 & A \end{pmatrix} = g_i q_i g_i^{-1}$ for some $q_i \in Q$. Therefore,

$$\begin{pmatrix} 1 & x \\ 0 & A \end{pmatrix} g_i e_1 = g_i q_i e_1,$$

and $A u_i = u_i$. This implies that $\psi(u_i) = e_\phi u_i = 0$ for each i, and thus $\psi \in \Phi$. Therefore, the conjugate $\begin{pmatrix} 1 & e_\phi A \\ 0 & I_{n-1} \end{pmatrix}$ lies in T, which proves that T is normal. □

Proof of Theorem 1.1 By Lemma 4.2, any intersection $Q \cap \bigcap_{i=1}^{n-2} g_i Q g_i^{-1}$ contains an infinite normal amenable subgroup T. Therefore, all the L^2–Betti numbers of any intersection $Q \cap \bigcap_{i=1}^{k} g_i Q g_i^{-1}$ vanish for $k \le n-2$ considering Lemma 3.2. We have that $b_i^{(2)}(E_n(R)) = 0$ for any $0 \le i \le n-2$ by Lemma 2.2. □

Proof of Corollary 1.2 When $n = 2$, it is clear that both $\mathrm{GL}_n(R)$ and $\mathrm{SL}_n(R)$ are infinite, since $E_2(R)$ is an infinite subgroup. Thus $b_0^{(2)}(\mathrm{GL}_2(R)) = b_0^{(2)}(\mathrm{SL}_2(R)) = 0$. We have already proved that $b_i^{(2)}(E_n(S)) = 0$ for infinite commutative noetherian ring S and $0 \le i \le n-2$, since the ring S would be k–rigid for any integer k by Proposition 3.7. If S is a finite subring of R, the group $E_n(S)$ is also finite. Therefore, we still have $b_i^{(2)}(E_n(S)) = 0$ for $1 \le i \le n-2$. Note that every commutative ring R is the directed colimit of its subrings S that are finitely generated as $\mathbb{Z}$–algebras (noetherian rings by the Hilbert basis theorem). Since the group $E_n(R)$ is the union of the directed system of subgroups $E_n(S)$, we get that

$$b_i^{(2)}(E_n(R)) = 0$$

for $0 \le i \le n-2$; see [14, Theorem 7.2(3)] and its proof. When R is commutative and $n \ge 3$, a result of Suslin [19] says that the group $E_n(R)$ is a normal subgroup of $\mathrm{GL}_n(R)$ and $\mathrm{SL}_n(R)$. Lemma 2.3 implies that $b_i^{(2)}(\mathrm{GL}_n(R)) = b_i^{(2)}(\mathrm{SL}_n(R)) = 0$ for each $i \in \{0, \dots, n-2\}$. □

We follow [2] to define the elementary subgroups of symplectic groups and orthogonal groups. Let E_{ij} denote the $n \times n$ matrix with 1 in the $(i, j)^{\text{th}}$ position and zeros elsewhere. Then for $i \neq j$, the matrix $e_{ij}(a) = I_n + aE_{ij}$ is an elementary matrix, where I_n is the identity matrix of size n. With n fixed, for any integer $1 \leq k \leq 2n$, set $\sigma k = k + n$ if $k \leq n$ and $\sigma k = k - n$ if $k > n$. For $a \in R$ and $1 \leq i \neq j \leq 2n$, we define the elementary unitary matrices $\rho_{i,\sigma i}(a)$ and $\rho_{ij}(a)$ with $j \neq \sigma i$ as follows:

- $\rho_{i,\sigma i}(a) = I_{2n} + aE_{i,\sigma i}$ with $a \in R$.
- Fix $\varepsilon = \pm 1$. We define $\rho_{ij}(a) = \rho_{\sigma j,\sigma i}(-a') = I_{2n} + aE_{ij} - a'E_{\sigma j,\sigma i}$ with $a' = a$ when $i, j \leq n$; $a' = \varepsilon a$ when $i \leq n < j$; $a' = a\varepsilon$ when $j \leq n < i$; and $a' = a$ when $n + 1 \leq i, j$.

When $\varepsilon = -1$, we have the elementary symplectic group

$$\mathrm{ESp}_{2n}(R) = \langle \rho_{i,\sigma i}(a),\ \rho_{ij}(a) \mid a \in R,\ i \neq j,\ i \neq \sigma j \rangle.$$

When $\varepsilon = 1$, we have the elementary orthogonal group

$$\mathrm{EO}(n,n)(R) = \langle \rho_{ij}(a) \mid a \in R,\ i \neq j,\ i \neq \sigma j \rangle.$$

Note that for the orthogonal group, each matrix $\rho_{i,\sigma i}(a)$ is not in $\mathrm{EO}(n,n)(R)$.

There is an obvious embedding

$$\mathrm{Sp}_{2n}(R) \to \mathrm{Sp}_{2n+2}(R), \quad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & \beta \\ 0 & 0 & 1 & 0 \\ 0 & \gamma & 0 & \delta \end{pmatrix}.$$

Denote the image of $A \in \mathrm{Sp}_{2n}(R)$ by $I \oplus A \in \mathrm{Sp}_{2n+2}(R)$. Let

$$Q_1 = \left\langle (I \oplus A) \cdot \prod_{i=1}^{2n} \rho_{1i}(a_i) \,\middle|\, a_i \in R,\ A \in \mathrm{Sp}_{2n-2}(R),\ I \oplus A \in \mathrm{ESp}_{2n}(R) \right\rangle$$

and

$$S_1 = \left\langle \prod_{i=1}^{2n} \rho_{1i}(a_i) \,\middle|\, a_i \in R \right\rangle.$$

Similarly, we can define

$$Q_2 = \left\langle (I \oplus A) \prod_{\substack{1 \leq i \leq 2n \\ i \neq n+1}} \rho_{1i}(a_i) \,\middle|\, a_i \in R,\ A \in O(2n-2, 2n-2)(R),\ I \oplus A \in \mathrm{EO}(n,n)(R) \right\rangle$$

and

$$S_2 = \left\langle \prod_{\substack{1 \leq i \leq 2n \\ i \neq n+1}} \rho_{1i}(a_i) \,\middle|\, a_i \in R \right\rangle.$$

Since S_i is abelian and normal in Q_i, all the L^2–Betti numbers of Q_i vanish for $i = 1, 2$.

Proof of Theorem 1.3 We prove the theorem by induction on n. When $n=2$, both $\mathrm{Sp}_{2n}(R)$ and $O(n,n)(R)$ are infinite, and therefore we have

$$b_0^{(2)}(\mathrm{Sp}_4(R)) = b_0^{(2)}(O(4,4)(R)) = 0.$$

The subgroup $\mathrm{ESp}_{2n}(R)$ is normal in $\mathrm{Sp}_{2n}(R)$ when $n \geq 2$, and the subgroup $\mathrm{EO}(n,n)(R)$ is normal in $O(n,n)(R)$ when $n \geq 3$; see [3, Corollary 3.10]. It suffices to prove the vanishing of Betti numbers for $G = \mathrm{ESp}_{2n}(R)$ and $\mathrm{EO}(n,n)(R)$.

We check the condition of Lemma 2.2 for $Q = Q_1$ (resp. Q_2) as follows. Note that

$$Q = \{g \in G \mid ge_1 = e_1\}.$$

Let $g_1, g_2, \ldots, g_k$ ($g_0 = I_{2n}$, $k \leq n-2$) be any k elements in G and

$$K = \langle g_0 e_1, g_1 e_1, \ldots, g_k e_1 \rangle$$

the submodule in R^{2n} generated by all $g_i e_1$. Recall that the symplectic (resp. orthogonal) form $\langle -,- \rangle\colon R^{2n} \times R^{2n} \to R$ is defined by $\langle x,y \rangle = x^T \varphi_n y$ (resp. $\langle x,y \rangle = x^T \psi_n y$). Let

$$C := \{v \in R^{2n} \mid \langle v, g_i e_1 \rangle = 0 \text{ for each } i = 0, \ldots, k-1\}.$$

Let $\varepsilon = -1$ for $\mathrm{ESp}_{2n}(R)$ and $\varepsilon = 1$ for $\mathrm{EO}(n,n)(R)$. For each $r \in R$, set $\delta_\varepsilon^r = r$ if $\varepsilon = -1$ and $\delta_\varepsilon^r = 0$ if $\varepsilon = 1$. For each $u, v \in C$ with $\langle u,u \rangle = \langle u,v \rangle = \langle v,v \rangle = 0$, define the transvections in G (see [20, page 287], Eichler transformations in [10, pages 214, 223–224])

$$\tau(u,v)\colon R^{2n} \to R^{2n}, \quad x \mapsto x + \varepsilon u \langle v,x \rangle - v \langle u,x \rangle,$$
$$\tau_{v,r}\colon R^{2n} \to R^{2n}, \quad x \mapsto x - \delta_\varepsilon^r v \langle v,x \rangle.$$

Note that $\tau_{v,r}$ is nonidentity only in $\mathrm{ESp}_{2n}(R)$. We have

$$\tau(u,v)(g_i e_1) = \tau_{v,r}(g_i e_1) = g_i e_1$$

for each i. Therefore, the transvections $\tau(u,v), \tau_{v,r} \in \bigcap_{i=0}^k g_i Q g_i^{-1}$. Let

$$T = \langle \tau(u,v), \tau_{v,r} \mid u,v \in C,\ \langle u,u \rangle = \langle u,v \rangle = \langle v,v \rangle = 0,\ r \in R \rangle$$

be the subgroup generated by the transvections in G. For any $g \in \bigcap_{i=0}^k g_i Q g_i^{-1}$, we have $g g_i e_1 = g_i e_1$, and thus

$$\langle gu, g_i e_1 \rangle = \langle gu, g g_i e_1 \rangle = \langle u, g_i e_1 \rangle = 0.$$

This implies that $g\tau(u,v)g^{-1} = \tau(gu, gv) \in T$ and $g\tau_{v,r}g^{-1} = \tau_{gv,r} \in T$. Therefore, the subgroup T is a normal subgroup in $\bigcap_{i=0}^k g_i Q g_i^{-1}$.

When R is a PID, the submodule K and the complement C are free of smaller ranks.

Case (i) $\boldsymbol{K \cap C = 0}$ Since $R^{2n} = K \oplus C$ (note that each $g_i e_1$ is unimodular), the symplectic (resp. orthogonal) form on R^{2n} restricts to a nondegenerate symplectic (resp. orthogonal) form on C. Let $T < G$ be as defined before. It is known that the transvections generate the elementary subgroups [10, pages 223–224], and thus $T \cong \mathrm{ESp}_{2m}(R)$ (resp. $\mathrm{EO}(m,m)(R)$) for $m = \mathrm{rank}(C) \le n-2$. Since $k \le n-2$, we have $m \ge 4$. By induction,

$$b_s^{(2)}\Bigl(\bigcap_{i=0}^{k} g_i Q g_i^{-1}\Bigr) = b_s^{(2)}(T) = 0$$

for $s \le \frac{1}{2}\,\mathrm{rank}(C) - 2$. When $s + k \le n-2$, we have that $s \le \frac{1}{2}\,\mathrm{rank}(C) - 2$ since $\mathrm{rank}(C) \ge 2n-(k+1)$. Therefore, $b_s^{(2)}\bigl(\bigcap_{i=0}^{k} g_i Q_1 g_i^{-1}\bigr) = 0$, and Lemma 2.2 implies that for any $i \le n-2$,

$$b_i^{(2)}(G) = 0.$$

Case (ii) $\boldsymbol{K \cap C \neq 0}$ For any $u, v \in K \cap C$ and any $g \in \bigcap_{i=0}^{k} g_i Q g_i^{-1}$, we have that $gu = u$, $gv = v$ and

$$g\tau(u,v)g^{-1} = \tau(gu, gv) = \tau(u,v).$$

This implies that $\tau(u,v)$ lies in the center of $\bigcap_{i=0}^{k} g_i Q g_i^{-1}$. Note that when $G = \mathrm{ESp}_{2n}(R)$, the transvection $\tau(u,u)$ is not trivial for any $u \in K \cap C$. When $G = \mathrm{EO}(n,n)(R)$ and $\mathrm{rank}(K \cap C) \ge 2$, the transvection $\tau(u,v)$ is not trivial for any linearly independent $u, v \in K \cap C$. Moreover, for two elements r, s with $r^2 \neq s^2$, we have $\tau(ru, rv) \neq \tau(su, sv)$ when $\tau(u,v) \neq I_{2n}$ (take note that for $G = \mathrm{ESp}_{2n}(R)$, we can just let $u = v$ from above). The infinite PID R contains infinitely many square elements. In summary, as $K \cap C$ is a free R–module, the subgroup

$$T' = \langle \tau(u,v) \mid v \in K \cap C \rangle < G$$

is an infinite abelian normal subgroup of $\bigcap_{i=0}^{k} g_i Q g_i^{-1}$. Therefore,

$$b_s^{(2)}\Bigl(\bigcap_{i=0}^{k} g_i Q g_i^{-1}\Bigr) = b_s^{(2)}(T') = 0$$

for each integer $s \ge 0$. Therefore, for any $i \le n-2$, we have that $b_i^{(2)}(G) = 0$ by Lemma 2.2.

The remaining situation is that $G = \mathrm{EO}(n,n)(R)$ and $\mathrm{rank}(K \cap C) = 1$. Choose the decomposition $C = (K \cap C) \oplus C_1$. The orthogonal form restricts to a nondegenerate orthogonal form on C_1. (Suppose that for some $x \in C_1$, we have $\langle x, y\rangle = 0$ for any $y \in C_1$. Since $\langle x, k\rangle = 0$ for any $k \in K$, we know that $\langle x, y\rangle = 0$ for any $y \in C$. This implies $x \in K$, which gives $x = 0$.) Since $k \leq n-2$, the even number $\operatorname{rank}(C_1) \geq 4$. A similar argument as in case (i) finishes the proof. □

Remark 4.3 Let T be the normal subgroup of $\bigcap_{i=0}^{k} g_i Q g_i^{-1}$ constructed in the proof of Theorem 1.3. We do not know whether the L^2–Betti numbers $b_i^{(2)}(T)$ vanish for a general infinite $(2n-1)$–rigid commutative ring R when $i \leq n-2-k$. If so, Theorem 1.3 would hold for any general infinite commutative ring by a similar argument as in the proof of Corollary 1.2.

Proof of Theorem 1.4 Note that when R is commutative, the elementary subgroups $E_n(R), \mathrm{ESp}_{2n}(R)$ and $\mathrm{EO}(n,n)(R)$ are normal in $\mathrm{SL}_n(R), \mathrm{Sp}_{2n}(R)$ and $O(n,n)(R)$, respectively; see [19; 3, Corollary 3.10]. Therefore, it is enough to prove the failure of acylindrically hyperbolicity for elementary subgroups. We prove (i) first. If R is finite, all the groups will be finite and thus not acylindrically hyperbolic. If R is infinite, then it is 2–rigid, and the subgroup Q is s–normal by Lemma 4.1. Suppose that $E_n(R)$ is acylindrically hyperbolic. Lemma 2.4 implies that both Q and S are acylindrically hyperbolic. However, the subgroup S is infinite abelian, which is a contradiction to the second part of Lemma 2.4.

For (ii), we may also assume that R is infinite since any finite group is not acylindrically hyperbolic. It suffices to prove that Q_1 (resp. Q_2) is s–normal in $\mathrm{ESp}_{2n}(R)$ (resp. $\mathrm{EO}(n,n)(R)$). (Note that Q_1 and Q_2 contain the infinite normal subgroups S_1 and S_2, respectively. If G is acylindrically hyperbolic, the infinite abelian subgroup S_1 or S_2 would be acylindrically hyperbolic. This is a contradiction to the second part of Lemma 2.4.) By definition, this is equivalent to proving that for any $g \in G$, the intersection $Q \cap g^{-1} Q g$ is infinite for $Q = Q_1$ and $Q = Q_2$. Let $ge_1 = (x_1, \dots, x_n, y_1, \dots, y_n)^T$. Let

$$t_A = \prod_{1 \leq i < j \leq n} \rho_{i,n+j}(a_{ij}) = \begin{pmatrix} I_n & A \\ 0 & I_n \end{pmatrix} \in G,$$

where $A = (a_{ij})$ is an $n \times n$ matrices with entries in R. Note that $a_{ji} = a_{ij}$ if $G = \mathrm{ESp}_{2n}(R)$ and $a_{ij} = -a_{ij}$ if $G = \mathrm{EO}(n,n)(R)$. Moreover, we have $\rho_{i,n+i}(a) \notin \mathrm{EO}(n,n)(R)$ and $\rho_{i,n+i}(a) \in \mathrm{ESp}_{2n}(R)$ for any $a \in R$. Direct calculation shows that $t_A(ge_1) - ge_1 = ((y_1, \dots, y_n)A^T, 0, \dots, 0)^T$. When $n \geq 4$ and $G = \mathrm{EO}(n,n)(R)$, the map $f\colon R^{n(n-1)/2} \to R^n$ defined by

$$(a_{ij})_{1 \leq i < j \leq n} \mapsto A(y_1, \dots, y_n)^T$$

has an infinite kernel $\ker f$ by 2–rigidity of infinite commutative rings. This implies that $\langle t_A \mid (a_{ij})_{1\le i<j\le n} \in \ker f\rangle < Q \cap g^{-1}Qg$ is infinite. When $n \geq 2$ and $G = \mathrm{ESp}_{2n}(R)$, the map $f\colon R^{n(n+1)/2} \to R^n$ defined by $(a_{ij})_{1\le i\le j\le n} \mapsto A(y_1, \dots, y_n)^T$ has an infinite kernel $\ker f$, and $Q \cap g^{-1}Qg$ is infinite by a similar argument. □

Proof of Corollary 1.5 By Proposition 3.7, Example 3.9 and Lemma 3.3, all these rings are n–rigid for any $n \geq 1$. The corollary follows Theorems 1.1 and 1.4. □

Acknowledgements The authors would like to thank the referee for pointing out a gap in a previous version of this paper. Ye is supported by Jiangsu Natural Science Foundation (No. BK20140402) and NSFC (No. 11501459).

References

[1] **U Bader**, **A Furman**, **R Sauer**, *Weak notions of normality and vanishing up to rank in L^2–cohomology*, Int. Math. Res. Not. 2014 (2014) 3177–3189 MR

[2] **A Bak**, *K–theory of forms*, Annals of Mathematics Studies 98, Princeton Univ. Press (1981) MR

[3] **A Bak**, **N Vavilov**, *Normality for elementary subgroup functors*, Math. Proc. Cambridge Philos. Soc. 118 (1995) 35–47 MR

[4] **M E B Bekka**, **A Valette**, *Group cohomology, harmonic functions and the first L^2–Betti number*, Potential Anal. 6 (1997) 313–326 MR

[5] **A Borel**, *The L^2–cohomology of negatively curved Riemannian symmetric spaces*, Ann. Acad. Sci. Fenn. Ser. A I Math. 10 (1985) 95–105 MR

[6] **F Dahmani**, **V Guirardel**, **D Osin**, *Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces*, Mem. Amer. Math. Soc. 1156, Amer. Math. Soc., Providence, RI (2017) MR

[7] **W Dicks**, **P A Linnell**, *L^2–Betti numbers of one-relator groups*, Math. Ann. 337 (2007) 855–874 MR

[8] **B Eckmann**, *Lattices, l_2–Betti numbers, deficiency, and knot groups*, Enseign. Math. 50 (2004) 123–137 MR

[9] **M Ershov**, **A Jaikin-Zapirain**, *Property (T) for noncommutative universal lattices*, Invent. Math. 179 (2010) 303–347 MR

[10] **A J Hahn**, **O T O'Meara**, *The classical groups and K–theory*, Grundl. Math. Wissen. 291, Springer (1989) MR

[11] **A V Jategaonkar**, *Integral group rings of polycyclic-by-finite groups*, J. Pure Appl. Algebra 4 (1974) 337–343 MR

[12] **Y Kida**, *The mapping class group from the viewpoint of measure equivalence theory*, Mem. Amer. Math. Soc. 916, Amer. Math. Soc., Providence, RI (2008) MR

[13] **T Y Lam**, *Lectures on modules and rings*, Graduate Texts in Mathematics 189, Springer (1999) MR

[14] **W Lück**, *L^2–invariants: theory and applications to geometry and K–theory*, Ergeb. Math. Grenzgeb. 44, Springer (2002) MR

[15] **M Mimura**, *Superrigidity from Chevalley groups into acylindrically hyperbolic groups via quasi-cocycles*, preprint (2015) arXiv to appear in J. Eur. Math. Soc.

[16] **M Olbrich**, *L^2–invariants of locally symmetric spaces*, Doc. Math. 7 (2002) 219–237 MR

[17] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368 (2016) 851–888 MR

[18] **J Peterson**, **A Thom**, *Group cocycles and the ring of affiliated operators*, Invent. Math. 185 (2011) 561–592 MR

[19] **A A Suslin**, *The structure of the special linear group over rings of polynomials*, Izv. Akad. Nauk SSSR Ser. Mat. 41 (1977) 235–252 MR In Russian; translated in Math. USSR-Izv. 11 (1977) 221–238

[20] **L N Vaseršteĭn**, *Stabilization for the classical groups over rings*, Mat. Sb. 93(135) (1974) 268–295 MR In Russian; translated in Math. USSR-Sb. 22 (1974) 271–303

Infinitus, Nanyang Technological University
Singapore

Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University
Jiangsu, China

jifeng@ntu.edu.sg, shengkui.ye@xjtlu.edu.cn

https://yeshengkui.wordpress.com/

Received: 15 August 2016 Revised: 18 February 2017

msp
Algebraic & Geometric Topology 17 (2017) 2841–2861

Klein-four connections and the Casson invariant for nontrivial admissible $U(2)$ bundles

CHRISTOPHER SCADUTO
MATTHEW STOFFREGEN

Given a rank-2 hermitian bundle over a 3–manifold that is nontrivial admissible in the sense of Floer, one defines its Casson invariant as half the signed count of its projectively flat connections, suitably perturbed. We show that the 2–divisibility of this integer invariant is controlled in part by a formula involving the mod 2 cohomology ring of the 3–manifold. This formula counts flat connections on the induced adjoint bundle with Klein-four holonomy.

57M27

1 Introduction

Let E be a $U(2)$ bundle over a closed, oriented and connected 3–manifold Y with the property that $w_2(E) \equiv c_1(E) \pmod 2$ has no torsion lifts to $H^2(Y;\mathbb{Z})$. Following Floer [4], we call such bundles *nontrivial admissible*. Floer defined the instanton homology $I_*(Y,E)$, which is an abelian group that is $\mathbb{Z}_2$–graded. Define $\lambda(Y,E)$ to be half the Euler characteristic of the instanton homology:

$$\lambda(Y,E) = \tfrac{1}{2}\chi[I_*(Y,E)].$$

This number is a signed count of suitably perturbed projectively flat connections on E. It is well known that $\lambda(Y,E)$ is an integer. Define the subset of triples

$$V_Y = \big\{\{a,b,c\} \subset H^1(Y;\mathbb{Z}_2) : a+b+c=0\big\}.$$

This set is naturally in correspondence with the set of subspaces of the $\mathbb{Z}_2$–vector space $H^1(Y;\mathbb{Z}_2)$ of dimension at most two. Write $b_1(2)$ for the $\mathbb{Z}_2$–dimension of $H_1(Y;\mathbb{Z}_2)$. Define for any given $x \in H^2(Y;\mathbb{Z}_2)$ the following nonnegative integer:

$$v_Y(x) = \Big|\big\{\{a,b,c\} \in V_Y : ab+bc+ac = x\big\}\Big|.$$

For the case in which $x = w_2(E)$ we simply write $v_Y(E)$.

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2841

Theorem 1.1 *Suppose E is a nontrivial admissible $U(2)$ bundle over a closed, oriented, connected 3–manifold Y with $b_1(2) \geq 3$. Then $\lambda(Y, E)$ is divisible by $2^{b_1(2)-3}$. Furthermore, we have*

$$2^{3-b_1(2)}\lambda(Y, E) \equiv v_Y(E) \pmod 2. \tag{1}$$

If $b_1(2) = 2$, this congruence also holds, implying that $v_Y(E)$ is even. If $b_1(2) = 1$, then the integer $v_Y(E)$ is zero. In these two cases $v_Y(E) \pmod 2$ yields no information about $\lambda(Y, E)$.

Note that Y supports a nontrivial admissible bundle if and only if $b_1(Y) \geq 1$, where $b_1(Y)$ denotes the rank of $H_1(Y;\mathbb{Z})$. In general we have $b_1(2) \geq b_1(Y)$, with strict inequality if and only if $H_1(Y;\mathbb{Z})$ has 2–torsion. Theorem 1.1 and its proof are generalizations of a rather simple idea due to Ruberman and Saveliev [13]. Their result is the case of Theorem 1.1 when $H_1(Y;\mathbb{Z})$ is free abelian of rank 3, ie when Y is a homology 3–torus. To obtain their statement, one identifies $v_Y(E)$ with the triple cup product modulo 2, which for a homology 3–torus is a simple computation. (More generally, see Corollary 1.6.) Our adaptation of Ruberman and Saveliev's argument is summarized, modulo perturbations, as follows.

The invariant $\lambda(Y, E)$ is one half of a signed count of projectively flat connections on the bundle E. There is an action of $H^1(Y;\mathbb{Z}_2)$ on this set of connections, and the quotient is identified with flat connections on the adjoint SO(3) bundle induced by E. The only possible stabilizers of this action are $\{1\}$, $\mathbb{Z}_2$ and V_4, the Klein-four group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. Further, the connections with stabilizer V_4 are flat connections with holonomy group V_4. The number $v_Y(E)$ is the number of connections on the induced SO(3) bundle with holonomy V_4, up to gauge equivalence. The proof of Theorem 1.1 follows from counting the $H^1(Y;\mathbb{Z}_2)$–orbits with stabilizer V_4.

Vanishing conditions, and relation to Lescop's invariant The right-hand quantity $v_Y(E) \pmod 2$ of congruence (1) is often, but not always, equal to zero. The parity also turns out to be independent of our choice of nontrivial admissible bundle E. To state the result:

$$k(Y) := \dim_{\mathbb{Z}_2}\{a \in H^1(Y;\mathbb{Z}_2) : a^2 = 0\} = \dim_{\mathbb{Z}_2}\ker(\beta^1).$$

Here β^1 is the Bockstein homomorphism defined on $H^1(Y;\mathbb{Z}_2)$ associated to the coefficient exact sequence $0 \to \mathbb{Z}_2 \to \mathbb{Z}_4 \to \mathbb{Z}_2 \to 0$. As is well known, $\beta^1(a) = a^2$. We note that if $H_1(Y;\mathbb{Z})$ is written as a direct sum of prime-power-order cyclic summands and copies of $\mathbb{Z}$, then $k(Y)$ is just the number of $\mathbb{Z}_{2^k}$ summands with $k > 1$, plus the number of $\mathbb{Z}$ summands. In particular, $k(Y) \geq b_1(Y)$.

Theorem 1.2 *Let Y be a closed, oriented and connected 3–manifold with $k(Y) \geq 1$. Let $x \in H^2(Y;\mathbb{Z}_2)$ be any element that is not a cup-square. Then $v_Y(x) \pmod 2$ is independent of the choice of such x. If furthermore $k(Y) \geq 4$ then $v_Y(x) \equiv 0 \pmod 2$.*

Note that the statement holds for a larger class of elements $x \in H^2(Y;\mathbb{Z}_2)$ than those just coming from admissible bundles. The conditions are best understood through the following examples, which are surgeries on the Borromean rings; see Figure 1. These examples have $b_1(Y) = 0$ and $b_1(2) = 3$.

Example 1.3 Consider the 3–manifold Y obtained by performing $(2,2,4)$ surgery on the Borromean rings. Such a manifold has first homology group isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_4$. Then $k(Y) = 1$. The rank-3 vector space $H^1(Y;\mathbb{Z}_2)$ has a basis formed by a, b, c, classes that are Poincaré dual to the meridians of the surgery loops. By intersecting homology classes and using Poincaré duality we obtain

$$c^2 = 0, \quad a^2 = bc, \quad b^2 = ac,$$

where ab, bc, ac form a basis of $H^2(Y;\mathbb{Z}_2)$. Now, ab is not a square, as are not $ab+bc$, $ab+ac$ or $ab+ac+bc$. All four of these elements have $v_Y(x) = 1 \in \mathbb{Z}$. On the other hand, all other elements in $H^2(Y;\mathbb{Z}_2)$ have $v_Y(x) \in \{0,2,4\}$. This illustrates the necessity of the nonsquare condition on x.

Example 1.4 Next, consider $(2,4,4)$ surgery on the Borromean rings. The $\mathbb{Z}_2$–cohomology ring is much the same as before, except now $b^2 = 0$, and $k(Y) = 2$. All nonzero $x \in H^2(Y;\mathbb{Z}_2)$ have $v_Y(x)$ odd. In fact, if $x \neq 0$, then $v_Y(x) = 1$, while $v_Y(a^2) = 5$ and $v_Y(0) = 4$. Here a^2 is a cup-square, but does not have a different parity from the other nonzero elements.

Example 1.5 Finally, $(4,4,4)$ surgery on the Borromean rings has the same $\mathbb{Z}_2$–cohomology ring as that of the 3–torus. Here $k(Y) = 3$, and $v_Y(x) = 1$ for $x \neq 0$, all nonsquares, while $v_Y(0) = 8$.

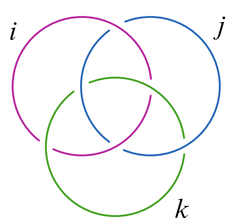


Figure 1: Surgery on the Borromean rings with framings (i,j,k) on the three components. When i, j, k are either 0 or various powers of 2, these surgeries yield nonvanishing examples of the congruence in Theorem 1.1, in which $v_Y(E) \equiv 1 \pmod 2$ and $k(Y) = 1,2,3$.

To make use of Theorem 1.1, one can replace the 4–framings in the above three examples by 0–framings, to get manifolds with the same $\mathbb{Z}_2$–cohomology rings but $b_1(Y) > 0$, ensuring that they support nontrivial admissible bundles.

In what follows, we describe how to deduce Theorem 1.2 using Theorem 1.1 and related results of Poudel [11] and Turaev [16]. By Poudel [11], the Casson invariant $\lambda(Y, E)$ may be identified with Lescop's invariant of [9], slightly modified. The proof utilizes Floer's exact triangle for instanton homology and Dehn surgery techniques à la Lescop [9]. As a result, the parity of $v_Y(E)$ is independent of E, the choice of nontrivial admissible bundle. After some substitutions, the congruences resulting from Theorem 1.1 and [11] may be summarized as follows.

Corollary 1.6 *Suppose $x \in H^2(Y;\mathbb{Z}_2)$ has no torsion lifts to $H^2(Y;\mathbb{Z})$. Then,* mod 2,

$$v_Y(x) \equiv \begin{cases} 2^{2-b_1(2)}\Delta''_Y(1) & \text{if } b_1(Y)=1,\\ 2^{3-b_1(2)}(\#(\gamma\cap F)) & \text{if } b_1(Y)=2,\\ 2^{3-b_1(2)}N\cdot(a\cup b\cup c)[Y] & \text{if } b_1(Y)=3,\\ 0 & \text{if } b_1(Y)\geq 4,\end{cases} \tag{2}$$

where N is the cardinality of Tor $H_1(Y;\mathbb{Z})$ *and other terms are defined below. In particular, if $b_1(Y)=3$ and $H_1(Y;\mathbb{Z})$ has an order-4 element, then $v_Y(x)\equiv 0 \pmod 2$.*

The right-hand sides are defined as follows. First, for $b_1(Y) = 1$, $\Delta_Y(t)$ is the Alexander polynomial of Y, normalized so that $\Delta_Y(1) = 1$ and $\Delta_Y(t) = \Delta_Y(t^{-1})$. If Y is 0–surgery on a knot K in an integral homology 3–sphere Σ, then $\Delta_Y(t)$ is just the Alexander polynomial $\Delta_{K\subset\Sigma}(t)$. Next, suppose $b_1(Y) = 2$. Take two oriented surfaces in Y that generate $H_2(Y;\mathbb{Q})$. Let γ be their intersection, and γ' the curve parallel to γ that induces the trivialization of the tubular neighborhood of γ given by the surfaces. Then $N\cdot\gamma'$ has a Seifert surface F in Y, and $\#(\gamma\cap F)$ is the count of intersection points, in general position. Finally, in the $b_1(Y) = 3$ case, the triple a, b, c generates $H^1(Y;\mathbb{Z})$ up to torsion, and $[Y]$ is the fundamental class of Y.

The vanishing implications of Corollary 1.6 look rather similar to those of Theorem 1.2, except that the role of $k(Y)$ is weakened to that of $b_1(Y)$. In other words, the role of counting summands of the form $\mathbb{Z}$ and $\mathbb{Z}_{2^k}$ for $k > 1$ is replaced by that of just counting $\mathbb{Z}$ summands. From the perspective of the $\mathbb{Z}_2$–cohomology ring, these kinds of summands are all the same. With this thought in mind, it is a rather straightforward task to establish Theorem 1.2 from Corollary 1.6 using realization results for the $\mathbb{Z}_2$–cohomology structure of 3–manifolds due to Turaev. See Section 7. We remark that, a posteriori, the divisibility properties of the quantities listed in Corollary 1.6 should imply Theorem 1.2. However, the authors prefer to mostly argue with the $\mathbb{Z}_2$–cohomology ring structure, in line with the definition of $v_Y(x)$.

Some more examples For any finitely generated abelian group H containing an element of order 4 or ∞, there is a 3–manifold Y with $H_1(Y;\mathbb{Z})$ isomorphic to H and $v_Y(x)=0$, in which x is any element that is not a cup-square. For this, just consider integer-framed surgeries on unlinks. Note also that the integer $v_Y(x)$ is stable under connect sums with $\mathbb{RP}^3$, which increases $b_1(2)$ by 1 while fixing $k(Y)$. This operation, applied to the three Borromean surgeries examples above, gives examples where $v_Y(x)\equiv 1 \pmod 2$ for any pair $b_1(2)$, $k(Y)$ such that $b_1(2)\geq 3$ and $k(Y)\in\{1,2,3\}$. In fact, it is straightforward to produce nonvanishing examples with $H_1(Y;\mathbb{Z})$ any isomorphism class of finitely generated abelian group with those same two constraints. We also have examples from Seifert-fibered spaces, with orientable base orbifold:

Proposition 1.7 *Let Y be a Seifert-fibered space with Seifert invariants given by $(g,b,(\alpha_1,\beta_1),\dots,(\alpha_r,\beta_r))$, where g is the genus of the base orbifold. Suppose that $x\in H^2(Y;\mathbb{Z}_2)$ is not a square. Then $v_Y(x)\equiv 1 \pmod 2$ if and only if $g=1$, all α_i are odd, and $b+\sum\beta_i\equiv 0 \pmod 2$.*

We note that such Seifert-fibered spaces have $b_1(Y)\in\{2,3\}$ and $b_1(2)=3$. Included in this list is of course the 3–torus. This proposition is easily proven using the description of the mod 2 cohomology ring of a Seifert-fibered space given in Aaslepp, Drawe, Hayat-Legrand, Sczesny and Zieschang [1]. See Section 8.

We mention that the Seifert-fibered spaces considered here for genus $g=0$ are double branched covers of Montesinos links. However, by Proposition 1.7 the relevant invariant $v_Y(x)$ in these cases is always even. In Section 8 we give an example of a double branched cover for which Theorem 1.1 has a nonvanishing congruence.

Discussion The integers $v_Y(x)$, and not just their parities, are interesting in the context of SO(3) gauge theory. Indeed, as is evident in the sequel, the V_4–connection classes counted by $v_Y(E)$ are persistent (unmoved) under a large class of perturbations. As such, they form a distinguished set of generators in the instanton Floer chain complex for the pair (Y,E), defined using any such perturbation. Klein-four connections also play a pivotal role in the SO(3) instanton homology for webs of Kronheimer and Mrowka [8] and its relation to the four-color theorem.

The authors did not see how to provide a general algebraic proof of Theorem 1.2, but we believe it can be done. Our main purpose in this article is to exhibit how the congruence in Theorem 1.1 requires hardly any work, once the picture for the relevant moduli spaces is established.

Finally, it should be mentioned that although we refer to the invariant $\lambda(Y,E)$ as a "Casson invariant", we are using the interpretation of Taubes [15] of Casson's invariant for integral homology 3–spheres, applied to nontrivial admissible bundles.

Outline In Section 2 we review the notion of nontrivial admissibility and the suitable generalization which motivates the hypotheses of Theorem 1.2. Sections 3 and 4 provide the background for the main argument of Theorem 1.1, which was sketched above and is presented concisely in Section 5. The issue of perturbations is ignored here, and then taken up in Section 6. In Section 7 we complete the proof of Theorem 1.2. Finally, in Section 8 we prove Proposition 1.7, record a connected sum formula for the parity of $v_Y(x)$, and discuss double branched covers.

Acknowledgements The authors would like to thank Danny Ruberman and Nikolai Saveliev for helpful discussions. Scaduto was supported by NSF grant DMS-1503100.

2 Nontrivial admissible bundles

Here we briefly discuss Floer's nontrivial admissibility condition. A good reference for this material is [2]. As in the introduction, we let Y be a closed, oriented and connected 3–manifold. An SO(3) bundle over Y is *nontrivial admissible* if its second Stiefel–Whitney class $x \in H^2(Y;\mathbb{Z}_2)$ satisfies the following three equivalent conditions; see [2, Lemma 1.1]:

- The image of x under $h\colon H^2(Y;\mathbb{Z}_2) \to \mathrm{Hom}(H_2(Y;\mathbb{Z}),\mathbb{Z}_2)$ is nonzero.
- There is an *orientable* surface $\Sigma \subset Y$ such that $\langle x,[\Sigma]\rangle \not\equiv 0$.
- The element $x \in H^2(Y;\mathbb{Z}_2)$ has no torsion lifts to $H^2(Y;\mathbb{Z})$.

One then defines a $U(2)$ bundle to be nontrivial admissible if its induced adjoint SO(3) bundle is nontrivial admissible. The definition is motivated by the fact that a nontrivial admissible $U(2)$ bundle admits no reducible flat connections. This avoids complications in instanton Floer theory. Using that h is surjective, and the fact that SO(3) bundles over a 3–manifold are characterized by the second Stiefel–Whitney class, we count the number of nontrivial admissible SO(3) bundles:

$$(2^{b_1(Y)}-1)2^{b_1(2)-b_1(Y)}.$$

According to Theorem 1.1 and Poudel's result mentioned in the introduction, the parity of $v_Y(E)$ is the same for all nontrivial admissible bundles E. However, Theorem 1.2 indicates that the parity of $v_Y(E)$ is invariant under a larger collection of bundles. Such bundles are characterized by having a second Stiefel–Whitney class $x \in H^2(Y;\mathbb{Z}_2)$ that satisfies the following equivalent conditions:

- The image of x under $g\colon H^2(Y;\mathbb{Z}_2) \to \mathrm{Hom}(\mathrm{PD}(\ker(\beta^1)),\mathbb{Z}_2)$ is nonzero.
- There is a surface $\Sigma \subset Y$ such that $\langle x,[\Sigma]\rangle \not\equiv 0$ and $\Sigma\cdot\Sigma \equiv 0 \in H_1(Y;\mathbb{Z}_2)$.
- The element $x \in H^2(Y;\mathbb{Z}_2)$ has no order-2 lifts to $H^2(Y;\mathbb{Z})$.
- The element $x \in H^2(Y;\mathbb{Z}_2)$ is not the cup-square of an element from $H^1(Y;\mathbb{Z}_2)$.

Note that here Σ is not necessarily orientable, and PD: $H^1(Y;\mathbb{Z}_2) \to H_2(Y;\mathbb{Z}_2)$ is the Poincaré duality isomorphism. We briefly remark on the equivalence of these conditions, leaving the details to the reader. The first two bullets are equivalent because $\mathrm{PD}(\ker(\beta^1)) \subset H_2(Y;\mathbb{Z}_2)$ is spanned by the classes $[\Sigma]$ with the stated conditions. The equivalence of the third and fourth conditions follow from understanding the Bockstein homomorphisms in this setting — see eg [5, Section 3.E] — and the remaining equivalences make use of the nondegeneracy of Poincaré duality. These conditions are the natural extensions of the prior three conditions when one wants to treat $\mathbb{Z}$ summands and $\mathbb{Z}_{2^k}$ summands for $k > 1$ the same. We note that the ring $H^*(Y;\mathbb{Z}_2)$ cannot see the difference between such summands. Since g is surjective, the number of SO(3) bundles of this more general type is

$$(2^{k(Y)} - 1)2^{b_1(2)-k(Y)}.$$

The most basic example of such a bundle that is not nontrivial admissible is the nontrivial SO(3) bundle over the lens space $L(4,1)$.

3 Configuration spaces and stabilizers

Fix a connection A_0 on $\det(E)$, and let $\mathcal{C}_E$ be the space of connections A on E with determinant connection $\mathrm{Tr}(A) = A_0$. Let $\mathcal{G}_E$ be the gauge transformation group consisting of smooth unitary automorphisms of E that are determinant 1. The configuration space is the quotient $\mathcal{B}_E = \mathcal{C}_E/\mathcal{G}_E$. The nontrivial admissibility of E implies that all projectively flat points in $\mathcal{B}_E$ are irreducible, meaning that the $\mathcal{G}_E$–stabilizer of every such connection $A \in \mathcal{C}_E$ is as small as possible:

$$\mathrm{Stab}_{\mathcal{G}_E}(A) = \{\pm 1\}.$$

The $U(2)$ bundle E induces an SO(3) bundle $\mathfrak{su}(E)$, which may be defined as the subbundle of $\mathrm{End}(E)$ consisting of trace-free, skew-hermitian endomorphisms. We let $\mathcal{G}_{\mathfrak{su}(E)}$ denote the full SO(3) gauge transformation group of $\mathfrak{su}(E)$. Any $A \in \mathcal{C}_E$ induces a connection $A_{\mathrm{ad}} \in \mathcal{C}_{\mathfrak{su}(E)}$, and this induces a bijection between $\mathcal{C}_E$ and $\mathcal{C}_{\mathfrak{su}(E)}$. Indeed, any $U(2)$ connection A on E is uniquely determined by $\mathrm{Tr}(A)$ on $\det(E)$ and A_{ad} on $\mathfrak{su}(E)$. The condition that A be projectively flat is equivalent to A_{ad} being flat. In contrast to the $U(2)$ case, however, when A_{ad} is flat we have

$$\mathrm{Stab}_{\mathcal{G}_{\mathfrak{su}(E)}}(A_{\mathrm{ad}}) \in \big\{\{1\}, \mathbb{Z}_2, V_4\big\}.$$

Indeed, the difference between the determinant-1 unitary gauge group and the SO(3) gauge group is described by an action of $H^1(Y;\mathbb{Z}_2)$ on $\mathcal{B}_E$ that gives $\mathcal{B}_{\mathfrak{su}(E)}$ as its quotient space. The action is as follows: $H^1(Y;\mathbb{Z}_2)$ parametrizes the isomorphism classes of flat complex line bundles (with connection) χ with holonomy $\{\pm 1\}$. Then

$[\chi]$ acts on $[A]\in\mathcal{B}_E$ by tensoring the bundle-with-connection (E,A) with χ. See eg [3, Section 5.6]. We then have the more precise statement that $\mathrm{Stab}_{\mathcal{G}_{\mathfrak{su}(E)}}(A_{\mathrm{ad}})$ is naturally a subspace of $H^1(Y;\mathbb{Z}_2)$, with the constraint that

$$\dim_{\mathbb{Z}_2}\mathrm{Stab}_{\mathcal{G}_{\mathfrak{su}(E)}}(A_{\mathrm{ad}})\in\{0,1,2\}.$$

In summary, we see that even though any projectively flat connection in $\mathcal{B}_E$ is irreducible, its image in $\mathcal{B}_{\mathfrak{su}(E)}$ may not be irreducible. Flat connections on $\mathfrak{su}(E)$ with $\mathcal{G}_{\mathfrak{su}(E)}$–stabilizer isomorphic to $\mathbb{Z}_2$ are exactly those whose holonomy is contained in $O(2)$, but not in an SO(2) or Klein-four subgroup. Equivalently, these are flat connections that are compatible with a splitting

$$\mathfrak{su}(E)=\lambda\oplus L,$$

where λ is a nontrivial real line bundle and L is an unoriented real 2–plane bundle, and for which the connection on L is irreducible. Connections with stabilizer V_4 are those whose holonomy is also isomorphic to V_4. Equivalently, these are flat connections compatible with a splitting

$$\mathfrak{su}(E)=\lambda_1\oplus\lambda_2\oplus\lambda_3$$

into a sum of three nontrivial real line bundles. We write $\mathcal{B}^{V_4}_{\mathfrak{su}(E)}\subset\mathcal{B}_{\mathfrak{su}(E)}$ for the subset of flat connections on $\mathfrak{su}(E)$ with V_4–stabilizer, which we henceforth call *Klein-four connections*.

Remark 3.1 If the assumption of nontrivial admissibility is removed, three other kinds of stabilizers in the SO(3)–gauge group can occur: SO(2), $O(2)$ and SO(3).

4 Klein-four connections

The subset of Klein-four connections in $\mathcal{B}_{\mathfrak{su}(E)}$ is a finite, discrete set. As the elements are characterized by having holonomy V_4, a finite group, they must all be flat, as a simple continuity argument shows. Alternatively, each splitting $\mathfrak{su}(E)=\lambda_1\oplus\lambda_2\oplus\lambda_3$ into nontrivial real line bundles supports a unique compatible connection, which of course must be flat. Let us consider the larger set

$$\mathcal{B}^{\geq V_4}=\{\text{connections over } Y \text{ on any SO(3) bundle with holonomy inside a } V_4\}/\text{gauge}.$$

Then $\mathcal{B}^{\geq V_4}$ is parametrized by SO(3) bundles of the form $\lambda_1\oplus\lambda_2\oplus\lambda_3$ over Y. Noting that $w_1(\mathfrak{su}(E))=0$, sending such a bundle to the triple $\{w_1(\lambda_1),w_1(\lambda_2),w_1(\lambda_3)\}$ sets up a bijection

$$\mathcal{B}^{\geq V_4}\overset{1:1}{\longleftrightarrow}\bigl\{\{a,b,c\}\subset H^1(Y;\mathbb{Z}_2)\text{ with } a+b+c=0\bigr\}=:V_Y.$$

Yet another description of $\mathcal{B}^{\geq V_4}$ is as the set of homomorphisms $\mathrm{Hom}(\pi_1(Y), V_4)$ modulo the action of $S_3 = \mathrm{Aut}(V_4)$. A simple counting argument shows that $\mathcal{B}^{\geq V_4}$ has cardinality

$$2^{b_1(2)-1} + \tfrac{1}{6}(4^{b_1(2)} + 2).$$

Now, the elements of $\mathcal{B}^{\geq V_4}$ that live on $\mathfrak{su}(E)$ are the ones with

$$w_2(E) = w_2(\lambda_1 \oplus \lambda_2 \oplus \lambda_3) = a_1a_2 + a_2a_3 + a_1a_3, \quad a_i = w_1(\lambda_i).$$

Thus we have the following bijection describing Klein-four connections on $\mathfrak{su}(E)$:

$$\mathcal{B}^{V_4}_{\mathfrak{su}(E)} \overset{1:1}{\longleftrightarrow} \big\{\{a,b,c\} \in V_Y \text{ with } ab + bc + ac = w_2(E)\big\}.$$

We see now that $v_Y(E) = |\mathcal{B}^{V_4}_{\mathfrak{su}(E)}|$, and the statement of Theorem 1.1 is the congruence

$$\lambda(Y,E) \equiv 2^{b_1(2)-3} \cdot |\mathcal{B}^{V_4}_{\mathfrak{su}(E)}| \pmod{2^{b_1(2)-2}}. \tag{3}$$

5 The argument modulo perturbations

We now prove Theorem 1.1 under the assumption that all moduli spaces to follow are nondegenerate, so that no perturbations are needed. The argument uses the most basic information we have from the $H^1(Y;\mathbb{Z}_2)$–action. Consider the moduli space of projectively flat connections on E:

$$\mathcal{M}_E := \big\{[A] \in \mathcal{B}_E : F_A = \tfrac{1}{2}F_{A_0} \cdot \mathrm{id}_E\big\}.$$

This is a finite set, and each of its points is irreducible. This moduli space is invariant under the $H^1(Y;\mathbb{Z}_2)$–action, and its quotient is the space of flat connections on $\mathfrak{su}(E)$:

$$\mathcal{M}_{\mathfrak{su}(E)} := \{[B] \in \mathcal{B}_{\mathfrak{su}(E)} : F_B = 0\}.$$

We need the following observation. An element $w \in H^1(Y;\mathbb{Z}_2)$ affects the relative mod 8 Floer grading $\mathrm{gr}[A]$ of $[A] \in \mathcal{M}_E$ by (see [2, Propositions 1.9 and 1.13])

$$\mathrm{gr}(w \cdot [A]) - \mathrm{gr}[A] \equiv 4(w_2(E)w + w^3)[Y] \pmod 8,$$

so the $H^1(Y;\mathbb{Z}_2)$–action preserves the $\mathbb{Z}_2$–gradings. Here $[Y]$ is the fundamental class of Y. In particular, each $H^1(Y;\mathbb{Z}_2)$–orbit lies in a single $\mathbb{Z}_2$–grading. The proof is now completed by counting orbit sizes. Each connection in $\mathcal{M}_{\mathfrak{su}(E)}$ with stabilizer at most $\mathbb{Z}_2$ gives an orbit of size either

$$|H^1(Y;\mathbb{Z}_2)| = 2^{b_1(2)} \quad \text{or} \quad |H^1(Y;\mathbb{Z}_2)/\mathbb{Z}_2| = 2^{b_1(2)-1},$$

lying upstairs in $\mathcal{M}_E$. Thus $2^{b_1(2)-1}$ divides the *signed* count of $\mathcal{M}_E$, with the prior observation about gradings in mind. The remaining connections downstairs in $\mathcal{M}_{\mathfrak{su}(E)}$

are Klein-four connections, and so in fact are given by the set $\mathcal{B}^{V_4}_{\mathfrak{su}(E)}$. Each point in this set contributes an orbit of size

$$|H^1(Y;\mathbb{Z}_2)/V_4| = 2^{b_1(2)-2}$$

upstairs in $\mathcal{M}_E$. Recalling that $\lambda(Y,E)$ is *half* the signed count of points in $\mathcal{M}_E$, we recover the congruence (3), proving Theorem 1.1 under the assumption of nondegeneracy.

6 Including holonomy perturbations

In general, the moduli space $\mathcal{M}_E$ is degenerate and we need to perturb the projectively flat equation to achieve the transversality we want. Henceforth we assume that our 3–manifold Y is equipped with a Riemannian metric. The standard class of perturbations used are known as *holonomy perturbations* [6; 13]. The input for such a perturbation is an embedding $\Gamma = \{\gamma_k\}_{k=1}^m$ into Y of solid tori $\gamma_k\colon S^1\times D^2 \to Y$. We require that the embedded tori γ_k have a common normal disk, meaning that the image of $\{1\}\times D^2$ under γ_k is the same for all k. We also require that the images of the core loops $S^1\times\{0\}$ are disjoint away from the normal disk. Fix a trivialization of $\det(E)$ over the image of Γ, which is homotopically a wedge (bouquet) of circles. This allows us to consider the holonomy around the γ_k as living in $\mathrm{SU}(2)$. Let $f\colon \mathrm{SU}(2)^m \to \mathbb{R}$ be a conjugation invariant function, ie

$$f(ga_1g^{-1},\dots,ga_mg^{-1}) = f(a_1,\dots,a_m) \quad \text{for all } g\in \mathrm{SU}(2).$$

We also choose a smooth 2–form μ on D^2 with compact support in the interior and integral 1. From this data one constructs a holonomy perturbation h, given as follows:

$$h(A) = \int_{D^2} f(\mathrm{Hol}_{\gamma_{1,z}}(A),\dots,\mathrm{Hol}_{\gamma_{m,z}}(A))\,\mu(z).$$

Here $\gamma_{k,z}$ is the loop $t\mapsto \gamma_k(t,z)$ in Y. Fixing only the data Γ, we define $\mathcal{H}_\Gamma$ to be the space of perturbations constructed as above. Each $h\in\mathcal{H}_\Gamma$ yields a well-defined function $h\colon \mathcal{B}_E\to\mathbb{R}$.

One way to guarantee that the perturbation h is $H^1(Y;\mathbb{Z}_2)$–equivariant is to require that each loop $\mathrm{im}(\gamma_k)$ is zero as a class in $H_1(Y;\mathbb{Z}_2)$. We call such Γ *mod-*2 *trivial*, following [13], where this condition is introduced. We record their observation:

Lemma 6.1 *If Γ is mod-2 trivial, then each $h\in\mathcal{H}_\Gamma$ is $H^1(Y;\mathbb{Z}_2)$–equivariant.*

Now, the perturbed $U(2)$ moduli space $\mathcal{M}^h_E$ is the set of critical points of the perturbed Chern–Simons functional $\mathrm{CS}+h$. Specifically, for a suitable normalization of CS, we obtain

$$\mathcal{M}^h_E = \big\{[A]\in\mathcal{B}_E : F_A - \tfrac12 F_{A_0}\cdot \mathrm{id}_E + \star\,\nabla h(A) = 0\big\}.$$

If Γ is mod-2 trivial, this perturbed moduli space inherits the $H^1(Y;\mathbb{Z}_2)$–action from $\mathcal{B}_E$, and its quotient space is the perturbed SO(3) moduli space for $\mathfrak{su}(E)$. We also record the following:

Lemma 6.2 *Suppose Γ is mod-2 trivial. For any $h \in \mathcal{H}_\Gamma$, Klein-four connections are unmoved in the SO(3) moduli space. More precisely, we always have the relation*

$$\mathcal{M}^h_{\mathfrak{su}(E)} \cap \mathcal{B}^{V_4}_{\mathfrak{su}(E)} = \mathcal{B}^{V_4}_{\mathfrak{su}(E)}.$$

As such perturbations are $H^1(Y;\mathbb{Z}_2)$–equivariant, a similar statement holds for the connections in the $U(2)$ moduli space $\mathcal{M}^h_E$ lying above Klein-four connections. In fact, the lemma clearly follows from this latter case, which is justified as follows. First, the $H^1(Y;\mathbb{Z}_2)$–equivariance of our perturbations imply that Klein-four connections in $\mathcal{B}_E$ are always perturbed to Klein-four connections. Second, we recall that the space of Klein-four connection classes is a finite discrete set. Important here is our earlier observation that any connection with Klein-four stabilizer is in fact flat. In particular, the gradient of our perturbation is a Klein-four invariant vector $v \in T\mathcal{B}_E$, which must be the 0 vector by discreteness of the set of Klein-four connections.

Our goal is to find a mod-2 trivial Γ such that for small, generic $h \in \mathcal{H}_\Gamma$ the moduli space $\mathcal{M}^h_E$ is nondegenerate. Section 5 of [13] shows that this can be achieved if Γ is *abundant* at each projectively flat $[A] \in \mathcal{M}_E$. We need to slightly generalize the definition of abundancy given in [13], which only considers stabilizers isomorphic to $\{1\}$ and $\mathbb{Z}_2$. To begin, note that $H^1(Y;A_{\mathrm{ad}})$, the Zariski tangent space to $[A]$ in $\mathcal{M}_E$, carries an action by the stabilizer, denoted

$$S_A := \mathrm{Stab}_{H^1(Y;\mathbb{Z}_2)}[A] = \mathrm{Stab}_{\mathcal{G}_{\mathfrak{su}(E)}}(A_{\mathrm{ad}}). \tag{4}$$

We remark that the second equality in (4) is not true in general, and is contingent upon the nontrivial admissibility of E. Recall that S_A is one of $\{1\}$, $\mathbb{Z}_2$ or V_4. Now, decompose the tangent space into its S_A–invariant subspace V_A, and the S_A–equivariant orthogonal complement to V_A:

$$H^1(Y;A_{\mathrm{ad}}) = V_A \oplus V_A^{\perp}.$$

The space V_A is the Zariski tangent space of $[A]$ internal to the stratum of $\mathcal{M}_E$ consisting of connection classes with stabilizer isomorphic to S_A. The complement $V_A^{\perp}$ is the Zariski normal bundle fiber in $\mathcal{M}_E$ at $[A]$ relative to the aforementioned stratum. For a vector space W we write $\mathrm{Sym}(W)$ for the space of symmetric bilinear forms on W. If W has a linear G–action by some group G, we write $\mathrm{Sym}(W)^G$ for the forms that are G–invariant.

Definition 6.3 A mod-2 trivial Γ is *abundant* at a projectively flat $[A] \in \mathcal{M}_E$ if there exist perturbations $\{h_i\}_{i=1}^n \subset \mathcal{H}_\Gamma$ and some k such that $Dh_i(A) = 0$ for $k+1 \le i \le n$, and such that the following map that is defined from $\mathbb{R}^n$ to $\mathrm{Hom}(V_A, \mathbb{R}) \oplus \mathrm{Sym}(V_A^\perp)^{S_A}$ is surjective:

$$(x_1, \dots, x_n) \mapsto \left(\sum_{i=1}^k x_i \, Dh_i(A), \sum_{i=k+1}^n x_i \, \mathrm{Hess}\, h_i(A) \right). \tag{5}$$

Note that if S_A is trivial, then V_A accounts for the entire tangent space, and in particular $V_A^\perp = 0$. Thus only the left-hand factor of the map (5) is relevant. This is the condition of "first-order abundancy", and is sufficient to achieve nondegeneracy for small, generic perturbations when there are no other (lower) strata to consider. At the other extreme, when S_A is isomorphic to V_4, we have $V_A = 0$. In this case (5) reduces to a condition purely of "second-order abundancy".

If S_A is isomorphic to $\mathbb{Z}_2$, then V_A and $V_A^\perp$ are the $+1$ and -1 eigenspaces of the $\mathbb{Z}_2$–action, respectively, and are V_+ and V_- in the notation of [13]. In this case $\mathrm{Sym}(V_A^\perp)^{S_A}$ is the same as $\mathrm{Sym}(V_-)$. Our choice of $\mathrm{Sym}(V_A^\perp)^{S_A}$ in Definition 6.3 is sufficient for the arguments of Section 5 in [13] to go through in part because a generic element therein is nondegenerate; see the proof of Proposition 5.4 in [13]. When S_A is isomorphic to $\{1\}$ or $\mathbb{Z}_2$, our definition agrees with that of [13].

We are left with producing a mod-2 trivial Γ which is abundant for all $[A] \in \mathcal{M}_E$. To this end, the work of Ruberman and Saveliev implies the following:

Lemma 6.4 [13, Proposition 5.2] *There exists a mod-2 trivial Γ that is abundant for all connections in $\mathcal{M}_E$ that do not descend to $\mathrm{SO}(3)$ Klein-four connections.*

This allows us to focus on the situations in which S_A is isomorphic to V_4, the case in which A_{ad} is a Klein-four connection. We have the following facts, used in [13, Section 5.5], stated informally:

- If Γ is abundant, and Γ' is close to Γ, then Γ' is abundant.
- If Γ is abundant and $\Gamma \subset \Gamma'$, then Γ' is abundant.

In these situations, we are assuming that Γ and Γ' have the same fixed normal disk with basepoint. Now suppose we can show, for each A with S_A isomorphic to V_4, the existence of a mod-2 trivial Γ abundant at $[A]$. Then it is straightforward to conclude, using these two facts and Lemma 6.4, that there exists a mod-2 trivial Γ' abundant at all $[A] \in \mathcal{M}_E$. Thus the following lemma completes the proof of Theorem 1.1:

Lemma 6.5 *There is an abundant mod-2 trivial Γ for any $[A] \in \mathcal{M}_E$ that descends to an $\mathrm{SO}(3)$ Klein-four connection.*

Proof We follow the method used in [13] of passing to a finite cover. Let A be a projectively flat connection on E with stabilizer S_A isomorphic to V_4. The SO(3) connection A_{ad} is compatible with a splitting $\mathfrak{su}(E) = \lambda_1 \oplus \lambda_2 \oplus \lambda_3$ in which the λ_i are nontrivial and distinct real line bundles. The stabilizer S_A is given explicitly by

$$S_A = \{0, a_1, a_2, a_3\} \subset H^1(Y; \mathbb{Z}_2), \quad a_i = w_1(\lambda_i).$$

Here, a_i corresponds to the gauge transformation of $\mathfrak{su}(E)$ that simultaneously reflects λ_{i+1} and λ_{i+2}, while fixing λ_i, where indices are taken mod 3. Define a homomorphism $\pi_1(Y) \to S_A$ by

$$\gamma \mapsto a_1(\gamma)a_1 + a_2(\gamma)a_2 + a_3(\gamma)a_3.$$

Let $p\colon Y' \to Y$ be the covering space corresponding to this homomorphism. Under this covering A_{ad} pulls back to a trivial connection, denoted A'_{ad}; see [13, Lemma 5.6]. In particular, each of λ_i pulls back under p to a trivial real line bundle λ'_i. Note that the covering transformation group of $Y' \to Y$ is the Klein-four group S_A.

It is known [6, Proposition 67 and Lemma 58] that there is some Γ', a collection of embedded solid tori in Y', that is abundant at the trivial connection A'_{ad} in the following sense: there exist perturbations $\{h_i\}_{i=1}^n \subset \mathcal{H}_{\Gamma'}$ such that the map from $\mathbb{R}^n$ to $\mathrm{Sym}(H^1(Y'; A'_{\mathrm{ad}}))^{\mathrm{SO}(3)}$ given by

$$(x_1, \dots, x_n) \mapsto \sum_{i=1}^n x_i \,\mathrm{Hess}\, h_i(A'_{\mathrm{ad}}) \tag{6}$$

is surjective. The appearance of the SO(3) here is the gauge stabilizer of the connection A'_{ad}. Let Γ be the image of Γ' under p, slightly perturbed in Y so that it is of the form described at the beginning of this section. By construction, Γ is mod-2 trivial. Consider the following map:

$$\mathrm{Sym}(H^1(Y'; A'_{\mathrm{ad}}))^{\mathrm{SO}(3)} \to \mathrm{Sym}(H^1(Y; A_{\mathrm{ad}}))^{V_4}. \tag{7}$$

Here the V_4 refers to S_A. The map (5) is the composition of (6) with (7). Thus, to show abundancy of Γ at A, it suffices to show that (7) is surjective. The map (7) is induced by the pull-back map:

$$V_4 \circlearrowright H^1(Y; A_{\mathrm{ad}}) \xrightarrow{\;p^*\;} H^1(Y'; A'_{\mathrm{ad}}) \circlearrowleft \mathrm{SO}(3) \tag{8}$$

This map is equivariant with respect to the indicated gauge stabilizer actions, upon considering V_4 as a subgroup of $\mathrm{SO}(3)$. More precisely, V_4 refers to the $\mathcal{G}_{\mathfrak{su}(E)}$–stabilizer of A_{ad}, while SO(3) refers to the $\mathcal{G}_{p^*\mathfrak{su}(E)}$–stabilizer of A'_{ad}.

To show that (7) is surjective, consider the following two decompositions:

$$H^1(Y;A_{\mathrm{ad}}) = \bigoplus_{i=1}^{3} H^1(Y;\lambda_i), \quad H^1(Y';A'_{\mathrm{ad}}) = H^1(Y';\mathbb{R}) \otimes \mathbb{R}^3. \tag{9}$$

Implicit here is a trivialization for each λ'_i, and the $\mathbb{R}^3$ should be thought of as coming from the induced trivialization of $\lambda'_1 \oplus \lambda'_2 \oplus \lambda'_3$. The map (8) respects these decompositions. In the left-hand decomposition of (9), the V_4 action is as follows: a_i acts as -1 on $H^1(Y;\lambda_{i+1}) \oplus H^1(Y;\lambda_{i+2})$, and $+1$ on $H^1(Y;\lambda_i)$. In the tensor product appearing in (9), the SO(3)–action on $\mathbb{R}^3$ is standard, and is trivial on $H^1(Y';\mathbb{R})$. From these descriptions, it is straightforward to verify that these decompositions induce identifications between the domain and codomain of (7) with $\mathrm{Sym}(H^1(Y';\mathbb{R}))$ and $\bigoplus_{i=1}^3 \mathrm{Sym}(H^1(Y;\lambda_i))$, respectively. The map (7) can then be seen as the map

$$\mathrm{Sym}(H^1(Y';\mathbb{R})) \to \bigoplus_{i=1}^{3} \mathrm{Sym}(H^1(Y;\lambda_i)), \tag{10}$$

in which each of the three components is the map induced by pull-back, after trivializing λ'_i. Now, (10) is surjective because the three relevant pull-back maps are injective, and their three images pairwise intersect at 0. This is evident from the decomposition

$$H^1(Y';\mathbb{R}) = H^1(Y;\mathbb{R}) \oplus H^1(Y;\lambda_1) \oplus H^1(Y;\lambda_2) \oplus H^1(Y;\lambda_3),$$

which is induced by the covering transformation group S_A acting on $H^1(Y';\mathbb{R})$. This action should not to be confused with the gauge stabilizer action of S_A on $H^1(Y;A_{\mathrm{ad}})$ which was used above. The summand $H^1(Y;\mathbb{R})$ is the invariant subspace under this action, while $H^1(Y;\lambda_i)$ is the complement of $H^1(Y;\mathbb{R})$ inside the invariant subspace for the subgroup $\{0, a_i\}$. □

Remark 6.6 For a discussion of some of the technical assumptions used here, see Section 5.6 of [13]. For a detailed study of the abundancy of holonomy perturbations in the context of the equivariant Kuranishi method, see [7].

7 Establishing the vanishing result

Here we complete the proof of Theorem 1.2. The remaining step is to use a realization result for the $\mathbb{Z}_2$–cohomology ring due to Turaev in conjunction with Corollary 1.6. Recall that for a closed, oriented and connected 3–manifold we have the triple cup product form

$$u_Y\colon H^1(Y;\mathbb{Z}_2) \otimes H^1(Y;\mathbb{Z}_2) \otimes H^1(Y;\mathbb{Z}_2) \to \mathbb{Z}_2, \quad u_Y(a,b,c) = (a \cup b \cup c)[Y].$$

The trilinear form u_Y determines the $\mathbb{Z}_2$–cohomology ring of Y. It was originally proven by Postnikov that any symmetric trilinear form satisfying $u(a,a,b) = u(b,b,a)$ is realized by a closed, oriented and connected 3–manifold. Recall also that we have the linking form

$$L_Y\colon \operatorname{Tor} H_1(Y;\mathbb{Z}) \otimes \operatorname{Tor} H_1(Y;\mathbb{Z}) \to \mathbb{Q}/\mathbb{Z},$$

which is a nondegenerate symmetric bilinear form. The linking form interacts with the $\mathbb{Z}_2$–cohomology ring in the following way. Let $\psi\colon \mathbb{Z}_2 \to \mathbb{Q}/\mathbb{Z}$ be the injection defined by $\psi(k \pmod 2) = k/2$. Then for all $a, b \in H^1(Y;\mathbb{Z}_2)$ we have the relation

$$\psi(u_Y(a,a,b)) = L_Y(a^\dagger, b^\dagger), \tag{11}$$

where for any $a \in H^1(Y;\mathbb{Z}_2)$ the element $a^\dagger \in \operatorname{Tor} H_1(Y;\mathbb{Z})$ is defined by the condition that $L_Y(a^\dagger, c) = \psi(a(c))$ for all $c \in \operatorname{Tor} H_1(Y;\mathbb{Z})$. Here we are of course identifying $H^1(Y;\mathbb{Z}_2)$ with $\operatorname{Hom}(H_1(Y;\mathbb{Z}), \mathbb{Z}_2)$. An implication of Turaev's work is the following result (see also [10; 14] for related results):

Theorem 7.1 [16] *Let H be a finitely generated abelian group, and let*

$$u\colon \operatorname{Hom}(H, \mathbb{Z}_2)^{\otimes 3} \to \mathbb{Z}_2$$

be a symmetric trilinear form. There exists a closed, orientable and connected 3–manifold Y such that the pair (H,u) is equivalent to $(H_1(Y;\mathbb{Z}), u_Y)$ if and only if there exists a nondegenerate symmetric bilinear form $L\colon \operatorname{Tor} H^{\otimes 2} \to \mathbb{Q}/\mathbb{Z}$ such that (11) holds with $u_Y = u$ and $L_Y = L$.

Proof of Theorem 1.2 Let Y be such that $k(Y) \geq 4$, and suppose that x is not a cup-square. Equivalently, x has no order-2 lift to $H^2(Y;\mathbb{Z})$. Our goal is show that $v_Y(x) \equiv 0 \pmod 2$. We choose an isomorphism

$$H_1(Y;\mathbb{Z}) \simeq \bigoplus_{i=1}^{4} A_i \oplus B,$$

where A_i is an abelian group of the form $\mathbb{Z}_{2^k}$ for $k > 1$ or a copy of $\mathbb{Z}$. Make these choices so that x has a lift to $H^2(Y;\mathbb{Z})$ with support in A_1, not of order 2, which can be done by our assumption on x. Recall that $\operatorname{Tor} H_1(Y;\mathbb{Z})$ is the torsion of $H^2(Y;\mathbb{Z})$ by the universal coefficients theorem. Now define H by replacing the A_i summands with copies of $\mathbb{Z}$:

$$H := \bigoplus_{i=1}^{4} A_i' \oplus B, \quad A_i' := \mathbb{Z}$$

With our identifications we have a natural isomorphism between $H^1(Y;\mathbb{Z}_2)$ and $\operatorname{Hom}(H, \mathbb{Z}_2)$, and with this understood we set $u := u_Y$. Also, noting that $\operatorname{Tor} H$

is simply Tor $H_1(Y;\mathbb{Z})$ with some summands possibly thrown away, we define L to be the restriction of L_Y. With our identifications, the terms appearing in (11) are unchanged. Thus Theorem 7.1 implies the existence of a closed, oriented and connected 3–manifold Z with first homology and triple cup product form given by (H,u). By our choices, x has no torsion lifts, and is thus equal to $w_2(E)$ for a nontrivial admissible $U(2)$ bundle E over Z. Now Poudel's result in the guise of Corollary 1.6 says $v_Z(x) \equiv 0 \pmod 2$, since $b_1(Z) \geq 4$. Since the $\mathbb{Z}_2$–cohomology rings of Y and Z are the same, we then get $v_Y(x) \equiv 0 \pmod 2$. The independence of x as a choice having no order-2 lift to $H^2(Y;\mathbb{Z})$ is established in much the same way as the vanishing. □

8 Examples and properties

In this section we prove Proposition 1.7, which yields examples of $v_Y(x) \pmod 2$ for Seifert-fibered spaces. We then produce a connected sum formula for the parity of $v_Y(x)$. Finally, we illustrate how to compute $v_Y(x)$ for double branched covers of links.

Seifert-fibered spaces Let Y be a Seifert-fibered 3–manifold over an oriented base orbifold, with Seifert invariants $(g,b,(\alpha_1,\beta_1),\dots,(\alpha_r,\beta_r))$. Here g is the genus of the base orbifold. The mod 2 cohomology ring of Y is completely described in [1].

Lemma 8.1 *Suppose $x \in H^2(Y;\mathbb{Z}_2)$ is not a square. If any of the α_i are even, or if all α_i are odd and $b+\sum\beta_i \equiv 1 \pmod 2$, then $v_Y(x)=0$.*

Proof We begin with the following easily verified observation. In general, we have

$$\{a^2 : a \in H^1(Y;\mathbb{Z}_2)\} \subset \{ab : a,b \in H^1(Y;\mathbb{Z}_2)\}. \tag{12}$$

When these sets are equal, then $v_Y(x)=0$. For if the triple $\{a,b,a+b\} \in V_Y$ had $a^2+b^2+ab=x$, then x would in fact be a square, a contradiction. Now we appeal to [1, Theorem 2.9]. When there is some even α_i ("case $n=0$" in [1]), we easily check that these two sets in (12) are equal. This is particularly immediate when there is an α_i divisible by 4, and the mod 2 cohomology ring of Y is isomorphic to that of a connect sum of some copies of $\mathbb{RP}^3$ and some copies of $S^1 \times S^2$. Finally, if all α_i are odd and $b+\sum\beta_i \equiv 1 \pmod 2$, then the ring is isomorphic to that of a connect sum of $2g$ copies of $S^1 \times S^2$, whence by the same reasoning $v_Y(x)=0$. □

Proof of Proposition 1.7 First, since $b_1(Y)$ is equal to either $2g$ or $2g+1$, the integer $v_Y(x)$ is even by Corollary 1.6 unless $g=1$. By the above lemma, it remains to check that $v_Y(x) \equiv 1 \pmod 2$ when $g = 1$ and all α_i are odd and $b + \sum \beta_i \equiv 0 \pmod 2$. One can conclude from [1] that $H^1(Y;\mathbb{Z}_2)$ has a basis a, b, c with $a^2 = b^2 = 0$ and nonzero products ab, bc, ac, the three of which provide a basis for $H^2(Y;\mathbb{Z}_2)$. Depending on some divisibility conditions on the β_i, either $c^2 = 0$ or $c^2 = ab$. The element ac, for one, is never a square, so we set $x = ac$. In either case we compute $v_Y(x) = 1$. □

Connected sums Now let x be any element of $H^2(Y;\mathbb{Z}_2)$. Recall that V_Y may be viewed as $\mathrm{Hom}(\pi_1(Y), V_4)$ modulo the action of $S_3 = \mathrm{Aut}(V_4)$. As such, it makes sense to keep track of the S_3–stabilizers of the orbits. For a set X with S_3–action we define the triple $\check{v}(X) = (\check{v}_1, \check{v}_2, \check{v}_3)$ where $\check{v}_1$, $\check{v}_2$, $\check{v}_3$ are the numbers of orbits with stabilizers of orders 1, 2, 6, respectively. For two such sets X_1 and X_2 with S_3–actions we have

$$\check{v}(X_1 \times_{S_3} X_2) = \check{v}(X_1) \times \check{v}(X_2),$$

where we define the product $\times$ on triples as follows:

$$\check{v} \times \check{u} := (6\check{v}_1\check{u}_1 + 3\check{v}_1\check{u}_2 + 3\check{v}_2\check{u}_1 + \check{v}_1\check{u}_3 + \check{v}_3\check{u}_1 + \check{v}_2\check{u}_2,\ \check{v}_2\check{u}_2 + \check{v}_2\check{u}_3 + \check{v}_3\check{u}_2,\ \check{v}_3\check{u}_3).$$

Define the norm of a triple to be the L^1–norm: $|\check{v}| = \check{v}_1 + \check{v}_2 + \check{v}_3$. Write $\check{v}_Y(x)$ for the triple $\check{v}(X)$, with X the subset of $\mathrm{Hom}(\pi_1(Y), V_4)$ that lives on an SO(3) bundle E with $x = w_2(E)$. Thus X/S_3 is the subset of $\{a, b, c\} \in V_Y$ such that $ab + bc + ac = x$. With our new notation, we have

$$v_Y(x) = |\check{v}_Y(x)|.$$

Now, given $x_i \in H^2(Y_i;\mathbb{Z}_2)$ it is easy to verify the connect sum relation

$$v_{Y_1 \# Y_2}(x_1 + x_2) = |\check{v}_{Y_1}(x_1) \times \check{v}_{Y_2}(x_2)|.$$

Note also that if x is not a cup-square, then $\check{v}_Y(x)$ has the form

$$\check{v}_Y(x) = (\check{v}_1, 0, 0).$$

In general, the third entry $\check{v}_3$ is equal to 1 if and only if $x = 0$, and is otherwise 0. Also, the second entry $\check{v}_2$ is the number of nontrivial cup-square-roots of x:

$$\check{v}_2 = |\{a \in H^1(Y;\mathbb{Z}_2) : a \neq 0, a^2 = x\}|, \quad \text{where } \check{v}_Y(x) = (\check{v}_1, \check{v}_2, \check{v}_3).$$

In particular, the sum $\check{v}_2 + \check{v}_3$ is either zero or the cardinality of the kernel of the Bockstein map $H^1(Y;\mathbb{Z}_2) \to H^2(Y;\mathbb{Z}_2)$, which is by definition $2^{k(Y)}$. Putting these observations together, and using our freedom to choose x that is not a square (below choose $x_2 = 0$), we compute the following:

Proposition 8.2 *Suppose $x_i \in H^2(Y_i;\mathbb{Z}_2)$ and that x_1 is not a cup-square. Then*

$$v_{Y_1\#Y_2}(x_1+x_2) \equiv \begin{cases} v_{Y_1}(x_1) \pmod 2 & \text{if } k(Y_2)=0, \\ 0 \pmod 2 & \text{otherwise.} \end{cases}$$

In particular, we recover the fact (mod 2) that $v_Y(x)$ is stable under connect summing with $\mathbb{RP}^3$. More generally, these statements clearly hold when the decompositions are only algebraic, instead of geometric: for example, if there is a decomposition $H^1(Y;\mathbb{Z}_2) = A \oplus B$ where $A \cup B = 0$ and B has an element of order 4 or ∞, then $v_Y(x) \equiv 0 \pmod 2$ for any x not a cup-square.

Double branched covers The above Seifert-fibered examples for genus $g=0$ are double branched covers of Montesinos links, but in all of those cases $v_Y(x)$ vanishes (mod 2) for nonsquares x. Here we compute a nonvanishing example in which Y is a double branched cover $\Sigma(L)$ of a link L in S^3. First, we describe the $\mathbb{Z}_2$ cohomology rings of such manifolds. Let L be a link with components $L_1,\dots,L_n$, and let S_i be a Seifert surface for L_i. Then S_i lifts to a closed surface F_i in the branched cover $\Sigma(L)$. Write $a_i \in H^1(\Sigma(L);\mathbb{Z}_2)$ for the Poincaré dual of $[F_i]$.

Proposition 8.3 *Let L be an n–component link. The vector space $H^1(\Sigma(L);\mathbb{Z}_2)$ has dimension $n-1$, and it is generated by the n classes a_i subject to the one relation*

$$a_1 + \cdots + a_n = 0. \tag{13}$$

The triple cup product form on $H^1(\Sigma(L);\mathbb{Z}_2)$ is determined by the values

$$(a_i \cup a_j \cup a_k)[\Sigma(L)] \equiv \begin{cases} \sum_{\ell\neq i} \operatorname{lk}(L_i,L_\ell) \pmod 2 & \text{for } i=j=k, \\ \operatorname{lk}(L_i,L_k) \pmod 2 & \text{for } i=j\neq k, \\ 0 \pmod 2 & \text{for } i,j,k \text{ distinct.} \end{cases}$$

This proposition is proved for two-component links in [12, Proposition 9.2], and the proof easily generalizes. We sketch the argument. To begin, we mention that $H_1(\Sigma(L);\mathbb{Z}_2)$ is in bijection with the subsets of $\{1,\dots,n\}$ of *even* cardinality:

$$H_1(\Sigma(L);\mathbb{Z}_2) \xleftrightarrow{1:1} \big\{S \subset \{1,\dots,n\} : |S| \equiv 0 \pmod 2\big\}. \tag{14}$$

The bijection goes as follows. Given such a subset, pair off elements. For the pair $\{i,j\}$, draw an arc in S^3 between components L_i and L_j, otherwise missing L. Lift the arcs to a union of loops in $\Sigma(L)$ to obtain a class in $H_1(\Sigma(L);\mathbb{Z}_2)$. Now, assume the F_i are transverse to one another. Then it is not hard to see, when $i \neq j$, that $F_i \cap F_j$ is mod 2 homologous to

$$\operatorname{lk}(L_i,L_j)\cdot\{i,j\},$$

where we view $\{i,j\}$ as an element of $H_1(\Sigma(L);\mathbb{Z}_2)$ via the above bijection. Upon

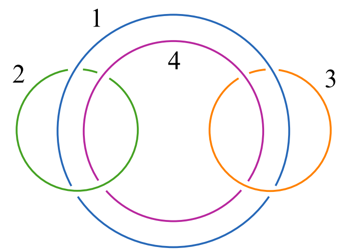


Figure 2: The link $L=$ L8n8 with its four components labeled by $\{1,2,3,4\}$. This link has determinant zero and thus its branched double cover supports nontrivial admissible bundles.

taking Poincaré duals, this yields the proposition. We note that addition on the subsets appearing on the right side of (14) is the symmetric difference of sets.

Let $Y=\Sigma(L)$, and let f be the function from V_Y to $H_1(Y;\mathbb{Z}_2)$ that sends a flat Klein-four connection class to the Poincaré dual of its second Stiefel–Whitney class:

$$f\{a,b,c\}=\mathrm{PD}(ab+bc+ac).$$

Let L be the four-component link L8n8 depicted in Figure 2, and let a_i be the classes described in Proposition 8.3 for L, so that a_i is dual to the lifted Seifert surface of L_i. In particular, a_1, a_2, a_3 form a basis for $H^1(Y;\mathbb{Z}_2)$. For illustration, using Proposition 8.3 we compute

$$\mathrm{PD}(a_1^2)=\mathrm{PD}\left(a_1(a_2+a_3+a_4)\right)=\sum_{i=2}^{4}\mathrm{lk}(L_1,L_i)\cdot\{1,i\}=\{1,2\}+\{1,3\}=\{2,3\}.$$

The bijection (14) is implicit in our notation, aligning subsets of $\{1,2,3,4\}$ of even size with elements of $H_1(Y;\mathbb{Z}_2)$. We then compute f on all fifteen of the Klein-four connection classes in V_Y:

$$\begin{aligned}
f\{0,0,0\}&=0, & f\{a_1,a_2,a_1{+}a_2\}&=\underline{\{3,4\}},\\
f\{a_1,a_1,0\}&=\{2,3\}, & f\{a_1,a_3,a_1{+}a_3\}&=\underline{\{2,4\}},\\
f\{a_2,a_2,0\}&=\{1,4\}, & f\{a_2,a_3,a_2{+}a_3\}&=0,\\
f\{a_3,a_3,0\}&=\{1,4\}, & f\{a_1,a_2{+}a_3,a_1{+}a_2{+}a_3\}&=0,\\
f\{a_1{+}a_2,a_1{+}a_2,0\}&=\{1,2,3,4\}, & f\{a_2,a_1{+}a_3,a_1{+}a_2{+}a_3\}&=\underline{\{1,3\}},\\
f\{a_1{+}a_3,a_1{+}a_3,0\}&=\{1,2,3,4\}, & f\{a_3,a_1{+}a_2,a_1{+}a_2{+}a_3\}&=\underline{\{1,2\}},\\
f\{a_2{+}a_3,a_2{+}a_3,0\}&=0, & f\{a_1{+}a_2,a_1{+}a_3,a_2{+}a_3\}&=0,\\
f\{a_1{+}a_2{+}a_3,a_1{+}a_2{+}a_3,0\}&=\{2,3\}.
\end{aligned}$$

We find that the cup-squares form a 2–dimensional subspace of $H^2(Y;\mathbb{Z}_2)$, appearing as the outputs of the left-hand column. Thus $k(Y)=1$. We have four nonsquares, appearing as the nonzero (underlined) entries in the right-hand column. Each has one Klein-four class, and so $v_Y(x)\equiv 1 \pmod 2$ when x is not a cup-square. The link L has determinant zero, ie $b_1(Y)>0$, so Y has a nontrivial admissible $U(2)$ bundle E. By Theorem 1.1 we conclude

$$\lambda(Y,E)\equiv 1 \pmod 2.$$

Proposition 8.3 similarly computes the parity of $2^{4-n}\lambda(Y,E)$, when $\det(L)=0$, from only knowing the mod 2 linking matrix of L.

References

[1] **K Aaslepp**, **M Drawe**, **C Hayat-Legrand**, **C A Sczesny**, **H Zieschang**, *On the cohomology of Seifert and graph manifolds*, Topology Appl. 127 (2003) 3–32 MR

[2] **P J Braam**, **S K Donaldson**, *Floer's work on instanton homology, knots and surgery*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 195–256 MR

[3] **S K Donaldson**, *Floer homology groups in Yang–Mills theory*, Cambridge Tracts in Mathematics 147, Cambridge Univ. Press (2002) MR

[4] **A Floer**, *Instanton homology and Dehn surgery*, from "The Floer memorial volume" (H Hofer, C H Taubes, A Weinstein, E Zehnder, editors), Progr. Math. 133, Birkhäuser, Basel (1995) 77–97 MR

[5] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR

[6] **C M Herald**, *Legendrian cobordism and Chern–Simons theory on 3–manifolds with boundary*, Comm. Anal. Geom. 2 (1994) 337–413 MR

[7] **C M Herald**, *Transversality for equivariant exact 1–forms and gauge theory on 3–manifolds*, Adv. Math. 200 (2006) 245–302 MR

[8] **P B Kronheimer**, **T S Mrowka**, *Tait colorings, and an instanton homology for webs and foams*, preprint (2015) arXiv

[9] **C Lescop**, *Global surgery formula for the Casson–Walker invariant*, Annals of Mathematics Studies 140, Princeton Univ. Press (1996) MR

[10] **M M Postnikov**, *The structure of the ring of intersections of three-dimensional manifolds*, Doklady Akad. Nauk. SSSR 61 (1948) 795–797 MR In Russian

[11] **P Poudel**, *Lescop's invariant and gauge theory*, J. Knot Theory Ramifications 24 (2015) art. id. 1550050, 16 pages MR

[12] **P Poudel**, *Link homology and equivariant gauge theory*, PhD thesis, University of Miami (2015) MR Available at `https://search.proquest.com/docview/1687760309`

[13] **D Ruberman**, **N Saveliev**, *Rohlin's invariant and gauge theory, I: Homology 3–tori*, Comment. Math. Helv. 79 (2004) 618–646 MR

[14] **D Sullivan**, *On the intersection ring of compact three manifolds*, Topology 14 (1975) 275–277 MR

[15] **C H Taubes**, *Casson's invariant and gauge theory*, J. Differential Geom. 31 (1990) 547–599 MR

[16] **V G Turaev**, *Cohomology rings, linking coefficient forms and invariants of spin structures in three-dimensional manifolds*, Mat. Sb. 120(162) (1983) 68–83 MR In Russian; translated in Math. USSR-Sb. 48 (1984) 65–79

Department of Mathematics, Brandeis University
Waltham, MA, United States

Department of Mathematics, University of California
Los Angeles, CA, United States

`cscaduto@scgp.stonybrook.edu`, `mstoffregen@math.ucla.edu`

Received: 28 August 2016 Revised: 2 June 2017

msp

Algebraic & Geometric Topology 17 (2017) 2863–2891

Infinite order corks via handle diagrams

ROBERT E GOMPF

The author recently proved the existence of an infinite order cork: a compact, contractible submanifold C of a 4–manifold and an infinite order diffeomorphism f of ∂C such that cutting out C and regluing it by distinct powers of f yields pairwise nondiffeomorphic manifolds. The present paper exhibits the first handle diagrams of this phenomenon, by translating the earlier proof into this language (for each of the infinitely many corks arising in the first paper). The cork twists in these papers are twists on incompressible tori. We give conditions guaranteeing that such twists do not change the diffeomorphism type of a 4–manifold, partially answering a question from the original paper. We also show that the "δ–moves" recently introduced by Akbulut are essentially equivalent to torus twists.

57N13, 57R55

1 Introduction

The failure of high-dimensional topology to apply to smooth 4–manifolds led to the notion of a *cork twist*. As originally formulated, this consists of changing the diffeomorphism type of a closed 4–manifold X by removing a compact, contractible, smooth submanifold C from X and regluing it by an involution f of ∂C. The first example of a cork twist was published by Akbulut [1] in 1991. A few years later, Curtis, Freedman, Hsiang and Stong [8] and Matveyev [16] showed that any two homeomorphic, simply connected (smooth) 4–manifolds are related by a cork twist. (See Gompf and Stipsicz [14] and Gompf [13] for more history.) The question was immediately raised of whether higher order corks may exist—and in particular, whether there was such a pair $C \subset X$ and an infinite order diffeomorphism f of ∂C such that the $\mathbb{Z}$–indexed family of homeomorphic manifolds obtained by regluing using all powers of f were pairwise nondiffeomorphic. In weaker form, can there even be a contractible 4–manifold C with a boundary diffeomorphism for which no nonzero power extends to a self-diffeomorphism of C? No progress was made on these questions until recently. Corks of all finite orders were constructed in 2016 by Tange [19] and Auckly, Kim, Melvin and Ruberman [7]. A withdrawn 2014 posting of Akbulut [3] attempted to construct infinite order corks of the weaker sort, using handle calculus. In 2016 [13], the author of the present paper constructed an infinite family

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2863

of examples $C(r,s;m) \subset X$ (for $r,s > 0 > m$), each one affirmatively answering the stronger question, using an entirely different plan of attack and no handle diagrams. This raised the question of how the paper translates into the language of these diagrams. Section 3 of the present paper gives the result, in a form independent of, but clearly derived from, [13]. This section shows boundary diffeomorphisms of contractible 4–manifolds, and how they can provide infinitely many diffeomorphism types of ambient 4–manifolds, as presented in the currently preferred language of the subject. We also find conditions under which such diffeomorphisms necessarily do extend over a given 4–manifold, partially answering a question from [13]. A recent paper of Akbulut [4] seeks to generalize the twists of Section 3; our final section shows that his viewpoint is equivalent to ours.

After a quick exposition of the relevant 3–manifold diffeomorphisms in Section 2, this paper proceeds with three independent sections, beginning with our translation of the proofs in [13] into handle calculus (Section 3). We sketch the correspondence between the proofs as we go along. We exhibit a cork C by a diagram (Figure 4) constructed from the existence proof of [13], and show it is diffeomorphic to $C(r,s;m)$ as exhibited by Figure 3. Then we embed C in a family of larger manifolds $Z_k(r,s;m)$, related (as k varies in $\mathbb{Z}$, with the other variables fixed) by powers of a twist parallel to an incompressible torus in ∂C (Figure 11). Finally, we show that these manifolds embed in a family of closed manifolds X_k related by the same cork twists, and distinguish these using the same method as [13]: we show that they are obtained by the Fintushel–Stern knot construction on elliptic surfaces.

Section 4 gives our criterion guaranteeing that cutting and regluing does not change the diffeomorphism type of a 4–manifold. We partially answer a question from [13]: The nontrivial cork twists of that paper were diffeomorphisms of $\partial C(r,s;m)$ twisting along an incompressible torus parallel to its longitude. It was asked whether twisting parallel to the meridian was also nontrivial. We show that the answer is no for a family of manifolds including each $C(r,s;-1)$. Thus, while the potential torus twists are indexed by $H_1(T^2) \cong \mathbb{Z} \oplus \mathbb{Z}$, only one $\mathbb{Z}$–summand is useful for producing exotic 4–manifolds. The question remains open for $m < -1$, but we also see that for the specific embedding $C(r,s;m) \subset X$ used in [13], only one $\mathbb{Z}$–summand (the longitude) affects the resulting diffeomorphism type. Thus, a nontriviality proof for meridian twisting would at least require a different setup. Coupling our meridian twist criterion with a recent observation of Ray and Ruberman, we see that there are contractible manifolds C for which every boundary diffeomorphism extends over C even though ∂C contains an incompressible torus.

Our final section concerns the recent paper [4]. Starting from the preliminary version of Section 3 of the present paper, Akbulut sketched the proof of the simplest case

$C(1,1;-1)$. The apparent motivation was to introduce the notion of *δ–moves* as an alternative to torus twists. Such moves depend on a choice of auxiliary band (and other data not specified in that paper), so appear to provide additional generality. We show, however, that δ–moves are essentially equivalent to torus twists under broad hypotheses, for example, on irreducible homology spheres (Corollary 5.5). Since incompressible tori are tightly constrained in 3–manifolds, δ–moves are harder to find than they initially seem to be, and might be most easily located using the well-developed theory of incompressible tori (as Akbulut implicitly did in obtaining the proof he posted from our Section 3). To obtain our equivalence, we must first address foundational issues such as well-definedness of δ–moves. We also observe technical difficulties that need to be addressed whenever δ–moves (or torus twists) are used for diagrammatically cutting and pasting 4–manifolds.

Remarks (a) It is known that a cork twist cannot change the homeomorphism type of a 4–manifold, since every boundary diffeomorphism f of a contractible 4–manifold C extends over it homeomorphically. For a short proof, use f to glue two copies of C along their boundary, obtaining a homotopy 4–sphere that automatically bounds a contractible 5–manifold W (via a smooth h-cobordism to S^4, or by working topologically and observing that ∂W is homeomorphic to S^4 by Freedman [11]). We can view W as a topological h-cobordism of C with a fixed product structure over ∂C realizing f. Freedman's h-cobordism theorem [11] extends the product structure, and projecting to C extends f.

(b) Some authors require corks to be Stein domains by definition. This seems to be an entirely separate issue from that of changing diffeomorphism types by twisting, although Akbulut and Matveyev [6] showed that corks, in the original sense where f is an involution, can always be modified to admit Stein structures. The author has made no attempt to address the Stein condition in this paper or its predecessor. It remains an interesting question whether any of these corks are (or can be modified to be) Stein domains.

We work in the smooth category throughout the paper. For simplicity, we assume (unless otherwise indicated) that all 3–manifolds are orientable and closed, and all 4–manifolds are orientable and compact (allowing boundary).

2 Torus twists

We begin with a quick exposition of the 3–manifold diffeomorphisms that will be central to this paper. Let $T \subset M$ be a torus embedded in a 3–manifold. Identify a tubular neighborhood of T with $S^1 \times S^1 \times I$, and let α and β denote $S^1 \times \{\zeta\} \times \{0\}$ and $\{\zeta\} \times S^1 \times \{0\}$, respectively, for some $\zeta \in S^1$.

Definition 2.1 The *torus twist* on T parallel to α is the diffeomorphism $f\colon M \to M$ obtained from $f(\theta, \phi, t) = (\theta + 2\pi t, \phi, t)$ by extending as the identity on the rest of M and smoothing.

Informally, we cross a Dehn twist on the annulus $\alpha \times I$ with the identity on β. This is a well-known, classical diffeomorphism of M, sometimes called a *Dehn twist* on T. Since T can be identified with $S^1 \times S^1$ so that α represents any preassigned primitive homology class of T, every element of $H_1(T)$ determines some power of a torus twist. More formally, T is contained, by Lie group multiplication, in the group $\mathrm{Diff}_+(T)$ of orientation-preserving self-diffeomorphisms of T, inducing an isomorphism $\pi_1(T) \to \pi_1(\mathrm{Diff}_+(T))$. This descends by torus twisting to a homomorphism $\mathbb{Z} \oplus \mathbb{Z} \cong \pi_1(T) \to \pi_0(\mathrm{Diff}_+(M))$. For example, when M is a torus bundle, all self-diffeomorphisms fixing one fiber pointwise arise by twisting on another fiber. For most irreducible 3–manifolds M, torus twists on all incompressible tori together generate the group $\pi_0(\mathrm{Diff}_+(M))$ up to a finite extension. (See the last corollary of Waldhausen [22].) When T is compressible and M is irreducible, T either lies in a ball (bounded by the compressed torus) or bounds a solid torus over which the action of T extends, so all twists on T are isotopic to the identity. Note that irreducibility is necessary: For a connected sum, we expect nontrivial slide diffeomorphisms constructed by dragging the site of the sum around a loop γ in one summand. Such a diffeomorphism can be described as a twist about the compressible torus bounding a tubular neighborhood of γ.

We can similarly define twists on Klein bottles. We will see in Section 5 that these are less useful than torus twists, but we introduce them for completing the discussion there of δ–moves. If $K \subset M$ is an embedded Klein bottle, we identify K as a bundle over a circle β with fiber α. The previous description can still be applied over intervals in β, since the monodromy around β reverses orientations of both α and I (by orientability of M), hence, commutes with the Dehn twist.

Both kinds of twists have a convenient surgery description. First, draw a framed link diagram of M so that α is an unknot in the ambient S^3 with the torus or Klein bottle T inducing the 0–framing. Such a diagram can be obtained from an arbitrary diagram, in which α appears as a framed knot, by blowing up to change some crossings and adjust its framing. (One can also simultaneously control β so that the two curves bound disks in the ambient S^3 with disjoint interiors.) To realize the twist, blow up a ± 1–framed curve γ at α, slide it around T as in Figure 1 until it returns to its original position,

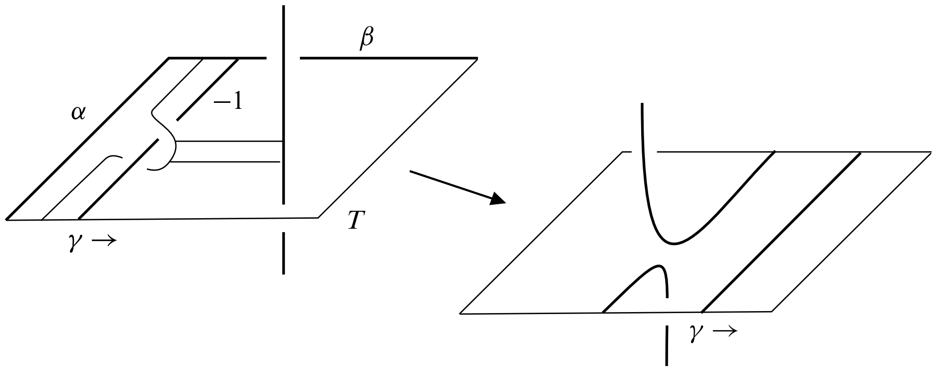


Figure 1: Torus twisting

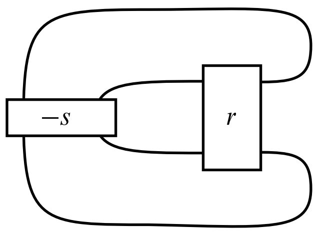


Figure 2: The double twist knot $\kappa(r,-s)$

and blow it back down. Any curve intersecting T must slide over γ as it passes, creating the required twist. To verify that this gives a well-defined diffeomorphism, it is not necessary to see that T is embedded, only that γ returns to its original position after sliding around M (to allow conjugation by the diffeomorphism from M to its blowup.)

3 Diagrams of corks

The first examples of infinite order corks were constructed in [13]. Let $E(r,s)$ denote the complement of the double twist knot $\kappa(r,-s)$ shown in Figure 2 (where the boxes count full twists). The manifold $C(r,s;m)$ obtained from $I \times E(r,s)$ by adding a 2–handle along an m–framed meridian in $\{1\} \times E(r,s)$ is contractible. Its boundary has an obvious incompressible torus T, namely $\{0\} \times \partial E(r,s)$. (In fact there is a pair of incompressible tori, exhibited by moving the 2–handle to the middle level $\frac{1}{2}$ and taking the tori at levels 0, 1. These are interchanged by an orientation-reversing symmetry of the construction, and they are only parallel when $|m| = 1$.) Let $f\colon \partial C(r,s;m) \to \partial C(r,s;m)$ be the torus twist on T parallel to the longitude λ of T (the curve bounding a Seifert surface in $\{0\} \times E(r,s)$). In [13] it was shown that for fixed $r,s,n > 0 > m$,

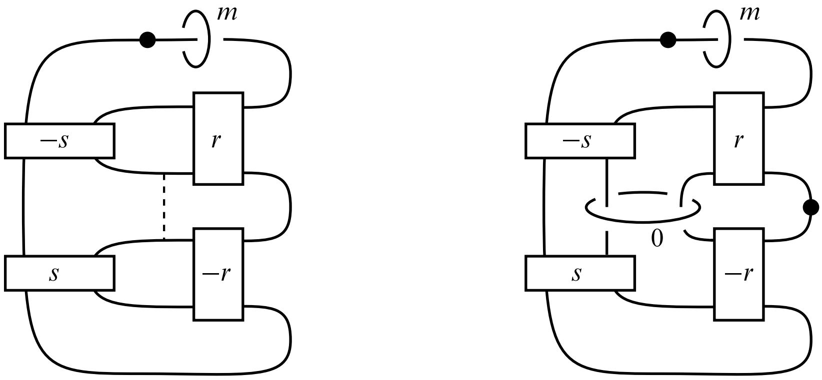


Figure 3: The cork $C(r,s;m)$

there is a canonical embedding $C(r,s;m) \hookrightarrow X'_0 = E(n) \# (r+s-m-3)\overline{\mathbb{CP}^2}$ into a blown up elliptic surface. It was also shown that the manifold X'_k obtained from X'_0 by cutting out $C(r,s;m)$ and regluing it using f^k is the corresponding blowup of the manifold X_k obtained from $E(n)$ by the Fintushel–Stern construction on the knot $\kappa(k,-1)$. (The blowups can usually be avoided.) Since these latter manifolds for $k \in \mathbb{Z}$ are known to be pairwise nondiffeomorphic [9], [10], each choice of $r,s > 0 > m$ yields an infinite order cork. In particular, the self-diffeomorphisms f^k are not related to each other by any self-diffeomorphism of the cork, or to the identity unless $k=0$. (This last sentence applies to a larger range of r,s,m, due to the obvious symmetries $C(-r,-s;m) = C(r,s;m) = C(s,r;m)$ and the reflection reversing the sign of m. However, it is crucial that $rs > 0$, ie the twists in Figure 2 have opposite handedness. Otherwise, $C(1,-1;-1)$ is a counterexample; see Corollary 4.4 and its preceding discussion.) The existence proof of an infinite order cork producing the family $\{X_k\}$ did not use any handle diagrams, and recognizing the corks required only some simple 3–dimensional surgery. Since it seems useful to understand the proofs using diagrams, we now provide their translations. The resulting proofs are independent of [13] and almost entirely handle-theoretic, but seem unlikely to have been conceived without benefit of the abstract version.

To draw diagrams of the manifolds in [13], we need diagrams of products of various 3–manifolds with I and S^1. We use a method that was pioneered by Akbulut and Kirby; see eg [5; 2]. A detailed exposition is given in [14, Section 6.2], particularly the solved Exercise 6.2.5(b) (and Example 4.6.8 for products with S^1). We illustrate with the diagrams of $C(r,s;m)$ in Figure 3. Ignoring the m–framed meridian in each diagram, we obtain $I \times E(r,s)$. To understand the resulting diagram on the left, consider the horizontal plane of reflection. Adding its point at infinity and then thickening, we obtain an embedding of $I \times S^2$ in S^3 that represents the lateral boundary of $I \times B^3$.

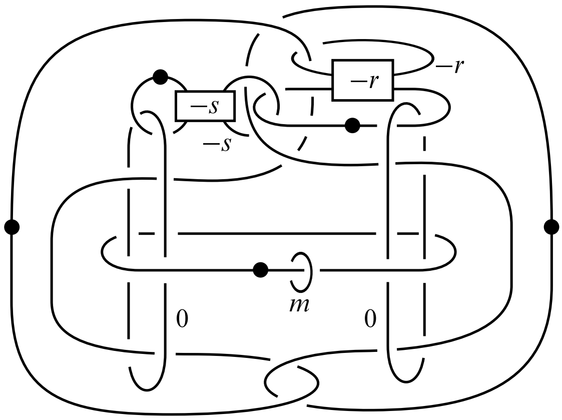


Figure 4: The cork $C \approx C(r, s; m)$

The upper and lower complementary 3–balls in the diagram represent $\{1\} \times B^3$ and $\{0\} \times B^3$, respectively. The dotted knot $\kappa(r, -s) \# -\kappa(r, -s)$ in the diagram is the boundary of the ribbon disk $I \times \kappa(r, -s) \subset I \times B^3$ (where we interpret the knot as a tangle in B^3). This disk, which can also be interpreted as the half-spin of the tangle in B^4, should be deleted from B^4 to obtain $I \times E(r, s)$ (as the dot indicates). The dashed arc represents the ribbon move transforming the dotted knot into a 2–component unlink, exhibiting the ribbon disk with a pair of local minima and a saddle point. If we use this decomposition to build a handle diagram of the complement, we obtain the diagram on the right, with the dotted unlink representing the pair of local minima and the 0–framed 2–handle arising when the saddle is submerged into the 4–ball. The m–framed meridian in each diagram is the 2–handle specified in the definition of $C(r, s; m)$. To see the torus T in the left diagram, note that the bisecting 2–sphere in S^3 intersects the knot in two points. Remove these intersections by ambient surgery, using a tube comprising the boundary of a tubular neighborhood of the lower knot $-\kappa(r, -s) = \kappa(-r, s)$. The resulting torus in $\partial C(r, s; m)$ is $T = \{0\} \times \partial E(r, s)$ (seen in the boundary orientation inherited from $C(r, s; m)$). This torus is also visible in the right picture, running twice over the 0–framed 2–handle (corresponding to the two intersections of T with the dashed arc in the left picture).

To exhibit infinite order corks, we need a very different description of $C(r, s; m)$ and its boundary diffeomorphism. The 4–manifold C shown in Figure 4 arises from the proof of the main theorem of [13]. We discuss how this figure arises and then show that C is diffeomorphic to $C(r, s; m)$. (This entire discussion could be excised to leave a complete but mysterious proof that C is an infinite order cork. Note that it is obviously contractible, being simply connected with Euler characteristic 1.) In [13], the cork C was constructed from $Y = I \times \Sigma \times S^1$, where Σ is a punctured torus, by adding three 2–handles and then drilling out the cores of two of them (connected by annuli to the far boundary of the product with I). If we modify Figure 4 by removing all four circles passing through twist boxes, as well as the m–framed meridian, what remains is this Y. We can see this by unwinding the two large dotted circles from each other, but it is more instructive to view the picture as a product with S^1: Starting from a trivial proper embedding $\Sigma \subset B^3$ with $\lambda = \partial\Sigma$ on the equator of ∂B^3, we obtain its tubular neighborhood $I \times \Sigma$ as the complement of a clasped pair of arcs, with the boundary of each in a single hemisphere. By the method used in the previous paragraph, this picture becomes the clasp of the two large dotted circles in the top center of the figure, and its mirror image at the bottom. Thus, we have $I \times \Sigma \times I$ exhibited as a 0–handle and two 1–handles. The algorithm for changing a product with I to a product with S^1 introduces a $(k+1)$–handle for each k–handle of the original diagram. In this case, we obtain a new 1–handle (the central dotted circle) and the two 0–framed 2–handles. We can think of the 1–handle as connecting the top and bottom boundaries $I \times \Sigma \times \{0,1\}$ to each other, and each 2–handle connects a meridian of a dotted circle (essentially the core of the 1–handle) with its mirror image on the other boundary component.

We complete the analysis of Figure 4 by restoring the remaining curves to get C. At each twist box, we have a rationally canceling handle pair that represents a 2–handle added along a generator of Σ, with its core drilled out (the dotted circle). The m–framed circle represents the undrilled handle attached to a product circle. Note that the diagram can be simplified by canceling the m–framed meridian, and when $r = s = 1$, there is further cancellation at the twist boxes. (When $m = -1$ also, Figure 1 of Akbulut's recent preprint [4] shows the result.) According to the construction in [13], the cork twist on C is a twist on the torus $T = \{0\} \times \partial\Sigma \times S^1$ parallel to $\lambda = \{0\} \times \partial\Sigma \times \{\theta\}$. Interpreting λ as the equator of ∂B^3 as before, we draw it as in Figure 5. (It encircles the clasp in $I \times \Sigma \times I$, but is drawn at one side to make room for the additional handles of Y.) Since the dual curve $\mu \subset T$ is a product circle in Y, the pair appears as in that same figure. Thickening these to annuli using the 0–framing, we obtain a punctured torus T_0 whose union with an embedded disk is T. We will verify directly from the diagrams that $\delta = \partial T_0$ bounds an embedded disk in ∂C with interior disjoint from T_0, and that the λ–twist on the resulting torus has the required properties, but first we identify C:

Proposition 3.1 *The 4–manifold C in Figure 4 is diffeomorphic to $C(r,s;m)$ in Figure 3.*

Note that this gives an independent check that C is made from a double twist knot whose twists have opposite handedness. In [13], this handedness was determined by a delicate inspection of 3–manifold orientations.

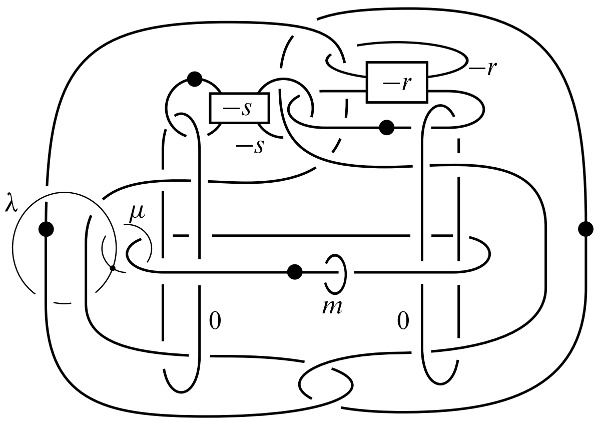


Figure 5: The punctured torus T_0 in ∂C is obtained by 0–framed thickening of $\lambda \cup \mu$. Its boundary is $\delta = \partial T_0$.

Proof Starting from Figure 4, unwind the clasps of the two large dotted circles by moving the leftmost strand in the diagram through the large right circle and back to its place. After shortening the two 0–framed curves by an isotopy, we obtain Figure 6. Next, we perform two double 1–handle slides as indicated. That is, each arrow represents two strands of a dotted circle being slid across another one. Throughout this proof, the link consisting of all dotted circles will be an unlink, so we are sliding 1–handles in the classical sense (ie, no nontrivial dotted ribbon links appear). We encounter a notational technicality: We slide using 0–framed parallel copies of the small dotted circles. These pass through negative twist boxes, so to restore the 0–framings we must add compensating positive twist boxes. The result, after the two obvious handle pair cancellations, is Figure 7. Next, simplify the two large dotted curves by pulling the clasps through all twist boxes as indicated by the arrows, dragging along the dotted circle with the m–framed meridian. Raise the lowermost strand of the latter dotted circle so that it is positioned between the $-s$–framed 2–handle and the $\pm r$ twist boxes, then eliminate its self-crossing by flipping over the clasp running through the $\pm r$ twist boxes. The rightmost dotted circle can then be shrunk into the middle of the figure, which should then be recognizable as Figure 8. We next wish to cancel the $-s$–framed 2–handle. Since we cannot slide a dotted circle over a 2–handle, we first introduce a canceling 1–2 pair as in Figure 9, then double slide the new 2–handle as indicated and cancel the $-s$–framed handle, obtaining Figure 10. (Note that after the double slide, the new 2–handle initially runs twice through the $-s$–twist box, consistent with sliding over a $-s$–framed curve. However, we can immediately pull it down through the twist box to its position in Figure 10.) Finally, slide the $-r$–twist box across the circle with the rightmost dot. This is a standard 3–manifold move (see Figure 18 in

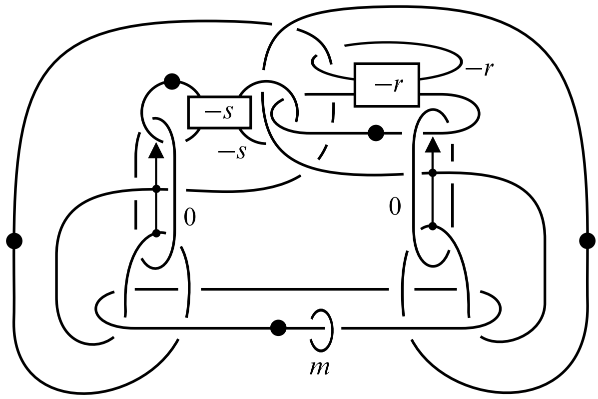


Figure 6: Double 1–handle slides on a diagram of C obtained from Figure 4 by isotopy

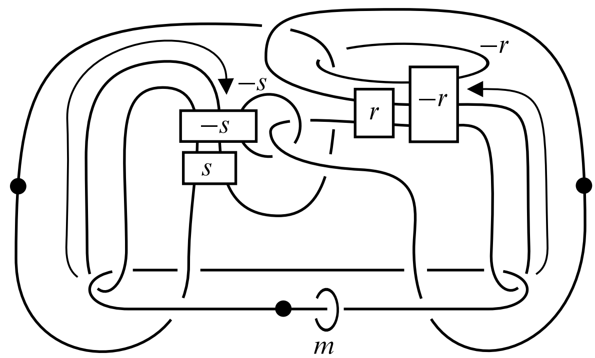


Figure 7: Further simplifying C

Section 5), obtained by repeatedly blowing up a $+1$–framed curve encircling the twist box, sliding it across the dotted circle, and blowing it back down. One way to interpret this move 4–dimensionally is to think of the dotted circles as representing $\#3\, S^1 \times S^2$, containing a framed link. Perform the move on this 3–manifold, then uniquely fill in the 4–manifold $\natural 3\, S^1 \times D^3$. This move changes the framing on one 2–handle from $-r$ to 0, and that handle immediately cancels a dotted circle. The result is isotopic to the right-hand diagram in Figure 3. $\square$

Proposition 3.2 *The circle $\delta = \partial T_0 \subset \partial C$ in Figures 5 and 11 bounds a disk D in ∂C with interior disjoint from T_0. The resulting torus twist parallel to λ changes k by 1 in Figure 11 while otherwise preserving all curves in the figure (and their orientations).*

There are various approaches to the proof. A simple way to exhibit D disjoint from the fine curves is by handle sliding δ to get an unknot, as in Figure 12. (Although this

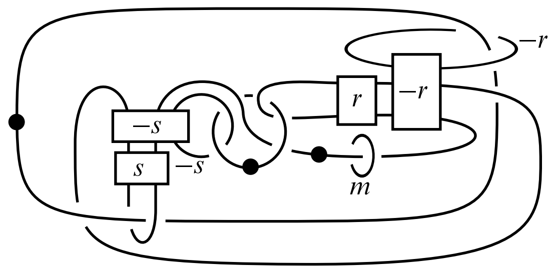


Figure 8: An isotopic simplification of Figure 7

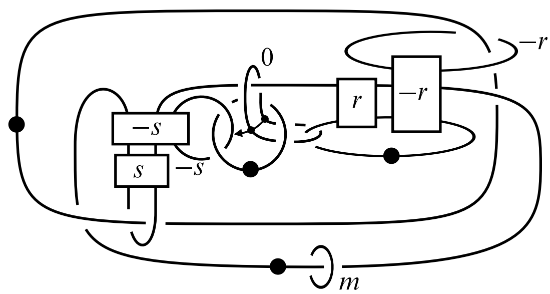


Figure 9: A new 1–2 pair and a double slide

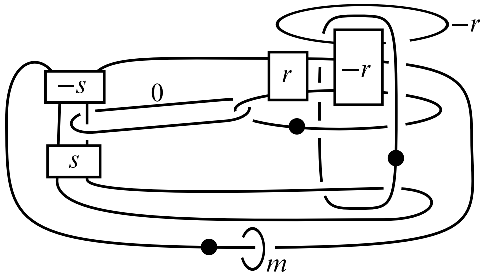


Figure 10: Transferring $-r$ twists yields $C(r,s;m)$

is unnecessary for proving our main theorem, an interested reader can verify that δ is isotopic to δ_1. Two double handle slides change this to δ_2, and two more yield an unknot. The boundary orientation of δ induced by the counterclockwise orientation of T_0 in Figure 11 is shown as an aid.) We can make D disjoint from $\operatorname{int} T_0$ by 3–manifold theory (see proof of Theorem 5.4), but this could create intersections with the fine curves, resulting in their unexpected movement during the torus twist (see last paragraph of Section 5.) This issue can presumably be dealt with, but we instead prove the theorem definitively with a direct approach.

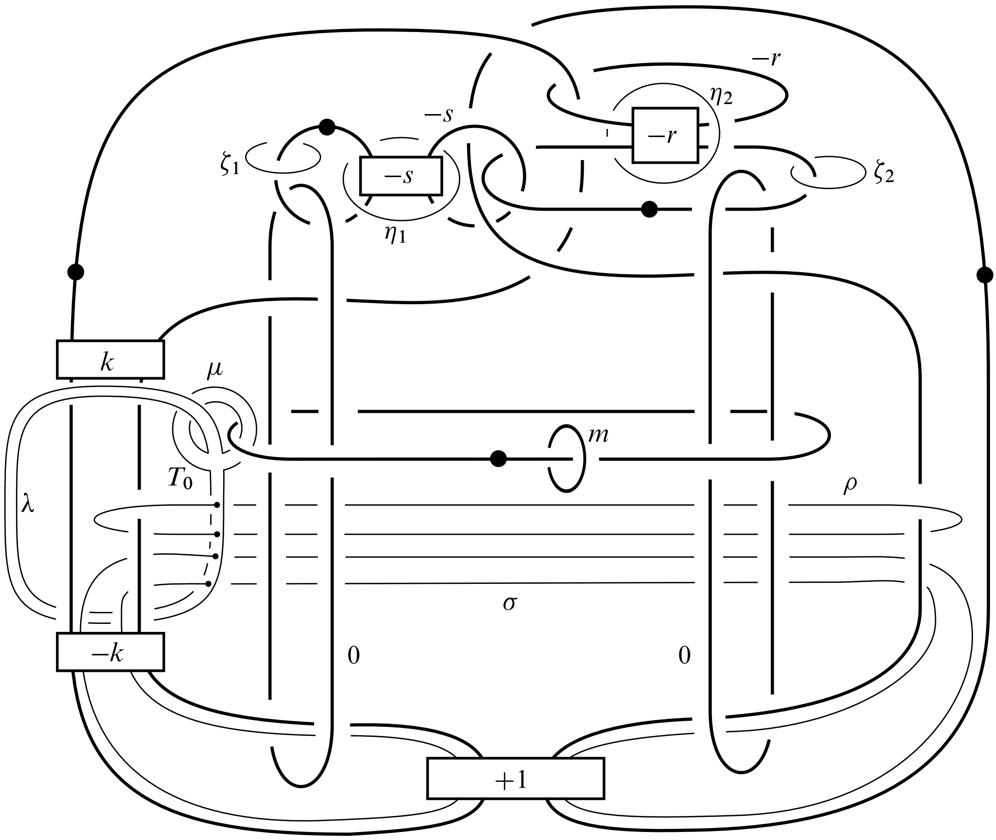


Figure 11: The 4–manifold $Z_k(r,s;m)$ is obtained from $C(r,s;m)$ by adding 0–framed 2–handles along ζ_1, ζ_2 and ρ, then a 3–handle (and ignoring the other fine curves). Note that some curves intersect the punctured torus T_0, which is drawn explicitly as a thickening of the wedge $\lambda \cup \mu$. Its boundary δ is unlabeled.

Proof Set $k=0$, and drag T_0 and all auxiliary curves in Figure 11 simultaneously through the computation of the previous proof. This is routine but tedious; details are left to the intrepid reader. (One can treat T_0 as a framed wedge of circles, as long as its intersections with the auxiliary curves are handled carefully. These intersections will eventually be dragged through the $-s$–twist box, in the downward direction. Note that the curves η_i will remain closely encircling the negative twist boxes, so need not be carefully tracked; the curves ζ_i will be similarly rooted to the positive twist boxes as soon as these boxes appear.) The result is Figure 13, where the curves η_i and ζ_i encircling the twist boxes are suppressed, and T_0 is the obvious 0–framed thickening of $\lambda \cup \mu$. (We have drawn the thickening near where T_0 intersects the other curves.) We

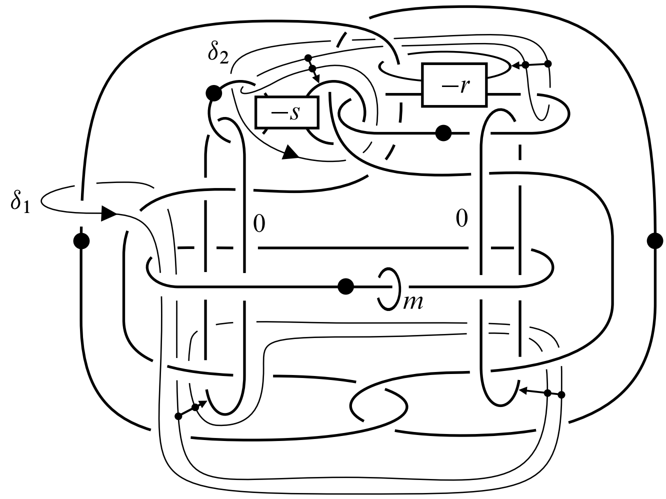


Figure 12: A proof that δ bounds a disk disjoint from the fine curves of Figure 11

see that T_0 is as originally described in $C(r,s;m)$, with λ the longitude of $\kappa(-r,s)$ and μ a meridian, and the disk D easily visualized. The remaining curves can be explicitly seen to avoid T except for the original intersections of T with ρ and σ. The torus twist f wraps these curves parallel to λ at the intersections, and fixes all curves elsewhere. When we transport this description back to Figure 11 with $k=0$, we can undo the wrapping caused by f^k by an isotopy that restores the twist boxes to their original values $\pm k$. □

Figure 11 was drawn so that the case $r=s=-m=1$ exhibits Akbulut's "δ–curve" [4, Figure 4] as the boundary of T_0. We will show in Section 5 that under broad hypotheses, every δ–curve arises from a torus in this manner. The author's first diagrammatic proof used a different approach: Blow up a (-1)–framed unknot parallel to λ in the lower half of Figure 11 (encircling the $-k$–twist box), slide it around T to get a curve encircling the $+k$–twist box using a diagram similar to Figure 12, then blow it back down. While this is the same torus twist (cf also Figure 1 and its discussion in the last paragraph of Section 2), it is less clear from this method that the implicitly described torus is embedded. However, the diffeomorphism is still well-defined by this procedure and has the required properties. This easier method is strong enough for all of our subsequent discussion, since we do not need to exhibit the diffeomorphism as a twist on an embedded torus.

Now consider the 4–manifold obtained from C by attaching 0–framed 2–handles along the fine curves ζ_1, ζ_2 and ρ shown in Figure 11. We will show that its boundary contains a nonseparating 2–sphere. Add a 3–handle along this sphere and call the result

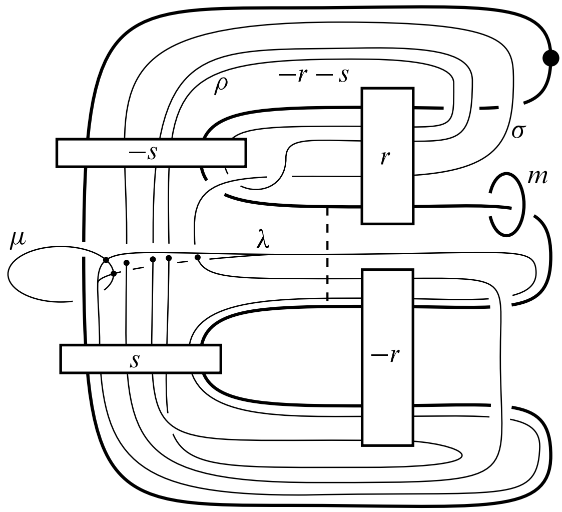


Figure 13: A complete picture of the torus $T \subset \partial C$, with the fine curves of Figure 11. The generating circles of T are μ and λ, the latter of which has been partially thickened in T to show the intersections of T with the fine curves. The rest of T is given by the horizontal plane of symmetry (and point at infinity), surgered by a tube following the lower half of the dotted circle so that T contains μ and λ.

$Z_k(r,s;m)$. This is independent of the choice of 2–sphere (by Trace [20], for example), and canonically contains a copy of C. Another picture of $Z_k(r,s;m)$ is given by Figure 14, where we have switched to dotted ribbon knot notation and canceled the new 2–handles ζ_1, ζ_2 and ρ against 1–handles. (Ignore the curve σ in Figure 14 but include the other fine curves, which come from 2–handles in Figure 11, and the 3–handle.) Canceling ρ has joined the two large dotted circles, forming the knot $K_k \# -K_k$, where K_k is the twist knot $\kappa(k,-1)$. This dotted ribbon knot represents $I \times E(k,1)$; cf Figure 3. (The comparison with $I \times E(r,s) \subset C(r,s;m)$ is superficial.) Without the fine curves, Figure 14 represents the manifold $W_k = S^1 \times E(k,1)$. (Each handle of $I \times E(k,1)$ generates an additional handle of index higher by one in W_k (cf Figure 4), with the canceled 2–handle ρ generating the 3–handle.) The fine curves in Figure 14, two of which are isotopic, represent the product S^1 (m–framed), meridians of K_k ($-r$–framed and $-s$–framed), and its longitude σ in $\{0\} \times E(k,1)$. Unlike previously, this recognition of W_k is crucial to our proof, so we provide a check:

Proposition 3.3 *The manifolds $Z_k(r,s;m)$ and W_k given by the diagrams are well-defined (ie the relevant 3–manifolds have a nonseparating sphere for the 3–handle). The 3–manifolds ∂W_k are all diffeomorphic, preserving the fine curves of Figure 14 and*

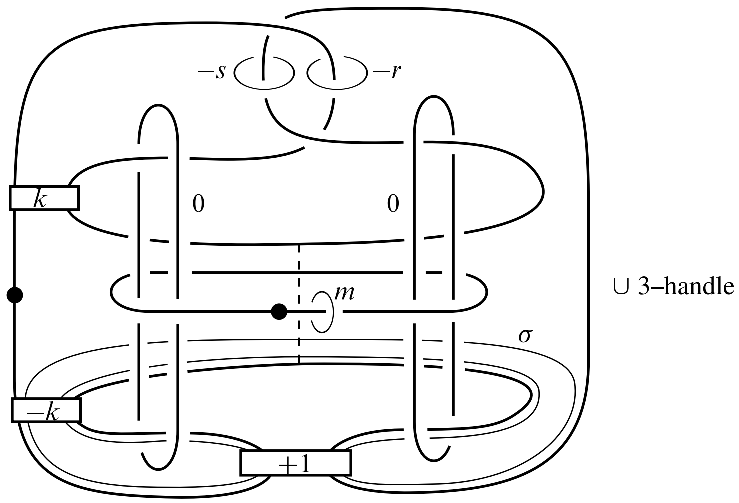


Figure 14: Another picture of $Z_k(r,s;m)$, with extra curve σ

their orientations. The 4–manifold W_0 is diffeomorphic to $T^2 \times D^2$, with σ bounding the essential disk, and the $-r$– and m–framed curves arising as factors of a product decomposition $T^2 = S^1 \times S^1$.

Proof Interpreting Figure 14 as a 3–manifold (ignoring the 3–handle), we can eliminate the clasps from the dotted ribbon knot by blowing up a $+1$–framed unknot as in Figure 15, sliding this unknot over the tall unknots as shown, and blowing back down. We can now cancel the $\pm k$ twist boxes by twisting one tall unknot k times about its long axis. This identifies each ∂W_k with ∂W_0 (and similarly for $Z_k(r,s;m)$) and reduces well-definedness to the $k=0$ case. Now consider Figure 14 to be a 4–manifold. When $k=0$ the $\pm k$–twist boxes can be erased, so we can pull the outer strand of σ through the $+1$–twist box and unwind the outer strands of the large dotted circle as in Figure 16. To get this figure, we also swing both tall curves to the inner rear of the large dotted circle. They are then parallel, so one can be slid over the other to become a 0–framed unknot unlinked from the rest of the diagram. This exhibits the nonseparating sphere in ∂W_k. Canceling this unknot with the 3–handle, we obtain Figure 16, which is the Borromean rings with fine meridians of each component. This has the required interpretation. □

We can now identify the manifolds $Z_k(1,1;-1)$ for all $k \in \mathbb{Z}$. First, $Z_0(r,s;m)$ is obtained from $W_0 = T^2 \times D^2$ by adding three 2–handles along embedded circles in copies of $T^2 \times \{p\}$, as shown in Figure 17. (The 1–handles and 0–framed 2–handle exhibit $T^2 \times D^2$ so that the trivial torus bundle structure on its boundary is easily

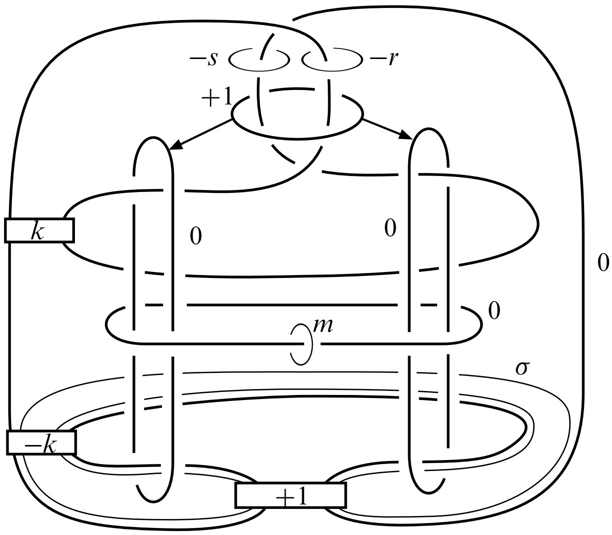


Figure 15: Simplifying ∂W_k

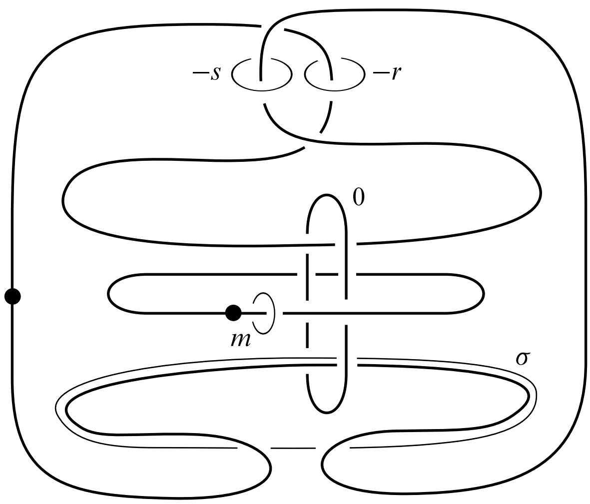


Figure 16: Identifying W_0 as $T^2 \times D^2$

visible.) When $r = s = 1$ and $m = -1$, this diagram is a well-known description of an elliptic fibration over a disk, a cusp neighborhood with an extra vanishing cycle; see eg [14, Section 8.2]. Thus, $Z_0(1,1;-1)$ naturally embeds in the elliptic surface $E(n)$ for any fixed $n > 0$, and is easily seen in link diagrams of the latter. The curve σ is a section of the induced torus bundle structure on $\partial Z_0(1,1;-1)$ (as is again visible in Figure 17; see [14]). For general k, $Z_k(r,s;m)$ is obtained from $Z_0(r,s;m)$ by

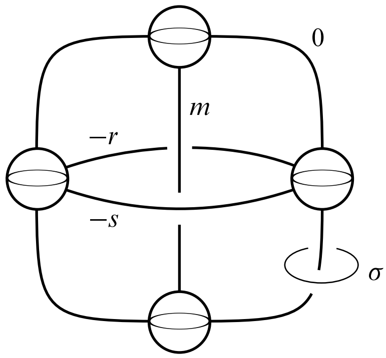


Figure 17: $Z_0(r,s;m)$ showing elliptic fibration of $Z_0(1,1;-1)$ with section σ

replacing $W_0 = T^2 \times D^2$ by W_k, preserving the longitude σ. This is precisely the Fintushel–Stern knot construction, using the knot K_k, and (when $r = s = 1$ and $m = -1$) using a regular fiber of the elliptic fibration on $Z_0(1,1;-1)$.

To prove our main theorem, we need one last routine lemma:

Lemma 3.4 *A self-diffeomorphism φ of the pair $(\partial Z_0(1,1;-1),\sigma)$ that preserves the orientation of σ must be isotopic to the identity (through self-maps of the pair).*

The control of σ is necessary, in order to rule out twists on fiber tori. The proof rules out horizontal tori.

Proof We can identify each fiber of the torus bundle with $\mathbb{R}^2/\mathbb{Z}^2$ so that σ is zero in each fiber. Then the monodromy is an element $A \in \mathrm{SL}(2,\mathbb{Z})$. If we use the obvious basis for $\mathbb{R}^2$ in Figure 17, then A is given by $\pi/2$ rotation. (This is both well-known and routine to verify in the figure. See also [14].) Choose a fiber F and assume its image $\varphi(F)$ is transverse to it. Since F is incompressible, each circle of intersection is trivial in F if and only if it is trivial in $\varphi(F)$. Each innermost circle in $\varphi(F)$ also bounds a disk in F. The two disks together bound a ball B. If one disk intersects σ (necessarily in a unique point), then so does the other, and $\sigma \cap B$ is an unknotted arc in B. (Otherwise, the complement of a lift of σ to the universal cover of $\partial Z_0(1,1;-1)$ would have nonabelian fundamental group.) Either way, we can eliminate trivial circles by isotoping φ pairwise until all circles (if any remain) are essential. These must be parallel to each other in both of the tori F and $\varphi(F)$, cutting each into annuli. Since ± 1 is not an eigenvalue of A, no such annulus of $\varphi(F)$ can surject onto the base circle. Thus, if the intersection is nonempty, $\varphi(F)$ has more than one annulus, and we can choose one that does not contain the intersection point with σ. This fits together with an annular region in F that we choose disjoint from σ, to form a nullhomologous torus. This torus is compressible (as seen, for example, in the $\mathbb{Z}$–cover of $\partial Z_0(1,1;-1)$), so it bounds a solid torus, with the annuli bounded by longitudes of it. We can now reduce the number of intersection circles until $\varphi(F)$ is disjoint from F. Cutting along F, we see $\varphi(F)$ as an incompressible torus in $F \times I$. Applying standard theory to the complement of σ (eg Waldhausen [21, Proposition 3.1]) we can arrange $\varphi(F)$ to be a fiber, and then isotope φ so that it covers the identity map on the base circle. It is easily checked that the only elements of $\mathrm{SL}(2,\mathbb{Z})$ that commute with A are powers of A. Thus, on each fiber, φ restricts to A^j for a fixed j. We can then change φ to the identity by a fiber-preserving isotopy covering a $2\pi j$–rotation on the base. □

We can now prove our main theorem. For $k \in \mathbb{Z}$ and fixed $r, s, n > 0 > m$, let X_k be the 4–manifold obtained from $E(n)$ by the Fintushel–Stern construction on a fiber, using the twist knot K_k (so $X_0 = E(n)$). We have embeddings

$$C(r,s;m) \subset Z_0(r,s;m) \subset X_0 \# N\overline{\mathbb{CP}^2},$$

where $N = r+s-m-3 \geq 0$, and the last embedding is obtained from the simplest case $r = s = -m = 1$ by blowing up meridians of the three negatively framed 2–handles of Figure 17 to suitably lower their framings. Let X_k^* be the manifold obtained from $X_0 \# N\overline{\mathbb{CP}^2}$ by cutting out $C(r,s;m)$ and regluing it via the torus twist f^k.

Theorem 3.5 [13] *For each k, the manifold X_k^* is diffeomorphic to $X_k \# N\overline{\mathbb{CP}^2}$. In particular, the manifolds X_k^* for $k \in \mathbb{Z}$ are pairwise nondiffeomorphic, so (for each fixed choice of r, s, m as above) $(C(r,s;m), f)$ is an infinite order cork.*

Proof First consider the simplest case $r = s = -m = 1$. Starting from the embedding $Z_0(1,1;-1) \subset X_0$, we can cut out a regular neighborhood W_0 of a fiber inside $Z_0(1,1;-1)$ and replace it by W_k, obtaining $Z_k(1,1;-1) \subset X_k$ with the embedding preserving σ (Proposition 3.3 and below). Alternatively, we can cut out the cork $C(1,1;-1)$ and reglue it by f^k, obtaining $Z_k(1,1;-1) \subset X_k^*$, again preserving σ (Proposition 3.2). Clearly, the complements of $Z_k(1,1;-1)$ in these two closed manifolds are identified (preserving σ). But the two embeddings of $Z_k(1,1;-1)$ are related by a diffeomorphism preserving σ and its orientation, so by Lemma 3.4 we can assume they agree on the identified boundaries of the complements. Thus, the diffeomorphisms fit together as required.

For the general case, we blow up to obtain embeddings

$$Z_0(r,s;m) \subset Z_0(1,1;-1) \# N\overline{\mathbb{CP}^2} \subset X_0 \# N\overline{\mathbb{CP}^2}.$$

The first embedding is obtained from Figure 11 by adding -1–framed meridians to the curves with framings $-r$, $-s$ and m so that blowing down changes all three framings to -1. After the two 2–handles ζ_i of $Z_0(r,s;m)$ cancel their 1–handles, two sets of these -1–framed meridians can be drawn as parallel copies of the curves η_i. Since the torus twist does not disturb η_1, η_2 or the m–framed meridian, it gives embeddings $Z_k(r,s;m) \subset Z_k(1,1;-1) \# N\overline{\mathbb{CP}^2} \subset X_k^*$. The theorem now follows from Lemma 3.4 as in the previous case. □

In principle, there should be a direct proof of the theorem, by drawing X_k^* and $X_k \# N\overline{\mathbb{CP}^2}$, and exhibiting an explicit diffeomorphism. A link diagram of X_k was drawn by Akbulut, then independently produced as [14, Figure 10.2] (discussion on pages 407–408), using the technique of [2]. This diagram is obtained from Figure 11 of $Z_k(1,1;-1)$ by adding some 2–handles and a 4–handle. One 2–handle is attached along σ with framing $-n$. The others are -1–framed and attached along parallel copies of the circles with framing m and $-r$ (or $-s$), but the two types of new curves are interleaved. A diagram of X_k^* can be similarly constructed by torus twisting $Z_0(1,1;-1)$. To show the diagrams are diffeomorphic, it suffices to connect the $-r$– and m–framed curves by a framed arc whose union with the two attached circles and σ is preserved (after handle slides) by the torus twist, since all the new 2–handles will be attached in a neighborhood of these. This project has not been attempted with sufficient intensity for success.

4 Twists that preserve 4–manifolds

Having explicitly exhibited infinite order cork twists, we now address the opposite issue, finding conditions under which twisting a contractible submanifold does not change the diffeomorphism type of a 4–manifold. Let $T \subset M$ be an embedded torus or Klein bottle in a 3–manifold, and let f be the twist on T parallel to a circle $\alpha \subset T$, as described in Section 2. Let W be the elementary cobordism built from $I \times M$ by adding a 2–handle h to $\{1\} \times M$ along a parallel copy γ of α, with framing ± 1 relative to T. Thus, the top boundary $\partial_+ W$ is obtained from M by surgery on γ.

Theorem 4.1 *The twist f on $\partial_- W = \{0\} \times M$ extends over W so that it is the identity on $\partial_+ W$.*

Proof Let g_t be an isotopy of the identity on M, supported in a tubular neighborhood of T, that preserves T setwise but rotates it once parallel to a circle β dual to α. Interpret the isotopy $g_t \circ f$ as a self-diffeomorphism of $I \times M$. We can assume that γ lies outside the support of this map in $\{1\} \times M$, then extend over the handle h by the identity. In $\partial_+ W$, Figure 1 (reversed) shows an isotopy rotating T back to its original position while undoing the twist produced by f. □

In the case where T is a torus, the above diffeomorphism of W is a manifestation of a *fishtail twist*. The latter has been used in various forms for some decades; see [12] for a recent discussion. If N denotes a tubular neighborhood of T in M, then $I \times N \approx T^2 \times D^2$, and $I \times N \cup h$ is a fishtail neighborhood. It is well known that the twist on $\{0\} \times T$ parallel to α extends over this neighborhood as the identity on the rest of its boundary. The main point is that the boundary is a torus bundle with monodromy given by a Dehn twist parallel to α, so the torus twist can be absorbed by a fiber-preserving isotopy covering a full rotation of the base.

As an application, we partially answer [13, Question 1.6]. Let $D \subset B^4$ be a slice disk for a composite slice knot $K = \partial D$. For example, D can be the obvious ribbon disk for any nontrivial knot of the form $\kappa \# -\kappa$, the case considered in [13]. Let $C = C(D, m)$ be the contractible 4–manifold obtained from the slice complement by adding a 2–handle along a meridian with framing $m \neq 0$, so that C is $C(r, s; m)$ in the case $\kappa = \kappa(r, -s)$. The boundary of C is the homology sphere obtained by $(-1/m)$–surgery on K. It is irreducible and has incompressible tori as in the previous section: Start with a sphere S in S^3 intersecting K in two points and splitting it nontrivially as a sum $K_0 \# K_1$. Remove the intersections by surgering S to a torus in $S^3 - K$, using a tube following K_1. Such a torus has an obvious product decomposition, with one factor a meridian of K and the other a 0–framed longitude of K_1. The cork twists of $C(r, s; m)$ in the previous section have this form for a longitudinal twist, on the unique incompressible torus if $m = -1$. It was asked in [13] whether twisting a fixed embedding of C by the full action of such a torus could give a family of distinct diffeomorphism types indexed by $\mathbb{Z} \oplus \mathbb{Z}$. Previously, a preliminary version of Akbulut's 2014 posting [3] unsuccessfully attempted to show that the meridian twist was an infinite order cork twist in the case of the obvious ribbon disk for the square knot with $m = -1$, that is, the manifold $C(1, -1; -1)$ in our present notation. However, both constructions are impossible when $m = \pm 1$:

Corollary 4.2 *Every torus twist parallel to the meridian of K extends over $C(D, \pm 1)$. In particular, for each $r, s \in \mathbb{Z}$, the meridian twist on $\partial C(r, s; \pm 1)$ extends over $C(r, s; \pm 1)$.*

Proof We find a cobordism $W \subset C$ as in the theorem with $\partial_- W = \partial C$. Since the diffeomorphism extends over W as the identity on $\partial_+ W$, we can extend as the identity over the rest of C. To construct W, begin with a collar of ∂C. The additional 2–handle h is obtained by thickening the cocore of the meridian 2–handle h^* of C.

The attaching circle of h is a 0–framed meridian to that of h^*. Interpreting the diagram as a 3–manifold and blowing down h^*, we realize h by a ∓ 1–framed meridian of K as required. □

It follows that for a fixed embedding and torus, cutting $C(D;\pm 1)$ out of a 4–manifold and regluing it by torus twists generates a family of 4–manifolds whose diffeomorphism types are indexed at most by $\mathbb{Z}$. The problem remains open when $|m| > 1$. However, distinguishing meridian twists of $C(r,s;m)$ would require a somewhat different approach, since the proof in [13] depends on an embedding in a 4–manifold X satisfying the hypothesis of the following for the meridian α:

Corollary 4.3 *Let $Y \subset X$ be a 4–manifold pair, and let f be a twist on a torus or Klein bottle $T \subset \partial Y$, parallel to some curve α. Suppose that $X - \operatorname{int} Y$ contains an embedded disk with boundary α, inducing framing ± 1 relative to T. Then cutting out Y and regluing it after twisting by a power of f yields a manifold diffeomorphic to X.*

Proof Observe the cobordism W in $X - \operatorname{int} Y$. Extend f^k outward from there by the identity. □

The same question of [13] asks about longitudinal twists for slice disks not covered by the main theorem of that paper. Ray and Ruberman [18] have recently observed that when K_1 is a torus knot, *every* twist on the torus determined by K_1 extends over $C(D,\pm 1)$. This is seen by combining Corollary 4.2 with the Seifert circle action on the complement of K_1, which shows that twisting on some (nonzero) longitude is isotopic to the identity. A closer look yields the first examples of contractible manifolds, including $C(1,-1;-1)$, that cannot be nontrivial corks even though their boundaries have incompressible tori:

Corollary 4.4 *Let $C = C(D,\pm 1)$ be obtained as above with $\partial D = \kappa \# -\kappa$, where κ is a torus knot. Then every diffeomorphism of ∂C extends over C.*

Proof The boundary of C is ∓ 1–surgery on $\kappa \# -\kappa$, which is made by gluing together the complements of κ and $-\kappa$ along their boundary tori T. (The surgery can be interpreted as a twist in the gluing map.) Since T is the unique incompressible torus in ∂C, it is preserved by any self-diffeomorphism (up to isotopy). We have seen that twists on T extend. The nonzero signature of κ rules out orientation-preserving diffeomorphisms of ∂C switching the two knot complements, and orientation-reversing switches are ruled out by the handedness of the gluing map. The complement of κ is Seifert fibered with two exceptional fibers, and any self-diffeomorphism preserves this structure, so we are left with only the involution of κ that reverses its string orientation. The induced involution of ∂C obviously extends. □

5 Torus twists and δ–moves

In this section, we give a careful definition of Akbulut's δ–moves, and almost entirely reduce them to twists on tori, Klein bottles and spheres. Torus and Klein bottle twists were introduced in Section 2. Twists on spheres are defined similarly but have order at most 2 in $\pi_0(\mathrm{Diff}_+(M))$ (since $\pi_1(\mathrm{SO}(3)) = \mathbb{Z}/2$). Klein bottle twists are of limited use: They only arise when M contains the I–bundle over the Klein bottle with orientable total space, a somewhat rare phenomenon. In particular, this does not occur for homology spheres, so there can be no Klein bottle cork twists. A tubular neighborhood of a Klein bottle K is bounded by a torus T double covering K. This is incompressible if and only if K is, since any compressing disk for K must be bounded by an orientation-preserving loop in K. It is easy to see that the square of a twist on K is a twist on T parallel to the same circle α.

We define δ–moves following Akbulut [4], with additional attention to detail in anticipation of the upcoming proofs. First, consider the standard 3–manifold diffeomorphism given by Figure 18, which can be obtained by blowing up a ± 1–framed unknot around one twist box, sliding this unknot over the 0–framed circle so that it surrounds the other twist box, then blowing back down. For a δ–move, start with a framed circle C in a 3–manifold M. Draw M as surgery on a link L in S^3 so that C appears as a 0–framed unknot in $S^3 - L$, spanned by a disk $\Delta \subset S^3$. Let $C_\pm$ be a pair of circles parallel to C in the diagram and disjoint from Δ. Connect these circles by a (possibly complicated) band b in the diagram, disjoint from Δ and from the interior of the annulus A bounded by $C_\pm$. (See Figure 19, ignoring the horizontal dashed curve.) The surface $T_0 = A \cup b$ is an embedded punctured torus or Klein bottle in $S^3 - L$, depending on the twisting of b. Let $\delta = \partial T_0$. Under the additional hypothesis that δ is unknotted in the 3–manifold M, we can add a suitably framed 2–handle to $I \times M$ along δ and cancel it with a 3–handle to recover $I \times M$. If T_0 is orientable, this 2–handle will be 0–framed in S^3 (since the normal to δ along T_0 will give the 0–framing in both S^3 and M). Otherwise, we hypothesize that its framing is 0 in S^3. Since C is unknotted in S^3 and b avoids A and Δ, we can then apply Figure 18 to change k to $k+1$ in Figure 19. Canceling the new twists by an isotopy in S^3, we return to the original diagram (ignoring the dashed curve). Assuming the 3–handle can be suitably controlled (an issue we discuss below), the net effect is a self-diffeomorphism of M. To see that this diffeomorphism may be nontrivial, note that it wraps the dashed curve twice around A parallel to C.

Definition 5.1 The above diffeomorphism (when defined) is called a *δ–move* [4]. The corresponding link diagram of M, with δ unknotted in M (inducing the 0–framing in S^3) and drawn as in Figure 19 for an explicit choice of b, will be called a *δ–move*

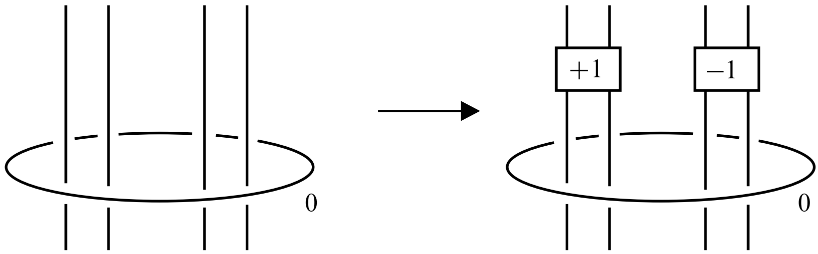


Figure 18: Twisting along a 0–framed unknot

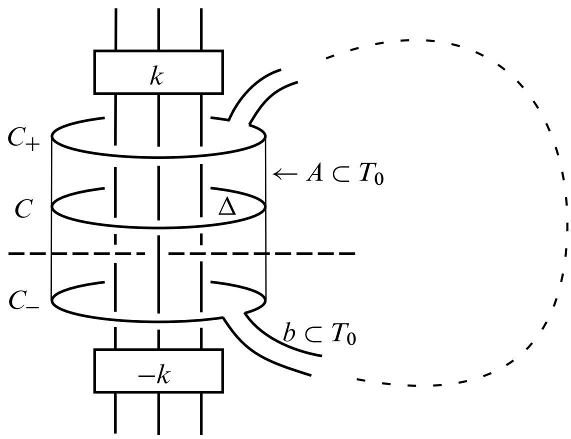


Figure 19: A δ–move, where $\delta = \partial T_0$ is the band-sum of C_+ and C_- along the band b

diagram. A δ–move diagram will be called *orientable* or *nonorientable* according to whether T_0 is orientable. It will be called *compressible* if there is a disk $d \subset M$ such that $d \cap T_0 = \partial d$ is neither trivial nor boundary-parallel in T_0. It will be called *incompressible* otherwise.

The relation between surface twists and δ–moves begins with the following:

Proposition 5.2 *Every torus (resp. Klein bottle) twist on a 3–manifold M is isotopic to a δ–move with an orientable (resp. nonorientable) diagram.*

Proof Let $T \subset M$ be a torus or Klein bottle, containing circles α and β as in Definition 2.1 (where β is a section of the circle bundle in the Klein bottle case). Choose a surgery diagram of M in which α is given by an unknot in $S^3 - L$ whose 0–framing is given by the normal vectors to α in T, and whose spanning disk is disjoint from β. Then the subset $\alpha \cup \beta$ has a neighborhood in T that can be identified with T_0 in the definition of a δ–move, with α identified with C. The boundary δ of this T_0 explicitly bounds a disk $D \subset T \subset M$, so is unknotted in M as required, and correctly framed. (Even a nonorientable T_0 induces the 0–framing on δ in S^3, as seen by using Δ to surger it to a disk.) To reinterpret the twist on T parallel to α as a δ–move, we attach a 2–handle to δ with framing 0, then cancel it with a 3–handle, whose attaching sphere can be chosen to be D capped with the core of the 2–handle. The δ–move is realized by blowing up a ± 1–framed circle γ at C_+, sliding it over the 2–handle at δ to C_-, blowing it back down, and canceling the twists as in Figure 19. If we cancel the 2–3 handle pair, the slide over the 2–handle becomes an isotopy across the disk $D \subset T \subset M$. Thus, the slide appears in M as an isotopy dragging γ from C_+ to C_- around T in the direction that avoids the intervening annulus A. (In S^3, we see a handle slide each time γ follows D over a handle.) To show that this δ–move is the twist on T, it suffices to work in a tubular neighborhood of T containing the support of the diffeomorphisms and check that the δ–move diverts any curve in M that crosses T, parallel to α. This is true for curves intersecting A, as Figure 19 shows. Other curves through T will be suitably modified as in Figure 1 when γ collides with them. □

To make progress on a converse to this proposition, we must understand the extent to which a δ–move is well-defined in general. Attaching the 2–handle along δ caps off the surface T_0 to an embedded torus or Klein bottle T. However, this lives not in M, but rather in the manifold $M_\# = M \# S^1 \times S^2$ obtained from M by 0–surgery on the unknot δ. If the disk D in M along which the 3–handle is attached is disjoint from $\operatorname{int} T_0$, we can eliminate the difficulty by canceling the 2–3 pair, obtaining a torus or Klein bottle twist on M as in the previous proof. However, a proposed advantage of δ–moves is their apparent additional generality, so we should consider what happens when $\operatorname{int} D$ is allowed to intersect T_0 or other surfaces in the construction. (For a specific example, start with a twist on a separating sphere, surger the sphere at its poles to an immersed Klein bottle, and interpret this as a δ–move with a nonorientable diagram.) In this generality, we have a torus or Klein bottle twist $f_\#$ in $M_\#$ that we wish to interpret as a diffeomorphism of M. We recover M from $M_\#$ by surgering out the attaching sphere $S \subset M_\#$ of the 3–handle, which is obtained by capping D with the core of the new 2–handle. The first difficulty we encounter if $\operatorname{int} D$ intersects T_0 is that $f_\#$ may move S. Thus, to have a well-defined diffeomorphism of M, we must isotope $f_\#(S)$ back to S in $M_\#$ before surgering back to M. This is not always possible. For example, starting from a torus twist exhibited as in the previous proof, we can obtain D from the obvious disk by tubing it together with an essential sphere along an arc that intersects A. Then $f_\#$ can change the arc by a nontrivial element of $\pi_1(M)$ so that $f_\#$ changes the class of S in $\pi_2(M)$. If we know that $f_\#$ is isotopic to a diffeomorphism preserving S, then it does extend over the 3–handle, so restricts to a diffeomorphism on M. However, this diffeomorphism need not be unique: Starting again from a torus twist, with the torus bounding a solid torus in M, construct a new M' by connected sum with another 3–manifold. If the sum occurs outside the solid torus, the twist of M' is still trivial. If it occurs inside, we can obtain a slide diffeomorphism with infinite order in $\pi_0(\mathrm{Diff}_+(M'))$ (detected by its effect on $\pi_2(M')$). Thus, a δ–move depends in general on the particular choice of auxiliary disk D capping δ in M. This can be difficult to specify explicitly in a diagram. To make matters worse, it is a nontrivial problem to understand the extent to which the choice of isotopy from $f_\#(S)$ back to S affects the resulting diffeomorphism of M. Fortunately, the issue can be resolved through work of Hatcher and McCullough [15].

Theorem 5.3 *Every δ–move diagram for an irreducible 3–manifold M determines a unique δ–move diffeomorphism up to isotopy. On a reducible manifold M, a δ–move diagram, together with a choice of auxiliary disk $D \subset M$ spanning δ (up to isotopy rel boundary) determines at most one diffeomorphism up to isotopy and elements of order 2 in $\pi_0(\mathrm{Diff}_+(M))$. The latter are composites of twists on a fixed collection of disjoint spheres.*

Proof Given a δ–move diagram and a fixed choice of spanning disk $D \subset M$ for δ, let $S \subset M_\#$ be the associated surgery sphere. Given two isotopies of $f_\#(S)$ to S in $M_\#$, we wish to relate the corresponding diffeomorphisms of M. Before the surgery is reversed, these are related by composition with a diffeomorphism of the pair $(M_\#, S)$ that is isotopic (not preserving S) to the identity. By [15, Lemma 3.4] (with $n = 0$ and $S_0 = S$), such a diffeomorphism, up to isotopy, comes from a composite of sphere twists on the manifold M_1 made by cutting $M_\#$ along S. We reverse the surgery by capping off the boundary components of M_1 with balls. If M is irreducible, the spheres in question all bound balls in M, so their twists are isotopic to the identity. Otherwise, McCullough [17, Section 3] shows that the sphere twists of M generate a normal subgroup $\mathcal{R}(M)$ of $\pi_0(\mathrm{Diff}_+(M))$ isomorphic to $\oplus_r \mathbb{Z}_2$ for some finite r. We can surger M on 2–spheres to get a connected sum of irreducible manifolds, and for any such presentation, the sum spheres and surgery spheres together can be assumed disjoint and comprise a generating set for $\mathcal{R}(M)$. (Thus, r is at most the number of prime summands of M, with equality only when $M = \#r S^1 \times S^2$.) The reducible case of the theorem follows immediately, since any isotopy of D rel boundary results in an isotopy of the corresponding diffeomorphisms of M. For the remaining case, suppose M is irreducible. Existence follows since S lies in the unique isotopy class of nonseparating spheres in $M_\#$, and uniqueness follows since the disk spanning δ in M is unique up to isotopy rel boundary. □

Because of the difficulty of tracking isotopy classes of spanning disks in diagrams, it is natural either to assume that M is irreducible or to allow the spanning disk to vary. A

δ–move diagram may represent more than one diffeomorphism in the reducible case (although the curve δ itself is held fixed by the diagram). We show that under broad hypotheses, every diagram represents a diffeomorphism, which can be taken to be a torus or Klein bottle twist.

Theorem 5.4 *Every δ–move diagram represents a δ–move that is isotopic to a torus or Klein bottle twist parallel to C, provided that the 3–manifold M has no $\mathbb{R}P^3$ summand, or that the diagram is orientable or incompressible. (The case of a Klein bottle only arises if the diagram is nonorientable.) If the diagram is compressible, the resulting twist is isotopic to the identity, provided that the diagram is orientable or M is irreducible (and not $\mathbb{R}P^3$).*

Corollary 5.5 *Every δ–move diagram for a homology sphere M represents a δ–move that is isotopic to a torus twist. If M is also irreducible, then δ–moves and torus twists comprise the same subset of $\pi_0(\mathrm{Diff}_+(M))$.* □

Proof of Theorem 5.4. We begin with a δ–move diagram, whose curve δ bounds an embedded disk $D \subset M$ by definition. We wish to modify D so that its interior becomes disjoint from T_0. Then $T_0 \cup D$ is an embedded torus or Klein bottle, and the proof of Proposition 5.2 shows that the resulting twist is realized up to isotopy by the diagram. Recall that the surfaces T_0 and D induce the same framing on their common boundary δ, so we can assume their interiors intersect in a finite collection of circles. We can eliminate all circles bounding disks in T_0 by successively replacing disks in D by innermost disks in T_0. If any innermost circle of D is then boundary-parallel in T_0, the required new version of D is obtained by joining the corresponding innermost disk to an annulus parallel to a boundary collar of T_0. Otherwise, either there are no remaining circles and we are done, or an innermost disk d of D exhibits the diagram as compressible. In the latter case, if T_0 is orientable, the required disk is obtained by surgering a parallel copy of T_0 along d. The resulting torus T is exhibited as the boundary of a solid torus. Thus, the diagram represents a twist on the boundary of a solid torus, which is in turn isotopic to the identity. If T_0 is nonorientable, ∂d cannot bound a Möbius band in T_0, or else we could construct an embedded projective plane, whose tubular neighborhood would be an $\mathbb{R}P^3$ summand violating our hypotheses. Thus, ∂d is the unique nonseparating circle in T_0 with orientable complement, namely the circle C generating the δ–move. Now we change tactics, modifying T_0: An isotopy of M rotating d by a full turn untwists the δ–move near C at the expense of adding twists on a pair of parallel copies of d. This isotopes the Klein bottle twist $f_\#$ in $M_\#$ to a twist on the sphere $S^* \subset M_\#$ made from T_0 by surgering on d. The disk D, and hence the surgery sphere S, can easily be made disjoint from S^*, so that they are not moved by the sphere twist in $M_\#$. Thus, $f_\#$ only changes S by an isotopy. It follows immediately that the original diagram, together with this S (or D) and isotopy, determines a δ–move, and it is isotopic to the twist on the sphere in M descending from S^* by surgery on S. We can further surger this sphere in M along a tube connecting its poles, obtaining a torus twist. Alternatively, if M is irreducible, the sphere bounds a ball over which the twist extends, so the twist is isotopic to the identity. □

The proof also gives a more general result about the compressible case:

Corollary 5.6 *If M has no $\mathbb{R}P^3$ summand, then every compressible δ–move diagram represents an element of order at most 2 in $\pi_0(\mathrm{Diff}_+(M))$ (that is the identity in the orientable case).*

Proof The orientable case is given by the theorem. Its proof shows that the nonorientable case can be reduced to a twist on a sphere. □

Corollary 5.7 *Suppose M is atoroidal with no $\mathbb{R}P^3$ summand. Then every δ–move diagram represents an element of order at most 2 in $\pi_0(\mathrm{Diff}_+(M))$. If the diagram is orientable or M is irreducible, it represents the identity.*

Proof By definition, M has no incompressible tori, and hence no incompressible Klein bottles. The proof of Theorem 5.4 generates such a surface from any incompressible δ–move diagram, so the previous corollary applies. The last sentence follows from the statement of the theorem. □

Akbulut's motivation for introducing δ–moves was to study cork twisting by diagrams as in Section 3, starting from a pair $Y \subset X$ and regluing Y. This raises a subtle technical issue. From the viewpoint of the definition of δ–moves, the issue centers on the isotopy from $f_\#(S)$ to S for surgery reversal. We start with a link diagram of $M = \partial Y$, and then add additional handles along a framed link $L' \subset M$ to get X. Given a δ–move diagram for M, we must understand the effects of a resulting move on L'. To introduce the canceling 2–3 handle pair without moving L', this link must be disjoint from the disk D where the 3–handle attaches, a condition that can be routinely checked. The effects of the resulting torus twist $f_\#$ on L' in $M_\#$ are easy to see: We can assume that L' intersects the torus only in the annulus A in the diagram, and then L' is only changed at these intersections, by twisting parallel to C. However, we must then compose $f_\#$ with a diffeomorphism g (isotopic to $\mathrm{id}_{M_\#}$) returning S to its original position for surgery. Note that g can be complicated; for example, any sheets of S intersecting T_0 in parallel copies of β will be dragged by $f_\#$ over the 2–handle attached to δ. The effect of g on $f_\#(L')$ is an unspecified isotopy in $M_\#$ that could cause band-summing with parallel copies of δ. Reversing this isotopy could cause intersections of the link with S that would prevent it from surviving the surgery. Thus, it is not clear how the δ–move affects the auxiliary link L' without a more careful analysis.

An analogous problem arises from the viewpoint of expressing δ–moves as torus twists. If we start from a δ–move diagram, we can see where T_0 intersects L'. Given a procedure for seeing that δ is unknotted in M, it is routine to verify if the resulting disk D is disjoint from L'. If D is not directly visible in the diagram, however, we must assume it intersects $\operatorname{int} T_0$ and apply the method of Theorem 5.4. This replaces D by a new disk D' disjoint from $\operatorname{int} T_0$, and it is not generally clear whether D' intersects L'. Since D' is constructed in a neighborhood of $T_0 \cup D$, it avoids every link component disjoint from T_0. However, the diffeomorphism is only interesting when L' has nontrivial intersection with T_0, in which case further analysis is needed to determine whether f causes unexpected movement of L'. This is why we exhibited L' and the entire torus T simultaneously in the same diagram for Proposition 3.2. As mentioned there, there are other approaches, notably drawing T as an isotopy of a circle and applying Figure 1.

References

[1] **S Akbulut**, *A fake compact contractible* 4*–manifold*, J. Differential Geom. 33 (1991) 335–356 MR

[2] **S Akbulut**, *Scharlemann's manifold is standard*, Ann. of Math. 149 (1999) 497–510 MR

[3] **S Akbulut**, *An infinite order cork*, preprint (2014) arXiv

[4] **S Akbulut**, *On infinite order corks*, preprint (2016) arXiv To appear in Proc. Gökova Geom. Topol. Conf. (2016)

[5] **S Akbulut**, **R Kirby**, *Branched covers of surfaces in* 4*–manifolds*, Math. Ann. 252 (1979/80) 111–131 MR

[6] **S Akbulut**, **R Matveyev**, *A convex decomposition theorem for* 4*–manifolds*, Internat. Math. Res. Notices (1998) 371–381 MR

[7] **D Auckly**, **H J Kim**, **P Melvin**, **D Ruberman**, *Equivariant corks*, Algebr. Geom. Topol. 17 (2017) 1771–1783 MR

[8] **C L Curtis**, **M H Freedman**, **W C Hsiang**, **R Stong**, *A decomposition theorem for h–cobordant smooth simply-connected compact* 4*–manifolds*, Invent. Math. 123 (1996) 343–348 MR

[9] **R Fintushel**, **R J Stern**, *Knots, links, and* 4*–manifolds*, Invent. Math. 134 (1998) 363–400 MR

[10] **R Fintushel**, **R J Stern**, *Double node neighborhoods and families of simply connected 4–manifolds with* $b^+ = 1$, J. Amer. Math. Soc. 19 (2006) 171–180 MR

[11] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR

[12] **R E Gompf**, *More Cappell–Shaneson spheres are standard*, Algebr. Geom. Topol. 10 (2010) 1665–1681 MR

[13] **R E Gompf**, *Infinite order corks*, Geom. Topol. 21 (2017) 2475–2484 MR

[14] **R E Gompf**, **A I Stipsicz**, *4–manifolds and Kirby calculus*, Graduate Studies in Math. 20, Amer. Math. Soc., Providence, RI (1999) MR

[15] **A Hatcher**, **D McCullough**, *Finite presentation of 3–manifold mapping class groups*, from "Groups of self-equivalences and related topics" (R A Piccinini, editor), Lecture Notes in Math. 1425, Springer (1990) 48–57 MR

[16] **R Matveyev**, *A decomposition of smooth simply-connected* h*–cobordant 4–manifolds*, J. Differential Geom. 44 (1996) 571–582 MR

[17] **D McCullough**, *Twist groups of compact 3–manifolds*, Topology 24 (1985) 461–474 MR

[18] **A Ray**, **D Ruberman**, *4–dimensional analogues of Dehn's lemma*, J. London Math. Soc. 96 (2017) 111–132

[19] **M Tange**, *Finite order corks*, Internat. J. Math. 28 (2017) art. id. 1750034, 26 pages MR

[20] **B Trace**, *On attaching 3–handles to a 1–connected 4–manifold*, Pacific J. Math. 99 (1982) 175–181 MR

[21] **F Waldhausen**, *On irreducible 3–manifolds which are sufficiently large*, Ann. of Math. 87 (1968) 56–88 MR

[22] **F Waldhausen**, *Recent results on sufficiently large 3–manifolds*, from "Algebraic and geometric topology, Part 2" (R J Milgram, editor), Proc. Sympos. Pure Math. XXXII, Amer. Math. Soc., Providence, RI (1978) 21–38 MR

Department of Mathematics, The University of Texas
Austin, TX, United States

gompf@math.utexas.edu

Received: 7 September 2016 Revised: 11 March 2017

msp

Algebraic & Geometric Topology 17 (2017) 2893–2914

Detecting essential surfaces as intersections in the character variety

MICHELLE CHU

We describe a family of hyperbolic knots whose character variety contain exactly two distinct components of characters of irreducible representations. The intersection points between the components carry rich topological information. In particular, these points are nonintegral and detect a Seifert surface.

57M25; 20C15, 57M27, 57M50

1 Introduction

The $SL_2\mathbb{C}$ character varieties of the fundamental groups of hyperbolic 3–manifolds carry a lot of topological information. In particular, Culler and Shalen [4] developed a technique to detect embedded essential surfaces in a 3–manifold that arise from nontrivial actions of the fundamental group on a tree arising from ideal points in the $SL_2\mathbb{C}$ character variety. The $SL_2\mathbb{C}$ character variety of a hyperbolic knot group contains multiple components, including the canonical component, which contains the character of a holonomy representation, and a component containing the characters of reducible representations. We address the following question:

Question 1.1 *How do multiple components in the $SL_2\mathbb{C}$ character variety interact? In particular, what can we say about the characters in the intersection between multiple components?*

In this paper we consider a family of two-bridge knots whose character varieties contain two distinct curves containing characters of irreducible representations. For this family, the existence of multiple curves was known to Ohtsuki [12] and the existence of exactly two distinct curves was shown by Macasieb, Peterson, and van Luijk [9]. The main result of this paper is the following theorem.

Theorem 1.2 *There exist infinitely many two-bridge knots having two distinct algebraic curve components of irreducible representations in their character varieties and whose intersection points detect a Seifert surface.*

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2893

As is well known, components of the character varieties of two-bridge knots have the structure of algebraic curves which lie naturally in $\mathbb{C}P^2$ (see Section 3.1). As such, Bezout's theorem guarantees finitely many points of intersection between any two curves. Some of these points are ideal, so that, following the methods in Culler and Shalen [4], they detect essential surfaces. It turns out that for this family, affine intersection points determine characters of algebraic nonintegral, irreducible representations and also give interesting topological information. We once again obtain nontrivial actions on a tree, and hence also detect essential surfaces (see Section 2.5).

Interestingly, the characters in the intersection are nonintegral over the prime 2. In addition to these, one can check by explicit computation that the character varieties for the two-bridge knots 7_7, 8_{11}, 9_6, 9_{10}, 9_{17}, 10_5, 10_9 and 10_{32} contain exactly two distinct curves of irreducible representations. It is also true in these examples that affine intersections between multiple curves are algebraic nonintegral, correspond to irreducible representations, and, furthermore, the trace of the meridian in these representations fails to be integral by a prime over 2.

There appears to be no algebrogeometric reason as to why these affine intersection points are nonintegral, and in particular nonintegral by a prime over 2. For instance, computed examples of affine intersection points between curves of characters for two different knots were sometimes integral and other times not. This data suggests the following questions.

Question 1.3 *Suppose K is a hyperbolic two-bridge knot with multiple components of characters of irreducible representations in its character variety. When are intersection points between these components algebraic nonintegral? When are they nonintegral over the prime 2? What slopes are detected? What happens for general knots?*

1.1 Outline

The paper is organized as follows. In Section 2 we give the necessary background on character varieties, two-bridge knots, boundary slopes and actions on trees associated to algebraic nonintegral representations. We also introduce the family of two-bridge knots of interest in this paper. Section 3 builds on the work of Macasieb, Petersen and van Luijk [9]. We construct the character variety, define the smooth variety birationally equivalent to the character variety introduced in [9], and then use the birational equivalence to describe the points of intersection between components. In Section 4 we state and prove a precise version of Theorem 1.2 and determine the surfaces detected by affine intersection points. In Section 5 we describe in detail the intersection points for the first two knots in the family. In Section 6 we make some final remarks on the existence of multiple components and describe examples of other two-bridge knots with multiple components.

Acknowledgments The author thanks her advisor Alan W Reid for his generous guidance, support, feedback and encouragement. The author also thanks Daryl Cooper, Cameron Gordon, Darren Long, Matthew Stover and Anh Tran for insightful conversations and suggestions. Finally, we thank the anonymous referee for their comments and in particular for pointing out that the knots in question do not have a unique Seifert surface and for suggesting Lemma 4.4.

2 Preliminaries

2.1 Character varieties

We begin with some background on representation varieties and in particular character varieties. For more on this material see [4].

Let Γ be a finitely generated group. The $\mathrm{SL}_2\mathbb{C}$–representation variety of Γ is the set $\mathrm{R}(\Gamma) = \mathrm{Hom}(\Gamma, \mathrm{SL}_2\mathbb{C})$ and has the structure of an affine algebraic set over $\mathbb{Q}$ with coordinates given by the matrix entries of the images of the generators of Γ.

The $\mathrm{SL}_2\mathbb{C}$ character variety of Γ is the set $\widetilde{X}(\Gamma) = \{\chi_\rho : \rho \in \mathrm{R}(\Gamma)\}$, where the character $\chi_\rho\colon \Gamma \to \mathbb{C}$ is the map defined by $\chi_\rho(\gamma) = \mathrm{tr}(\rho\gamma)$ for all $\gamma \in \Gamma$. For all $\gamma \in \Gamma$ define the map $t_\gamma\colon \mathrm{R}(\Gamma) \to \mathbb{C}$ by $t_\gamma(\rho) = \chi_\rho(\gamma)$. The ring R generated by 1 and the maps t_γ for $\gamma \in \Gamma$ turns out to be finitely generated by, say, $\{t_{\gamma_1}, \dots, t_{\gamma_m}\}$ for some elements $\gamma_1, \dots, \gamma_m \in \Gamma$. It follows that a character $\chi_\rho \in \widetilde{X}(\Gamma)$ is determined by its values on the finitely many elements $\gamma_1, \dots, \gamma_m \in \Gamma$. We get that $\widetilde{X}(\Gamma)$ has the structure of an affine algebraic set in $\mathbb{C}^m$ with coordinate ring R. Different sets of generators for R give different models for $\widetilde{X}(\Gamma)$ which are all isomorphic over $\mathbb{Z}$.

An $\mathrm{SL}_2\mathbb{C}$ representation $\rho \in \mathrm{R}(\Gamma)$ is reducible if, up to conjugation, $\rho(\gamma)$ is upper triangular for every γ, and otherwise irreducible. An $\mathrm{SL}_2\mathbb{C}$ representation $\rho \in \mathrm{R}(\Gamma)$ is abelian if its image is abelian, and otherwise is nonabelian. Every irreducible representation is nonabelian. However, there exist reducible nonabelian representations. The set of characters of abelian representations $X_{\mathrm{ab}}(\Gamma)$ is itself a variety. Let $X_{\mathrm{na}}(\Gamma)$ be the Zariski closure of $\widetilde{X}(\Gamma) - X_{\mathrm{ab}}(\Gamma)$ and denote it by $X(\Gamma)$.

If two representations ρ, $\rho' \in \mathrm{R}(\Gamma)$ are conjugate, then $\chi_\rho = \chi_{\rho'}$. Also if $\chi_\rho = \chi_{\rho'}$ and ρ is irreducible then ρ and ρ' are conjugate. Therefore, when considering irreducible representations, we may think of $X(\Gamma)$ as the space of irreducible representations modulo conjugation.

Whenever Γ is the fundamental group of an orientable, complete hyperbolic 3–manifolds of finite volume, there is an irreducible component of $X(\Gamma)$ containing the character of a holonomy representation of the 3–manifold. This component is called the *canonical component*.

One can also define the $\mathrm{PSL}_2\mathbb{C}$–character variety (see [3, Section 3; 8, Section 2.1; 9, Section 2.1.2]). In the case of Γ a knot group, the $\mathrm{PSL}_2\mathbb{C}$–character variety $\widetilde{Y}(\Gamma)$ is the quotient $\widetilde{X}(\Gamma)/\mathrm{Hom}(\Gamma, \pm 1)$, where ± 1 is the kernel of the homomorphism $\mathrm{SL}_2\mathbb{C} \to \mathrm{PSL}_2\mathbb{C}$. It has as coordinate ring the subring of R of elements invariant under ± 1. Let $Y(\Gamma)$ denote the image of $X(\Gamma)$ in $\widetilde{Y}(\Gamma)$.

2.2 Two-bridge knots

Two-bridge knots are those nontrivial knots admitting a knot diagram with two maxima. Every two-bridge knot is associated to a two-bridge normal form (p,q), where p and q are integers with p odd and $0 < q < p$. Whenever $q \neq 1$, the associated knot is hyperbolic. Two knots with two-bridge normal forms (p,q) and (p',q') are equivalent if and only if $p = p'$ and either $q = q'$ or $qq' \equiv \pm 1 \bmod p$.

The knot group corresponding to the two-bridge normal form (p,q) has a presentation $\langle a, b : aw = wb \rangle$, where a and b are meridians and $w = a^{\epsilon_1} b^{\epsilon_2} \cdots a^{\epsilon_{p-2}} a^{\epsilon_{p-1}}$ with $\epsilon_i = (-1)^{\lfloor iq/p \rfloor}$ and $\lfloor \cdot \rfloor$ the floor function (see [13, Proposition 1; 10, Proposition 1]).

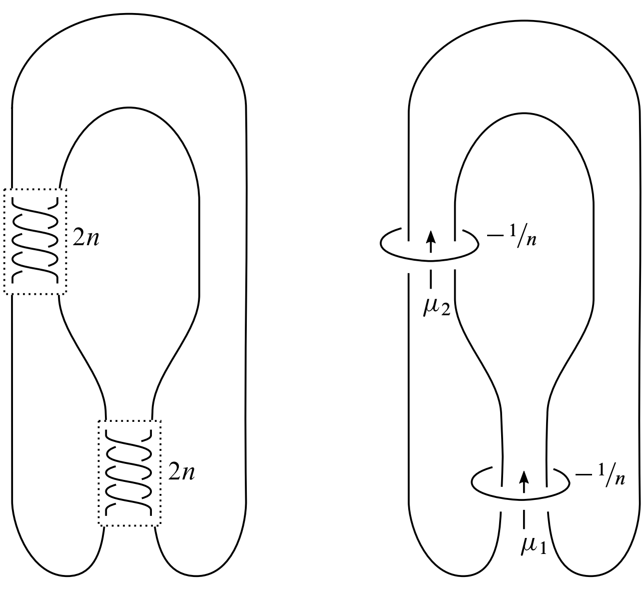


Figure 1: The two-bridge knot $J(2n, 2n)$ (left) and the Borromean rings (right)

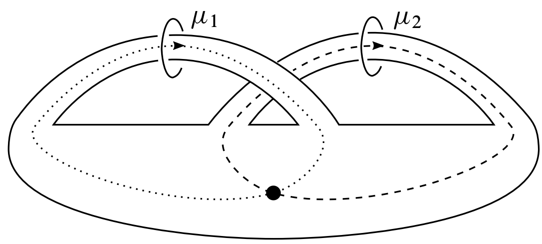


Figure 2: Borromean rings after isotopy

For this presentation the preferred meridian is given by a and the corresponding preferred longitude is given by $ww^*a^{-2e(w)}$, where w^* is w written backwards and $e(w) = \sum \epsilon_i$, so that the total exponent sum of the longitude is 0 (see [6, Section 2]).

2.3 A family of two-bridge knots

The knots to be considered in this paper are the members of the family of hyperbolic two-bridge knots $J(2n, 2n)$ for $n \geq 2$ with two-bridge normal form $(4n^2-1, 4n^2-2n-1)$. Note that this form is equivalent to $(4n^2-1, 2n)$. These have knot diagrams as shown in Figure 1, left, and are obtained as $-\frac{1}{n}$ and $-\frac{1}{n}$ surgeries on two components of the Borromean rings as in Figure 1, right. The first knot in this family, $J(4,4)$, is the knot 7_4 in the knot tables with two-bridge normal form $(15, 11)$.

The knot group Γ_n for $J(2n, 2n)$ can be computed as in [6, Proposition 1]. It has presentation

$$\Gamma_n = \pi_1(S^3 \setminus J(2n, 2n)) = \langle a, b : aw^n = w^n b \rangle, \tag{2-1}$$

where $w = (ab^{-1})^n(a^{-1}b)^n$. As in Section 2.2, the preferred meridian is a with corresponding preferred longitude $(w^n)(w^n)^*$.

These knots have an orientable Seifert surface of genus 1 whose fundamental group is generated by the images of the meridians μ_1 and μ_2 after the two Dehn surgeries (see

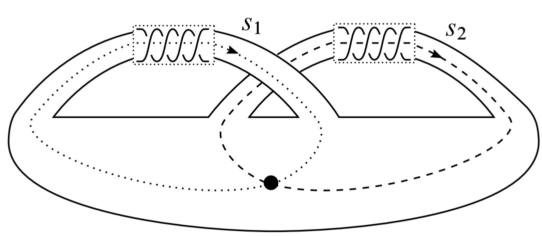


Figure 3: A Seifert surface for the knot $J(2n, 2n)$

Figures 2 and 3). From the proof of [6, Proposition 1], these correspond to

$$s_1 = ((ab^{-1})^n (a^{-1}b)^n)^n \quad \text{and} \quad s_2 = (ab^{-1})^n. \tag{2-2}$$

These two elements generate a free subgroup and their commutator $s_1 s_2^{-1} s_1^{-1} s_2$ corresponds to the preferred longitude. We note that this is not a unique Seifert surface. In fact, it can be shown following [5] that there are two nonisotopic Seifert surfaces.

2.4 Boundary slopes

An essential surface in a 3–manifold is a properly embedded orientable incompressible surface which is not boundary parallel. Let $V := V(K)$ denote the exterior of the knot K. Any embedded essential surface S with nonempty boundary in V will have nonempty boundary $\partial S = S \cap \partial V$, a collection of disjoint circles on the torus boundary of V. Since these circles are disjoint, they represent the same element in the fundamental group of the boundary torus. We identify $\pi_1(\partial V)$ with the group $\mathbb{Z} \times \mathbb{Z}$, where the factors are generated by the preferred meridian and the preferred longitude. Therefore, these circles in ∂S correspond to a class $(p, q) \in \mathbb{Z} \times \mathbb{Z}$, where p and q are relatively prime. We call p/q the slope of S, and represent it in $\pi_1(V)$ by the element $M^p L^q$, where M is the meridian and L the longitude. We say that p/q is a boundary slope for K if there is an essential surface S in V with slope p/q. We call the class $(0, 1)$ the 0–slope and the class $(1, 0)$ the ∞–slope.

Note that two-bridge knots have small exteriors, that is, they do not contain closed embedded essential surfaces (see [5]).

2.5 Algebraic nonintegral representations and actions on the tree

The topics in this section can be found in [17, Section II.1; 4, Sections 1 and 2.3; 18, Section 3; 16, Section 1]. The description of the tree and the action follows from [18, Section 3].

Let H be a number field with a discrete valuation $v\colon H^* \to \mathbb{Z} \cup \{\infty\}$. There is a canonical way to construct a simplicial tree $T_{H,v}$ on which $\mathrm{SL}_2(H)$ acts without inversion. This construction was described by Serre in this form, but was previously discovered by Bruhat and Tits. Let $\mathcal{O}_v$ be the valuation ring and let π be a choice of uniformizer. Define the graph $T_{H,v}$ with vertices given by the homothety classes of lattices in H^2 and an edge between two vertices if there exist representative lattices Λ_0 and Λ_1 and a linear automorphism M of H^2 of determinant β with $v(\beta) = 1$ which maps Λ_0 onto Λ_1. It turns out that $T_{H,v}$ is a tree and $\mathrm{SL}_2(H)$ acts on it simplicially and without inversions by the action induced from the action on H^2.

When the fundamental group $\Gamma = \pi_1(V)$ has a representation ρ into the group $\mathrm{SL}_2(H)$, there is an induced action of Γ on the tree $T_{H,v}$ via the representation ρ. If the action of Γ on $T_{H,v}$ is nontrivial, it induces a splitting of $\Gamma = \pi_1(V)$ along an edge stabilizer. By Culler and Shalen, there is an essential surface associated to this action (see [4, Theorems 2.2.1 and 2.3.1]). The fundamental group of this associated essential surface is contained in an edge stabilizer. We say such an essential surface is detected by the representation ρ.

Whenever there is an element $\gamma \in \Gamma$ with $v(\mathrm{tr}(\rho\gamma)) < 0$, the action of Γ on $T_{H,v}$ is nontrivial. In particular, consider a representation $\rho\colon \Gamma \to \mathrm{SL}_2(H)$ where H is an algebraic number field and such that there is some element $\gamma \in \Gamma$ with $\mathrm{tr}(\rho\gamma)$ not an algebraic integer. We call such a representation an *algebraic nonintegral representation*. Then there is some prime ideal $\mathcal{P}$ in $\mathcal{O}_v$ such that $v_{\mathcal{P}}(\mathrm{tr}(\rho(\gamma))) < 0$.

The following lemma is a restatement of Corollary 3 of [16]. It describes how to determine the slope detected by a representation with an algebraic nonintegral character.

Lemma 2.1 *Let V be a hyperbolic knot exterior and $\rho\colon \pi_1(V) \to \mathrm{SL}_2(k)$ a representation, where k is a number field. If $\chi_\rho(\gamma)$ is not an algebraic integer for some slope $\gamma \in \partial V$ but $\chi_\rho(\delta)$ is an algebraic integer for another slope $\delta \in \pi_1(\partial V)$, then χ_ρ detects an essential surface S in V with boundary slope δ.*

3 Character variety of $J(2n, 2n)$

3.1 Character varieties: the standard model

Consider the knot $J(2n, 2n)$ with knot group presentation as in (2-1). The generators a and b are conjugate in the group, so $t_a = t_b = t_{b^{-1}}$. We may take t_a and $t_{ab^{-1}}$ as the generators for the ring R defined in Section 2.1 and also as coordinates in $X(\Gamma_n)$ (see [4, Proposition 1.4.1]).

A nonabelian representation $\rho_0 \in R(\Gamma_n)$ with $t_a(\rho_0) = x$ and $t_{ab^{-1}}(\rho_0) = r$ is conjugate in $\mathrm{SL}_2\mathbb{C}$ to a representation ρ with $A =: \rho(a)$ and $B =: \rho(b)$ given by

(3-1) $$A = \begin{pmatrix} \mu & 1 \\ 0 & \mu^{-1} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \mu & 0 \\ 2-r & \mu^{-1} \end{pmatrix}.$$

This ρ_0 is reducible exactly when $r = 2$.

We set $x = \mathrm{tr}(A)$ and $r = \mathrm{tr}(AB^{-1})$. Choose μ for which $x = \mu + \mu^{-1}$ and let $W_n = (AB^{-1})^n (A^{-1}B)^n$. Then an assignment of μ and r extends to a representation if and only if $AW_n^n = W_n^n B$. This results in four equations in μ and r, one for each matrix coordinate. However, the vanishing set of these four equations can be defined by a single equation in x and r which is independent of the choice between μ and μ^{-1} (see [14, Theorem 1]).

Definition 3.1 Let $f_0(u) = 0$, $f_1(u) = 1$ and define $f_{j+1}(u) = u \cdot f_j(u) - f_{j-1}(u)$. Define also $g_j(u) = f_j(u) - f_{j-1}(u)$.

The variety $X(\Gamma_n)$ is defined as the vanishing set of the polynomial

$$f_n(t)(f_n(r)g_n(r)(-x^2+2+r)-1)+f_{n-1}(t), \tag{3-2}$$

where

$$t = \operatorname{tr}(W_n) = (2-r)(x^2-2-r)f_n^2(r)+2 \tag{3-3}$$

and $W_n = (AB^{-1})^n(A^{-1}B)^n$ (see [9, Proposition 3.8]).

The variety $X(\Gamma_n)$ is an affine algebraic curve. Also, it is the double cover of the variety $Y(\Gamma_n)$ with variables (r, y) via the covering map

$$X(\Gamma_n) \to Y(\Gamma_n), \quad (r, x) \mapsto (r, x^2-2), \tag{3-4}$$

(see [9, Section 2.2.2]).

Affine algebraic curves may be completed naturally into projective curves by homogenizing their defining polynomials. Therefore, we may think of $X(\Gamma_n)$ and $Y(\Gamma_n)$ as projective curves in $\mathbb{C}P^2$ composed of an affine part and finitely many points of completion, that is, ideal points.

3.2 Character varieties: the smooth model

The varieties $X(\Gamma_n)$ and $Y(\Gamma_n)$ for $J(2n, 2n)$ are not smooth at infinity. To get around this, a new projective model $D(\Gamma_n)$ was introduced in [9]. This new model is birationally equivalent to $Y(\Gamma_n)$ and each of its irreducible components is smooth.

Let $D(\Gamma_n)$ be the projective closure of the affine variety in the coordinates $r = t_{ab^{-1}}$ and $t = t_w$. It is the vanishing set of the polynomial

$$g_{n+1}(r)g_n(t) - g_n(r)g_{n+1}(t). \tag{3-5}$$

Theorem 3.2 *For $n \geq 2$ in $\mathbb{Z}$ the following statements hold:*

(1) *$D(\Gamma_n)$ is birationally equivalent to $Y(\Gamma_n)$ via the map*

$$Y(\Gamma_n) \to D(\Gamma_n), \quad (r, y) \mapsto (r, (2-r)(y-r)f_n^2(r)+2). \tag{3-6}$$

(2) *$Y(\Gamma_n)$ and $D(\Gamma_n)$ are isomorphic outside a finite number of points (r, y) in $Y(\Gamma_n)$ given by $(r-2) f_n(r) = 0$.*

(3) *$D(\Gamma_n)$ consists of two irreducible components: the component D_0 defined by the line $r - t$ and the component D_1 defined as the projective closure of the complement of D_0. Furthermore, each component is smooth.*

(4) *$Y(\Gamma_n)$ has two irreducible components: the canonical component Y_0 and the component Y_1. Furthermore, Y_0 is birationally equivalent to D_0 and Y_1 to D_1.*

(5) *$X(\Gamma_n)$ has two irreducible components: the canonical component X_0 and the component X_1. Furthermore, X_0 is the double cover of Y_0 and X_1 the double cover of Y_1* (see (3-4)).

Proof These statements are given in Propositions 4.4 and 4.6 of [9]. □

3.3 Intersections between components

In this section we describe a polynomial G_n which determines the r–coordinate of the intersection points of D_0 and D_1. We show G_n also determines the r–coordinate in the affine intersection points of Y_0 and Y_1 via the birational equivalence, and thus also for the affine intersection points of X_0 and X_1.

Definition 3.3 Let $g_i' = dg_i/du$ and define

(3-7) $$G_j = g_{j+1}' g_j - g_{j+1} g_j'.$$

Lemma 3.4 *For $j \geq 2$ in $\mathbb{Z}$ the following statements hold:*

(1) *f_j is monic, separable and of degree $n-1$.*

(2) *$(u+2)G_j = f_{2j} + 2j$.*

(3) *G_j is monic and of degree $2j-2$.*

(4) *G_j and f_j do not share a root.*

(5) *$f_{2j} = uf_j^2 - 2f_j f_{j-1}$.*

(6) *$f_j(2) = j$.*

Proof Proofs for (1), (2) and (3) are found in [9, Lemmas 3.3 and 5.4] but we gather the necessary information here for convenience. Recall that $f_0 = 0$, $f_1 = 1$ and $f_j = u \cdot f_{j-1} - f_{j-2}$. It follows by induction on j that f_j is monic and of degree $j-1$. To prove separable consider the ring $\mathbb{Z}[u][s]/(s^2 - us + 1) \cong \mathbb{Z}[s, s^{-1}]$. Here

$u = s + s^{-1}$ and it follows that $f_j(u) = (s^j - s^{-j})/(s - s^{-1})$ by induction on j. Consider the following polynomial in $\mathbb{Z}[u][s]/(s^2 - us + 1)$:

$$(s^{j+1} - s^{j-1}) f_j = (s^{j+1} - s^{j-1}) \frac{s^j - s^{-j}}{s^1 - s^{-1}} = (s^{2j} - 1) \frac{s^1 - s^{-1}}{s^1 - s^{-1}} = s^{2j} - 1.$$

Since $s^{2j} - 1$ is separable, so is f_j, and this proves (1).

The identity in (2) follows directly from the definition of G_j by considering its image in the ring $\mathbb{Z}[u][s]/(s^2 - us + 1)$.

By (1), f_{2j} is monic and of degree $2j - 1 \geq 3$. Using the identity in (2), G_j is also monic and is of degree $(2j - 1) - 1 = 2j - 2$, proving (3).

Suppose ω is a root of G_j and pick $\sigma \in \mathbb{C}^*$ such that $\omega = \sigma + \sigma^{-1}$. By (2), $0 = f_{2j}(\omega) + 2j$, so then

$$-2j = f_{2j}(\omega) = \frac{\sigma^{2j} - \sigma^{-2j}}{\sigma - \sigma^{-1}} = (\sigma^j + \sigma^{-j}) \frac{\sigma^j - \sigma^{-j}}{\sigma - \sigma^{-1}} = (\sigma^j + \sigma^{-j}) f_j(\omega).$$

Then (4) follows since this implies $f_j(\omega) \neq 0$.

The identity in part (5) follows by considering $f_j(u) = (s^j - s^{-j})/(s + s^{-1})$ in the ring $\mathbb{Z}[u][s]/(s^2 - us + 1)$ to get

$$f_{2j} = f_j(f_{j+1} - f_{j-1})$$

and using the recursive definition for f_{j+1}.

The identity in part (6) follows from induction on j. □

We can now describe the points of intersection between D_0 and D_1.

Lemma 3.5 *If the point $P = (r_0, t_0)$ is in the intersection of $D_0(\Gamma_n)$ and $D_1(\Gamma_n)$, then P satisfies $G_n(r_0) = G_n(t_0) = 0$.*

Proof A similar statement is included in [9, Lemma 5.5]. We include a complete proof for the relevant case.

In the ring $\mathbb{Z}[u][s]/(s^2 - us + 1)$,

$$f_n^2 - f_{n-1} f_{n+1} = \frac{(s^n - s^{-n})^2}{(s - s^{-1})^2} - \frac{(s^{n-1} - s^{1-n})(s^{n+1} - s^{-n-1})}{(s - s^{-1})^2} = 1.$$

As a polynomial, we have that $f_n^2 - f_{n-1} f_{n+1} = 1$.

Recall $F = g_{n+1}(r)g_n(t) - g_n(r)g_{n+1}(t)$ is the defining polynomial for $D(\Gamma_n)$. Since P is a point in the intersection of two components, it is a singular point of $D(\Gamma_n)$. Therefore $F_r := \partial F/\partial r(P) = 0$ and $F(P) = 0$. We can then easily check that

$$0 = g_n(r_0)F_r(P) - g_n'(r_0)F(P) = g_n(t_0)G_n(r_0)$$

and

$$0 = g_{n+1}(r_0)F_r(P) - g_{n+1}'(r_0)F(P) = g_{n+1}(t_0)G_n(r_0).$$

If $G_n(r_0) \neq 0$ then $g_n(t_0) = g_{n+1}(t_0) = 0$. However,

$$f_n g_n - f_{n-1} g_{n+1} = f_n^2 - f_{n-1} f_{n+1} = 1$$

implies g_n and g_{n+1} are relatively prime polynomials and contradicts $g_n(t_0) = g_{n+1}(t_0) = 0$. Thus $G_n(r_0) = 0$ and also $G_n(t_0) = 0$. □

Lemma 3.6 *The affine parts of X_0 and X_1 are smooth. Furthermore, their affine intersection points correspond to irreducible representations and are determined 2-to-1 with the intersection points of D_0 and D_1.*

Proof From Theorem 3.2, D_0 and D_1 are smooth and isomorphic to Y_0 and Y_1 outside of the points with $(r-2)f_n(r) = 0$. This implies that all the affine points of Y_0 and Y_1, and equivalently of X_0 and X_1 outside of the points with $(r-2)f_n(r) = 0$ are smooth. The affine part of $Y(\Gamma)$ does not contain points (r, y) satisfying $f_n(r) = 0$. Similarly, the affine part of $X(\Gamma)$ does not contain points (r, x) satisfying $f_n(r) = 0$.

Let $F = f_n(t(r,x))(f_n(r)g_n(r)(-x^2+2+r)-1) + f_{n-1}(t(r,x))$ with $t(r,x) = (2-r)(x^2-2-r)f_n^2(r)+2$, the defining polynomial for X as in (3-2). Suppose (r_0, x_0) is a point in the affine part of X with $r_0 = 2$. Then $t(r_0, x_0) = 2$ and $x_0^2 = (4n^2-1)/n^2$. In particular, $x_0 \neq 0$. Let $F_x = \partial F/\partial x$. Then

$$F_x = \frac{\partial f_n(t)}{\partial t}\frac{dt}{dx}(f_n(r)g_n(r)(-x^2+2+r)-1) + f_n(t)f_n(r)g_n(r)(-2x) + \frac{\partial f_{n-1}(t)}{\partial t}\frac{dt}{dx}$$

with

$$\frac{dt}{dx} = (2-r)f_n^2(r)(x).$$

Evaluating at (r_0, x_0) we get $dt/dx(r_0, x_0) = 0$ and

$$\begin{aligned} F_x(r_0, x_0) &= 0 + f_n(2)f_n(2)g_n(2)(-2x_0) + 0 \\ &= (n)(n)(n-n+1)(-2x_0) \qquad \text{by Lemma 3.4(6)} \\ &= -2n^2 x_0 \\ &\neq 0. \end{aligned}$$

Therefore (r_0, x_0) is a smooth point in X. In particular, (r_0, x_0) is not an intersection point of X_0 and X_1.

If (r_1, x_1) is a point in the affine intersection of X_0 and X_1 then $r_1 \neq 2$, so it corresponds to an irreducible representation. This point (r_1, x_1) maps to a point in the intersection of D_0 and D_1 via the map $X(\Gamma_n) \to Y(\Gamma_n) \to D(\Gamma_n)$ (the covering map composed with the birational equivalence). □

4 Proof of the main result

Theorem 1.2 will follow from Theorems 4.1 and 4.2.

4.1 Surface detection

In this section we show that the intersection points detect essential surfaces.

Theorem 4.1 *Every intersection point in $X(\Gamma_n)$ detects an essential surface in the complement of $J(2n, 2n)$ in S^3.*

Proof The work of Culler and Shalen (see [4, Theorem 2.2.1 and Proposition 2.3.1]) shows that any ideal point in the character variety will give rise to an embedded essential surface in the knot exterior. Thus we need only consider the affine intersection points.

By Lemma 3.6, any affine intersection point (r_0, x_0) of X_0 and X_1 maps to the point (r_0, r_0) in the intersection of D_0 and D_1, since D_0 is defined by the line $r - t$ (see Theorem 3.2(3)). Therefore, by Lemma 3.5, $G_n(r_0) = 0$. Notice that the defining polynomials for D_0 and D_1 have degree 1 and $2n-2$, respectively, so by Bezout's theorem for smooth algebraic curves, they have $2n-2$ distinct intersections points. Since G_n is of degree $2n-2$ (see Lemma 3.4(3)), it must be that the roots of G_n exactly determine the intersection points of D_0 and D_1. We will now show that the x_0 at each intersection point (r_0, x_0) is not an algebraic integer. It will then follow that the affine intersection points detect essential surfaces (see Section 2.5).

Let $H = \mathbb{Q}(r_0, \mu_0)$ and v a valuation on H with $v(\pi) = 1$ and for some uniformizer π over the prime 2 and μ_0 an eigenvalue of A. Let $\mathfrak{p}$ be the prime associated with v and $\mathbb{F}_{\mathfrak{p}}$ its residue field. Then $\mathbb{F}_{\mathfrak{p}}$ has characteristic 2. Combining Lemmas 3.4(2) and 3.4(5), we get

$$(u+2)G_n = u f_n^2 - 2 f_n f_{n-1} + 2n.$$

The reduction of this equation to $\mathbb{F}_\mathfrak{p}$ shows $G_n = f_n^2$ over $\mathbb{F}_\mathfrak{p}$. Evaluating at r_0 gives $0 = f_n^2(r_0)$ in $\mathbb{F}_\mathfrak{p}$. Therefore $v(f_n^2(r_0)) > 0$, so $f_n^2(r_0)$ is not a unit and thus $1/f_n^2(r_0)$ is not an algebraic integer. Combining (3-3) with $t_0 = r_0$ we get that

$$x_0^2 = 2 + r_0 - \frac{1}{f_n^2(r_0)} \tag{4-1}$$

is not an algebraic integer. □

4.2 Detected slope

In this section we determine the slope of the detected surfaces and prove the following theorem.

Theorem 4.2 *The affine intersection points in $X(\Gamma_n)$ detect a Seifert surface.*

The following trace identities for $M_1, M_2 \in \mathrm{SL}_2\mathbb{C}$ follow from Cayley–Hamilton:

$$\mathrm{tr}(M_1) = \mathrm{tr}(M_1^{-1}), \tag{4-2}$$

$$\mathrm{tr}(M_1 M_2) = \mathrm{tr}(M_2 M_1), \tag{4-3}$$

$$\begin{aligned} \mathrm{tr}(M_1 M_2) &= (\mathrm{tr} M_1)(\mathrm{tr} M_2) - \mathrm{tr}(M_1^{-1} M_2) \\ &= (\mathrm{tr} M_1)(\mathrm{tr} M_2) - \mathrm{tr}(M_1 M_2^{-1}). \end{aligned} \tag{4-4}$$

The following identities follows from the previous identities by induction:

$$\begin{aligned} \mathrm{tr}(M_1^k) &= \mathrm{tr}(M_1) f_k(\mathrm{tr}(M_1)) - 2 f_{k-1}(\mathrm{tr}(M_1)) \\ &= f_{k+1}(\mathrm{tr}(M_1)) - f_{k-1}(\mathrm{tr}(M_1)), \end{aligned} \tag{4-5}$$

$$\begin{aligned} \mathrm{tr}[M_1, M_2] &= \mathrm{tr}(M_1 M_2 M_1^{-1} M_2^{-1}) \\ &= \mathrm{tr}^2(M_1) + \mathrm{tr}^2(M_2) + \mathrm{tr}^2(M_1 M_2) - \mathrm{tr}(M_1)\mathrm{tr}(M_2)\mathrm{tr}(M_1 M_2) - 2. \end{aligned} \tag{4-6}$$

Lemma 4.3 *Let S_1 and S_2 be the images of s_1 and s_2 at a representation corresponding to a point (r, x) in $X(\Gamma_n)$. The trace of $S_1 S_2^{-1}$ is given by*

$$f_n(r)(f_n(t)\delta_{1,1} - r f_{n-1}(t)) - f_{n-1}(r)(f_{n+1}(t) - f_{n-1}(t)) \tag{4-7}$$

with

$$t = (2 - r)(x^2 - 2 - r) f_n^2(r) + 2$$

and

$$\delta_{1,1} = \mathrm{tr}(W_n B A^{-1}) = (2 - r) f_{n-1}(r) f_n(r)(x^2 - 2 - r) + r.$$

Proof Recall that $\operatorname{tr}(S_1) = \operatorname{tr}(W_n^n)$, where $\operatorname{tr}(W_n) = t$ and $\operatorname{tr}(S_2^{-1}) = \operatorname{tr}((BA^{-1})^n)$, where $\operatorname{tr}(BA^{-1}) = \operatorname{tr}(AB^{-1}) = r$.

Set $\delta_{d,e} = \operatorname{tr}(W_n^d(BA^{-1})^e)$ and

$$\gamma_{d,e} = f_e(r)(f_d(t)\delta_{1,1} - rf_{d-1}(t)) - f_{e-1}(r)(f_{d+1}(t) - f_{d-1}(t)).$$

The statement of the lemma is equivalent to $\delta_{d,e} = \gamma_{d,e}$ in the case $d = n$ and $e = n$.

We have

$$\begin{aligned}
\delta_{d,0} &= \operatorname{tr}(W_n^d) \\
&= \operatorname{tr}(W_n) f_d(\operatorname{tr}(W_n)) - 2f_{d-1}(\operatorname{tr}(W_n)) \quad \text{by (4-5)} \\
&= tf_d(t) - 2f_{d-1}(t) \\
&= f_{d+1}(t) - f_{d-1}(t) \\
&= \gamma_{d,0}
\end{aligned}$$

and

$$\begin{aligned}
\delta_{0,e} &= \operatorname{tr}((BA^{-1})^e) \\
&= \operatorname{tr}(BA^{-1}) f_e(\operatorname{tr}(BA^{-1})) - 2f_{e-1}(\operatorname{tr}(BA^{-1})) \quad \text{by (4-5)} \\
&= \operatorname{tr}(AB^{-1}) f_e(\operatorname{tr}(AB^{-1})) - 2f_{e-1}(\operatorname{tr}(AB^{-1})) \quad \text{by (4-2)} \\
&= rf_e(r) - 2f_{e-1}(r) \\
&= \gamma_{0,e}.
\end{aligned}$$

Clearly $\gamma_{1,1} = \delta_{1,1}$, which is given by

$$\begin{aligned}
\delta_{1,1} &= \operatorname{tr}(W_n BA^{-1}) \\
&= \operatorname{tr}(BA^{-1} W_n) \quad \text{by (4-3)} \\
&= \operatorname{tr}(BA^{-1}(AB^{-1})^n(A^{-1}B)^n) \\
&= \operatorname{tr}((AB^{-1})^{n-1}(A^{-1}B)^{n-1}A^{-1}B) \\
&= \operatorname{tr}(W_{n-1}A^{-1}B) \\
&= (2-r)f_{n-1}(r)f_n(r)(x^2-2-r) + r \quad \text{by [9, Lemma 3.6].}
\end{aligned}$$

Notice that $\delta_{d,1}$ satisfies the recursion

$$\delta_{d,1} = \operatorname{tr}(W_n^d BA^{-1}) = \operatorname{tr}(W_n)\operatorname{tr}(W_n^{d-1}BA^{-1}) - \operatorname{tr}(W_n^{d-2}BA^{-1}) = t\delta_{d-1,1} - \delta_{d-2,1},$$

as does

$$\begin{aligned}
\gamma_{d,1} &= f_d(t)\delta_{1,1} - rf_{d-1}(t) \\
&= tf_{d-1}(t)\delta_{1,1} - f_{d-2}(t)\delta_{1,1} - rtf_{d-2}(t) + rf_{d-3}(t) \\
&= t(f_{d-1}(t)\delta_{1,1} - rf_{d-2}(t)) - (f_{d-2}(t)\delta_{1,1} + rf_{d-3}(t)) \\
&= t\gamma_{1,1} - \gamma_{2,1}.
\end{aligned}$$

Also notice that $\delta_{d,e}$ satisfies the recursion

$$\begin{aligned}\delta_{d,e} &= \operatorname{tr}(W_n^d(BA^{-1})^e)\\ &= \operatorname{tr}(W_n^d(BA^{-1})^{e-1}BA^{-1})\\ &= \operatorname{tr}(W_n^d(BA^{-1})^{e-1})\operatorname{tr}(BA^{-1}) - \operatorname{tr}(W_n^d(BA^{-1})^{e-2})\\ &= r\delta_{d,e-1} - \delta_{d,e-2},\end{aligned}$$

as does

$$\begin{aligned}\gamma_{d,e} &= f_e(r)(f_d(t)\delta_{1,1} - rf_{d-1}(t)) - f_{e-1}(r)(f_{d+1}(t) - f_{d-1}(t))\\ &= (rf_{e-1}(r) - f_{e-2}(r))(f_d(t)\delta_{1,1} - rf_{d-1}(t))\\ &\qquad\qquad - (rf_{e-2}(r) - f_{e-3}(r))(f_{d+1}(t) - f_{d-1}(t))\\ &= r\gamma_{d,e-1} - \gamma_{d,e-2}.\end{aligned}$$

Since the equivalence is satisfied for $\delta_{0,0}$, $\delta_{1,0}$, $\delta_{0,1}$ and $\delta_{1,1}$, this completes the proof. □

Lemma 4.4 *If S is a connected essential surface in the exterior of $J(2n,2n)$ with slope zero, then S is a genus 1 Seifert surface.*

Proof Recall from Section 2.3 that the knots $J(2n,2n)$ have two-bridge normal form $(4n^2-1,2n)$. Using the language of [5], the unique continued fraction expansion for the knot $J(2n,2n) = K_{2n/(4n^2-1)}$ of the form $[a_1,-a_2,a_3,\dots,\pm a_k]$ as in [5, Figure 5] and the proceeding paragraph is given by $[2n-1,-1,2n-1]$. By [5, Theorem 1(c)] and the remarks on [5, page 229] and the top of [5, page 230], any essential surface is carried by a branched surface corresponding to a minimal edge path involving only the heavy lines in [5, Figure 5]. Following the remarks at the end of [5, page 229], there are four minimal edge paths. These correspond to the continued fraction expansions

$$[\underbrace{-2,\dots,-2}_{2n-2},-3,\underbrace{-2,\dots,-2}_{2n-2}],\quad [\underbrace{-2,\dots,-2}_{2n-1},2n-1],\quad [2n-1,\underbrace{-2,\dots,-2}_{2n-1}],\quad [2n,2n].$$

By [5, Proposition 2], the branched surfaces associated to these continued fraction expansions will carry essential surfaces of slopes determined solely by the continued fraction expansion. The corresponding slopes are $2-8n$, $-4n$, $-4n$ and 0, respectively.

Any connected surface of slope zero is therefore carried by the branched surface $\Sigma[2n,2n]$. By [5, Proposition 1(1)] and the remark directly following it, this surface is a single-sheeted orientable surface. Such an essential connected surface of slope zero is then isotopic to either $S_1(0)$ or $S_1(1)$ as constructed in [5, page 227]. It is easy to see from the construction that $S_1(0)$ and $S_1(1)$ are nonseparating surfaces. Therefore, a connected essential surface of slope zero in the exterior of $J(2n,2n)$ is a Seifert surface.

By [5, Corollary to Proposition 1], all essential Seifert surfaces for a two-bridge knot have the same genus. Since the Seifert surface described in Section 2.3 has genus 1, all essential Seifert surfaces for $J(2n,2n)$ also have genus 1. □

We can now prove Theorem 4.2.

Proof of Theorem 4.2 Suppose that (r_0,x_0) is a point in the affine intersection of X_0 and X_1 and recall that $t_0=r_0$ (see Theorem 3.2(3)). Let S_1 and S_2 be the images of s_1 and s_2 of a representation corresponding to (r_0,x_0). Recall from Section 2.3 that the preferred longitude is given by $(W_n)^n(W_n^*)^n = S_1S_2^{-1}S_1^{-1}S_2$.

By (4-6), $S_1S_2^{-1}S_1^{-1}S_2$ has trace

$$\begin{aligned}\mathrm{tr}(S_1S_2^{-1}S_1^{-1}S_2) &= \mathrm{tr}^2(S_1)+\mathrm{tr}^2(S_2)+\mathrm{tr}^2(S_1S_2^{-1})-\mathrm{tr}(S_1)\mathrm{tr}(S_2)\mathrm{tr}(S_1S_2^{-1})-2\\ &= t_0^2+r_0^2+\mathrm{tr}^2(S_1S_2^{-1})-t_0r_0\mathrm{tr}(S_1S_2^{-1})-2\\ &= 2r_0^2+(1-r_0^2)\mathrm{tr}^2(S_1S_2^{-1})-2.\end{aligned}$$

Since $\mathrm{tr}(W_n)=t_0=r_0$, $S_1=W_n^n$ has trace $r_0f_n(r_0)-2f_{n-1}(r_0)$ and, since $\mathrm{tr}(AB^{-1})=r_0$, also $S_2=(AB^{-1})^n$ has trace $r_0f_n(r_0)-2f_{n-1}(r_0)$. From Lemma 4.3, $S_1S_2^{-1}$ has trace

$$(2-r_0)(x_0^2-2-r_0)f_{n-1}f_n^3+r_0f_n^2-r_0f_{n-1}f_n-f_{n-1}f_{n+1}+f_{n-1}^2$$

evaluated at r_0. Since r_0 is an algebraic integer (see Lemmas 3.5 and 3.4(3)), it suffices to show that $x_0^2f_n^2(r_0)$ is an algebraic integer, guaranteeing the integrality of $\mathrm{tr}(S_1S_2^{-1})$. Applying (4-1), we get

$$x_0^2f_n^2(r_0)=\left(2+r_0-\frac{1}{f_n^2(r_0)}\right)f_n^2(r_0)=(2+r_0)f_n^2(r_0)-1.$$

which is an algebraic integer. The theorem now follows from Lemmas 2.1 and 4.4. □

Remark There are finitely many characters of reducible representations in $X(\Gamma_n)$. These are contained in X_0 and also detect the slope zero. To see this, let $(r_0,x_0)\in X$ correspond to a reducible representation ρ. Then $r_0=2$. Substituting $r_0=2$ at (3-3) we get $t_0=2$ and at (3-2) we get

$$x_0^2=\frac{4n^2-1}{n^2}, \tag{4-8}$$

which is not an algebraic integer. We may conjugate ρ so that $\rho(\Gamma)$ is generated by

$$\begin{pmatrix} \mu & 1 \\ 0 & \mu^{-1} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{pmatrix}.$$

Then the character ρ is the same as a character of a diagonal representation, which is abelian. Therefore the traces of the images of s_1, s_2 and $s_1 s_2^{-1}$ are all the same as the trace of the image of the identity, which is the integer 2.

5 Two examples

We consider in detail the first two knots in the family $J(2n, 2n)$, namely 7_4 $(n = 2)$ and $11a_{363}$ $(n = 3)$.

5.1 The first knot

The first knot in the family $J(2n, 2n)$ is the knot 7_4 of two-bridge normal form $(15, 11)$ with knot group

$$\Gamma_2 = \langle a, b : aw^2 = w^2 b \rangle,$$

where $w = ab^{-1}ab^{-1}a^{-1}ba^{-1}b$. The variety $X(\Gamma_2)$ is defined by the polynomial

$$(-1 + 2r^2 + r^3 - r^2x^2)(1 + 4r - 4r^2 - r^3 + r^4 - 2rx^2 + 3r^2x^2 - r^3x^2),$$

where the first factor defines the canonical component X_0 and the second factor defines the component X_1. These two curves intersect at 20 points, counting multiplicities. However, 16 of these correspond to 2 ideal points (each with multiplicity 8). The affine intersections points (r, x) are

$$\Big(1-i, \sqrt{3 - \tfrac{3}{2}i}\Big), \quad \Big(1-i, -\sqrt{3 - \tfrac{3}{2}i}\Big), \quad \Big(1+i, \sqrt{3 + \tfrac{3}{2}i}\Big), \quad \Big(1+i, -\sqrt{3 + \tfrac{3}{2}i}\Big),$$

each with multiplicity 1. The x–coordinates of these points are the four roots of the polynomial $4x^4 - 24x^2 + 45$. These algebraic nonintegral numbers determine the traces of the meridian.

Consider the representation $\rho\colon \Gamma_2 \to \mathrm{SL}_2\mathbb{C}$ given by

$$\rho(a) = \begin{pmatrix} \mu & 1 \\ 0 & \mu^{-1} \end{pmatrix} \quad \text{and} \quad \rho(b) = \begin{pmatrix} \mu & 0 \\ 1+i & \mu^{-1} \end{pmatrix}$$

corresponding to the point $\big(1-i, \sqrt{3 - \frac{3}{2}i}\big)$, with $\mu = \frac{1}{2}\big(\sqrt{-1 - \frac{3}{2}i} + \sqrt{3 - \frac{3}{2}i}\big)$. The image of the longitude is the matrix

$$\begin{pmatrix} 7 + 12i + 2\sqrt{-24 + 42i} & -8\sqrt{-3 - 6i} \\ 0 & 7 + 12i - 2\sqrt{-24 + 42i} \end{pmatrix}$$

with trace $14 + 24i$, an algebraic integer.

Remark It is easy to see that for the representations given by these affine intersection points, the restriction of the peripheral subgroup is faithful. The meridian and longitude are mapped to loxodromics with the same axis. However, since one has nonintegral trace and the other integral trace, these generate a nondiscrete $\mathbb{Z}^2$. This leads us to ask the following question: Could these representations be faithful? A positive answer would imply that the noncanonical component contains faithful representations, and in particular do not come from a quotient.

5.2 The second knot

The second knot in the family $J(2n,2n)$ is the knot $11a_{363}$ of two-bridge normal form $(35,29)$ with knot group

$$\Gamma_2 = \langle a, b : aw^3 = w^3 b\rangle,$$

where $w = ab^{-1}ab^{-1}ab^{-1}a^{-1}ba^{-1}ba^{-1}b$. The variety $X(\Gamma_3)$ is defined by the polynomial

$$\begin{aligned}&(1+r-4r^2-2r^3+2r^4+r^5-x^2+2r^2x^2-r^4x^2)\\ &\times\bigl(1+8r-40r^2-46r^3+110r^4+71r^5-113r^6-43r^7+54r^8+11r^9-12r^{10}-r^{11}+r^{12}\\ &\quad-8x^2-8rx^2+60r^2x^2+21r^3x^2-130r^4x^2-7r^5x^2+118r^6x^2-16r^7x^2\\ &\quad-46r^8x^2+12r^9x^2+6r^{10}x^2-2r^{11}x^2+4x^4-19r^2x^4+5r^3x^4+32r^4x^4\\ &\quad-15r^5x^4-22r^6x^4+15r^7x^4+4r^8x^4-5r^9x^4+r^{10}x^4\bigr),\end{aligned}$$

where the first factor defines the canonical component X_0 and the second factor defines the component X_1. These two curves intersect at 84 points, counting multiplicities. However, 76 of these correspond to 2 ideal points (with multiplicities 24 and 52). There are 8 affine intersections points (r,x), each with multiplicity 1. The r–coordinates are the four roots of the polynomial r^4-2r^3+3. The x–coordinates are the eight roots of the polynomial $144x^8-1424x^6+5160x^4-8400x^2+6125$. These algebraic nonintegral numbers determine the traces of the meridian.

Consider the representation $\rho\colon \Gamma_3 \to \mathrm{SL}_2\mathbb{C}$ given by

$$\rho(a) = \begin{pmatrix}\mu & 1\\ 0 & \mu^{-1}\end{pmatrix} \quad\text{and}\quad \rho(b) = \begin{pmatrix}\mu & 0\\ s & \mu^{-1}\end{pmatrix}$$

corresponding to one of the intersection points, with $\mu \approx 0.44228+0.601587i$ (an algebraic number of degree 8 over $\mathbb{Q}$) and $s \approx 2.60504+0.835079i$ (an algebraic integral of degree 4 over $\mathbb{Q}$). The image of the longitude has trace a root of the polynomial

$$\ell^4-212\ell^3+15768\ell^2-385360\ell+8647328$$

($\approx 95.247+42.4755i$), an algebraic integer.

6 Final remarks

6.1 Multiple components

Riley [15] describes three cases in which a noncanonical component of characters of irreducible representations can arise in the character variety of two-bridge knots. One way we get a noncanonical component is if there exists an epimorphism from the knot group onto another knot group. However, this is not the case for the knots $J(2n, 2n)$.

Claim 6.1 *There is no epimorphism from Γ_n onto another knot group.*

Proof The knot $J(2n, 2n)$ has Alexander polynomial

$$\Delta_n(t) = n^2t^2 + (1-2n^2)t + n^2.$$

Since its quadratic discriminant, $1-4n^2$, is negative, Δ_n is an irreducible integral polynomial.

Denote the knot $J(2n, 2n)$ by K and suppose there exists an epimorphism from Γ_n onto the knot group Γ' for some other knot K'. The Alexander polynomial of K' must divide Δ_n (see eg Remark (3) of [2, Proposition 1.11]) and, furthermore, K' is necessarily a two-bridge knot [2, Corollary 1.3]. However, two-bridge knots have nontrivial Alexander polynomials. Therefore K' must have the same Alexander polynomial $\Delta_n(t)$.

Let $\widetilde{M}$ and $\widetilde{M}'$ denote the infinite cyclic covers of $S^3 - J(2n, 2n)$ and $S^3 - K'$, respectively. Mayland [10] expressed the derived groups $\gamma(M)$ and $\gamma(M')$ of $\widetilde{M}$ and $\widetilde{M}'$ for any two-bridge knots as a union of parafree groups in such a way that [1, Proposition 2.1] applies to show $\gamma(M)$ and $\gamma(M')$ are residually torsion-free nilpotent. That is, $\gamma(M)_\omega \cong 1 \cong \gamma(M')_\omega$, where G_ω is the ω–term in the lower central series and ω is the first infinite cardinal.

Since the knots share the same Alexander polynomial, $H_1(\widetilde{M}) \cong H_1(\widetilde{M}')$. We can now apply a theorem of Stallings [19, Theorem 3,4] to the epimorphism $h\colon \pi_1(\widetilde{M}) \to \pi_1(\widetilde{M}')$ to conclude h is an isomorphism. Therefore, Γ_n and Γ' are isomorphic. □

Note that Claim 6.1 was also proved in [11, Proposition 3.1].

Another way in which noncanonical components of characters of irreducible representations can arise in the character variety is when the knot has a certain nice symmetry described by Ohtsuki. In particular, whenever a two-bridge knot has two-bridge normal form (α, β) with $\beta^2 \equiv 1 \bmod \alpha$ and $\beta \neq 1$, there is a diagram from which one

knot	(p,q)	detected slope	fibered	(p,q)–symmetry	epimorphism
7_4	$(15,11)$	0	no	yes	no
7_7	$(21,13)$	6	yes	yes	no
8_{11}	$(27,19)$	6	no	no	no
9_6	$(27,5)$	18	no	no	3_1
9_{17}	$(39,25)$	10	yes	yes	no

Table 1: Knots with two components of irreducible representations

can see an orientation-preserving involution. This involution induces a nontrivial action on the character variety. However, it fixes a neighborhood of the character of a holonomy representation. Therefore, there exists a noncanonical component containing characters of irreducible representations (see [12, Proposition 5.5]). Notice that the knots studied in this paper satisfy these conditions. They have two-bridge normal form $(4n^2-1, 4n^2-2n-1)$.

6.2 Other examples of two-bridge knots with two components

In Table 1 we list knots with crossing number at most 9 whose character variety contain exactly two components of irreducible components. For all of these, the intersection points are Galois conjugates and detect the same slope. The table includes the 2–bridge normal form, the detected slope, whether or not the knot is fibered or has the (p,q)–symmetry described in Section 6.1, and if there is an epimorphism from the knot group to another knot group. Whenever a knot is fibered, a Seifert surface cannot be detected by ideal nor by algebraic nonintegral points in the character variety.

In addition to these, the knot groups for the knots 10_5, 10_9 and 10_{32} are known to have epimorphisms onto the trefoil knot group. Indeed, the two-bridge knots 9_6, 10_5, 10_9 and 10_{32} are the only knots up to 10 crossings whose knot groups have epimorphisms to another two-bridge knot (see [7, Theorem 1.1]). The knot groups surject to the trefoil knot group in such a way that the peripheral subgroup is sent to the peripheral subgroup of the trefoil knot group. Since the noncanonical component of the character variety corresponds to the canonical component of the trefoil character variety, the detected slopes correspond to detected slopes of the trefoil knot. As a fibered knot, the only detected slope of the trefoil knot is 6, so the detected slopes for 9_6, 10_5, 10_9 and 10_{32} are multiples of 6.

6.3 Two-bridge knots with three components

One may also want to consider two-bridge knots with three distinct components of irreducible representations in the character variety. Two examples of these are the knots

9_{23} and 10_{40} with two-bridge normal forms $(45, 19)$ and $(75, 29)$, both of which satisfy the symmetry condition described above and provide epimorphisms to the trefoil group. These two knots each have character varieties with a canonical component, a distinct component corresponding to the symmetry condition, and a distinct component corresponding to the canonical component of the character variety of the trefoil (to see that the knot group has an epimorphism to the trefoil knot group, refer to [7, Theorem 1.1]). All pairwise intersection points between these three components are algebraic nonintegral with the trace of the meridian nonintegral by a prime over 2, and correspond to irreducible representations. We note that the character variety of the knot 10_{40} has triple intersection points between these three components.

References

[1] **G Baumslag**, *Groups with the same lower central sequence as a relatively free group, II: Properties*, Trans. Amer. Math. Soc. 142 (1969) 507–538 MR

[2] **M Boileau**, **S Boyer**, **A W Reid**, **S Wang**, *Simon's conjecture for two-bridge knots*, Comm. Anal. Geom. 18 (2010) 121–143 MR

[3] **S Boyer**, **X Zhang**, *On Culler–Shalen seminorms and Dehn filling*, Ann. of Math. 148 (1998) 737–801 MR

[4] **M Culler**, **P B Shalen**, *Varieties of group representations and splittings of 3–manifolds*, Ann. of Math. 117 (1983) 109–146 MR

[5] **A Hatcher**, **W Thurston**, *Incompressible surfaces in 2–bridge knot complements*, Invent. Math. 79 (1985) 225–246 MR

[6] **J Hoste**, **P D Shanahan**, *A formula for the A–polynomial of twist knots*, J. Knot Theory Ramifications 13 (2004) 193–209 MR

[7] **T Kitano**, **M Suzuki**, *A partial order in the knot table*, Experiment. Math. 14 (2005) 385–390 MR

[8] **D D Long**, **A W Reid**, *Commensurability and the character variety*, Math. Res. Lett. 6 (1999) 581–591 MR

[9] **M L Macasieb**, **K L Petersen**, **R M van Luijk**, *On character varieties of two-bridge knot groups*, Proc. Lond. Math. Soc. 103 (2011) 473–507 MR

[10] **E J Mayland, Jr**, *Two-bridge knots have residually finite groups*, from "Proceedings of the Second International Conference on the Theory of Groups" (M F Newman, editor), Lecture Notes in Math. 372, Springer (1974) 488–493 MR

[11] **F Nagasato**, **M Suzuki**, **A T Tran**, *On minimality of two-bridge knots*, Internat. J. Math. 28 (2017) art. id. 1750020 MR

[12] **T Ohtsuki**, *Ideal points and incompressible surfaces in two-bridge knot complements*, J. Math. Soc. Japan 46 (1994) 51–87 MR

[13] **R Riley**, *Parabolic representations of knot groups, I*, Proc. London Math. Soc. 24 (1972) 217–242 MR

[14] **R Riley**, *Nonabelian representations of 2–bridge knot groups*, Quart. J. Math. Oxford Ser. 35 (1984) 191–208 MR

[15] **R Riley**, *Algebra for Heckoid groups*, Trans. Amer. Math. Soc. 334 (1992) 389–409 MR

[16] **S Schanuel**, **X Zhang**, *Detection of essential surfaces in 3–manifolds with* SL_2*–trees*, Math. Ann. 320 (2001) 149–165 MR

[17] **J-P Serre**, *Trees*, Springer (2003) MR

[18] **P B Shalen**, *Representations of 3–manifold groups*, from "Handbook of geometric topology" (R J Daverman, R B Sher, editors), North-Holland, Amsterdam (2002) 955–1044 MR

[19] **J Stallings**, *Homology and central series of groups*, J. Algebra 2 (1965) 170–181 MR

Department of Mathematics, The University of Texas
Austin, TX, United States

mchu@math.utexas.edu

Received: 21 September 2016 Revised: 14 April 2017

msp

Algebraic & Geometric Topology 17 (2017) 2915–2960

The surgery exact triangle in Pin(2)–monopole Floer homology

FRANCESCO LIN

We prove the existence of an exact triangle for the Pin(2)–monopole Floer homology groups of three-manifolds related by specific Dehn surgeries on a given knot. Unlike the counterpart in usual monopole Floer homology, only two of the three maps are those induced by the corresponding elementary cobordism. We use this triangle to describe the Manolescu correction terms of the manifolds obtained by (± 1)–surgery on alternating knots with Arf invariant 1.

57M27

Introduction

The goal of this paper is to describe the relation between the Pin(2)–monopole Floer homology groups of three-manifolds which are obtained from a given one by Dehn surgery on a knot. Pin(2)–monopole Floer homology is a gauge-theoretic invariant of closed connected and oriented three-manifolds. It was introduced by the author in [19] as the analogue of Manolescu's Pin(2)–equivariant Seiberg–Witten Floer homology groups for rational homology spheres (see Manolescu [22]) in the context of Kronheimer and Mrowka's monopole Floer homology [12]. In particular, it can be used to give an alternative disproof of the long standing triangulation conjecture; see also Manolescu [21] for a nice survey.

Unlike Manolescu's construction, the definition in [19] works for every closed oriented connected three-manifold Y. It associates in a functorial way to each such Y three groups fitting in a long exact sequence

$$\cdots \xrightarrow{i_*} \widecheck{HS}_\bullet(Y) \xrightarrow{j_*} \widehat{HS}_\bullet(Y) \xrightarrow{p_*} \overline{HS}_\bullet(Y) \xrightarrow{i_*} \cdots \tag{1}$$

which are read respectively *H-S-to*, *H-S-from* and *H-S-bar*. These are also (relatively) graded topological modules over the graded ring

$$\mathcal{R} = \mathbb{F}[\![V]\!][Q]/(Q^3),$$

where V and Q have degrees respectively -4 and -1, and $\mathbb{F}$ is the field with two elements. The ring $\mathcal{R}$ should be thought as the completion (with reverse gradings) of

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2915

the cohomology of the classifying space of

$$\mathrm{Pin}(2) = S^1 \times jS^1 \subset \mathbb{H},$$

where $\mathbb{H}$ denotes the quaternions. These objects are obtained by studying the negative gradient flow of the Chern–Simons–Dirac functional on the three-manifold from a Floer-theoretic point of view. For a general spin^c structure $\mathfrak{s}$, the equations have an S^1 symmetry, but when $\mathfrak{s}$ is induced by a genuine spin structure, quaternionic geometry comes into play and the equations acquire a Pin(2) symmetry. Pin(2)–monopole Floer homology is then constructed by suitably exploiting this extra input.

In the present paper, we develop in this setting one of the essential features of Floer homology theories for three-manifolds, namely *surgery exact triangles*. Their construction dates back to Floer's original instanton invariants [8], and they turn out to be a key tool for topological applications of Floer theories to three-manifold topology. For example, the version for monopole Floer homology is introduced by Kronheimer, Mrowka, Ozsváth and Szabó [13], and is used to prove a conjecture of Gordon [11] regarding a surgery characterization of the unknot.

Suppose we are given a connected compact oriented three-manifold Z with torus boundary ∂Z, and let γ_i, $i = 1, 2, 3$, be oriented simple closed curves having intersection numbers

$$\gamma_1 \cdot \gamma_2 = \gamma_2 \cdot \gamma_3 = \gamma_3 \cdot \gamma_1 = -1.$$

Call Y_i the three-manifold obtained by Dehn filling ∂Z along γ_i. Associated to this data there is a canonical cobordism W_i from Y_i to Y_{i+1} given by a single 2–handle attachment along a suitably framed copy of the knot. The key observation for our purposes is that, among these three cobordisms, exactly two are spin, while the third is not. The typical examples of such triples are given by

$$\infty, p, p+1 \quad \text{and} \quad 0, 1/(q+1), 1/q$$

surgeries on a knot in the three-sphere, where both p and q are integers. In the first case, the nonspin cobordism is W_1 if p is odd and W_3 if p is even. In the second case, the nonspin cobordism is always W_2. The following is then main result of the paper.

Theorem 1 *Suppose that the nonspin cobordism is W_3. There exists a map*

$$\check{F}_3\colon \widecheck{HS}_{\bullet}(Y_3) \to \widecheck{HS}_{\bullet}(Y_1)$$

of $\mathcal{R}$–modules such that the triangle

$$\begin{array}{ccc} \widecheck{HS}_{\bullet}(Y_2) & \xrightarrow{\widecheck{HS}_{\bullet}(W_2)} & \widecheck{HS}_{\bullet}(Y_3) \\ {\scriptstyle \widecheck{HS}_{\bullet}(W_1)} \nwarrow & & \swarrow {\scriptstyle \check{F}_3} \\ & \widecheck{HS}_{\bullet}(Y_1) & \end{array}$$

is exact. The map $\check{F}_3$ *is uniquely defined for each pair of three-manifolds* Y_3*,* Y_1 *such that the latter is obtained by Dehn surgery in the former, and in which the corresponding elementary cobordism given by a 2–handle attachment is not spin. The same statement holds for the* from *and* bar *versions.*

The map $\check{F}_3$ in the statement above is genuinely different from the one induced by the cobordism W_3 as defined in [19]. In fact, already in simple examples (see Section 5), the triangle where all the maps are the ones induced by cobordisms might have nonzero composite maps. This is due to the fact that in Pin(2)–monopole Floer homology, there are interesting modulo four periodicity phenomena to take account of; see, for example, the blow-up formula in Section 2. Unfortunately, we cannot provide a more geometric interpretation of this map at the moment.

Before discussing examples of the surgery exact triangle, we compute the Pin(2)–monopole Floer homology groups of the homology spheres obtained by surgery on the trefoil knot. As these are (up to orientation reversal) the Brieskorn spheres $\Sigma(2,3,6n\pm 1)$, we recover in our setting the results in [22]. In the statement, we adopt the following notation. For any rational number d, let $\mathcal{V}_d^+$ be the $\mathbb{F}[\![V]\!]$–module $\mathbb{F}[V^{-1},V]\!]/V\mathbb{F}[\![V]\!]$, where the grading is shifted so that the element 1 has degree d. Multiplication by V has degree -4, and the double square brackets indicate that we are considering a quotient of the ring of Laurent power series. We will refer to these $\mathbb{F}[\![V]\!]$–modules simply as *towers*, and they will arise as the image of the map i_* in the long exact sequence (1). We also denote a trivial $\mathcal{R}$–summand of the form $\mathbb{F}^k$ all concentrated in degree d by $\mathbb{F}^k\langle d\rangle$.

Theorem 2 *We have for* $k\geq 0$ *the isomorphisms of graded* $\mathcal{R}$*–modules*

$$\widecheck{HS}_\bullet(\Sigma(2,3,12k+5))\cong \mathcal{V}_4^+\oplus\mathcal{V}_3^+\oplus\mathcal{V}_2^+\oplus\mathbb{F}^k\langle 1\rangle,$$
$$\widecheck{HS}_\bullet(\Sigma(2,3,12k+1))\cong \mathcal{V}_2^+\oplus\mathcal{V}_1^+\oplus\mathcal{V}_0^+\oplus\mathbb{F}^k\langle -1\rangle,$$

where the action of Q *(which has degree* -1*) is an isomorphism from the first tower to the second tower, an isomorphism from the second tower to the third tower, and zero otherwise. The direct sum of the three towers is the image of* i_**. Similarly, for* $k>0$*, we have*

$$\widecheck{HS}_\bullet(\Sigma(2,3,12k-1))\cong \mathcal{V}_2^+\oplus\mathcal{V}_1^+\oplus\mathcal{V}_4^+\oplus\mathbb{F}^{k-1}\langle 1\rangle,$$
$$\widecheck{HS}_\bullet(\Sigma(2,3,12k-5))\cong \mathcal{V}_0^+\oplus\mathcal{V}_{-1}^+\oplus\mathcal{V}_2^+\oplus\mathbb{F}^{k-1}\langle -1\rangle,$$

where the action of Q *(which has degree* -1*) is an isomorphism from the first tower to the second tower, maps the second tower onto the third tower, and is zero otherwise. Again the direct sum of the three towers is the image of* i_**.*

The idea behind the computation is that if the usual monopole Floer homology $\widecheck{HM}_\bullet$ is simple enough, the $\mathrm{Pin}(2)$ counterpart can be determined in a purely algebraic way from the Gysin exact sequence

$$\cdots \xrightarrow{\cdot Q} \widecheck{HS}_k(Y) \xrightarrow{\iota_*} \widecheck{HM}_k(Y) \xrightarrow{\pi_*} \widecheck{HS}_k(Y) \xrightarrow{\cdot Q} \widecheck{HS}_{k-1}(Y) \xrightarrow{\iota_*} \cdots \tag{2}$$

introduced in Section 4.3 of [19]. This granted, we will discuss how the Floer groups of these manifolds fit in the surgery exact triangle, and this will provide a model for more interesting computations. Similarly, we compute the $\mathrm{Pin}(2)$–monopole Floer homology groups of the homology spheres obtained by surgery on the figure-eight knot.

Theorem 3 *Denote by E_n the manifold obtained by $1/n$ surgery on the figure-eight knot. Let $\mathfrak{s}_0$ be the only self-conjugate spin^c structure on E_0. Then we have the isomorphism of graded $\mathcal{R}$–modules*

$$\widecheck{HS}_\bullet(E_0,\mathfrak{s}_0) \cong \mathcal{V}_1^+ \oplus \mathcal{V}_0^+ \oplus \mathcal{V}_{-1}^+ \oplus \mathcal{V}_2^+,$$

where the action of Q (which has degree -1) is an isomorphism from the first tower to the second, maps the third tower onto the fourth tower, and zero otherwise. The group for the other spin^c structures vanishes, and the map i_* is surjective. Furthermore, we have for $k \geq 0$,

$$\widecheck{HS}_\bullet(E_{2k+1}) \cong \mathcal{V}_0^+ \oplus \mathcal{V}_{-1}^+ \oplus \mathcal{V}_2^+ \oplus \mathbb{F}^k\langle -1\rangle,$$

and similarly for $k > 0$,

$$\widecheck{HS}_\bullet(E_{2k}) \cong \mathcal{V}_2^+ \oplus \mathcal{V}_1^+ \oplus \mathcal{V}_0^+ \oplus \mathbb{F}^k\langle -1\rangle.$$

Here the action of Q is the same as the analogous modules appearing in Theorem 2, and the image of i_* consists exactly of the direct sum of the three towers.

Recall more generally that for any rational homology sphere Y equipped with a self-conjugate spin^c structure $\mathfrak{s}$, the group $\widecheck{HS}_\bullet(Y,\mathfrak{s})$, considered as an $\mathbb{F}[\![V]\!]$–module, decomposes as a finite part and the sum of three towers

$$\mathcal{V}_c^+ \oplus \mathcal{V}_b^+ \oplus \mathcal{V}_a^+.$$

The action of Q sends the first tower onto the second and the second tower onto the third, and the union of the sum of the three towers is the image of the map i_*. Manolescu's correction terms are then defined to be the rational numbers

$$\alpha(Y) \geq \beta(Y) \geq \gamma(Y),$$

all of which have the same fractional part, such that

$$a = 2\alpha(Y), \quad b = 2\beta(Y)+1, \quad c = 2\gamma(Y)+2.$$

The inequalities between these quantities follow from the module structure. These are rational lifts of the Rokhlin invariant of $(Y,\mathfrak{s})$ which are invariant under spin homology cobordisms, and they are integers in the case of a genuine homology sphere. The direct sum of the three towers is an $\mathcal{R}$–submodule whose abstract isomorphism class as an absolutely graded $\mathcal{R}$–module will be denoted by $\mathcal{S}^+_{\alpha,\beta,\gamma}$. We call this a *standard* Pin(2)*–module*. For example, the four direct sums of towers appearing in the statement of Theorem 2 are respectively

$$\mathcal{S}^+_{1,1,1},\quad \mathcal{S}^+_{0,0,0},\quad \mathcal{S}^+_{2,0,0},\quad \mathcal{S}^+_{1,-1,-1}.$$

We can use the surgery exact triangle to determine the Manolescu correction terms of integral homology spheres obtained by (±1)–surgery on alternating knots with Arf invariant 1.

Theorem 4 *Let K be an alternating knot with signature $\sigma=\sigma(K)\le 0$ and Arf invariant 1. Then the value of α, β and γ of the (±1)–surgery on it is determined in the following table:*

(3)

σ	$(+1)$–*surgery*	(-1)–*surgery*
$-8k$	$-2k+1,-2k-1,-2k-1$	$1,-2k+1,-2k-1$
$-8k-2$	$-2k-1,-2k-1,-2k-1$	$1,-2k-1,-2k-1$
$-8k-4$	$-2k-1,-2k-1,-2k-1$	$1,-2k-1,-2k-1$
$-8k-6$	$-2k-1,-2k-1,-2k-3$	$1,-2k-1,-2k-1$

The main idea is again that the Floer homology of surgeries on alternating knots is simple enough so that the Pin(2)–case can be recovered from the usual one by means of the Gysin exact sequence. Our computation relies on the knowledge of the monopole Floer homology of these spaces. This follows from the results in the context of Heegaard Floer homology provided by Ozsváth and Szabó [25] and the isomorphism between the two theories (due to Kutluhan, Lee and Taubes [14; 15; 16; 17; 18] and Colin, Ghiggini and Honda [3; 4; 5; 6]).

A nice consequence of this computation is the existence of homology spheres not homology cobordant to any Seifert fibered space. In Stoffregen [28], an example consisting of the connected sum of two Seifert fibered spaces is provided. We use the following analogous obstruction, which we prove (as in [28]) using the results of Mrowka, Ozsváth and Yu [23].

Proposition 5 *For a Seifert fibered rational homology sphere Y equipped with a spin structure $\mathfrak{s}$, either $\alpha(Y,\mathfrak{s})=\beta(Y,\mathfrak{s})$ or $\beta(Y,\mathfrak{s})=\gamma(Y,\mathfrak{s})$.*

As a consequence of Theorem 4, we obtain the following.

Corollary 6 *For $k \geq 1$, the manifold obtained by (-1)–surgery on an alternating knot with signature $-8k$ and Arf invariant 1 is not homology cobordant to any Seifert fibered space.*

Along the way, we will discuss maps on spin cobordisms with $b_2^+ = 1, 2$. We record the following result as we expect it to have interesting topological applications.

Theorem 7 *Let W be a smooth spin cobordism between two spin rational homology spheres $(Y_0, \mathfrak{s}_0)$ and $(Y_1, \mathfrak{s}_1)$. If $b_2^+(W) = 1$, then the inequalities*

$$\alpha(Y_1, \mathfrak{s}_1) \geq \beta(Y_0, \mathfrak{s}_0) + \tfrac{1}{8}(b_2^-(W) - 1),$$
$$\beta(Y_1, \mathfrak{s}_1) \geq \gamma(Y_0, \mathfrak{s}_0) + \tfrac{1}{8}(b_2^-(W) - 1)$$

hold. If $b_2^+(W) = 2$, then the inequality

$$\alpha(Y_1, \mathfrak{s}_1) \geq \gamma(Y_0, \mathfrak{s}_0) + \tfrac{1}{8}(b_2^-(W) - 2)$$

holds.

This should be thought as a generalization of Donaldson's Theorems B and C (see Donaldson and Kronheimer [7]) regarding closed spin four-manifolds with $b_2^+ = 1, 2$, in the same way as Frøyshov's result is a generalization of Donaldson's Theorem A. It is analogous in spirit to Kronheimer's Seiberg–Witten theoretic proof that inspired Furuta's work on the $\frac{11}{8}$–conjecture [10].

Acknowledgements The author would like to thank his advisor Tom Mrowka for the support throughout the development of this project, Tye Lidman and Ciprian Manolescu for the useful comments, and the referee for carefully reading the manuscript and suggesting many improvements. Part of the work was carried over while the author was visiting the University of Pisa. He would like to express his gratitude to them, and in particular Bruno Martelli, for their hospitality. This work was partially supported by NSF grants DMS-0805841 and DMS-1005288.

1 Overview of Pin(2)–monopole Floer homology

In this section, we provide an overview of the formal properties of Pin(2)–monopole Floer homology and discuss the key steps of its construction as in Chapter 4 of [19]. For a more detailed introduction to the usual case, see the first three chapters of [12]. Throughout the paper, we will always work with coefficients in the field with two elements $\mathbb{F}$.

The formal structure Let Y be a closed connected oriented three-manifold. There is a natural action $\jmath$ of the set of spin^c structures $\mathrm{Spin}^c(Y)$ given by complex conjugation. We denote the orbits by $[\mathfrak{s}]$, and call the fixed points of this action *self-conjugate spin^c–structures*. The first Chern class of such a spin^c structure $\mathfrak{s}$ is always two-torsion. To each self-conjugate spin^c structure $[\mathfrak{s}]$, we associate the (completed) Pin(2)–monopole Floer homology groups

$$\widecheck{HS}_\bullet(Y,\mathfrak{s}), \quad \widehat{HS}_\bullet(Y,\mathfrak{s}), \quad \overline{HS}_\bullet(Y,\mathfrak{s}).$$

There are analogous cohomological versions. These groups carry a relative $\mathbb{Z}$ grading and an absolute $\mathbb{Q}$ grading. They also carry a structure of topological graded module over the ring

$$\mathcal{R} = \mathbb{F}[\![V]\!][Q]/(Q^3),$$

where the actions of V and Q have degrees respectively -4 and -1. These groups should be thought as computing the middle-dimensional homology of an infinite-dimensional manifold with boundary $\mathcal{B}^\sigma(Y,\mathfrak{s})/\jmath$. In particular, they compute the homology of the space, the homology relative to the boundary and the homology of the boundary. Indeed, they fit in the expected long exact sequence (1), and they satisfy a version Poincaré duality with respect to orientation reversal.

In the case in which $\mathfrak{s}$ is not self-conjugate, we define $\widecheck{HS}_\bullet(Y,[\mathfrak{s}])$ to be the usual monopole Floer homology groups

$$\widecheck{HM}_\bullet(Y,\mathfrak{s}) \equiv \widecheck{HM}_\bullet(Y,\bar{\mathfrak{s}}),$$

which are canonically isomorphic. This can be thought as an $\mathcal{R}$–module via the coefficient extension

$$\mathcal{R} \to \mathbb{F}[\![U]\!]$$

obtained by sending V to U^2 and Q to 0. It is convenient to work with the direct sum of all these groups (all but finitely many of which are trivial), and define

$$\widecheck{HS}_\bullet(Y) = \bigoplus_{[\mathfrak{s}]\in\mathrm{Spin}^c(Y)/\jmath} \widecheck{HS}_\bullet(Y,[\mathfrak{s}]),$$

and similarly for the other versions. This total group does not carry a relative $\mathbb{Z}$ grading anymore, but has a canonical $\mathbb{Z}/2\mathbb{Z}$ grading.

The basic computation is the case of the three sphere. Recall that we have discussed in the introduction the basic modules $\mathcal{V}_d^+$ and $\mathcal{S}_{\alpha,\beta,\gamma}^+$. We then have

$$\widecheck{HS}_\bullet(S^3) = \mathcal{S}^+_{0,0,0} = \mathcal{V}_2^+ \oplus \mathcal{V}_1^+ \oplus \mathcal{V}_0^+,$$

where in the second description the action of Q (which has degree -1) is an isomorphism from the first tower to the second and from the second tower to the third. Also, for a rational number d the module $\mathcal{V}_d$ is defined to be the ring of Laurent power series $\mathbb{F}[V^{-1}, V]\!]$ seen as an $\mathbb{F}[\![V]\!]$–module with the grading shifted so that the element 1 has degree d. We then have the isomorphism of $\mathcal{R}$–modules

$$\overline{HS}_\bullet(S^3) = \mathcal{V}_2 \oplus \mathcal{V}_1 \oplus \mathcal{V}_0. \tag{4}$$

The action of Q is an isomorphism from the first summand to the second and from the second summand to the third. We will denote this absolutely graded $\mathcal{R}$–module by $\mathcal{S}$. The map

$$i_*\colon \overline{HS}_\bullet(S^3) \to \widecheck{HS}_\bullet(S^3)$$

is surjective. Finally $\widehat{HS}_\bullet(S^3)$ can be identified with $\mathcal{R}\langle -1\rangle$ where again the braces indicate the grading shift.

These groups also satisfy functoriality properties. If W is a connected cobordism from Y_0 to Y_1, we obtain a homomorphism of $\mathcal{R}$–modules

$$\widecheck{HS}_\bullet(W)\colon \widecheck{HS}_\bullet(Y_0) \to \widecheck{HS}_\bullet(Y_1),$$

which also decomposes according to the pairs of conjugate spin^c structures on W. Furthermore, if W_0 is a cobordism from Y_0 to Y_1 and W_1 is a cobordism from Y_1 to Y_2, we have the composition law

$$\widecheck{HS}_\bullet(W_1 \circ W_0) = \widecheck{HS}_\bullet(W_1) \circ \widecheck{HS}_\bullet(W_0).$$

A more general functoriality property (following from the work of [2]) is the following. Suppose W is a cobordism with several incoming ends Y_- and $Y_1, \dots, Y_n$ and one outgoing end Y_+. Then there is an induced map

$$\widecheck{HS}_\bullet(Y_-) \otimes \widehat{HS}_\bullet(Y_0) \otimes \cdots \otimes \widehat{HS}_\bullet(Y_n) \to \widecheck{HS}_\bullet(Y_+).$$

This holds also when the two *to* groups are replaced by their *from* and *bar* counterparts. As a special case of this general construction, the cobordism $I \times Y$ with a ball removed induces a map

$$\widecheck{HS}_\bullet(Y) \otimes \mathcal{R}\langle -1\rangle \to \widecheck{HS}_\bullet(Y),$$

where we think of the additional boundary component as an incoming S^3 end, and we use this map as the definition of the $\mathcal{R}$–module structure.

As briefly discussed in the introduction, in the case that Y is a rational homology sphere equipped with a self conjugate spin^c structure $\mathfrak{s}$ the groups have a rather simple structure. In particular, we have the isomorphism of *relatively* graded $\mathcal{R}$–modules

$$\overline{HS}_\bullet(Y, \mathfrak{s}) \cong \overline{HS}_\bullet(S^3) \cong \mathcal{S}.$$

Analogously the group $\widecheck{HS}_\bullet(Y,\mathfrak{s})$ is zero for degrees negative enough and the map i_* is an isomorphism in degrees high enough. This implies that for some $\alpha \geq \beta \geq \gamma$ rational numbers with the same fractional part we have that

$$i_*(\overline{HS}_\bullet(Y,\mathfrak{s})) \cong \mathcal{S}^+_{\alpha,\beta,\gamma}$$

as absolutely graded $\mathcal{R}$–modules. A very interesting special case is when Y is an actual integral homology sphere. In that case, the correction terms are integral lifts of the Rokhlin invariant, and are invariant under homology cobordism. Furthermore,

$$\beta(-Y) = -\beta(Y),$$

where $-Y$ denotes the manifold with opposite orientation, and using these properties one can show that the long standing triangulation conjecture is false; see [22; 19].

Construction of monopole Floer homology We quickly describe the construction of monopole Floer homology. Equip Y with a riemannian metric. A spin^c structure on Y is given by a rank two hermitian bundle $S \to Y$ together with a Clifford multiplication

$$\rho\colon TY \to \mathrm{Hom}(S,S),$$

ie a bundle map with the property that

$$\rho(v)^2 = -\|v\|^2 \mathrm{Id}_S \quad \text{for each } v \in TY.$$

We can define the configuration space $\mathcal{C}(Y,\mathfrak{s})$ consisting of pairs (B,Ψ) where

- B is a spin^c connection on S, ie a connection compatible with the Clifford action;
- Ψ is a spinor, ie a section of S.

The group of automorphisms of the Clifford bundle $\mathcal{G}(Y,\mathfrak{s})$ is given by the set of maps u from Y to S^1, and it acts on the configuration space as

$$u \cdot (B,\Psi) = (B - u^{-1} du, u \cdot \Psi).$$

A configuration with $\Psi \neq 0$, called *irreducible*, has trivial stabilizer, while a configuration $(B,0)$, called *reducible*, has stabilizer S^1 given by the constant gauge transformations. The functional on the configuration space used to define the Floer chain complex is the Chern–Simons–Dirac functional, given after a choice of a base spin^c connection B_0 by

$$\mathcal{L}(B,\Psi) = -\frac{1}{8}\int_Y (B^t - B_0^t) \wedge (F_{B^t} + F_{B_0^t}) + \frac{1}{2}\int_Y \langle D_B\Psi, \Psi\rangle \, d\mathrm{vol}.$$

Here B^t denotes the connection induced on the determinant line bundle $\bigwedge^2 S$, and D_B is the Dirac operator associated to B. When $c_1(\mathfrak{s})$ is not torsion, this functional is invariant only under the identity component of the gauge group, and is well defined with values in $\mathbb{R}/2\pi^2\mathbb{Z}$.

From here, the early approaches to Seiberg–Witten Floer homology went on considering the flow of the L^2 formal gradient of the functional $\mathcal{L}$, given by the formula

$$\operatorname{grad}\mathcal{L}(B,\Psi) = \big(\big(\tfrac{1}{2} * F_{B^t} + \rho^{-1}(\Psi\Psi^*)_0\big) \otimes 1_S, D_B\Psi\big)$$

on the space of irreducible configurations modulo gauge. This caused invariance issues, and the loss of important information carried by the reducibles. Kronheimer and Mrowka's approach solved this by passing to the blown-up configuration space $\mathcal{C}^\sigma(Y,\mathfrak{s})$. This consists of triples (B,r,ψ) where

- B is again a spin^c connection;
- r is a nonnegative real number;
- ψ is a spinor with unit L^2 norm.

This operation is essentially passing to polar coordinates in the spinor component, and there is a canonical blow-down map π given by

$$(B,r,\psi) \mapsto (B, r\psi).$$

The action of the gauge group on this space is free, and the quotient $\mathcal{B}^\sigma(Y,\mathfrak{s})$ is an infinite-dimensional manifold with boundary $\partial\mathcal{B}^\sigma(Y,\mathfrak{s})$ consisting of the preimage of the reducibles under the blow-up map. Furthermore, the gradient of the Chern–Simons–Dirac functional naturally extends to the blown-up configuration space as the vector field

$$\begin{aligned}(5)\quad &(\operatorname{grad}\mathcal{L})^\sigma(B,r,\psi)\\ &\quad = \big(\big(*\tfrac{1}{2}F_{B^t} + r^2\rho^{-1}(\psi\psi^*)_0\big)\otimes 1_S, \Lambda(B,r,\psi)r, D_B\psi - \Lambda(B,r,\psi)\big),\end{aligned}$$

where $\Lambda(B,r,\psi)$ is the real valued function $\langle D_B\psi,\psi\rangle_{L^2(Y)}$. Monopole Floer homology is then defined by applying the usual ideas of Morse–Witten homology to this vector field on the space $\mathcal{B}^\sigma(Y,\mathfrak{s})$.

There are a few observations to be made. First of all, the vector field $(\operatorname{grad}\mathcal{L})^\sigma$ is not the gradient of any natural function on $\mathcal{C}^\sigma(Y,\mathfrak{s})$ with respect to any natural metric. Furthermore, it is tangent to the boundary $\partial\mathcal{B}^\sigma(Y,\mathfrak{s})$, so the latter is invariant under the flow. This implies that there are two kinds of reducible critical points: the *stable* ones, for which the Hessian in the direction normal to the boundary is positive, and the *unstable* ones. The unstable manifold of a stable critical point is entirely contained in the boundary, and similarly for the stable manifold of an unstable critical point. In

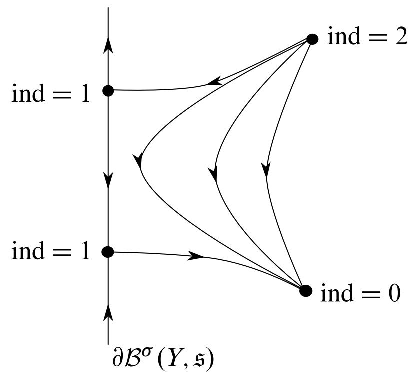


Figure 1: A one-dimensional family of trajectories limiting to a broken trajectory with three components

particular, a nonempty space of trajectories from a stable point to an unstable point is *never* transversely cut out. We say that such a pair is *boundary obstructed*. This implies for example that a one-dimensional family of unparametrized trajectories between two irreducible critical points may break into three components, the middle one being a boundary obstructed trajectory; see Figure 1.

Nevertheless, for generic perturbations one can define three chain complexes computing the three relevant homologies as follows. Denote by C^o, C^u and C^s the $\mathbb{F}$–vector spaces generated by the irreducible, unstable and stable critical points. One can define the linear maps

$$\partial^o_o\colon C^o \to C^o, \quad \partial^o_s\colon C^o \to C^s, \quad \partial^u_o\colon C^u \to C^o, \quad \partial^u_s\colon C^u \to C^s$$

obtained by counting irreducible trajectories in zero-dimensional moduli spaces between critical points of a specified type, and similarly

$$\bar\partial^s_s\colon C^s \to C^s, \quad \bar\partial^s_u\colon C^s \to C^u, \quad \bar\partial^u_s\colon C^u \to C^s, \quad \bar\partial^u_u\colon C^u \to C^u$$

obtained by counting reducible trajectories. Notice that the maps ∂^u_s and $\bar\partial^u_s$ count points in different moduli spaces. The three Floer chain complexes are then defined as the $\mathbb{F}$–vector spaces

$$\check{C}_* = C^o \oplus C^s, \quad \hat{C}_* = C^o \oplus C^u, \quad \bar{C}_* = C^s \oplus C^u$$

equipped respectively with the differentials

$$\check\partial = \begin{bmatrix} \partial^o_o & \partial^u_o\bar\partial^s_u \\ \partial^o_s & \bar\partial^s_s + \partial^u_s\bar\partial^s_u \end{bmatrix}, \quad \hat\partial = \begin{bmatrix} \partial^o_o & \partial^u_o \\ \bar\partial^s_u\partial^o_s & \bar\partial^s_u\partial^u_s \end{bmatrix}, \quad \bar\partial = \begin{bmatrix} \bar\partial^s_s & \bar\partial^s_u \\ \bar\partial^u_s & \bar\partial^u_u \end{bmatrix}.$$

The Floer homology groups are the homology groups of these complexes, and are independent (in a functorial way) of the choices made.

We can give a nice description of the critical points in the blowup in the generic case. Recall first that the compactness properties of the Seiberg–Witten equations imply that for a generic perturbation there are only finitely many critical points downstairs. The blow-down map is a diffeomorphism on the irreducibles, so we have a one-to-one correspondence in this case, so the irreducible critical points upstairs consist of a finite number of points. On the other hand, it is easy to see from (5) that each reducible critical point downstairs gives rise to a countable collection of stable and unstable critical points. Indeed, each of these corresponds to the quotient by the S^1–action of the unit sphere of an eigenspace of the Dirac operator. These are generically all (complex) one-dimensional, and the eigenvalues form a discrete sequence of real numbers which is infinite in both directions. The computation of the Floer groups for S^3 then readily follows because the positive scalar curvature implies that there are no irreducible critical points for perturbations small enough.

Finally, given a cobordism W between Y_0 and Y_1, the induced map is defined by counting solutions to the four-dimensional Seiberg–Witten equations (in the blowup of the configuration space) on the manifold W^* obtained by adding cylindrical ends. This generally involves infinitely many moduli spaces, and their sum only makes sense after a suitable completion of the groups with respect to some negative filtration, which we denote using the bullet. The map induced by W and the class $U^d \in \mathbb{F}[\![U]\!]$ is defined by considering the manifold W_p^* with an additional incoming end of the form $(-\infty, 0] \times S^3$ obtained by removing a ball (such that the metric is a product near the boundary) and adding a cylindrical end. We can choose the metric such that S^3 has positive scalar curvature, and the map is obtained by counting the solutions to the Seiberg–Witten equations on W_p^* that converge to the d^{th} unstable critical point at this additional end.

The case of a self-conjugate spin^c structure In the case that the spin^c structure is self-conjugate, or equivalently the spin^c structure is induced by a genuine spin structure, we have two extra features:

- a preferred base connection B_0 with B_0^t flat, the spin connection induced by the Levi–Civita connection;
- a quaternionic structure j on the Clifford bundle S, ie a complex antilinear automorphism such that $j^2 = -\text{Id}$.

The latter follows from the observation that $\text{Spin}(3)$ can be identified with $\text{SU}(2)$ and the spinor representation is just the usual action on $\mathbb{C}^2 = \mathbb{H}$. We will write the action of j and the multiplication by complex numbers from the right. These two features are compatible in the sense that the Dirac operator D_{B_0} is quaternionic linear. In this case, the configuration space $\mathcal{C}(Y,\mathfrak{s})$ comes with a diffeomorphism $\jmath$ given by

$$\jmath \cdot (B_0 + b, \Psi) = (B_0 - b, \Psi \cdot \jmath).$$

This induces an involution on the moduli space of configurations $\mathcal{B}(Y,\mathfrak{s})$. Its only fixed points are the equivalence classes $[B,0]$ with B the spin connection of a spin structure inducing the spin^c structure $\mathfrak{s}$. There are $2^{b_1(Y)}$ such spin structures: for example on $S^2 \times S^1$, the two spin structures both induce the only torsion spin^c structure. Similarly, there is an induced involution (still denoted by $\jmath$) on the blown-up moduli space of configurations $\mathcal{B}^\sigma(Y,\mathfrak{s})$ which is fixed point-free.

The main idea is then to do Floer theory in a $\jmath$–invariant fashion, and compute the homology of the quotient $\mathcal{B}^\sigma(Y,\mathfrak{s})/\jmath$. This is done by producing a (Morse–Bott) chain complex with a natural $\mathbb{Z}/2\mathbb{Z}$–action, and consider the homology of the invariant subchain complex. To do this we need to restrict to perturbations that preserve the symmetry of our set up. The generic picture will then be the following. Irreducible critical points are not fixed points of the action downstairs, so they will still constitute a finite number of points, and furthermore they come in pairs related by the action of $\jmath$. The same thing happens for reducible critical points for which the connection is not spin, with the action of $\jmath$ exchanging the two towers of reducible critical points. In the case when the connection is spin, the perturbed Dirac operator will still be quaternionic, hence the eigenspaces will always be even-dimensional over the complex numbers. In particular, the perturbation will not be regular in the usual sense. Nevertheless, generically the eigenspaces will all be two-dimensional, and we will obtain a copy of S^2 as critical submanifold for each of them. This is the quotient of the unit sphere of the eigenspace by the action of S^1, which is just a Hopf fibration. Finally, the action of $\jmath$ on S^2 is just the antipodal map.

For a generic perturbation, the critical manifolds described above will be Morse–Bott, meaning that the Hessian is nondegenerate in the normal directions. The three Morse–Bott chain complexes computing the homology introduced in Chapter 3 of [19] are defined following the framework of [9]. The underlying vector space is the direct sum of some variants of the singular chain complexes of the critical submanifolds. The differential combines the singular differential together with terms involving different critical submanifolds. In particular, given critical submanifolds $[\mathfrak{C}_\pm]$, there are evaluation maps on the compactified moduli spaces of trajectories connecting them,

$$\mathrm{ev}_\pm\colon \breve{M}^+([\mathfrak{C}_-],[\mathfrak{C}_+]) \to [\mathfrak{C}_\pm],$$

sending a trajectory to its limit points. Then a chain

$$f\colon \sigma \to [\mathfrak{C}_-]$$

gives rise (under suitable transversality hypotheses) to the chain

$$\mathrm{ev}_+\colon \sigma \times \breve{M}^+([\mathfrak{C}_-],[\mathfrak{C}_+]) \to [\mathfrak{C}_+],$$

where the underlying space is the fibered product under the maps f and ev_-, and is mapped to $[\mathfrak{C}_+]$ using the evaluation ev_+. The total differential of σ is then defined to be the sum of its singular differential and all these chains obtained via fibered products. The proofs in this new framework carry over with the same formulas as the usual one: identities relating zero-dimensional moduli spaces coming from boundaries of one-dimensional spaces are now identities (at the chain level) of chains arising as the codimension-one strata of fibered products as above. Of course, we need to consider chains σ in some class of geometric objects so that the fibered products with the moduli spaces remain in that class. This is a delicate point because in our context the compactified moduli spaces are not in general manifolds with corners, neither combinatorially nor topologically. In Section 3.1 of [19], we introduce a suitable class of such objects called δ–chains, and define a modified singular chain complex for a smooth manifold obtained by quotienting out a class of degenerate δ–chains. These degenerate chains are essentially chains whose image is contained in the image of a chain of strictly smaller dimension; see also [20]. This construction leads to a well defined homology theory for smooth manifolds satisfying the Eilenberg–Steenrod axioms.

These chain complexes come with a natural involution (given by composition with the involution $\jmath$), and the Pin(2)–monopole Floer homology groups are defined as the homologies of the invariant subcomplexes. Finally, by exploiting the positive scalar curvature we see that the Pin(2)–monopole Floer homology groups of the three-sphere are isomorphic (after grading reversal) to the claimed ones.

While transversality for the three-dimensional equations can be achieved using the equivariant counterpart of the three-dimensional perturbations used in [12], in the case of the map induced by a cobordism W new perturbations (defined only in the blown-up configuration space of the cobordism) need to be considered. This is because the three-dimensional perturbations always vanish at the fixed points of the involution $\jmath$. In particular, if we are only considering them, a spin connection on a cobordism always gives rise to a reducible solution, and this might not be regular simply for index reasons. In Section 4.2 of [19], we define Pin(2)–equivariant ASD–perturbations. These are maps

$$\hat{\omega}\colon \mathcal{C}^\sigma(X) \to L^2(X; i\mathfrak{su}(2)), \tag{6}$$

which are gauge invariant, $\jmath$–equivariant (where $\jmath$ acts on the right-hand side as multiplication by -1) and satisfy suitable analytical properties. Furthermore, they form a collection large enough so that we can achieve transversality while preserving the Pin(2)–symmetry.

2 Some additional background results

In this section, we discuss some aspects of Pin(2)–monopole Floer homology which have not been treated in [19] and will be central to the discussion in the present work. In particular, we discuss the blow-up formula and describe the reducible moduli spaces on special spin cobordisms with $b_2^+ = 1$.

Recall that the *blowup* of a smooth 4–manifold X (possibly with boundary) is the smooth four-manifold

$$\widetilde{X} = X \# \overline{\mathbb{C}P}^2.$$

We will denote by E the exceptional divisor, which is a smooth embedded sphere of self-intersection -1. We define $\mathfrak{s}_k$ to be the spin^c structure on $\overline{\mathbb{C}P}^2$ such that

$$\langle c_1(\mathfrak{s}_k), [E] \rangle = 2k - 1.$$

In particular, the conjugate of $\mathfrak{s}_k$ is $\mathfrak{s}_{1-k}$. Given a spin^c structure $\mathfrak{s}$ on X, we define $\mathfrak{s} \# \mathfrak{s}_k$ to be the spin^c structure on $\widetilde{X}$ that restricts to $\mathfrak{s}$ on X and to $\mathfrak{s}_k$ on $\overline{\mathbb{C}P}^2$. It is shown in Section 39.3 of [12] that in monopole Floer homology, we have the identity

$$\widetilde{HM}_\bullet(\widetilde{X}, \mathfrak{s} \# \mathfrak{s}_k) = \widetilde{HM}_\bullet(U^{k(k-1)/2} \mid X, \mathfrak{s}).$$

The aim of this section is to prove the counterpart of this in the case of Pin(2)–monopole Floer homology. It is important to notice that the blowup $\widetilde{X}$ is never a spin manifold, as it carries a homology class $[E]$ with odd self-intersection.

Proposition 8 *Let X be a cobordism. If $\mathfrak{s}$ a self-conjugate spin^c structure, then*

$$\widetilde{HS}_\bullet(\widetilde{X}, [\mathfrak{s} \# \mathfrak{s}_k]) = \begin{cases} 0 & \text{if } k \equiv 0, 1 \pmod 4, \\ \widetilde{HS}_\bullet(Q^2 V^{\lfloor k(k-1)/4 \rfloor} \mid X, [\mathfrak{s}]) & \text{otherwise.} \end{cases} \tag{7}$$

If the spin^c structure on X is not self-conjugate, then

$$\widetilde{HS}_\bullet(\widetilde{X}, [\mathfrak{s} \# \mathfrak{s}_k]) = 0$$

for every k. The same statement holds for the from *and* bar *versions.*

As it will be clear from the proof, it is natural to interpret the two cases in (7) by considering the parity of $\frac{1}{2}k(k-1)$: in the first case it is even, in the latter it is odd. This quantity will arise in the formula (8) for the dimension of moduli spaces of solutions on a punctured $\overline{\mathbb{C}P}^2$.

Proof The result can be proved by a neck stretching argument close to the ones in Section 3.2 of [19], and in particular the construction of the $\mathcal{R}$–module structure. We focus on the case of a self-conjugate spin^c structure $\mathfrak{s}$ first. Consider the separating three-sphere S^3 along which the connected sum is performed. We can suppose that the metric on this S^3 is the standard round one, and it is a product in a neighborhood. Hence we can define a one-dimensional family of metrics parametrized by $T \in [0,\infty)$ by adding longer and longer necks of the form $[0,T]\times S^3$. By considering the compactified moduli spaces of solutions parametrized by this family of metrics, we can construct a chain homotopy between the map defining $\widecheck{HS}_\bullet(\tilde{X},[\mathfrak{s}\#\mathfrak{s}_k])$ (which is the map corresponding to the stratum $T=0$), and a new chain map corresponding to an additional stratum at $T=\infty$ which we describe in detail. Intuitively, at $T=\infty$ the cobordism is decomposed in two pieces, and the solutions converge to solutions on each of the two pieces. On one hand we have the moduli spaces for the self conjugate spin^c structure on X with an additional cylindrical end $(0,\infty]\times S^3$ (denoted by X_p^* in [19]) with a Pin(2)–equivariant perturbation on it. Recall from Section 1 that the map induced by the cobordism X and an element x in $\mathcal{R}$ is given by considering the moduli spaces on X_p^* which are asymptotic on this additional end to a given representative of x in $\widehat{HS}_\bullet(S^3)$. The latter can be realized as a $\jmath$–invariant chain in a critical submanifold. Furthermore, we have the moduli spaces on a punctured $\overline{\mathbb{C}P}^2$ (which we denote by $(\overline{\mathbb{C}P}^2)^*$) and the perturbation on the cylindrical end is Pin(2)–equivariant. Denote the critical submanifold corresponding to the i^{th} negative eigenvalue by $[\mathfrak{C}_{-i}]$. The dimension of the moduli space of solutions in the spin^c structure $\mathfrak{s}_k$ converging to the critical submanifold $[\mathfrak{C}_{-i}]$ is given by

$$\dim M^+\big((\overline{\mathbb{C}P}^2)^*,[\mathfrak{C}_{\lfloor k(k-1)/4\rfloor-1}],\mathfrak{s}_k\big)=k(k-1)+4i-2, \tag{8}$$

see also Lemma 5.3 in [13]. As the critical submanifold is 2–dimensional, the only interesting moduli space for our purposes when dealing with the spin^c structure $\mathfrak{s}_k$ is

$$M_k=M^+\big((\overline{\mathbb{C}P}^2)^*,[\mathfrak{C}_{\lfloor k(k-1)/4\rfloor-1}],\mathfrak{s}_k\big),$$

as the others have dimension too big, so are degenerate, as discussed in Section 1. Following the proof of Frøyshov's theorem (Section 39.1 in [12]), as there are no L^2 anti-self-dual forms on $(\overline{\mathbb{C}P}^2)^*$ there is only one reducible connection. Furthermore, as the solutions converge on the cylindrical end to an unstable configuration they are necessarily reducible. If $k\equiv 0,1 \pmod 4$ then M_k is the projectivization of to the two-dimensional kernel of the corresponding Dirac equation. In particular, it is a two-dimensional sphere of reducible solutions, and the evaluation map is either constant or a diffeomorphism. On the other hand, if $k\not\equiv 0,1 \pmod 4$, then M_k is the projectivization of the one-dimensional kernel, so it consists of a single point.

Going back to the map induced by $\tilde{X}$ on $\widecheck{HS}_\bullet$, we see via the chain homotopy provided by the neck stretching argument that this is equivalent to the map induced by X together with a specified cohomology class in $\mathcal{R}$. In the case $k \equiv 0, 1 \pmod 4$, this is defined by the invariant chain given by the union $M_k \cup M_{1-k}$. As again the two chains are related by the action of $\jmath$, we have that either they both represent a generator of the top homology or they both represent the zero class. In any case, their union defines the zero class in the homology of the invariant subcomplex, so the induced map is zero. In the other case $k \not\equiv 0, 1 \pmod 4$, the cohomology class is given by $Q^2 V^{\lfloor k(k-1)/4 \rfloor}$, as a pair of antipodal points is a generator of the invariant subcomplex of the critical submanifold in dimension zero.

There is a subtlety regarding perturbations in this proof. Because the blowup $X \# \overline{\mathbb{C}P}^2$ is not spin, when performing the stretching argument we can use a nonequivariant perturbation on the neck while preserving equivariance. On the other hand, at the limit we need to use an equivariant perturbation on the manifold with cylindrical ends X_p^* in order for the map to make sense. From this, it is clear that in the case in which the spin^c structure on X is *not* self-conjugate, we can use nonequivariant perturbations all the way and obtain the second part of the result, as in this case the standard argument works. □

The second fact we want to point out is that unlike the usual counterpart in monopole Floer homology, the map $\overline{HS}_\bullet(W)$ induced by a cobordism W with $b_2^+ \geq 1$ is not necessarily zero. In fact, in a special case, we have the following characterization of the moduli spaces.

Proposition 9 *Let W be a cobordism between rational homology spheres with $b_1(W) = 0$ and $b_2^+ = 1$, and consider a self-conjugate spin^c structure $\mathfrak{s}_0$. Let A_0 be the corresponding spin^c connection. Suppose that for the fixed perturbations at the ends, there is only one reducible solution. Then there exists a regular perturbation on the cobordism such that for each reducible moduli space $M^+([\mathfrak{C}_-], [\mathfrak{C}_+])$ which is one-dimensional, the evaluation maps*

$$\mathrm{ev}_\pm \colon M^+([\mathfrak{C}_-], [\mathfrak{C}_+]) \to [\mathfrak{C}_\pm]$$

define $\jmath$–invariant chains generating the one-dimensional $\jmath$–invariant homologies.

If we restrict to Pin(2)–equivariant three-dimensional perturbations (which we recall are introduced in a collar $I \times \partial W$ of the boundary of W; see Chapter 24 in [12]), the reducible configuration in the blowdown $[A_0, 0]$ is always a solution, because the perturbations vanish at the fixed points of the involution. In the blowup, the reducible moduli spaces $M^+([\mathfrak{C}_-], [\mathfrak{C}_+])$ lying over $[A_0, 0]$ and such that the linearization of the equations has index 1 are in the best scenario a copy of $\mathbb{C}P^1$, given by the projectivization of the one-dimensional (over the quaternions) kernel of the perturbed Dirac operator $D^+_{A_0, \hat{\mathfrak{p}}}$. In particular, the moduli spaces cannot be transversely cut out.

Proof We will introduce a suitable regular perturbation for which we can explicitly describe the moduli spaces in play. As $b_2^+ = 1$, the operator d^+ does not have dense image, so we can find a smooth compactly supported self-dual form ω_0^+ not contained in it. We can suppose without loss of generality that the support of ω_0^+ is contained in W away from the collar of the boundary where three-dimensional perturbations are applied. This implies that the form ω_0^+ is not in the image of any of the linearizations of the perturbed anti-self-duality operators. We claim that there exists a smooth function on the (completion in the L^2_k norm of the) blown-up configuration space

$$\eta\colon \mathcal{C}_k^\sigma(W, \mathfrak{s}_W) \to \mathbb{R}$$

with the following properties:

(1) the map is gauge invariant and $\jmath$–equivariant where $\jmath$ acts on $\mathbb{R}$ as -1;

(2) the restriction of the map η to the reducible moduli space lying over $[A_0, 0]$ is transverse to zero;

(3) the $L^2(W; i\mathfrak{su}(2))$–valued map $\eta\cdot\rho_W(\omega_0^+)$ is, in the sense of Section 4.2 of [19], a Pin(2)–equivariant ASD–perturbation, so it can be legitimately used to perturb the equations.

The map η can be easily constructed from a given Pin(2)–equivariant ASD–perturbation $\hat{\omega}$ via the L^2 projection to the line spanned by a single configuration. The proofs in Section 4.2 in [19] carry over to show that the perturbation $\eta\cdot\rho(\omega_0^+)$ has the required analytic properties.

We can then describe the one-dimensional moduli spaces of reducible solutions as follows. The new equations for a reducible configuration $(A, 0, \phi)$ are (a small perturbation of)

$$D_A^+\phi = 0,$$
$$\rho(F_{A^t}^+) - \eta(A, 0, \phi)\omega_0^+ = 0.$$

As ω_0^+ generates the cokernel of d^+, the last equation implies $\eta(A, 0, \phi)$ is zero and $(A, 0, \phi)$ is a reducible solution for the equations without the additional perturbation. Hence the space of reducible solutions is identified as the zero locus of

$$\eta\colon \mathbb{C}P^1 \to \mathbb{R}.$$

This is transversely cut out by condition (2) and consists of an odd number of circles by condition (1), hence the result follows. □

Proof of Theorem 7 This result follows in the same way as the proof of monotonicity of Manolescu's invariants under negative definite cobordisms; see Section 4.4 of [19].

If $b_2^+ = 1$, after reducing to the case of $b_1 = 0$ via surgery, the description above tells us that for a self-conjugate spin^c structure $\mathfrak{s}$ the map

$$\overline{HS}_\bullet(W, \mathfrak{s})\colon \overline{HS}_\bullet(Y_0, \mathfrak{s}_0) \to \overline{HS}_\bullet(Y_1, \mathfrak{s}_1)$$

is as follows. After fixing isomorphisms of graded $\mathcal{R}$–modules

$$\overline{HS}_\bullet(Y_i, \mathfrak{s}_i) \cong \mathcal{S}\langle d_i \rangle,$$

where for some choice of d_i the angular brackets denote a global grading shift of d_i, the map is identified to be multiplication by QV^k where the integer k is determined by the intersection form of the cobordism, and the statement follows.

In the case $b_2^+ = 2$, an analogous characterization of the reducible moduli space holds: it consists of a number of points congruent to 2 modulo 4. The proof of this characterization follows that of Proposition 9, and we briefly sketch it here. The cokernel of d^+ now has dimension two, and we can construct an analogous $\jmath$ perturbation with values in the cokernel. When restricted to the reducible unperturbed moduli space, this has the form of

$$\eta\colon \mathbb{C}P^1 \to \mathbb{R}^2, \tag{9}$$

and the (transverse) zero set consists of a number of points congruent to 2 modulo 4 by $\jmath$ equivariance. The same argument as in Proposition 9 implies then that the reducible solutions are identified with the zero set of (9) and they are transversely cut out. This identifies the map on the bar version as the multiplication by Q^2V^k, and the result follows. □

3 The exact triangle

This section is dedicated to the proof of the main result of the paper, Theorem 1 in the introduction. We start by reviewing in detail the framework of the surgery exact triangle; see also [13; 12, Chapter 42]. Suppose we are given a knot K inside a three-manifold Y, and let Z be the manifold with torus boundary ∂Z obtained by removing a tubular neighborhood of it. Let μ_1, μ_2, μ_3 be closed simple curves on ∂Z with the property that the intersection numbers satisfy

$$\mu_1 \cdot \mu_2 = \mu_2 \cdot \mu_3 = \mu_3 \cdot \mu_1 = -1. \tag{10}$$

We can then obtain the three manifolds Y_1, Y_2 and Y_3 by Dehn filling along these curves. We extend this definition periodically so that, for example, $\mu_{n+3} = \mu_n$.

This construction behaves well with four-dimensional topology in the following sense. There is an elementary cobordism W_n from Y_n to Y_{n+1} obtained by attaching a single 2–handle $D^2 \times D^2$ to $[0,1] \times Y_n$ along $\{1\} \times K$ with framing μ_{n+1}. In this case, the knot $K \subset Y_{n+1}$ with framing μ_{n+2} can be identified with the boundary $\{0\} \times S^1$ of the cocore of the attached handle with framing -1 relative to the cocore $\{0\} \times D^2$. In W_n, there is a closed 2–cycle Σ_n defined as follows. On ∂Z there is the distinguished simple closed curve σ whose homology class generates the kernel of $H_1(\partial Z) \to H_1(Z)$, so in particular there is a surface $T \subset Z$ with boundary σ. The closed cycle Σ_n is obtained as the union of T, the core of the 2–handle and a piece in the solid torus $\partial D^2 \times D^2$.

Suppose we are in the case Z is a knot complement in S^3. Then there is canonical pair of curves m and l in ∂Z, namely the classical meridian and longitude of the knot. These are oriented so that $m \cdot l = -1$. In this case, any oriented simple closed curve μ can be described up to isotopy by its homology class

$$[\mu] = p[m] + q[l]$$

with (p,q) relatively prime. Forgetting about the orientation, this can be recorded by the ratio

$$r = p/q \in \mathbb{Q} \cup \{\infty\}.$$

Triples of curves satisfying the relations (10) come from triples of pairs

$$(p_1,q_1), \quad (p_2,q_2), \quad (p_3,q_3)$$

satisfying the conditions

$$-p_n q_{n+1} + p_{n+1} q_n = -1,$$

where as usual the subscripts are interpreted modulo 3. Very interesting cases are given by the slopes

$$r_1 = 0, \quad r_2 = 1/(q+1), \quad r_3 = 1/q$$

for $q \in \mathbb{Z}$ and

$$r_1 = p, \quad r_2 = p+1, \quad r_3 = \infty$$

for $p \in \mathbb{Z}$. In general, if none of the slopes is zero, up to cyclic permutation and change of sign we can suppose that our triple looks like

$$r_1 = \frac{p}{q}, \quad r_2 = \frac{p+p'}{q+q'}, \quad r_3 = \frac{p'}{q'}$$

with the properties

$$p, p' > 0, \quad -p'q + pq' = -1.$$

In particular, exactly two of p, p' and $p+p'$ are odd. Furthermore, in this case the cobordism with positive definite intersection form is W_3; see Section 42.3 in [12].

We now focus on the interactions of this construction with our invariants.

Lemma 10 *Among the three cobordisms W_1, W_2 and W_3, exactly two are spin.*

Proof Because the cobordism W_n is given by a two-handle attachment and three-manifolds are always spin, the fact that W_n is spin (ie its second Stiefel–Whitney class is zero) is equivalent to the fact that the cycle $[\Sigma_n]$ has even self-intersection. To see that this holds, we use the discussion above, which can be generalized to any manifold Z with torus boundary by taking $[l]$ to be a primitive element in the kernel of

$$H_1(\partial Z) \to H_1(Z),$$

and $[m]$ any other curve such that $[m] \cdot [l] = -1$. In particular, while the slope of a curve γ_n is not well defined (as it depends on the choice of the meridian $[m]$), the numerator p_n is. The self-intersection is up to sign just the product of the numerators $p_n p_{n+1}$, so the result follows because exactly two of them are odd. □

On the other hand, the composition $W_n \cup_{Y_{n+1}} W_{n+1}$ is never spin. In fact, it always contains a sphere E_n with self intersection -1. This is given by the union of the core of the 2–handle of W_{n+1} and the cocore of the 2–handle in W_n. From this description, it also follows that the cobordism

$$W_n \cup_{Y_{n+1}} W_{n+1}$$

between Y_n and Y_{n+2} is diffeomorphic to the opposite cobordism $\overline{W}_{n+2}$ blown up at a point. *Without loss of generality* we will suppose from now on that the nonspin cobordism is W_3.

We introduce the homological algebra needed for our purposes in an abstract setting. This is a slight variation of a standard triangle detection result in Floer homology; see Lemma 4.2 in [27] or Lemma 5.1 in [13]. Suppose we are given three chain complexes C_1, C_2 and C_3, and chain maps $f_1\colon C_1 \to C_2$ and $f_2\colon C_2 \to C_3$ such that the composition $f_2 \circ f_1$ is homotopic to zero via a nullhomotopy H_1. We can form the "iterated mapping cone" C whose underlying vector space is $C_3 \oplus C_2 \oplus C_1$ and whose differential is

$$\partial = \begin{pmatrix} \partial_3 & f_2 & H_1 \\ 0 & \partial_2 & f_1 \\ 0 & 0 & \partial_1 \end{pmatrix}.$$

This is a differential because f_1 and f_2 are chain maps and H_1 is a chain nullhomotopy for $f_2 \circ f_1$. The following is the key lemma in homological algebra we need.

Lemma 11 *Suppose the homology of the iterated mapping cone* $H_*(C,\partial)$ *is trivial. Then there is a map* $F_3\colon H_*(C_3)\to H_*(C_1)$ *such that the following triangle is exact:*

$$\begin{array}{ccc} H_*(C_2) & \xrightarrow{\quad (f_2)_* \quad} & H_*(C_3) \\ {\scriptstyle (f_1)_*}\nwarrow & & \swarrow{\scriptstyle F_3} \\ & H_*(C_1) & \end{array}$$

We will discuss the naturality properties of this construction (in our specific case) in detail in the proof of Theorem 1.

Proof The proof of this result follows closely that of the standard triangle detection lemma. First we form the mapping cone M_{f_1} of the chain map f_1, which is the chain complex with underlying vector space $C_1\oplus C_2$ and differential

$$d=\begin{pmatrix}\partial_2 & f_1\\ 0 & \partial_1\end{pmatrix}.$$

This is a differential because f_1 is a chain map. The short exact sequence of chain complexes

$$0\to C_2\xrightarrow{i} M_{f_1}\xrightarrow{p} C_1\to 0,$$

where the maps are respectively the inclusion and the quotient, induces an exact triangle:

$$\begin{array}{ccc} H_*(C_2) & \xrightarrow{\quad i_* \quad} & H_*(M_{f_1}) \\ {\scriptstyle (f_1)_*}\nwarrow & & \swarrow{\scriptstyle p_*} \\ & H_*(C_1) & \end{array}$$

Similarly the iterated mapping cone fits in the short exact sequence of chain complexes

$$0\to C_3\to C\to M_{f_1}\to 0.$$

In particular, we have a connecting homomorphism

$$\delta\colon H_*(M_{f_1})\to H_*(C_3),$$

which is induced by the chain map

$$(f_2+H_1)\colon M_{f_1}=C_2\oplus C_1\to C_3. \tag{11}$$

The fact that $H_*(C)$ is trivial is equivalent to δ being an isomorphism. So in the triangle above we can replace the homology of the mapping cone $H_*(M_{f_1})$ with $H_*(C_3)$ using this isomorphism. Finally the connecting homomorphism δ is given by (11), so we can identify the horizontal map as $(f_2)_*$. □

We now describe how our problem fits in the framework of Lemma 11. We denote by $\check{C}_i$ the chain complex $\check{C}_*(Y_i)$ computing the *to* version, and by

$$\check{f}_i \colon \check{C}_i \to \check{C}_{i+1}$$

the chain map defining the map induced by the cobordism W_i. We denote by $\check{M}$ the mapping cone of $\check{f}_1$. Recall that the composition $W_2 \circ W_1$ is $\overline{W}_3$ blown up at a point, and as $\overline{W}_3$ is not spin the induced map is zero by Proposition 8. There is in fact a natural chain homotopy to zero

$$\check{H}_1 \colon \check{C}_1 \to \check{C}_3$$

defined as follows. The chain homotopy is constructed by considering a one-dimensional family of metrics on the composite cobordisms $X_1 = W_2 \circ W_1$. There are two separating hypersurfaces Y_2 and S_1, the latter being the boundary of a neighborhood of the (-1)–sphere E_1. Choose a metric on X_1 such that S_1 has a metric obtained by flattening the round one near the Clifford torus $Y_2 \cap S_1$. We can construct then the family of metrics $Q(S_1, Y_2)$ parametrized by $T \in \mathbb{R}$ given by inserting a cylinder $[-T, T] \times S_1$ normal to S_1 for T negative, and a cylinder $[-T, T] \times Y_2$ normal to Y_2 for T positive. Following the notation of [13] we will use the letter Q to denote one-dimensional families of metrics. It will be clear from the context whether we are using this letter to indicate this or the element of $\mathcal{R}$. Also, we will denote by $\bar{Q}$ its compactification obtained by adding $\{\pm\infty\}$. We can define the moduli spaces parametrized by the family of metrics

$$M_z([\mathfrak{C}_-], X_1^*, [\mathfrak{C}_+])_Q.$$

This will be a smooth manifold for a generic choice of Pin(2)–equivariant perturbation. As in the proof of Proposition 8, this can be compactified to a moduli space

$$M_z^+([\mathfrak{C}_-], X_1^*, [\mathfrak{C}_+])_{\bar{Q}}$$

obtained by considering both broken trajectories and the fibered products of the compactified moduli spaces on the manifolds with (possibly more than two) cylindrical ends one obtains for $T = \pm\infty$, namely

$$W_1^* \amalg W_2^*$$

for $T = +\infty$ and

$$B_1^* \amalg Z_1^*$$

for $T = -\infty$. Here Z_1 is a neighborhood of the exceptional divisor, and B_1 can be identified with $\overline{W}_3$ with a ball removed. Similarly, one can consider the compactified moduli spaces consisting entirely of reducibles $M_z^{\mathrm{red}+}([\mathfrak{C}_-], X_1^*, [\mathfrak{C}_+])_{\bar{Q}}$. By taking fibered products with these moduli spaces one can define the $\jmath$–invariant linear map

$$H^o_o\colon C^o(Y_1)\to C^o(Y_3)$$

and similarly its companions $H^o_s, H^u_o, H^u_s, \bar{H}^s_s, \bar{H}^s_u, \bar{H}^u_s$ and $\bar{H}^u_u$. We then define the map

$$\check{H}_1=\begin{bmatrix} H^o_o & H^u_o\bar{\partial}^s_u+m^u_o(W_2)\bar{m}^s_u(W_1)+\partial^u_o\bar{H}^s_u \\ H^o_s & \bar{H}^s_s+H^u_s\bar{\partial}^s_u+m^u_s(W_2)\bar{m}^s_u(W_1)+\partial^u_s\bar{H}^s_u \end{bmatrix}$$

by the same formula that defines the chain map proving the composition formula; see Chapter 26 in [12] and Section 3.3 in [19]. Here the maps $m^*_*(W_*)$ are the components defining the chain maps $\check{f}_1$ and $\check{f}_2$.

In fact, the same construction applies to the other two composites to give rise to the maps $\check{H}_2$ and $\check{H}_3$, and we have the following result.

Lemma 12 *The map $\check{H}_1$ satisfies the identity*

$$\check{\partial}\circ\check{H}_1+\check{H}_1\circ\check{\partial}=\check{f}_2\circ\check{f}_1.$$

For $n=2,3$, the map $\check{H}_n$ is a chain homotopy between the composite $\check{f}_{n+1}\circ\check{f}_n$ and a chain map

$$\check{g}_n\colon \check{C}_\bullet(Y_n)\to\check{C}_\bullet(Y_{n+2})$$

computing the blown-up map as in Proposition 8 in Section 2.

Proof One just has to identify the contributions of the various codimension one strata of the moduli spaces $M^+_z([\mathfrak{C}_-],X^*_1,[\mathfrak{C}_+])_{\bar{Q}}$. The closure of the union of the codimension one strata of $M^+_z([\mathfrak{C}_-],X^*_1,[\mathfrak{C}_+])_T$ for T finite corresponds to the left-hand side. The moduli space $M^+_z([\mathfrak{C}_-],X^*_1,[\mathfrak{C}_+])_{+\infty}$ consists of the fibered products of the moduli spaces used to define the chain map on the right-hand side. The moduli spaces $M^+_z([\mathfrak{C}_-],X^*_1,[\mathfrak{C}_+])_{-\infty}$ define the zero map at the chain level. This is a manifestation at the chain level of the phenomenon underlying the proof of Proposition 8 in the non-self-conjugate case. The case of the other cobordisms is analogous and follows the proof of Proposition 8. □

We can then form as in Lemma 11 an iterated mapping cone $\check{C}$ of the three chain complexes $\check{C}_1,\check{C}_2,\check{C}_3$, the chain maps $\check{f}_1,\check{f}_2$ and the chain nullhomotopy $\check{H}_1$. As that lemma states, the main result underlying the existence of an exact triangle is the following.

Proposition 13 *The homology of the iterated mapping cone $\check{C}$ is zero.*

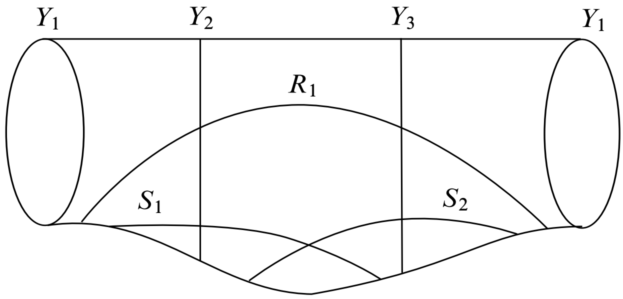


Figure 2: The five hypersurfaces in the triple composite

To prove this assertion, we will construct a chain map $\check{\varphi}$ from $\check{C}$ to itself homotopic to zero that induces an isomorphism at the homology level. Let V_1 be the manifold obtained as the triple composite

$$V_1 = W_1 \cup_{Y_2} W_2 \cup_{Y_3} W_3.$$

We introduce a two-dimensional family of (possibly degenerate) metrics parametrized by a pentagon as follows. This manifold contains five separating hypersurfaces, namely Y_2, Y_3, the two three-spheres S_1 and S_2 and the manifold R_1 homeomorphic to $S^1 \times S^2$ which is the boundary of a regular neighborhood of the (-1)–spheres E_1 and E_2 containing both S_1 and S_2. We can arrange them cyclically as Y_2, R_1, Y_3, S_2 and S_1 so that each of them intersects only its two neighbors; see Figure 2. For each pair S, S' of *non*intersecting hypersurfaces, we define the family of metrics parametrized by $\mathbb{R}^{>0} \times \mathbb{R}^{>0}$ by inserting cylinders $[-T_S, T_S] \times S$ and $[-T_{S'}, T_{S'}] \times S'$. This can be completed to a family of riemannian metrics over the "square"

$$\bar{P}(S, S') \cong [0, \infty] \times [0, \infty].$$

The five families obtained this way fit along their five edges corresponding to families of metrics in which only one of the T_S is nonzero. Hence we can obtain a family of metrics on the pentagon $\bar{P}$ obtained as their union; see Figure 3. For each hypersurface S, there is an edge $\bar{Q}_S$ of the pentagon (consisting of two of the edges of the squares) where $T_S = \infty$. One can arrange that the family of metrics is such that R_1, S_1 and S_2 have positive scalar curvature metrics.

One then considers the compactified moduli spaces of solutions parametrized by such a family, and uses them to construct maps between the chain complexes. The two strata corresponding to the edges $Q(Y_2)$ and $Q(Y_3)$ are exactly those that define the maps $\check{H}_2 \circ \check{f}_1$ and $\check{f}_3 \circ \check{H}_1$. Notice that unlike the case of [13], the sum of these two maps is not a chain map, as $\check{H}_2$ is not a chain homotopy between $\check{f}_3 \circ \check{f}_2$ and zero. The maps corresponding to the edges $Q(S_1)$ and $Q(S_2)$ correspond punctured cobordisms with an additional punctured $\overline{\mathbb{C}P}^2$ component.

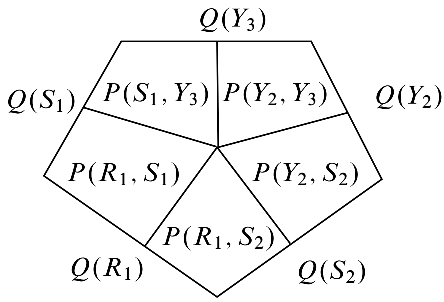


Figure 3: The family of metrics $\overline{P}$

The most interesting edge is the one given by $Q(R_1)$. The hypersurface R_1 is homeomorphic to $S^2 \times S^1$, hence because of positive scalar curvature the only interesting spin^c structure is the one with torsion first Chern class $\mathfrak{s}_0$. As in Section 4.4 of [19] we can fix a small regular perturbation with only two reducible solutions in the blowdown α_1 and α_0 corresponding to the spin connections B_1 and B_0 and no irreducible solutions. This is induced by a smooth $\jmath$–invariant Morse function on the one-dimensional torus $\mathbb{T}$ of flat connections

$$f\colon \mathbb{T} \to \mathbb{R}$$

with exactly two critical points via a gauge equivariant retraction of $\mathcal{B}(R_1, \mathfrak{s}_0)$ onto $\mathbb{T}$. We can suppose that α_1 is the maximum, so that there are exactly two trajectories connecting α_1 to α_0 in the blowdown. We call the critical submanifolds in the blowup $[\mathfrak{C}_i^{\mu}]$ for $\mu = 0, 1$ and $i \in \mathbb{Z}$. Here the superscript indicates the reducible solution on which the submanifold is lying over, and the index the eigenvalue it corresponds to. As usual, the index zero is for the first stable critical submanifold. In this case, the contributions of moduli spaces lying over the two trajectories connecting α_1 to α_0 in the blowdown cancel each other, so the homology is just the direct sum of the homologies of the critical submanifolds.

This hypersurface R_1 defines a decomposition of the triple composite as the union of two four-manifolds. The first, which we call U_1, has three boundary components and is the complement in $Y_1 \times [-1, 1]$ of a neighborhood of $K \times \{0\}$. The second one, which we call N_1, is the complement in $\overline{\mathbb{C}P}^2$ of an unknotted loop. Its second homology is generated by the two exceptional spheres E_1 and E_2, and spin^c structures $\mathfrak{s}_k$ that restricts to $\mathfrak{s}_0$ on the boundary is uniquely determined by

$$\langle c_1(\mathfrak{s}_k), [E_1] \rangle = \langle c_1(\mathfrak{s}_k), [E_2] \rangle = 2k - 1.$$

Here our notation slightly differs from [13]: what we call $\mathfrak{s}_k$ is denoted by them as $\mathfrak{t}_{k-1}$. On the manifold with one cylindrical end N_1^* we can consider the moduli spaces $M_k(N_1^*,[\mathfrak{C}_i^{\mu}])_{\bar{Q}}$ where $\bar{Q}$ is $\bar{Q}(R_1)$ relative to the spin^c structure $\mathfrak{s}_k$ of solutions converging to $[\mathfrak{C}_i^{\mu}]$. We have the following lemma; see Lemma 5.7 and following corollaries in [13].

Lemma 14 *The dimension of the moduli space $M_k(N_1^*,[\mathfrak{C}_i^{\mu}])_{\bar{Q}}$ is given by*

$$\dim M_k(N_1^*,[\mathfrak{C}_i^{\mu}])_{\bar{Q}} = \begin{cases} -\mu - k(k-1) - 4i, & i \geq 0, \\ -\mu - k(k-1) - 4i - 1, & i < 0. \end{cases}$$

In particular, the moduli spaces $M_k^{\mathrm{red}}(N_1^*,[\mathfrak{C}_i^{\mu}])_{\bar{Q}}$ are empty for all $i \geq 0$.

The moduli spaces above define chains in the critical submanifolds of R_1. In particular, we can consider σ^s, the one in the stable critical manifolds, σ_0^u in the unstable critical manifolds above α_0 and σ_1^u in the critical manifolds above α_1. We use the chains σ_μ^u in the unstable critical submanifolds to define the linear maps

$$L^o_{o,\mu}\colon C^o(Y_1) \to C^o(Y_1)$$

and its seven companions as the map induced by the manifold U_1 with three ends by fibered product with σ_μ^u on the R_1 end. We denote the sum of the two by dropping the μ and we can combine them in the maps

$$\check{L}_1 = \check{L}_{1,0} + \check{L}_{1,1}\colon \check{C}(Y_1) \to \check{C}(Y_1), \quad \check{L}_1 = \begin{bmatrix} L^o_o & L^u_o\bar{\partial}^s_u + \partial^u_o\bar{L}^s_u \\ L^o_s & \bar{L}^s_s + L^u_s\bar{\partial}^s_u + \partial^u_s\bar{L}^s_u. \end{bmatrix}.$$

Similarly, one can define the map

$$G^o_o\colon C^o(Y_1) \to C^o(Y_1)$$

and its seven companions induced by the fiber products with the moduli spaces on the triple composite parametrized by the pentagon of metrics $\bar{P}$, and the maps

$$\bar{r}^s_u\colon C^s(Y_1) \to C^u(Y_1),$$
$$\bar{r}^s_s\colon C^s(Y_1) \to C^s(Y_1)$$

obtained by fiber products of the moduli spaces on the manifold with three ends U_1^* with the chain σ^s in the critical stable manifolds of R_1. Finally we define

$$\check{G}_1\colon \check{C}_\bullet(Y_1) \to \check{C}_\bullet(Y_1), \quad \check{G}_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

where

$$a = G^o_o, \quad b = \partial^u_o\bar{G}^s_u + G^u_o\bar{\partial}^s_u + m^u_o\bar{H}^s_u + H^u_o\bar{m}^s_u + \partial^u_o\bar{r}^s_u,$$
$$c = G^o_s, \quad d = \bar{G}^s_s + \partial^u_s\bar{G}^s_u + G^u_s\bar{\partial}^s_u + m^u_s\bar{H}^s_u + H^u_s\bar{m}^s_u + \partial^u_s\bar{r}^s_u + \bar{r}^s_s.$$

Here again the m^*_* are the components of the corresponding maps $\check{f}_i$. The following is the analogue of Proposition 5.5 in [12]. We rephrase it in an alternative way because in our case the map $\check{f}_3 \circ \check{H}_1 + \check{H}_2 \circ \check{f}_1$ is not a chain map.

Lemma 15 *The map $\check{G}_1$ is a chain homotopy between the map arising as the sum of $\check{L}_1$ and the maps induced by the moduli spaces parametrized by $Q(S_1)$ and $Q(S_2)$ and the map $\check{f}_3 \circ \check{H}_1 + \check{H}_2 \circ \check{f}_1$.*

Proof The proof follows as usual by identifying the codimension-one strata of the moduli spaces involved in the definition of $\check{G}_1$, which are described by the same formulas as in [13]. In particular, the last map in the statement corresponds to the edges of the pentagon $Q(Y_2)$ and $Q(Y_3)$. □

We have the following key result, which is the analogue in our case of Lemma 5.10 in [13]. Because our critical submanifolds are two-dimensional, for our purposes we are only interested in the moduli spaces which have dimension at most two, as the others are degenerate as discussed in Section 1.

Lemma 16 *Suppose $k \geq 1$. Then the chain σ^u_1 consists of a generator of the top homology of the critical submanifold $[\mathfrak{C}^1_i]$ for $i = -\frac{1}{4}k(k-1)-1$, for k congruent to $0,1$ modulo four, while for $k = 2,3$ modulo four, it consists of an even number of points in $[\mathfrak{C}^1_i]$ for $i = -\frac{1}{4}k(k-1)-\frac{1}{2}$.*

Proof We first recall a result on the unperturbed anti-self-duality equations

$$F^+_{A^t} = 0$$

on the manifold N^*_1 from the proof of Lemma 5.10 in [13]. Given a metric g on N^*_1 which is standard on the end, there is a unique solution to such equation $A(k,g)$ with L^2 curvature for the spin^c structure $\mathfrak{s}_k$. This is because the manifold has no first homology and no self-dual, square integrable harmonic two forms (as the image of the relative second homology in the absolute one is zero). On the cylindrical end this connection $A(k,g)$ it is asymptotically flat so it defines a point

$$\theta_k(g) \in \mathbf{S},$$

where $\mathbf{S}$ is the circle of flat spin^c connections on $S^2 \times S^1$. On the family of metrics $\bar{Q}(R_1) = [-\infty,\infty]$ we have that $\theta_k(\pm\infty)$ is a spin connection. For example, at $-\infty$, the manifold decomposes in two pieces, one of which is a punctured $S^2 \times D^2$ with cylindrical ends, so it carries no L^2 harmonic forms. The key point in the proof of Lemma 5.10 in [13] is to show that the connection at the two ends of this family differ, so that without loss of generality we can assume $\theta_k(-\infty)=\alpha_1$ and $\theta_k(+\infty)=\alpha_0$. Furthermore, the map $\theta_{1-k}(g)$ is obtained by conjugation on the circle (with our convention: in that of [13] it is $\theta_{-1-k}(g)$).

We then consider the moduli spaces with asymptotics into $[\mathfrak{C}_i^1]$, which for dimensional reasons is interesting when it has dimension two or zero. In the first case, $k=0,1$ modulo 4 and $i=\frac{1}{4}k(k-1)-1$. We claim that for some choice of perturbations,

- the stratum $M_k(N_1^*,[\mathfrak{C}_i^1])_{-\infty}$ is a generator of the one-dimensional $\jmath$–invariant homology of the critical submanifold $[\mathfrak{C}_i^1]$;
- the stratum $M_k(N_1^*,[\mathfrak{C}_i^1])_{+\infty}$ is empty.

As we are only dealing with reducible solutions, the first claim follows in the same way as in Proposition 9. Indeed, our cobordism has $b_1=0$ and before adding the extra perturbation in the blowup of the configuration space of the cobordism the stratum over $-\infty$ consists of a two-dimensional sphere of reducibles lying over the spin connection. As in the case considered in Proposition 9, it is obstructed in codimension one, and the same construction of the additional perturbation carries over. Furthermore, as in that setting, the moduli spaces are already compact before we compactify them because there are no possible breaking points, as there are no trajectories from α_0 to α_1. The gluing results regarding our moduli spaces then imply that the union of the moduli spaces

$$M_k(N_1^*,[\mathfrak{C}_i^\mu])_{\bar{Q}}\cup M_{1-k}(N_1^*,[\mathfrak{C}_i^\mu])_{\bar{Q}}$$

is a $\jmath$–equivariant generator of the top homology of the critical submanifold, hence the claim. The second claim is clear as $\theta_k(+\infty)$ is α_0.

Finally, when $k=2,3$ modulo four, the strata $M_k(N_1^*,[\mathfrak{C}_i^1])_{\pm\infty}$ are both empty by transversality, hence the moduli spaces consist of an even number of points by symmetry. □

Before proving Proposition 13, we need to discuss the $\mathcal{R}$–module structure on the mapping cones we have defined.

Lemma 17 *The mapping cone $H_*(M_{\check{f}_1})$ of the chain map $\check{f}_1$ is an $\mathcal{R}$–module, and the mapping cone triangle for $\widecheck{HS}_\bullet(W_1)$ is an exact triangle of $\mathcal{R}$–modules. The coboundary map δ in Proposition 13 is a map of $\mathcal{R}$–modules.*

Proof The construction of the module structure follows the ideas in [1], and is described in Figure 4. Consider a point p which is not in the neighborhood of the knot K where the surgery operation is performed. In particular, we can consider p as

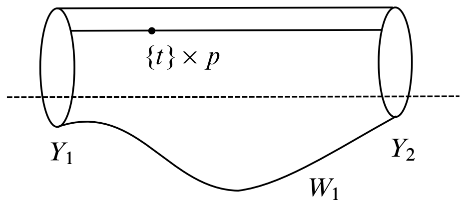


Figure 4: The module structure on the mapping cone. The cobordism is a product $(Y \setminus \mathrm{nbhd}(K)) \times [-1,1]$ above the dashed line.

a point in both Y_1 and Y_2, and use the same ball B_p embedded in $\mathbb{R}\times Y_i$ centered at p to compute the map induced (for example) by $V \in \mathcal{R}$ by looking at the moduli spaces that are asymptotic to the second unstable critical submanifold $[\mathfrak{C}_{-2}]$ on the additional incoming end. Here we assume that the metric has positive scalar curvature on the boundary of B_p and is a product near there. Call the associated chain maps $\check{V}_1$ and $\check{V}_2$. There is a natural chain homotopy $\check{\mathcal{H}}_1$ between $\check{f}_1 \circ \check{V}_1$ and $\check{V}_2 \circ \check{f}_1$ obtained by considering the compactifications of the moduli spaces of trajectories parametrized by the moving point $\{t\}\times p$. Indeed, we can identify the subset

$$\mathbb{R}\times (Y \setminus \mathrm{nbhd}(K)) \subset W_1^*,$$

and consider the union of the moduli spaces of trajectories on the cobordism with a puncture at $\{t\}\times p$ that are asymptotic to $[\mathfrak{C}_{-2}]$ on this additional end. As usual, we can compactify these moduli spaces and use them to define a map $\check{\mathcal{H}}_1$ satisfying

$$\check{\partial}_2 \circ \check{\mathcal{H}}_1 + \check{\mathcal{H}}_1 \circ \check{\partial}_1 = \check{f}_1 \circ \check{V}_1 + \check{V}_2 \circ \check{f}_1.$$

The endomorphism of $M_{\check{f}_1}$ defined by the matrix

$$\begin{pmatrix} \check{V}_2 & \check{\mathcal{H}}_1 \\ 0 & \check{V}_1 \end{pmatrix}$$

is then a chain map, and we define the induced map to be the action of V of $H_*(M_{\check{f}_1})$. The usual arguments show that this is well defined, and the maps in the triangle commute with this map.

The module structure on the iterated mapping cone $H_*(\check{C})$ is defined in an analogous way. Indeed, the same construction above applied to the cobordism W_2 leads to a chain homotopy $\check{\mathcal{H}}_2$ such that

$$\check{\partial}_3 \circ \check{\mathcal{H}}_2 + \check{\mathcal{H}}_2 \circ \check{\partial}_2 = \check{f}_2 \circ \check{V}_2 + \check{V}_3 \circ \check{f}_1,$$

where $\check{V}_3$ is the analogous chain map inducing the action of V on $\widecheck{HS}_\bullet(Y_3)$. We claim that there is a map $\check{\mathcal{G}}_1$ from $\check{C}_1$ to $\check{C}_3$ with the property that

$$\check{\partial}_3 \circ \check{\mathcal{G}}_1 + \check{\mathcal{G}}_1 \circ \check{\partial}_3 = \check{f}_2 \circ \check{\mathcal{H}}_1 + \check{\mathcal{H}}_2 \circ \check{f}_1 + \check{H}_1 \circ \check{V}_1 + \check{V}_3 \circ \check{H}_1, \tag{12}$$

so we have that

$$\begin{pmatrix} \check{V}_3 & \check{\mathcal{H}}_2 & \check{\mathcal{G}}_1 \\ 0 & \check{V}_2 & \check{\mathcal{H}}_1 \\ 0 & 0 & \check{V}_1 \end{pmatrix}$$

is a chain map. We use this map to define the action of V on $H_*(\check{C})$. The map $\check{\mathcal{G}}_1$ is constructed as follows. The cobordism with cylindrical ends attached $(W_2 \circ W_1)^*$ contains a copy of $(Y \setminus \mathrm{nbhd}(K)) \times \mathbb{R}$ hence in particular the line $p \times \mathbb{R}$, so we can consider the moduli space parametrized by $\mathbb{R} \times \mathbb{R}$ where the first component parametrizes the position of the point while the second parametrizes the family of metrics used to define the chain homotopy $\check{H}_1$. The map $\check{\mathcal{G}}_1$ is then defined by considering the moduli spaces of solutions parametrized by this family, and the identity (12) follows as usual by identifying the contributions of the four edges of the square $[-\infty, \infty] \times [-\infty, \infty]$. □

Proof of Proposition 13 Consider the map $\check{G}$ on the iterated mapping cone $\check{C}$ given by

$$\check{G} = \begin{pmatrix} \check{G}_3 & 0 & 0 \\ \check{H}_3 & \check{G}_2 & 0 \\ \check{f}_3 & \check{H}_2 & \check{G}_1 \end{pmatrix}.$$

Here $\check{G}_2$ and $\check{G}_3$ are the maps induced by the moduli spaces parametrized by the pentagon of metrics and perturbations on the other two triple composites. They satisfy the properties of $\check{G}_1$ that we have discussed above. We define the chain map

$$\check{\varphi}\colon \check{C} \to \check{C}$$

given by

$$\check{\varphi} = \check{\partial} \circ \check{G} + \check{G} \circ \check{\partial}.$$

The map $\check{\varphi}$ is nullhomotopic by definition, and our claim is that it also induces an isomorphism at the homology level. We have (for example) the identity

$$\check{V} \circ \check{\varphi} + \check{\varphi} \circ \check{V} = \check{\partial} \circ (\check{G} \circ \check{V} + \check{V} \circ \check{G}) + (\check{G} \circ \check{V} + \check{V} \circ \check{G}) \circ \check{\partial}, \tag{13}$$

so the map induced in homology is a map of $\mathcal{R}$–modules. Using the relations of Lemma 15 we can write the map $\check{\varphi}$ as the matrix

$$\check{\varphi} = \begin{pmatrix} \check{L}_3 + \check{h}_3 & \check{f}_2 \check{G}_2 + \check{G}_3 \check{f}_2 + \check{H}_1 \check{H}_2 & \check{H}_1 \check{G}_1 + \check{G}_3 \check{H}_1 \\ \check{g}_3 & \check{L}_2 + \check{h}_2 & \check{f}_1 \check{G}_1 + \check{G}_2 \check{f}_1 + \check{H}_3 \check{H}_1 \\ 0 & \check{g}_2 & \check{L}_1 + \check{h}_1 \end{pmatrix}.$$

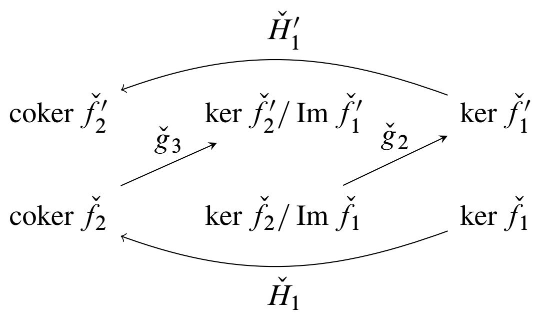
The lower diagonal terms $\check{g}_i$ are those appearing in Lemma 12, and the maps $\check{h}_i$ indicate the maps induced by the two edges in the pentagon corresponding to the blowups. For example, using the notation we have adopted throughout the section, the map $\check{h}_1$ is defined using the moduli spaces parametrized by the families of metrics $Q(S_1)$ and $Q(S_2)$. Notice that these are *not* chain maps. Unfortunately there is not a natural filtration respected by this map $\check{\varphi}$. Our strategy is to show that its mapping cone (which has a natural filtration) has trivial homology. In particular, we consider the chain complex whose underlying vector space is the sum of two copies of $\check{C}$ (where we distinguish the elements and groups in the first copy with the apostrophe)

$$\widetilde{C} = \check{C}' \oplus \check{C}$$

and differential

$$\widetilde{\partial} = \begin{pmatrix} \check{\partial}'_3 & \check{f}'_2 & \check{H}'_1 & \check{L}_3 + \check{h}_3 & * & * \\ 0 & \check{\partial}'_2 & \check{f}'_1 & \check{g}_3 & \check{L}_2 + \check{h}_2 & * \\ 0 & 0 & \check{\partial}'_1 & 0 & \check{g}_2 & \check{L}_1 + \check{h}_1 \\ 0 & 0 & 0 & \check{\partial}_3 & \check{f}_2 & \check{H}_1 \\ 0 & 0 & 0 & 0 & \check{\partial}_2 & \check{f}_1 \\ 0 & 0 & 0 & 0 & 0 & \check{\partial}_1 \end{pmatrix}.$$

This chain complex has a natural filtration induced by the upper triangular structure of the differential $\widetilde{\partial}$. Because the left lower entry of $\check{\varphi}$ vanishes, the E^1 page of the associated spectral sequence does not involve differentials between the corresponding subquotients of $\check{C}$ and $\check{C}'$. In particular, the E^2 page of the spectral sequence is given by



where by an abuse of the notation we are considering the maps induced on the subquotients by the indicated maps. On the other hand Lemma 12 tells us that the chain maps $\check{g}_2$ and $\check{g}_3$ induce at the homology level the maps $\check{f}_3 \circ \check{f}_2$ and $\check{f}_1 \circ \check{f}_3$. In particular, $\check{g}_2$ is zero on the kernel of $\check{f}_2$ and the image of $\check{g}_3$ is contained in the image of $\check{f}_1$, so the diagonal maps in the diagram above are both zero. So the subquotients of the two chain complexes $\check{C}$ and $\check{C}'$ do not interact at this page either, and the E^3 page is simply

$$\begin{array}{ccc} \operatorname{coker} \check{f}_2'/\operatorname{Im} \check{H}_1' & \ker \check{f}_2'/\operatorname{Im} \check{f}_1' & \ker \check{f}_1' \cap \ker \check{H}_1' \\ \check{L}_3+\check{h}_3 \Big\uparrow & \check{L}_2+\check{h}_2 \Big\uparrow & \check{L}_1+\check{h}_1 \Big\uparrow \\ \operatorname{coker} \check{f}_2/\operatorname{Im} \check{H}_1 & \ker \check{f}_2/\operatorname{Im} \check{f}_1 & \ker \check{f}_1 \cap \ker \check{H}_1 \end{array}$$

Our claim is that the vertical maps are isomorphisms, so that the E^4 page of the spectral sequence is zero, proving our claim that $\check{\varphi}$ is an isomorphism. From (13), it follows that the objects involved are $\mathcal{R}$–modules and the vertical arrows are maps of $\mathcal{R}$–modules. The maps $\check{h}_i$ above involve multiplication by the element Q^2 in $\mathcal{R}$, as they are defined via moduli spaces on manifolds parametrized by a family of metrics on which a blowup is already stretched to infinity (as for example in Proposition 8).

Recall from [19, Sections 3.3 and 4.4] that the group $\overline{HS}_\bullet(Y)$ is naturally a module over

$$\textstyle\bigwedge^*(H_1(Y;\mathbb{Z})/\mathrm{Tor}\otimes\mathbb{F})\otimes\mathcal{R}.$$

Indeed, consider a closed embedded loop γ in Y representing a given homology class x. A neighborhood of this loop has boundary $S^2\times S^1$ and we can suppose that the metric and perturbations on this are the same as we discussed above. In particular, we have that

$$\widehat{HS}_\bullet(S^2\times S^1)=(\mathcal{R}\oplus\mathcal{R}\langle -1\rangle)\langle -1\rangle,$$

where the first $\mathcal{R}$ summand corresponds to the critical submanifolds $[\mathfrak{C}_i^1]$ while the second one to the critical submanifolds $[\mathfrak{C}_i^1]$. Again here the brackets denote the grading shift. The cobordism $[-1,1]\times Y\setminus \mathrm{nbhd}(\{0\}\times\gamma)$ induces a map

$$\widecheck{HS}_\bullet(Y)\otimes(\mathcal{R}\oplus\mathcal{R}\{-1\})\{-1\}\to\widecheck{HS}_\bullet(Y).$$

When restricting to the elements of the first $\mathcal{R}$ summand, we recover the usual $\mathcal{R}$–module structure, while the elements in the second summand correspond to the action of the elements $x\otimes\mathcal{R}$. Because of Lemma 16, the term $\check{L}_{i,1}$ induces the sum of a power series in $\mathcal{R}$ with leading term 1. All the terms in the summand $\check{L}_{i,0}$ of $\check{L}_i$ also involve multiplication by the nilpotent element

$$[K]\in\textstyle\bigwedge^*(H_1(Y;\mathbb{Z})/\mathrm{Tor}\otimes\mathbb{F}).$$

Hence each of the vertical maps on the E^3 page is a sum of an isomorphism $\check{L}_{i,1}$ and a nilpotent map $\check{L}_{i,0}+\check{h}_i$ which commute because of the $\mathcal{R}$–module structure, so it is an isomorphism. □

From Proposition 13, we can conclude the main result of the present paper.

Proof of Theorem 1 The existence of the triangle follows from our discussion and Lemma 11. It is a triangle of $\mathcal{R}$–modules thanks to Lemma 17. It remains to show that the third map is well defined (ie independent of the choices we have made). The mapping cone construction satisfies the following naturality property: given two homotopic chain maps f and f' between chain complexes C_0 and C_1 there is an isomorphism between the two mapping cones such that the mapping cone exact triangles commute. In fact, if h is a chain homotopy between f and f', the canonical isomorphism is given by the matrix

$$M(h)=\begin{pmatrix}\mathrm{Id}_{C_0} & h\\ 0 & \mathrm{Id}_{C_1}\end{pmatrix}.$$

Furthermore, given another such chain homotopy h', if there exists a map

$$k\colon C_0\to C_1$$

such that $\partial'\circ k+k\circ\partial=h-h'$, then the induced isomorphism is the same, as the two maps $M(h)$ and $M(h')$ are homotopic via the map $\left(\begin{smallmatrix}0 & k\\ 0 & 0\end{smallmatrix}\right)$.

In our case, the two maps f and f' correspond to two different regular choices of metric and perturbation $(g_0,\mathfrak{p}_0)$ and $(g_1,\mathfrak{p}_1)$. In order to identify the mapping cones, we consider chain homotopy h constructed by considering the moduli spaces parametrized by a regular path $(g_t,\mathfrak{p}_t)$ for $t\in[0,1]$ connecting these two choices. Because the space of metrics and perturbations is contractible, any two such paths are homotopic via a generic homotopy $h_{s,t}$ (which we can think as a regular two-dimensional family of metrics and perturbations) relative to their endpoints. The map k is then constructed by considering the moduli spaces parametrized by this two-dimensional family.

The analogous construction carries over to show that the iterated mapping cone (hence the boundary map δ) is natural. Suppose we have two iterated mapping cones corresponding to triples f_1, f_2, H_1 and f_1', f_2', H_1'. Given, for $i=1,2$, chain homotopies h_i as above between f_i and f_i', we claim that there is a map

$$K\colon C_1\to C_3$$

satisfying the identity

$$\partial_3\circ K+K\circ\partial_1=f_2'\circ h_1+h_2\circ f_1+H_1+H_1',\tag{14}$$

so that the map

$$\begin{pmatrix}\mathrm{Id} & h_2 & K\\ 0 & \mathrm{Id} & h_1\\ 0 & 0 & \mathrm{Id}\end{pmatrix}$$

is an isomorphism between the two iterated mapping cones.

In our case (where we add checks to be consistent with our notation), the map $\check{K}$ is constructed by considering the moduli spaces parametrized by a pentagon of metrics (which are possibly degenerate) and perturbations. The five vertices of the pentagon correspond to the maps $\check{f}_2\circ\check{f}_1$, $\check{f}_2'\circ\check{f}_1$, $\check{f}_2'\circ\check{f}_1'$, and the two endpoints p and p' of the homotopies $\check{H}_1$ and $\check{H}_1'$ correspond to the blowup

$$W_2\circ W_1=\overline{W}_3\,\#\,\overline{\mathbb{C}P}^2$$

stretched to infinity. Four of the edges correspond to the four terms in the expression of $\check{\partial}_3\circ\check{K}+\check{K}\circ\check{\partial}_1$ in (14). The fifth edge is a path between p and p' through degenerate metrics for which the blowup is stretched to infinity. We can choose this path so that the copy of S^3 along which the connected sum is performed always has positive scalar curvature. The moduli spaces parametrized by this edge do not contribute to the boundary terms for the same reason that the composite map is zero; see Proposition 8. These five edges can be filled to a pentagon using again the contractibility of the space of metrics and perturbations, and the induced isomorphism is well defined for the same reason. □

4 Computations from the Gysin exact sequence

In this section, we show that when the usual monopole Floer homology of a three-manifold is very simple, the Pin(2)–monopole Floer homology can be recovered in purely algebraic terms from the Gysin exact sequence

$$\cdots\xrightarrow{\cdot Q}\widecheck{HS}_k(Y)\xrightarrow{\iota_*}\widecheck{HM}_k(Y)\xrightarrow{\pi_*}\widecheck{HS}_k(Y)\xrightarrow{\cdot Q}\widecheck{HS}_{k-1}(Y)\xrightarrow{\iota_*}\cdots. \tag{15}$$

For a rational number d, denote by $\mathcal{T}_d^+$ the graded $\mathbb{F}[\![U]\!]$–module $\mathbb{F}[U^{-1},U]\!]/U\mathbb{F}[\![U]\!]$, where 1 has degree d. In particular, $\mathcal{T}_0^+$ is isomorphic as a graded $\mathbb{F}[\![U]\!]$–module to the Floer homology group $\widecheck{HM}_\bullet(S^3)$.

It is useful to have the basic example of S^3 in mind, which we briefly recall. We have

$$\widecheck{HM}_k=\begin{cases}\mathbb{F} & \text{if } k\geq 0 \text{ even},\\ 0 & \text{otherwise},\end{cases}\qquad \widecheck{HS}_k=\begin{cases}\mathbb{F} & \text{if } k\geq 0,\ k\not\equiv 3 \pmod 4,\\ 0 & \text{otherwise},\end{cases} \tag{16}$$

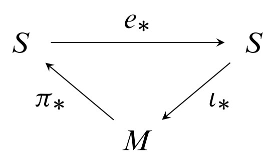
and the relevant Gysin exact sequence has the form

$$\begin{array}{ccccc}
\mathbb{F}_{4n+2} & \longrightarrow & \mathbb{F}_{4n+2} & & \mathbb{F}_{4n+2}\\
 & & \swarrow & & \\
\mathbb{F}_{4n+1} & & 0 & & \mathbb{F}_{4n+1}\\
 & & \swarrow & & \\
\mathbb{F}_{4n} & & \mathbb{F}_{4n} & \longrightarrow & \mathbb{F}_{4n}
\end{array} \tag{17}$$

for every $n \geq 0$. Here the indices denote the gradings, while the arrows are for the maps which are not trivial. Notice that these can be deduced directly by the group structure of the two groups and the exactness of the sequence. For a given rational homology sphere Y and a self conjugate spin^c structure $\mathfrak{s}$, we know that the Gysin sequence for $(Y, \mathfrak{s})$ looks like the one of (16) up to grading shift in degrees high enough. This motivates the following definition.

Definition 18 An *abstract Gysin sequence* $\mathcal{G}$ consists of the following data:

- an $\mathbb{F}[\![U]\!]$–module M and a $\mathcal{R}$–module S, both graded by a coset of $\mathbb{Z}$ in $\mathbb{Q}$ and bounded below;
- an exact triangle of $\mathcal{R}$–modules



  where the $\mathcal{R}$–module structure on M is given by Q acting trivially and V acting as U^2;
- the maps π_* and ι_* have degree zero while e_* has degree -1, and the triangle is isomorphic to the exact triangle (17) in degrees high enough.

In particular, the module M has a unique infinite-dimensional $\mathbb{F}[\![U]\!]$–submodule, and we denote by $2h(M)$ the minimum degree in which it is nontrivial. Because the exact triangle looks like (17) in degrees high enough, we have that the group S is trivial degrees $2h(M) + 4N - 1$ or $2h(M) + 4N + 1$ for N big enough. Of course, only one of the two possibilities is allowed.

Definition 19 If the group S is trivial in degrees $2h(M)+4N-1$ for N big enough, we say that the abstract Gysin sequence $\mathcal{G}$ is *even*, and we say that it is *odd* otherwise.

Here the terms *even* and *odd* refer to the grading modulo four in which S vanishes in degrees high enough, relative to $h(M)$. For example, the Gysin exact triangle for S^3 is even. We are ready to state the main result of the present section.

Proposition 20 *Suppose we are given an $\mathbb{F}[\![U]\!]$–module M of the form*

$$\mathcal{T}^+_{2k} \oplus \mathbb{F}^n \langle 2k-1 \rangle.$$

Then there exists a unique (up to isomorphism) abstract Gysin sequence in which M fits. If $n = 2m$ is even, then the sequence $\mathcal{G}$ is even, and

$$S \cong \mathcal{S}^+_{k,k,k} \oplus \mathbb{F}^m \langle 2k-1 \rangle,$$

while if $n = 2m+1$ is odd, the sequence $\mathcal{G}$ is odd, and

$$S \cong \mathcal{S}^+_{k+1,k-1,k-1} \oplus \mathbb{F}^{m+1}\langle 2k-1\rangle.$$

Suppose we are given an $\mathbb{F}[\![U]\!]$–module of the form

$$\mathcal{T}^+_{2k} \oplus \mathbb{F}^n\langle 2k\rangle.$$

Then there exists a unique up to isomorphism abstract Gysin sequence in which M fits. If $n = 2m+1$ is odd, then the sequence $\mathcal{G}$ is even, and

$$S \cong \mathcal{S}^+_{k,k,k} \oplus \mathbb{F}^{m+1}\langle 2k\rangle,$$

while if $n = 2m$ is even, the sequence $\mathcal{G}$ is odd, and

$$S \cong \mathcal{S}^+_{k+1,k+1,k-1} \oplus \mathbb{F}^{m+1}\langle 2k\rangle.$$

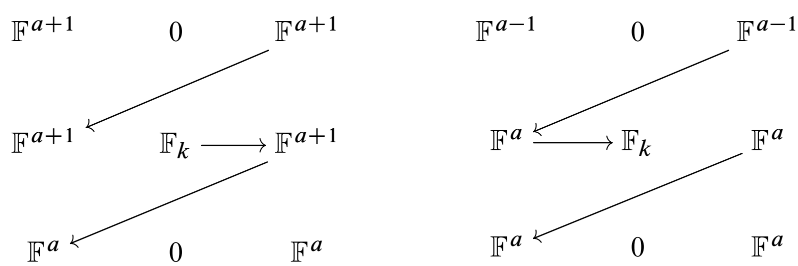
The key idea is the following easy observation which readily follows from the exactness of the Gysin exact sequence.

Lemma 21 *Given a Gysin exact sequence $\mathcal{G}$, suppose that for some k we have*

$$M_{k-1} = 0, \quad M_k = \mathbb{F}, \quad M_{k+1} = 0.$$

Then we have the two possibilities

(18)



where a is nonnegative in the left case and positive in the right case.

Definition 22 In the first situation, we say that the Gysin sequence at the level k is *increasing*, while in the second situation, we say that the Gysin sequence is *decreasing*.

Proof of Proposition 20 In the proof, we can assume without loss of generality that $k = 0$ after a grading shift. Suppose first that we are in the last case, so that $M_0 = \mathbb{F}^{2m+1}$. Clearly the suggested S fits in an abstract Gysin exact sequence, hence we need to prove uniqueness. Dimensional considerations and the exactness of the Gysin triangle imply that S is trivial in negative degrees while $\pi_*\colon M_0 \to S_0$ is surjective. Let the rank of the latter be $m+1+l$ for some $0\leq l\leq m$. To determine the structure of the whole group we can then use Lemma 21. In particular, because the rank of S_1 is odd the sequence implies that the rank of S_{3+4N} will be even for all $N\geq 0$, hence the Gysin sequence is even.

We claim that the Gysin triangle in degrees $4N\leq i\leq 4N+3$ contains a copy of the sequence (17) given by the image under a suitably high power of the map V. We prove this by induction, as it is of course true for N big enough by assumption. Denote by

$$v^{-N}, qv^{-N}, q^2v^{-N} \quad \text{and} \quad u^{-2N-1}, u^{-2N}$$

the respective generators of this copy. Of course

$$V\cdot u^{-2N-1}=u^{-2N+1},$$

so as ι_* maps v^{-N} to u^{-2N-1}, we see that $V\cdot v^{-N}$ is not zero. Denote this element by v^{-N+1}. This is mapped by ι_* to u^{-2N+1}, so this map is not zero and the sequence is decreasing at the level $4N-1$. This implies that $Q\cdot v^{-N+1}$ is not zero, and we call this element qv^{-N+1}. Similarly, we denote the image of this element under the action of Q by q^2v^{-N+1}. This is not zero because M_{4N+1} is trivial. Also the module structure implies that $\pi_*(u^{-2N+2})=q^2v^{-N+1}$.

This final observation implies that at each level $4N$ for $N\geq 1$ the sequence is increasing, so the ranks of S_{4N-1} form a nondecreasing sequence. As it has to be zero for N big enough, all these ranks are zero. By dimensional considerations, l has to be zero, and the result follows.

The proof in the other three cases is analogous. The only difference in the odd case is that one shows that the Gysin triangle in degrees $4N+2\leq i\leq 4N+5$ for $N\geq 0$ contains a copy of the standard one. □

Proof of Theorem 2 This follows readily by applying Proposition 20 to the case of the Brieskorn spheres $\Sigma(2,3,6n\pm 1)$. In particular, we have

$$\begin{aligned}
\widetilde{HM}_\bullet(-\Sigma(2,3,12k+5)) &= \mathcal{T}^+_{-2}\oplus\mathbb{F}^{2k}\langle -2\rangle,\\
\widetilde{HM}_\bullet(\Sigma(2,3,12k+1)) &= \mathcal{T}^+_{0}\oplus\mathbb{F}^{2k}\langle -1\rangle,\\
\widetilde{HM}_\bullet(-\Sigma(2,3,12k-1)) &= \mathcal{T}^+_{-2}\oplus\mathbb{F}^{2k-1}\langle -2\rangle,\\
\widetilde{HM}_\bullet(\Sigma(2,3,12k-5)) &= \mathcal{T}^+_{0}\oplus\mathbb{F}^{2k-1}\langle -1\rangle,
\end{aligned}$$

and the result for the given orientations follows from Poincaré duality. □

Remark 23 It is not surprising that in general the Pin(2)–monopole Floer homology cannot be recovered from the usual counterpart, as some ambiguities may arise. For example, if Y is the Brieskorn sphere $-\Sigma(2,3,11)$, we have

$$\widetilde{HM}_\bullet(Y \# Y) \cong \mathcal{T}^+_{-4} \oplus \mathbb{F}^3\langle -4\rangle \oplus \mathbb{F}\langle -3\rangle$$

as it can be computed by the connected sum formula in Heegaard Floer homology; see [26]. Then the two $\mathcal{R}$–modules

$$\mathcal{S}^+_{0,-2,-2} \oplus \mathbb{F}^2\langle -4\rangle \quad \text{and} \quad \mathcal{S}^+_{-2,-2,-2} \oplus \mathbb{F}^2\langle -4\rangle \oplus \mathbb{F}\langle -3\rangle$$

both fit in an abstract Gysin sequence. Notice that in this case we cannot recover Manolescu's correction terms either.

5 Examples

In this section, we discuss some simple computations of the Pin(2)–monopole Floer homology groups that can be done by applying the surgery exact triangle of Theorem 1. In order to get acquainted with the ideas, we first start with two examples of $\infty, 0, 1$ surgery on a knot in S^3 in which we already know all the groups involved in the computations. In general, we will label the maps by the surgery coefficient of the manifold corresponding to the domain. In this particular case, the nonspin cobordism is the one from Y_1 to Y_∞, so the map provided by Theorem 1 is $\overline{F}_1$.

Example 24 Suppose K is the unknot. Then

$$Y_0 = S^2 \times S^1 \quad \text{and} \quad Y_1 = S^3.$$

We know from the discussion in the previous section (see also Section 4.4 of [19] for more details) that for the unique self-conjugate spin^c structure $\mathfrak{s}_0$ we have the isomorphisms of graded $\mathcal{R}$–modules

$$\overline{HS}_\bullet(S^2 \times S^1, \mathfrak{s}_0) \cong \mathcal{S} \otimes (\mathbb{F} \oplus \mathbb{F}\langle -1\rangle),$$
$$\widetilde{HS}_\bullet(S^2 \times S^1, \mathfrak{s}_0) \cong \mathcal{S}^+_{0,0,0} \otimes (\mathbb{F} \oplus \mathbb{F}\langle -1\rangle).$$

In this case, for all the three-manifolds involved the map i_* is surjective. In particular, all the triangles are determined by the one for the bar version, and we will focus on the latter. The map

$$\overline{HS}_\bullet(W_\infty)\colon \overline{HS}_\bullet(S^3) \to \overline{HS}_\bullet(S^2 \times S^1, \mathfrak{s}_0),$$

which has degree -1, is an isomorphism onto the summand $\mathcal{S}\langle -1\rangle$ with lower degree, while the map

$$\overline{HS}_\bullet(W_0)\colon \overline{HS}_\bullet(S^2 \times S^1, \mathfrak{s}_0) \to \overline{HS}_\bullet(S^3),$$

which has degree zero, sends the lower summand $\mathcal{S}\langle -1\rangle$ to zero and is an isomorphism when restricted to the top summand $\mathcal{S}$. Indeed, the spin connection on W_∞ restricts to the spin connection B_0 on $S^2\times S^1$ which is the minimum, while the spin connection on W_0 restricts to the spin connection B_1. The corresponding moduli spaces of solutions are all copies of $\mathbb{C}P^1$ on which the evaluation maps diffeomorphically. This description can also be derived from the exact triangle for the usual monopole groups. It follows that the third map (which is called $\bar{F}_1$ in our case)

$$\bar{F}_1\colon \overline{HS}_\bullet(S^3)\to\overline{HS}_\bullet(S^3)$$

is zero. In particular, it is different from the map induced by the corresponding cobordism. In fact, the cobordism W_1 is a twice-punctured $\overline{\mathbb{C}P^2}$ so the induced map is the multiplication by a nonzero power series in which each term involves the multiplication by Q^2.

Example 25 Let K be the right-handed trefoil. Then $(+1)$–surgery is the Poincaré homology sphere (oriented as the boundary of the negative definite E_8 plumbing) and we computed in [19] that as $\mathcal{R}$–modules we have

$$\widetilde{HS}_\bullet(Y_1)\cong\mathcal{S}^+_{-1,-1,-1},$$

and the map i_* is surjective. Similarly, the 0–surgery is a flat torus bundle over the circle and we showed in Section 4.4 of [19] that as $\mathcal{R}$–modules,

$$\overline{HS}_\bullet(S^3_0(K),\mathfrak{s}_0)\cong(\mathcal{V}_1\oplus\mathcal{V}_0)\oplus(\mathcal{V}_{-1}\oplus\mathcal{V}_{-2}),$$
$$\widetilde{HS}_\bullet(S^3_0(K),\mathfrak{s}_0)\cong(\mathcal{V}^+_1\oplus\mathcal{V}^+_0)\oplus(\mathcal{V}^+_{-1}\oplus\mathcal{V}^+_{-2}).$$

In both cases, the action of Q (which has degree -1) is an isomorphism from the first summand to the second summand and from the third summand to the fourth summand. It is interesting to notice that (unlike in usual monopole Floer homology) the bar group is significantly different from the case of $S^2\times S^1$. This is another manifestation of the modulo four periodicity of the groups, and is related to the fact that the trefoil knot has Arf invariant 1. In more detail, in the blowdown the situation is analogous to that of $S^2\times S^1$, with two critical points α_1 and α_0 connected by two trajectories related by the action of $\jmath$. The key difference is that the family of Dirac operators has spectral flow $+1$ (so is in particular odd) along these paths, and in particular that the reducible solutions lying over α_0 are shifted up in degree by 2. This implies that the generator of the top homology of $[\mathfrak{C}^1_i]$ cancels with the generator of the bottom homology of $[\mathfrak{C}^0_i]$. For the remaining spin^c structures, the groups are zero by the adjunction inequality. Also in this case the i_* maps are surjective so it suffices to determine the reducible solutions and the bar version of the groups.

The maps $\overline{HS}_\bullet(W_\infty)$ and $\overline{HS}_\bullet(W_0)$ have degrees respectively -1 and 0, and the first map is given by multiplication by Q onto the first tower, while the latter is zero on the first tower and the identity on the second one. This follows from the same discussion of the moduli spaces on the cobordisms W_∞ and W_0 as above, the only difference being the cancellations happening in the chain complex of Y_0.

In particular, the map $\bar{F}_1$ is nonzero in this case. It is not straightforward to identify the map provided by the theorem in this case. Nevertheless we can say that the topmost homogeneous part of $\bar{F}_1$ lies in degree zero and that this part is an isomorphism in degree divisible by four and zero otherwise. This statement follows from the degrees of the *to* groups involved in the triangle and the module structure. Indeed, the generator of $\widecheck{HS}_0(Y_1)$ has to be mapped to the generator of $\widecheck{HS}_0(S^3)$ for degree reasons, so the degree zero part of $\bar{F}_1$ is an isomorphism. The module structure implies then that in general on the elements of degree $4k$ the map $\bar{F}_1$ is the product of the top homogeneous part by a fixed power series in V with leading coefficient 1.

Example 26 We discuss the case of $-1, 0$ and ∞–surgery on the right trefoil. Again, we know all the groups involved in the triangle, as $(-1/n)$–surgery on the trefoil is the Seifert fibered space $\Sigma(2,3,6n+1)$, but we take a more algebraic approach. Consider the following triangle:

$$\begin{array}{ccc} \widecheck{HS}_\bullet(\Sigma(2,3,7)) & \xrightarrow{\check{F}_{-1}} & \widecheck{HS}_\bullet(Y_0) \\ \check{F}_\infty \nwarrow & & \swarrow \check{F}_0 \\ & \widecheck{HS}_\bullet(S^3) & \end{array}$$

Again $\check{F}_{-1}$ and $\check{F}_0$ have degree respectively -1 and 0. From this, the maps are easily determined (using again the fact that the reduced Floer groups are trivial). In particular, the map

$$\check{F}_{-1}\colon \widecheck{HS}_{-1}(\Sigma(2,3,7)) \to \widecheck{HS}_{-2}(Y_0)$$

is an isomorphism for degree reasons. The module structure implies then that $\check{F}_{-1}$ in an isomorphism onto the image in degrees $4k$ and $4k-1$ for $k \geq 0$, and $\check{F}_0$ is an isomorphism onto the image in degrees $4k$ and $4k+1$ for $k \geq 0$. From this, as in the previous example, we see that $\check{F}_\infty$ has top degree zero, it is an isomorphism onto the image in degree $4k+2$, and zero otherwise.

The computation of the third map (the one corresponding to the nonspin cobordism) in the two examples we have just discussed ($\bar{F}_1$ and $\bar{F}_\infty$ respectively) only relies on the reducible solutions, so they hold in general for the map $\bar{F}_{1/(n+1)}$ in the $0, 1/(n+1), 1/n$ surgery triangle for a knot in a homology sphere. It is important to remark that there is a difference in the case n is even or odd related to relative grading of the reducibles.

As we have discussed both parities in our examples, we have proved the following result concerning the image of i_*.

Lemma 27 *Suppose the knot K has Arf invariant 1. Then in the setting as above, the map $\overline{F}_{1/(n+1)}$ corresponding to the nonspin cobordism is injective on the top tower and zero on the other two towers.*

The condition on the Arf invariant implies that the Rokhlin invariant of the $1/(n+1)$ and $1/n$ surgeries are different, so we are dealing with the cases of Examples 25 and 26.

Finally, we now show how to use the knowledge of the third map in order to provide a previously inaccessible computation. The same ideas will be used in the next section to compute the correction terms of (± 1)–surgery on alternating knots.

Proof of Theorem 3 As in [24], we have that $(+1)$–surgery on the figure-eight knot is $\Sigma(2,3,7)$. Furthermore, we know that for $\mathfrak{s} \neq \mathfrak{s}_0$ the Floer groups of E_0 vanish because of the adjunction inequality; see Corollary 40.1.2 in [12]. Using the fact that the reduced groups of S^3 and $\Sigma(2,3,7)$ are zero we can determine the Floer group $\widecheck{HS}_\bullet(E_0, \mathfrak{s}_0)$ as follows. In the triangle

$$\begin{array}{ccc} \widecheck{HS}_\bullet(S^3) & \xrightarrow{\quad \check{F}_\infty \quad} & \widecheck{HS}_\bullet(E_0) \\ {\scriptstyle \check{F}_1} \nwarrow & & \swarrow {\scriptstyle \check{F}_0} \\ & \widecheck{HS}_\bullet(\Sigma(2,3,7)) & \end{array}$$

the (top degree part) of the map $\check{F}_\infty$ is determined in light of Lemma 27. The result then follows as we know the degrees of $\check{F}_1$ and $\check{F}_0$. Finally, the other cases follow from Proposition 20. □

6 Surgery on alternating knots

In this final section, we show how to compute the Manolescu's correction terms of the homology spheres obtained by surgery on alternating knots. This relies on the computation of the usual monopole Floer homology groups provided in [25] (via the isomorphism between the theories due to Kutluhan, Lee and Taubes [14; 15; 16; 17; 18] and Colin, Ghiggini and Honda [3; 4; 5; 6] plus some additional considerations regarding absolute gradings) and some algebraic observations. We recall the main result from [25]. Given a knot K, its torsion coefficient $t_s(K)$ for an integer s is defined to be

$$t_s(K) = \sum_{j=1}^{\infty} j a_{|s|+j},$$

where the a_s are the coefficients of the symmetrized Alexander polynomial of K. For $\sigma \in 2\mathbb{Z}$ and an integer s, we define

$$\delta(\sigma, s) = \max\big(0, \big\lceil \tfrac{1}{4}(|\sigma| - 2|s|)\big\rceil\big).$$

Theorem 28 [25, Theorem 1.4] *Let K be an alternating knot oriented so that $\sigma = \sigma(K) \le 0$, and let $S^3_0(K)$ be the three-manifold obtained by zero surgery. Then, letting*

$$b_s = (-1)^{s+\sigma/2}(\delta(\sigma, s) - t_s(K)),$$

we have that:

- *for all $s > 0$ we have an $\mathbb{F}[\![U]\!]$ module isomorphism*
$$\widetilde{HM}_\bullet(S^3_0(K), \mathfrak{s}_s) \cong \mathbb{F}^{b_s} \oplus (\mathbb{F}[U]/U^{\delta(\sigma,s)})$$
with the first summand supported in degree $s + \frac{1}{2}\sigma \bmod 2$ while the second summand lies in odd degree;
- *there is an isomorphism of graded modules*
$$\widetilde{HM}_\bullet(S^3_0(K), \mathfrak{s}_0) \cong \mathcal{T}^+_{-1} \oplus \mathcal{T}^+_{-2\delta(\sigma,0)} \oplus \mathbb{F}^{b_0}\langle \tfrac{1}{2}\sigma - 1\rangle.$$

As briefly mentioned above it is important to notice that the isomorphism between monopole Floer homology and Heegaard Floer homology is only known to hold at the level of relatively graded groups. Nevertheless, in our simple case it can be seen to hold at the level absolutely graded groups thanks to the usual surgery exact triangle. Indeed, the maps in the triangle

$$\widetilde{HM}_\bullet(S^3) \to \widetilde{HM}_\bullet(S^3_0(K), \mathfrak{s}_0),$$
$$\widetilde{HM}_\bullet(S^3_0(K), \mathfrak{s}_0) \to \widetilde{HM}_\bullet(S^3_1(K))$$

have absolute degrees respectively -1 and 0 (recall that the absolute gradings in monopole Floer homology are shifted by $-\frac{1}{2}b_1(Y)$ with respect to those in Heegaard Floer homology), and the rank of the groups involved implies that the first map is an isomorphism onto the $\mathcal{T}^+_{-1}$ summand.

The last group in the statement of the result is the sum of two particularly simple modules, namely $\mathcal{T}^+_{-1}$ and $\mathcal{T}^+_{-2k} \oplus \mathbb{F}^{b_0}$, where the degree of the third summand is either $-2k$ or $-2k-1$.

We are now ready to provide the main computation in the present paper.

Proof of Theorem 4 The result follows from an application of the surgery exact triangle together with the result on the simple monopole Floer homology groups discussed above. Notice that we are only interested in the image of the map i_* (as a graded $\mathcal{R}$–module) as our goal is to compute Manolescu's correction terms. In particular, because of the module structure the map in the usual monopole Floer homology surgery exact triangle

$$\widetilde{HM}_\bullet(S^3) \to \widetilde{HM}_\bullet(S^3_0(K))$$

is an isomorphism onto the summand $\mathcal{T}^+_{-1}$ in the direct summand $\widetilde{HM}_\bullet(S^3_0(K), \mathfrak{s}_0)$, while is zero onto the others. Hence $\widetilde{HM}_\bullet(S^3_1(K))$ is isomorphic to a direct sum

$$\mathcal{T}^+_{-2\delta(\sigma,0)} \oplus \mathbb{F}^{b_0} \oplus \Bigl(\bigoplus_{c_1(\mathfrak{s})\neq 0} \widetilde{HM}_\bullet(S^3_0(K), \mathfrak{s})\Bigr).$$

The first two summands is a module of the form of Proposition 20. Furthermore, the Gysin sequence for the remaining summand is clear as the spin^c structures are conjugate in pairs and the sequence is functorial under cobordism maps. We can then apply Proposition 20 and determine $\widetilde{HS}_\bullet(S^3_1(K))$, and then reconstruct $i_*(\overline{HS}_\bullet(S^3_0(K), \mathfrak{s}_0))$ by applying the surgery exact triangle. The details in the case of a knot with Arf invariant 1 are analogous to those of the surgeries on the trefoil and figure-eight knot discussed in Section 5. From the knowledge of $i_*(\overline{HS}_\bullet(S^3_0(K), \mathfrak{s}_0))$, we can then easily compute Manolescu's correction terms by applying the surgery exact triangle. As an example, we focus on the case in which K has signature -8 (and Arf invariant 1) as it is particularly interesting in light of Corollary 6 in the introduction. In this case, we have that

$$i_*(\overline{HS}_\bullet(S^3_1(K))) \cong \mathcal{S}^+_{-1,-3,-3}.$$

This implies (as in the proof of Theorem 3 in the previous section) that as an absolutely graded $\mathcal{R}$–module we have

$$i_*(\overline{HS}_\bullet(S^3_0(K), \mathfrak{s}_0)) \cong (\mathcal{V}^+_1 \oplus \mathcal{V}^+_0) \oplus (\mathcal{V}^+_{-5} \oplus \mathcal{V}^+_{-2}),$$

where the Q–action maps the first tower onto the second tower and the third tower onto the fourth tower. Applying the surgery exact triangle again, we then obtain

$$i_*(\overline{HS}_\bullet(S^3_{-1}(K))) \cong \mathcal{S}^+_{1,-1,-3},$$

from which the result follows. □

We conclude by giving a proof of Proposition 5.

Proof of Proposition 5 It is shown in [23] that for a suitable choice of orientation of Y, there is a choice of metric and perturbation such that all the critical points (in the blowdown) have even degree. Suppose that Y is oriented in the same way as [23]. This implies that the generators of the two- and zero-dimensional homology of the reducible critical submanifolds are not affected by the differential in the chain complex computing $\widetilde{HS}_{\bullet}(Y,\mathfrak{s})$. Consider the maximum degree at which a generator of the one-dimensional homology of a reducible critical submanifold is involved in a nontrivial differential. As there are no irreducible critical points of odd degree, the module structure implies that the middle tower of $i_*(\overline{HS}_{\bullet}(Y,\mathfrak{s}))$ stops at that level. For the same reason, the bottom tower also stops at that level, hence we have that $\alpha(Y,\mathfrak{s})$ and $\beta(Y,\mathfrak{s})$ coincide.

Finally, if the orientation is the opposite, the same argument applied to $-Y$ implies that $\beta(Y,\mathfrak{s})$ and $\gamma(Y,\mathfrak{s})$ coincide. □

References

[1] **J M Bloom**, *A link surgery spectral sequence in monopole Floer homology*, Adv. Math. 226 (2011) 3216–3281 MR

[2] **J M Bloom**, *The combinatorics of Morse theory with boundary*, from "Proceedings of the Gökova Geometry-Topology Conference 2012" (S Akbulut, D Auroux, T Önder, editors), International Press, Somerville, MA (2013) 43–88 MR

[3] **V Colin**, **P Ghiggini**, **K Honda**, HF = ECH *via open book decompositions: a summary*, preprint (2011) arXiv

[4] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions, I*, preprint (2012) arXiv

[5] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions, II*, preprint (2012) arXiv

[6] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology, III: From hat to plus*, preprint (2012) arXiv

[7] **S K Donaldson**, **P B Kronheimer**, *The geometry of four-manifolds*, Clarendon, New York (1990) MR

[8] **A Floer**, *Instanton homology, surgery, and knots*, from "Geometry of low-dimensional manifolds, 1" (S K Donaldson, C B Thomas, editors), London Math. Soc. Lecture Note Ser. 150, Cambridge Univ. Press (1990) 97–114 MR

[9] **K Fukaya**, *Floer homology of connected sum of homology* 3*–spheres*, Topology 35 (1996) 89–136 MR

[10] **M Furuta**, *Monopole equation and the* $\frac{11}{8}$*–conjecture*, Math. Res. Lett. 8 (2001) 279–291 MR

[11] **C M Gordon**, *Dehn surgery on knots*, from "Proceedings of the International Congress of Mathematicians" (I Satake, editor), volume I, Math. Soc. Japan, Tokyo (1991) 631–642 MR

[12] **P Kronheimer**, **T Mrowka**, *Monopoles and three-manifolds*, New Mathematical Monographs 10, Cambridge Univ. Press (2007) MR

[13] **P Kronheimer**, **T Mrowka**, **P Ozsváth**, **Z Szabó**, *Monopoles and lens space surgeries*, Ann. of Math. 165 (2007) 457–546 MR

[14] **C Kutluhan**, **Y-J Lee**, **C H Taubes**, *HF=HM, I: Heegaard Floer homology and Seiberg–Witten Floer homology*, preprint (2010) arXiv

[15] **C Kutluhan**, **Y-J Lee**, **C H Taubes**, *HF=HM, II: Reeb orbits and holomorphic curves for the ech/Heegaard–Floer correspondence*, preprint (2010) arXiv

[16] **C Kutluhan**, **Y-J Lee**, **C H Taubes**, *HF=HM, III: Holomorphic curves and the differential for the ech/Heegaard Floer correspondence*, preprint (2010) arXiv

[17] **C Kutluhan**, **Y-J Lee**, **C H Taubes**, *HF=HM, IV: The Seiberg–Witten Floer homology and ech correspondence*, preprint (2011) arXiv

[18] **C Kutluhan**, **Y-J Lee**, **C H Taubes**, *HF=HM, V: Seiberg–Witten–Floer homology and handle addition*, preprint (2012) arXiv

[19] **F Lin**, *A Morse–Bott approach to monopole Floer homology and the triangulation conjecture*, preprint (2014) arXiv

[20] **M Lipyanskiy**, *Geometric homology*, preprint (2014) arXiv

[21] **C Manolescu**, *The Conley index, gauge theory, and triangulations*, J. Fixed Point Theory Appl. 13 (2013) 431–457 MR

[22] **C Manolescu**, Pin(2)*–equivariant Seiberg–Witten Floer homology and the triangulation conjecture*, J. Amer. Math. Soc. 29 (2016) 147–176 MR

[23] **T Mrowka**, **P Ozsváth**, **B Yu**, *Seiberg–Witten monopoles on Seifert fibered spaces*, Comm. Anal. Geom. 5 (1997) 685–791 MR

[24] **P Ozsváth**, **Z Szabó**, *Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary*, Adv. Math. 173 (2003) 179–261 MR

[25] **P Ozsváth**, **Z Szabó**, *Heegaard Floer homology and alternating knots*, Geom. Topol. 7 (2003) 225–254 MR

[26] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and three-manifold invariants: properties and applications*, Ann. of Math. 159 (2004) 1159–1245 MR

[27] **P Ozsváth**, **Z Szabó**, *On the Heegaard Floer homology of branched double-covers*, Adv. Math. 194 (2005) 1–33 MR

[28] **M Stoffregen**, Pin(2)*–equivariant Seiberg–Witten Floer homology of Seifert fibrations*, preprint (2015) arXiv

Department of Mathematics, Princeton University
Princeton, NJ, United States
fl4@math.princeton.edu

Received: 29 September 2016 Revised: 20 January 2017

Algebraic & Geometric Topology 17 (2017) 2961–3005

Odd knot invariants from quantum covering groups

SEAN CLARK

We show that the quantum covering group associated to $\mathfrak{osp}(1|2n)$ has an associated colored quantum knot invariant à la Reshetikhin–Turaev, which specializes to a quantum knot invariant for $\mathfrak{osp}(1|2n)$, and to the usual quantum knot invariant for $\mathfrak{so}(1+2n)$. In particular, this furnishes an "odd" variant of $\mathfrak{so}(1+2n)$ quantum invariants, even for knots labeled by spin representations. We then show that these knot invariants are essentially the same, up to a change of variables and a constant factor depending on the knot and weight.

17B37, 57M27

1 Introduction

1.1 Background

Quantum enveloping algebras associated to Kac–Moody Lie algebras are central objects in mathematics which have many remarkable connections to geometry, combinatorics, mathematical physics, and other areas. One such connection was produced by Reshetikhin and Turaev [23] and Turaev [27] by relating the representation theory of these quantum enveloping algebras to Laurent polynomial knot invariants, often referred to as quantum invariants, such as the (colored) Jones polynomial and the HOMFLYPT polynomial. The procedure for constructing these quantum invariants is quite general, and for instance has been generalized to construct quantum invariants from quantum enveloping superalgebras as well; see Blumen [3], Geer and Patureau-Mirand [12], Gould, Links and Zhang [13], Queffelec and Sartori [22], Sartori [25] and Zhang [30]. These super invariants are often equivalent to the nonsuper quantum invariants in some sense, but provide a novel perspective that can reveal additional features.

Many other connections have arisen from the categorification of quantum enveloping algebras and their representations; see Khovanov and Lauda [17] and Rouquier [24]. It was recently shown by Webster [28] that in fact, one can categorify all Reshetikhin–Turaev invariants using the machinery of categorified quantum enveloping algebras. This procedure generalizes Khovanov's homological categorification of the Jones polynomial [16]. We can summarize some of these connections in the picture in

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.2961

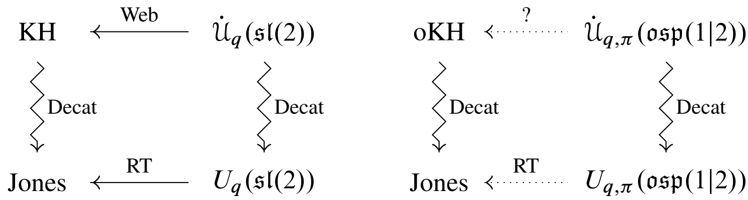


Figure 1

Figure 1 (left), where "Decat" refers to the appropriate decategorification, "RT" stands for the Reshetikhin–Turaev procedure for constructing the Jones polynomial from the standard quantum $\mathfrak{sl}(2)$ representation, and "Web" stands for Webster's categorification of RT which produces Khovanov homology.

This beautiful picture recently developed a twist with the discovery of "odd Khovanov homology", an alternate homological categorification of the Jones polynomial; see Ozsváth, Rasmussen and Szabó [21]. This discovery has spurred a program of "oddification": providing analogues of (categorified) quantum groups for this odd Khovanov homology by developing "odd" analogues of standard constructions; see Ellis and Lauda [11], Ellis, Khovanov and Lauda [10] and Mikhaylov and Witten [19]. In particular, one would like an "odd (categorified) $U_q(\mathfrak{sl}(2))$" which could produce odd Khovanov homology in a similar way to that described in Figure 1 (left). In particular, the decategorified "odd" quantum group should produce the Jones polynomial through some analogue of the Reshetikhin–Turaev procedure. It has been proposed (see [11] and Hill and Wang [14]) that such categorifications might naturally arise through categorifying the quantum covering group $U_{q,\pi}(\mathfrak{osp}(1|2))$, in other words, producing a diagram such as in Figure 1 (right).

This proposal has some heuristic evidence. Since the work of Zhang [29] connecting the representation theories of $\mathfrak{osp}(1|2n)$ and $\mathfrak{so}(1+2n)$, it has been expected that the associated knot invariants are essentially the same [30]. This expectation was partially verified by Blumen [2; 3], who showed that there is a link invariant associated to the two-dimensional quantum representation of $\mathfrak{osp}(1|2)$ (though no relation to the Jones polynomial was claimed), and that the $\mathfrak{osp}(1|2n)$ and $\mathfrak{so}(2n+1)$ invariants which are colored by the standard $(2n+1)$–dimensional representations are related up to some variable substitution, though the variable substitution has not been made explicit in general. However, to our knowledge, no proof of any relation between super and nonsuper colored type B knot invariants exists in the literature. The covering quantum group is a natural tool for filling this gap, as it has an explicit algebraic bridge between the super and nonsuper theories.

On the other hand, Mikhaylov and Witten [19] have produced candidates for "odd link homologies" categorifying $\mathfrak{so}(1{+}2n)$–invariants via topological quantum field theories using the orthosymplectic supergroups. This suggests that the conjecture represented by Figure 1 (right) should be generalized to include colored link invariants associated to $\mathfrak{osp}(1|2n)$ for any $n \geq 1$.

1.2 Results

A quantum covering group is an algebra $\boldsymbol{U}$ that marries the quantum enveloping superalgebra of an anisotropic Kac–Moody Lie superalgebra (eg $\mathfrak{osp}(1|2n)$) with the quantum enveloping algebra of its associated Kac–Moody Lie algebra, which is obtained by forgetting the parity in the root datum (eg $\mathfrak{so}(1+2n)$). This is done by introducing a new "half-parameter" π satisfying $\pi^2 = 1$, and substituting π everywhere a sign associated to the superalgebra braiding should appear; such algebras were defined and studied in detail in the series of papers by Clark, Fan, Hill, Li and Wang [4; 5; 6; 7; 8; 9].

These quantum covering groups retain the many nice properties of usual quantum groups such as a Hopf structure, a quasi-$\mathcal{R}$–matrix à la Lusztig [18, Chapter 4], a category $\mathcal{O}$, and even canonical bases. A key feature of a quantum covering group is that by specializing $\pi = 1$ (resp. $\pi = -1$), we obtain the quantum enveloping (super)algebra associated to the Kac–Moody Lie (super)algebra. Moreover, as discovered in [5], the quantum algebra and quantum superalgebra can be identified by a twistor map; that is, an automorphism of (an extension of) the covering quantum group which sends $\pi \mapsto -\pi$ and $q \mapsto \boldsymbol{t}^{-1}q$, where $\boldsymbol{t}^2 = -1$. This construction provides an algebraic realization of the connection between $\mathfrak{osp}(1|2n)$ and $\mathfrak{so}(1+2n)$ observed in [29].

In this paper, we use the machinery of covering quantum groups to construct "quantum covering knot invariants": knot invariants which arise from the representation theory of the finite type quantum covering groups à la Turaev [27].[1] To wit, consider the quantum covering group associated to the Lie superalgebra $\mathfrak{osp}(1|2n)$. We first associate a $\boldsymbol{U}$–module homomorphism to each elementary tangle (cups, caps, crossings) such that a straight strand is just the identity map, along with an interpretation of combining tangles (with joining top-to-bottom being composition of the associated maps, and placing along-side being tensor products of the maps). An arbitrary tangle can then be framed and associated with a $\boldsymbol{U}$–module homomorphism by "slicing" the diagram

[1] We do not need, and will not use, the usual additional ribbon structure of [23]. While the category of representations we consider is certainly rigid, the $\mathcal{R}$–matrix we use does not define a braiding in that the maps it defines fail to satisfy the hexagon identity in general. This can be fixed (see Remark 2.20), but it requires additional burdensome notation and analysis that we feel detracts more than it adds, particularly for the results of Section 4.

(that is, cutting it into vertical chunks containing at most one elementary diagram alongside any number of straight strands). Each slice corresponds to a $\boldsymbol{U}$–module homomorphism, and the tangle is sent to the composition of these maps. Note that a priori, this assignment is not unique, as many distinct slice diagrams and framings exist for an arbitrary tangle.

We then derive some identities with these maps that are versions of Turaev moves on the associated diagrams. These identities show that the map isn't dependent on the choice of slice diagram, but factors of π keep it from being an invariant of oriented framed tangles. In order to eliminate these factors, we need to expand our base ring to $\mathbb{Q}(q,\boldsymbol{t})^{\tau}$, where $\tau^2=\pi$, and renormalize the maps corresponding to certain elementary diagrams. Finally, a normalization factor (depending on the writhe of the tangle) yields a oriented tangle invariant; see Theorem 3.7.

In the rank-1 uncolored case, this invariant is simply the (unnormalized) Jones polynomial in the variable $\tau^{-1}q$; see Example 3.10. This suggests that the $\pi=-1$ (ie $\tau=\boldsymbol{t}$) specialization of the knot invariant, viewed as a function of q, should be related to the $\pi=1$ (ie $\tau=1$) specialization, viewed as a function of $\boldsymbol{t}^{-1}q$. To make this connection precise, we further develop the theory of twistors (see [4; 5]) to define a general operator on tensor powers of $\boldsymbol{U}$ and compatible operators on its representations. In particular, we show that the twistors $\mathfrak{X}$ on representations $\boldsymbol{t}$–commute with the maps S representing slices of tangles; that is, $\mathfrak{X}\circ S=\boldsymbol{t}^x S\circ\mathfrak{X}$ for some $x\in\mathbb{Z}$.

Once this is done, we obtain the following theorem (combining Theorems 3.7 and 4.24).

Main Theorem *Let K be any oriented knot and $\lambda\in X^+$ a dominant weight. There is a functor from the category $\mathcal{OTAN}$ of oriented tangles modulo isotopy to the category $\mathcal{O}$ of $\boldsymbol{U}$–module representations which sends K to a constant $J_K^\lambda(q,\tau)\in\mathbb{Q}(q,\boldsymbol{t})^{\tau}$, which we call the* ***covering knot invariant*** *of K. Moreover, let ${}_{\mathfrak{so}}J_K^\lambda(q)=J_K^\lambda(q,1)$ and ${}_{\mathfrak{osp}}J_K^\lambda(q)=J_K^\lambda(q,\boldsymbol{t})$ denote the specializations of the covering knot invariant to $\tau=1$ and $\tau=\boldsymbol{t}$. Then*

$$ {}_{\mathfrak{osp}}J_K^\lambda(q)=\boldsymbol{t}^{\star(K,\lambda)}\,{}_{\mathfrak{so}}J_K^\lambda(\boldsymbol{t}^{-1}q) $$

for some $\star(K,\lambda)\in\mathbb{Z}$.

In particular, this shows that, after extending scalars, there is indeed a map RT as in Figure 1 (right), and in fact such a map exists for all colored link invariants of any rank. In particular, this proves that the super and nonsuper colored knot invariants of type B are essentially the same. It remains to develop an analogue of the construction in [28] to complete the picture, though difficulties abound. For example, it is not necessarily clear how to extend the categorification to $\mathbb{Q}(q,\boldsymbol{t})^{\tau}$. Moreover, the categorification of covering algebra representations is not yet developed enough to produce the analogous machinery to [28]. We hope that our results will help cast light on these remaining questions.

1.3 Organization

The paper is organized as follows. In Section 2, we recall the definition of quantum covering $\mathfrak{osp}(1|2n)$, denoted by $\boldsymbol{U}$, and set our conventions. We also develop some additional facts about representations of $\boldsymbol{U}$, specifically about dual modules and (co)evaluation morphisms, and produce a universal-$\mathcal{R}$–matrix, which we will simply denote by $\mathcal{R}$, from the quasi-$\mathcal{R}$–matrix defined in [7]. In Section 3, we interpret the maps in terms of the usual graphical calculus define an associated knot invariant. More precisely, maps are represented by a finite number of labeled, nonintersecting oriented strands such that the $\mathcal{R}$–matrix is a positive crossing, the (co)evaluation morphisms are various cups and caps, and orientation is determined by whether the associated module in the domain/range is the dual module or not. We show that the maps satisfy identities such that the associated graphical calculus is almost an framed oriented tangle invariant, and is indeed an oriented tangle invariant after renormalizing these elementary diagrams by an integer power of τ and a factor depending on the writhe. Finally, in Section 4, we use the twistor maps introduced in [4; 5] to relate the morphisms in the $\pi = \pm 1$ cases. In particular, we develop some further details about the Hopf structure and representation theory of the enhanced quantum group $\widehat{\boldsymbol{U}}$, and construct twistors on tensor products of simple modules and their duals. We then show that these twistors almost commute (up to an integer power of $\boldsymbol{t}$) with the cups, caps, and crossings, allowing us to relate the $\mathfrak{so}$ and $\mathfrak{osp}$ knot invariants.

Acknowledgements I would like to thank David Hill for first suggesting to me this project constructing the quantum covering knot invariants, and Matt Hogancamp for helping me learn about quantum knot invariants at this project's inception. I would also like to thank Aaron Lauda, Weiqiang Wang, and Ben Webster for stimulating conversations about this project, and Sacha Blumen and several referees for very helpful comments.

2 Quantum covering $\mathfrak{osp}(1|2n)$

We begin by recalling the definition of the quantum covering algebra associated to $\mathfrak{osp}(1|2n)$ and setting our notation. We then elaborate on the representation theory of this algebra.

2.1 Root data

Let $I = I_0 \sqcup I_1$ with $I_0 = \{\underline{1}, \dots, \underline{n-1}\}$ and $I_1 = \{\underline{n}\}$, and define the parity $p(i)$ of $i \in I$ by $i \in I_{p(i)}$. For $1 \le r, s \le n$, we define

$$\underline{r} \cdot \underline{s} = \begin{cases} 2 & \text{if } r = s = n, \\ 4 & \text{if } r = s \neq n, \\ -2 & \text{if } r = s \pm 1, \\ 0 & \text{otherwise,} \end{cases} \qquad d_{\underline{r}} = \tfrac{1}{2} \underline{r} \cdot \underline{r},$$

and note that $p(\underline{r}) = d_{\underline{r}} \bmod 2$. Then $(I, \cdot)$ is a bar-consistent anisotropic super Cartan datum [7]. We extend $\cdot$ to a symmetric bilinear pairing on $\mathbb{Z}[I]$ and p to a parity function $p\colon \mathbb{Z}[I] \to \mathbb{Z}/2\mathbb{Z}$. Moreover, for $\nu = i_1 + \cdots + i_t \in \mathbb{N}[I]$, we set

$$\text{(2-1)} \qquad \operatorname{ht} \nu = t, \quad \boldsymbol{p}(\nu) = \sum_{1 \le r < s \le t} p(i_r) p(i_s), \quad \bullet(\nu) = \sum_{1 \le r < s \le t} i_r \cdot i_s.$$

Let $\Phi^+ \subset \mathbb{N}[I]$ denote the set of positive roots, and set

$$\text{(2-2)} \qquad \rho = \sum_{\alpha \in \Phi^+} \alpha = \sum_{i \in I} \rho_i i \in \mathbb{N}[I].$$

Note that we have $i \cdot \rho = i \cdot i$ for all $i \in I$.

Let $Y = \mathbb{Z}[I]$ be the root lattice and $X = \operatorname{Hom}(\mathbb{Z}[I], \mathbb{Z})$ be the weight lattice, and let $\langle \cdot, \cdot \rangle\colon Y \times X \to \mathbb{Z}$ be the natural pairing. We also identify $\mathbb{Z}[I]$ as a subspace of X so that $\langle \underline{r}, s \rangle = 2 \underline{r} \cdot \underline{s} / \underline{r} \cdot \underline{r}$. If $\nu = \sum_{i \in I} \nu_i i \in \mathbb{Z}[I]$, we set

$$\text{(2-3)} \qquad \widetilde{\nu} = \sum_{i \in I} d_i \nu_i i \in \mathbb{Z}[I],$$

and note $\langle \widetilde{\nu}, \mu \rangle = \nu \cdot \mu$ for any $\nu, \mu \in \mathbb{Z}[I]$; in particular, observe that for any $i \in I$,

$$\text{(2-4)} \qquad \langle \widetilde{\rho}, i \rangle = i \cdot i.$$

Then $((I, \cdot), X, Y, \langle \cdot, \cdot \rangle)$ is the root datum associated to $\mathfrak{osp}(1|2n)$, and forgetting the parity on the root datum yields the root datum associated to $\mathfrak{so}(1+2n)$. As usual, we define the dominant weights to be $X^+ = \{\lambda \in X \mid \langle i, \lambda \rangle \ge 0 \text{ for all } i \in I\}$.

Example 2.1 When $n = 1$, we identify $X = \mathbb{Z}$ where $\langle \underline{1}, k \rangle = k$ for $k \in \mathbb{Z}$. Then $Y = \mathbb{Z}\underline{1}$ can be identified with subset $2\mathbb{Z} \subset X$. We will freely use these identifications in later examples.

Note that the weight lattice X doesn't naturally have a parity grading compatible with that on $\mathbb{Z}[I]$. However, a parity grading on X can be defined as follows. First observe that X carries an action of the Weyl group W of type B_n, and that in particular, $\lambda - w\lambda \in \mathbb{Z}[I]$ for any $\lambda \in X$. Let w_0 denote the longest element of B_n. If $\lambda \in X$, then $w_0\lambda = -\lambda$ hence $2\lambda = \lambda - w_0\lambda \in \mathbb{Z}[I]$. We write $2\lambda = \sum_{i\in I}(2\lambda)_i i$ and define

$$P(\lambda) = p(2\lambda) \equiv (2\lambda)_{\underline{n}} \bmod 2. \tag{2-5}$$

This defines a parity grading on X, though it is obviously not compatible with the grading on $\mathbb{Z}[I]$ (indeed, for any $i \in I$ we have $P(i) = p(2i) = 2p(i) \equiv 0 \bmod 2$). In particular, P is constant on cosets $X/\mathbb{Z}[I]$. This parity can be expressed explicitly in terms of the rank and weight as follows.

Lemma 2.2 *Let the notation be as above. Then* $P(\lambda) \equiv n\langle \underline{n}, \lambda\rangle \bmod 2$.

Proof Let $1 \le s \le n-1$, and for convenience, set the notation $(2\lambda)_{\underline{0}} = 0$. We have

$$\langle \underline{s}, \lambda\rangle = \tfrac{1}{2}\langle \underline{s}, 2\lambda\rangle = \tfrac{1}{2}\sum_{i\in I}(2\lambda)_i\langle \underline{s}, i\rangle = (2\lambda)_{\underline{s}} - \tfrac{1}{2}((2\lambda)_{\underline{s}+1} + (2\lambda)_{\underline{s-1}}),$$
$$\langle \underline{n}, \lambda\rangle = (2\lambda)_{\underline{n}} - (2\lambda)_{\underline{n-1}}.$$

In particular, we see that $\frac{1}{2}((2\lambda)_{\underline{s+1}} + (2\lambda)_{\underline{s}-1}) = (2\lambda)_{\underline{s}} - \langle \underline{s}, \lambda\rangle \in \mathbb{N}$; thus $(2\lambda)_{\underline{s-1}} \equiv (2\lambda)_{\underline{s+1}} \bmod 2$ for all $1 \le s \le n-1$. Therefore, $(2\lambda)_{\underline{r}} \equiv (2\lambda)_{\underline{s}} \bmod 2$ whenever $r \equiv s \bmod 2$.

In particular, since $(2\lambda)_{\underline{0}} = 0$, we see that $(2\lambda)_{\underline{s}} \equiv 0 \bmod 2$ for each $s \equiv 0 \bmod 2$. If $n \equiv 0 \bmod 2$, then $P(\lambda) \equiv (2\lambda)_{\underline{n}} \equiv 0 \bmod 2$. If $n \equiv 1 \bmod 2$, then $(2\lambda)_{\underline{n}} = \langle \underline{n}, \lambda\rangle - (2\lambda)_{\underline{n-1}} \equiv \langle \underline{n}, \lambda\rangle \bmod 2$. □

Example 2.3 When $n = 1$, recall from Example 2.1 that we identify $X = \mathbb{Z}$. Then $P(k) \equiv (1)\langle \underline{1}, k\rangle \equiv k \bmod 2$ for any $k \in \mathbb{Z}$; hence our P–grading is just the natural parity grading on $\mathbb{Z}$.

Throughout, we will consider objects graded by $\widehat{X} = X \times (\mathbb{Z}/2\mathbb{Z})$. If M is $\widehat{X}$–graded and $m \in M$ is homogeneous, we let $\|m\|$ (resp. $|m|$, $p(m)$) denote its $\widehat{X}$–degree (resp. X–degree, $\mathbb{Z}/2\mathbb{Z}$–degree or parity). Further, for $\zeta = (\lambda, \epsilon) \in \widehat{X}$, we will set $|\zeta| = \lambda$, $p(\zeta) = \epsilon$, and $P(\zeta) = P(\lambda)$. (Note that $P(\zeta)$ is not the same as $p(\zeta)$ in general! They are independent quantities.)

For $\lambda \in X$, let $\widehat{\lambda} = (\lambda, 0) \in \widehat{X}$. The set $\{(\nu, p(\nu)) \mid \nu \in \mathbb{Z}[I]\} \subset \widehat{X}$ can and will be identified with $\mathbb{Z}[I]$ via $\nu \mapsto (\nu, p(\nu))$. In particular, if $\zeta = (\lambda, \epsilon) \in \widehat{X}$ and $\nu \in \mathbb{Z}[I]$, then

$$\zeta + \nu = (\lambda + \nu, \epsilon + p(\nu)) \in \widehat{X}. \tag{2-6}$$

With that in mind, the action of W on X generalizes naturally to $\widehat{X}$ by setting

$$s_i(\lambda,\epsilon) = (\lambda,\epsilon) - \langle i,\lambda\rangle i = (\lambda - \langle i,\lambda\rangle i, \epsilon - \langle i,\lambda\rangle p(i)), \tag{2-7}$$

where $i \in I$ and s_i is the corresponding simple reflection.

Lastly, we have the parity swap function $\Pi\colon \widehat{X} \to \widehat{X}$ defined by

$$\Pi((\lambda,\epsilon)) = (\lambda, 1-\epsilon). \tag{2-8}$$

2.2 Parameters

Let $\boldsymbol{t} \in \mathbb{C}$ such that $\boldsymbol{t}^2 = -1$. Let q be a formal parameter and let τ be an indeterminate such that

$$\tau^4 = 1.$$

For convenience, we will also define

$$\pi = \tau^2.$$

If R is a commutative ring with 1, we let

$$R^\tau = R[\tau]/(\tau^4 = 1), \quad R^\pi = R[\pi]/(\pi^2 = 1). \tag{2-9}$$

Throughout, our base ring will be $\mathbb{Q}(q,\boldsymbol{t})^\tau$, though occasionally we will also refer to the subring generated by $\mathbb{Q}(q)$ and π, which we identify with $\mathbb{Q}(q)^\pi$.

We denote by $\bar{\cdot}\colon \mathbb{Q}(q,\boldsymbol{t})^\tau \to \mathbb{Q}(q,\boldsymbol{t})^\tau$ the $\mathbb{Q}(\boldsymbol{t})^\tau$–algebra automorphism satisfying $\bar{q} = \pi q^{-1}$. We also define the $\mathbb{Q}(\boldsymbol{t})$–algebra automorphism $\mathfrak{X}$ given by $\mathfrak{X}(q) = \boldsymbol{t}^{-1} q$ and $\mathfrak{X}(\tau) = \boldsymbol{t}\tau$. We caution the reader that $\bar{\cdot}$ and $\mathfrak{X}$ will be used later to denote extensions of these algebra automorphisms which are defined on $\mathbb{Q}(q,\boldsymbol{t})^\tau$–algebras and $\mathbb{Q}(q,\boldsymbol{t})^\tau$–modules.

Given an $\mathbb{Q}(q,\boldsymbol{t})^\tau$–module (or algebra) M and $x \in \{\pm 1, \pm\boldsymbol{t}\}$, the $\mathbb{Q}(q,\boldsymbol{t})$–module (or algebra) $M|_{\tau=x} = \mathbb{Q}(q,\boldsymbol{t})_x \otimes_{\mathbb{Q}(q,\boldsymbol{t})^\tau} M$, where $\mathbb{Q}(q,\boldsymbol{t})_x = \mathbb{Q}(q,\boldsymbol{t})$ is viewed as a $\mathbb{Q}(q,\boldsymbol{t})^\tau$–module on which τ acts as multiplication by x. We call this the *specialization of M at $\tau = x$*. Moreover, $\mathbb{Q}(q,\boldsymbol{t})^\tau$ has orthogonal idempotents

$$\varepsilon_{\boldsymbol{t}^k} = \tfrac{1}{4}(1 + \boldsymbol{t}^k\tau + (\boldsymbol{t}^k\tau)^2 + (\boldsymbol{t}^k\tau)^3), \quad 0 \le k \le 3, \tag{2-10}$$

such that $\mathbb{Q}(q,\boldsymbol{t})^\tau = \mathbb{Q}(q,\boldsymbol{t})\varepsilon_1 \oplus \mathbb{Q}(q,\boldsymbol{t})\varepsilon_{\boldsymbol{t}} \oplus \mathbb{Q}(q,\boldsymbol{t})\varepsilon_{-1} \oplus \mathbb{Q}(q,\boldsymbol{t})\varepsilon_{-\boldsymbol{t}}$. In particular, since $\tau\varepsilon_x = x\varepsilon_x$, we see that for any $\mathbb{Q}(q,\boldsymbol{t})^\tau$–module M,

$$M|_{\tau=x} \cong \varepsilon_x M.$$

For $k \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{Z}$, the (q,π)–quantum integers, along with quantum factorial and quantum binomial coefficients, are defined as follows (see [7]):

$$[n]_{q,\pi} = \frac{(\pi q)^n - q^{-n}}{\pi q - q^{-1}}, \qquad [n]^!_{q,\pi} = \prod_{l=1}^{n} [l]_{q,\pi},$$
(2-11)
$$\begin{bmatrix} n \\ k \end{bmatrix}_{q,\pi} = \frac{\prod_{l=n-k+1}^{n}((\pi q)^l - q^{-l})}{\prod_{m=1}^{k}((\pi q)^m - q^{-m})}.$$

If $\nu = \sum_{i \in I} \nu_i i \in \mathbb{Z}[I]$, we write

$$q_\nu = \prod_{i \in I} q^{\nu_i d_i}, \quad \tau_\nu = \prod_{i \in I} \tau^{\nu_i d_i}, \quad \pi_\nu = \prod_{i \in I} \pi^{\nu_i d_i} = \pi^{p(\nu)}, \quad \boldsymbol{t}_\nu = \prod_{i \in I} \boldsymbol{t}^{\nu_i d_i}.$$

In particular, note that $q_i = q^{d_i}$ and $\pi_i = \pi^{d_i} = \pi^{p(i)}$, and set

$$[n]_i = [n]_{q_i,\pi_i}, \quad [n]^!_i = [n]^!_{q_i,\pi_i}, \quad \begin{bmatrix} n \\ k \end{bmatrix}_i = \begin{bmatrix} n \\ k \end{bmatrix}_{q_i,\pi_i}.$$

2.3 The covering quantum group

The covering quantum group associated to $\mathfrak{osp}(1|2n)$ (as well as some variants) was introduced and studied in the series of papers starting with [7]. We will recall the necessary definitions and elementary facts now.

Remark 2.4 Note that contrary to [7] and further papers in that series, we will take coefficients in the larger ring $\mathbb{Q}(q,\boldsymbol{t})^\tau \supset \mathbb{Q}(q)^\pi$. Nevertheless, all of the results until Section 3 are essentially statements over $\mathbb{Q}(q)^\pi$ which remain true after extending scalars to $\mathbb{Q}(q,\boldsymbol{t})^\tau$, so the reader may effectively ignore τ and $\boldsymbol{t}$ for the present.

Definition 2.5 [7] The half-quantum covering group $\boldsymbol{f}$ associated to the datum $(I,\cdot)$ is the $\mathbb{N}[I]$–graded $\mathbb{Q}(q,\boldsymbol{t})^\tau$–algebra on the generators θ_i for $i \in I$ with $|\theta_i| = i$, satisfying the relations

(2-12) $$\sum_{k=0}^{b_{ij}} (-1)^k \pi^{\binom{k}{2} p(i) + k p(i) p(j)} \begin{bmatrix} b_{ij} \\ k \end{bmatrix}_i \theta_i^{b_{ij}-k} \theta_j \theta_i^k = 0 \quad (i \neq j),$$

where $b_{ij} = 1 - \langle i, j \rangle$.

The algebra $\boldsymbol{f}$ carries a nondegenerate bilinear form $(\cdot,\cdot)$ which satisfies

(2-13) $$(1,1) = 1, \quad (\theta_i,\theta_i) = \frac{1}{1 - \pi_i q_i^{-2}}, \quad (\theta_i x, y) = (\theta_i,\theta_i)(x, {}_i r(y)),$$

where ${}_ir\colon \boldsymbol{f} \to \boldsymbol{f}$ is the $\mathbb{Q}(q,\boldsymbol{t})^\tau$–linear map satisfying ${}_ir(1)=0$, ${}_ir(\theta_j)=\delta_{ij}$, and ${}_ir(xy)={}_ir(x)y+\pi^{p(i)p(x)}q^{i\cdot|x|}x\ {}_ir(y)$. (Here, and henceforth, $\delta_{x,y}$ is set to be $\delta_{x,y}=1$ if $x=y$ and 0 otherwise.) We define the $\mathbb{Q}(\boldsymbol{t})^\tau$–linear bar involution $\bar{\ }$ on $\boldsymbol{f}$ by

$$\overline{\theta_i}=\theta_i,\quad \overline{q}=\pi q^{-1}.$$

We also define the $\mathbb{Q}(q,\boldsymbol{t})^\tau$–linear anti-involution σ on $\boldsymbol{f}$ by

$$\sigma(\theta_i)=\theta_i,\quad \sigma(xy)=\sigma(y)\sigma(x),$$

and the divided powers

$$\theta_i^{(n)}=\theta_i^n/[n]_i^!.$$

Definition 2.6 [7] The quantum covering group $\boldsymbol{U}$ associated to the root datum $((I,\cdot),\ Y,\ X,\ \langle\cdot,\cdot\rangle)$ is the $\mathbb{Q}(v)^\pi$–algebra with generators E_i, F_i, K_μ, and J_μ subject to the following relations for $i,j\in I$ and $\mu,\nu\in Y$:

(2-14) $J_\mu J_\nu=J_{\mu+\nu},\quad K_\mu K_\nu=K_{\mu+\nu},\quad K_0=J_0=J_\nu^2=1,\quad J_\mu K_\nu=K_\nu J_\mu,$

(2-15) $J_\mu E_i=\pi^{\langle\mu,i\rangle}E_iJ_\mu,\quad J_\mu F_i=\pi^{-\langle\mu,i\rangle}F_iJ_\mu,$

(2-16) $K_\mu E_i=q^{\langle\mu,i\rangle}E_iK_\mu,\quad K_\mu F_i=q^{-\langle\mu,i\rangle}F_iK_\mu,$

(2-17) $E_iF_j-\pi^{p(i)p(j)}F_jE_i=\delta_{ij}\dfrac{J_{d_ii}K_{d_ii}-K_{-d_ii}}{\pi_iq_i-q_i^{-1}},$

(2-18) $\displaystyle\sum_{k=0}^{b_{ij}}(-1)^k\pi^{\binom{k}{2}p(i)+kp(i)p(j)}\begin{bmatrix}b_{ij}\\k\end{bmatrix}_{q_i,\pi_i}E_i^{b_{ij}-k}E_jE_i^k=0\quad(i\neq j),$

(2-19) $\displaystyle\sum_{k=0}^{b_{ij}}(-1)^k\pi^{\binom{k}{2}p(i)+kp(i)p(j)}\begin{bmatrix}b_{ij}\\k\end{bmatrix}_{q_i,\pi_i}F_i^{b_{ij}-k}F_jF_i^k=0\quad(i\neq j).$

Since $Y=\mathbb{Z}[I]$ in this case, we note that $\boldsymbol{U}$ is actually generated by E_i, F_i, K_i, J_i for $i\in I$. For notational convenience, we set $\widetilde{J}_\nu=J_{\widetilde{\nu}}$ and $\widetilde{K}_\nu=K_{\widetilde{\nu}}$ so that (2-17) becomes

$$E_iF_j-\pi^{p(i)p(j)}F_jE_i=\delta_{ij}\frac{\widetilde{J}_i\widetilde{K}_i-\widetilde{K}_i^{-1}}{\pi_iq_i-q_i^{-1}}.$$

We also equip $\boldsymbol{U}$ with a bar involution $\bar{\cdot}\colon \boldsymbol{U}\to\boldsymbol{U}$ extending that on $\mathbb{Q}(q,\boldsymbol{t})^\tau$ by setting $\overline{E_i}=E_i$, $\overline{F_i}=F_i$, $\overline{K_\mu}=J_\mu K_{-\mu}$, and $\overline{J_\mu}=J_\mu$.

The algebras $\boldsymbol{U}$ and $\boldsymbol{f}$ are related in the following way. Let $\boldsymbol{U}^-$ be the subalgebra generated by F_i with $i\in I$, $\boldsymbol{U}^+$ the subalgebra generated by E_i with $i\in I$, and $\boldsymbol{U}^0$ the subalgebra generated by K_ν and J_ν for $\nu\in Y$. There is an isomorphism $\boldsymbol{f}\to\boldsymbol{U}^-$

(resp. $\boldsymbol{f} \to \boldsymbol{U}^+$) defined by $\theta_i \mapsto \theta_i^- = F_i$ (resp. $\theta_i \mapsto \theta_i^+ = E_i$). As shown in [7], there is a triangular decomposition

$$\boldsymbol{U} \cong \boldsymbol{U}^- \otimes \boldsymbol{U}^0 \otimes \boldsymbol{U}^+ \cong \boldsymbol{U}^+ \otimes \boldsymbol{U}^0 \otimes \boldsymbol{U}^-.$$

There is also a root space decomposition

$$\boldsymbol{U} = \bigoplus_{\nu \in \mathbb{Z}[I]} \boldsymbol{U}_\nu, \quad \boldsymbol{U}_\nu = \{x \in \boldsymbol{U} \mid J_\mu K_\xi m = \pi^{\langle \mu, \nu \rangle} q^{\langle \xi, \nu \rangle} m\}.$$

The root space decomposition induces a parity grading via $p(u) = p(|u|)$; hence in particular, $\boldsymbol{U}$ is $\widehat{X}$–graded.

We say an algebra is a "Hopf covering algebra" if it is a $\mathbb{Z}/2\mathbb{Z}$–graded algebra over $\mathbb{Z}^\pi$ with a coproduct, antipode, and counit satisfying the usual axioms of a Hopf superalgebra, but with the braiding replaced by $x \otimes y \mapsto \pi^{p(x)p(y)} y \otimes x$. Then the algebra $\boldsymbol{U}$ is a Hopf covering algebra under the coproduct $\Delta\colon \boldsymbol{U} \to \boldsymbol{U} \otimes \boldsymbol{U}$ satisfying

$$\Delta(E_i) = E_i \otimes 1 + K_i \otimes E_i, \qquad \Delta(K_\nu) = K_\nu \otimes K_\nu,$$
$$\Delta(F_i) = F_i \otimes \widetilde{K}_i^{-1} + 1 \otimes F_i, \qquad \Delta(J_\nu) = J_\nu \otimes J_\nu;$$

the antipode $S\colon \boldsymbol{U} \to \boldsymbol{U}$ satisfying $S(xy) = \pi^{p(x)p(y)} S(y)S(x)$ for $x, y \in \boldsymbol{U}$ and

$$S(E_i) = -\widetilde{J}_i^{-1} \widetilde{K}_i^{-1} E_i, \quad S(F_i) = -F_i \widetilde{K}_i, \quad S(K_\nu) = K_\nu^{-1}, \quad S(J_\nu) = J_\nu^{-1};$$

and the counit $\epsilon\colon \boldsymbol{U} \to \mathbb{Q}(q, \boldsymbol{t})^\tau$ satisfying

$$\epsilon(E_i) = \epsilon(F_i) = 0, \quad \epsilon(K_\nu) = \epsilon(J_\nu) = 1.$$

Moreover, for $x \in \boldsymbol{f}$, we have that

$$\begin{aligned} S^{\pm 1}(x^+) &= (-1)^{\operatorname{ht} \nu} \pi^{\boldsymbol{p}(\nu)} q^{\nu \cdot \nu/2} q_{\mp \nu} \widetilde{J}_{-\nu} \widetilde{K}_{-\nu} \sigma(x)^+, \\ S^{\pm 1}(x^-) &= (-1)^{\operatorname{ht} \nu} \pi^{\boldsymbol{p}(\nu)} q^{-\nu \cdot \nu/2} q_{\pm \nu} \sigma(x)^- \widetilde{K}_\nu. \end{aligned} \tag{2-20}$$

2.4 U–modules

In this paper, a weight $\boldsymbol{U}$–module is a $\boldsymbol{U}$–module M with a $\widehat{X}$–grading compatible with the grading on $\boldsymbol{U}$, such that

$$M = \bigoplus_{\lambda \in X} M_{\lambda,0} \oplus M_{\lambda,1}, \quad M_{\lambda,s} = \{m \in M \mid p(m) = s,\ J_\mu K_\nu m = \pi^{\langle \mu, \lambda \rangle} q^{\langle \nu, \lambda \rangle} m\},$$

and each $M_{\lambda,s}$ is a free $\mathbb{Q}(q, \boldsymbol{t})^\tau$–module of finite rank. For $\lambda \in X$, let $M_\lambda = M_{\lambda,0} \oplus M_{\lambda,1}$. We also define the parity-swapped module ΠM to be M as a vector space with the same action of $\boldsymbol{U}$, but with $\Pi M_{\lambda,s} = M_{\lambda,1-s}$. We let $\mathcal{O}_{\text{fin}}$ be the category of weight $\boldsymbol{U}$–modules of finite rank over $\mathbb{Q}(q,\boldsymbol{t})^\tau$. Henceforth, we shall *always* assume our $\boldsymbol{U}$–modules are in $\mathcal{O}_{\mathrm{fin}}$.

We define the (restricted) linear dual of a $\boldsymbol{U}$–module M:

$$M^* = \bigoplus_{\lambda\in X} (M_{\lambda,0})^* \oplus (M_{\lambda,1})^*, \quad (M_{\lambda,s})^* = \mathrm{Hom}_{\mathbb{Q}(q,\boldsymbol{t})^\tau}(M_{\lambda,s}, \mathbb{Q}(q,\boldsymbol{t})^\tau).$$

This is again a free $\mathbb{Q}(q,\boldsymbol{t})^\tau$–module, which has a $\mathbb{Z}/2\mathbb{Z}$–grading induced by that of V: namely, $p(f)=0$ if $f(v)=0$ for $p(v)=1$, and vice-versa. Moreover, the Hopf superalgebra structure of $\boldsymbol{U}$ induces an action of $\boldsymbol{U}$: for $f\in V^*$ and $x\in\boldsymbol{U}$, we define $xf\in V^*$ by $xf(v)=\pi^{p(f)p(x)}f(S(x)v)$. In particular, note that V^* is a $\boldsymbol{U}$–module with $(V^*)_{\lambda,s}=(V_{-\lambda,s})^*$. While V^*_λ is therefore ambiguous, we will always take it to denote $(V^*)_\lambda$. (In other words, our convention is that taking duals has precedence over taking weight spaces.)

We can construct the $\boldsymbol{U}$–module $V\otimes W = V\otimes_{\mathbb{Q}(q,\boldsymbol{t})^\tau} W$, for any $\boldsymbol{U}$–modules V and W, via the coproduct. In particular, we have $\boldsymbol{U}$–modules $V^*\otimes V$ and $V\otimes V^*$, both of which contain a copy of the trivial module $V(0)=\mathbb{Q}(q,\boldsymbol{t})^\tau$ as a direct summand. As the following lemma shows, there are natural projection and inclusion maps to a copy of the trivial module. We borrow notation from [26].

Lemma 2.7 *Fix a $\boldsymbol{U}$–module V and recall the definition of ρ from* (2-2).

(1) *Let $\mathrm{ev}_V\colon V^*\otimes V\to\mathbb{Q}(q,\boldsymbol{t})^\tau$ be the $\mathbb{Q}(q,\boldsymbol{t})^\tau$–linear map defined by $v^*\otimes w\mapsto v^*(w)$. Then ev_V is a $\boldsymbol{U}$–module epimorphism.*

(2) *Let $\mathrm{qtr}_V\colon V\otimes V^*\to\mathbb{Q}(q,\boldsymbol{t})^\tau$ be the $\mathbb{Q}(q,\boldsymbol{t})^\tau$–linear map defined by $v\otimes w^*\mapsto \pi^{p(v)p(w)}q^{-\langle\tilde\rho,|v|\rangle}w^*(v)$. Then qtr_V is a $\boldsymbol{U}$–module epimorphism.*

(3) *Let $\mathrm{coev}_V\colon \mathbb{Q}(q,\boldsymbol{t})^\tau\to V^*\otimes V$ be the $\mathbb{Q}(q,\boldsymbol{t})^\tau$–linear map defined by $1\mapsto \sum_{b\in B}\pi^{p(b)}q^{\langle\tilde\rho,|b|\rangle}b^*\otimes b$ for some homogeneous $\mathbb{Q}(q,\boldsymbol{t})^\tau$–basis B of V. Then coev_V is a $\boldsymbol{U}$–module monomorphism.*

(4) *Let $\mathrm{coqtr}_V\colon \mathbb{Q}(q,\boldsymbol{t})^\tau\to V\otimes V^*$ be the $\mathbb{Q}(q,\boldsymbol{t})^\tau$–linear map defined by $1\mapsto \sum_{b\in B} b\otimes b^*$ for some homogeneous $\mathbb{Q}(q,\boldsymbol{t})^\tau$–basis B of V. Then coqtr_V is a $\boldsymbol{U}$–module monomorphism.*

Proof This is elementary to verify. □

2.5 Simple modules and their duals

Let $\lambda\in X^+$ and recall from [7] that $V(\lambda)$ is the simple $\boldsymbol{U}$–module of highest weight λ such that the highest weight space has even parity. Then $V(\lambda)$ has finite rank and has the same character as the $\mathfrak{so}(2n+1)$ module of highest weight λ. In particular, the lowest weight vector has weight $w_0\lambda = -\lambda$; hence the parity of the lowest weight vector of $V(\lambda)$ is $P(\lambda)$. Using standard arguments (for example, analogues of [15, Sections 5.3 and 5.16]), and considering the above analysis, we obtain the following lemma.

Lemma 2.8 *For each $\lambda \in X^+$, there is an isomorphism $V(\lambda)^* \cong \Pi^{P(\lambda)} V(\lambda)$ and a natural isomorphism $V(\lambda)^{**} \to V(\lambda)$.*

Example 2.9 In the case $n = 1$, the module $V = V(m)$ for $m \in \mathbb{Z}_{\geq 0}$ has basis $v_{m-2k} = F^{(k)} v_m$ with $0 \leq k \leq m$, where v_m is a choice of highest weight vector. Note that by convention, $p(v_m) = 0$, so $p(v_{m-2k}) \equiv k \bmod 2$. The dual module $V(m)^*$ has a dual basis v^*_{m-2k}, $0 \leq k \leq m$, and the actions of E and F are given by

$$E v^*_{m-2k} = -\pi^k (\pi q)^{m-2k} [n+1-k] v^*_{m-2(k+1)},$$
$$F v^*_{m-2k} = -\pi^k (\pi q)^{m-2k+2} [k] v^*_{m-2(k-1)}.$$

In particular, this is a simple module generated by the highest weight vector v^*_{-m}, where $|v^*_{-m}| = -|v_{-m}| = m$ and $p(v^*_{-m}) = p(v_{-m}) \equiv m \bmod 2$; hence we have an isomorphism $V(m)^* \cong \Pi^m V(m)$.

For convenience, we will use the notation

$$V(-\lambda) = V(\lambda)^*, \quad \lambda \in X^+. \tag{2-21}$$

We denote the maps in Lemma 2.7 in the case $V = V(\lambda)$ with the subscript λ instead of $V(\lambda)$; for instance, $\mathrm{ev}_\lambda = \mathrm{ev}_{V(\lambda)}$. Note that

$$\mathrm{ev}_\lambda \circ \mathrm{coev}_\lambda = \sum_{\nu \in \mathbb{N}[I]} \mathrm{rank}_{\mathbb{Q}(q,t)^\tau}(V_{\lambda-\nu}) \pi^{p(\nu)} q^{\langle \widetilde{\rho}, \lambda-\nu \rangle} = \pi^{P(\lambda)} \,\mathrm{qtr}_\lambda \circ \mathrm{coqtr}_\lambda .$$

Example 2.10 For $n = 1$, we have $\rho = \widetilde{\rho} = \underline{1}$; hence for $\lambda = m$, $\langle \widetilde{\rho}, \lambda \rangle = m$. Then

$$\mathrm{ev}_m \circ \mathrm{coev}_m = q^m + \pi q^{m-2} + \cdots + \pi^m q^{-m} = \pi^m [m+1] = \pi^m \,\mathrm{qtr}_m \circ \mathrm{coqtr}_m .$$

2.6 Further properties of the quasi-$\mathcal{R}$–matrix

Let us recall the quasi-$\mathcal{R}$–matrix from [7, Section 4].

Theorem 2.11 [7] *Let $\boldsymbol{B}$ be any $\mathbb{Q}(q,\boldsymbol{t})^\tau$–basis of $\boldsymbol{f}$ such that $\boldsymbol{B}_\nu = \boldsymbol{B} \cap \boldsymbol{f}_\nu$ is a basis of $\boldsymbol{f}_\nu$ for any $\nu \in \mathbb{N}[I]$, with $\boldsymbol{B}_0 = \{1\}$. Let $\boldsymbol{B}^* = \{b^* \mid b \in \boldsymbol{B}\}$ be the basis of $\boldsymbol{f}$ dual to $\boldsymbol{B}$ under $(\cdot,\cdot)$. Define*

$$\Theta_\nu = (-1)^{\operatorname{ht}\nu}\pi^{\boldsymbol{p}(\nu)}\pi_\nu q_\nu \sum_{b\in\boldsymbol{B}_\nu} b^- \otimes (b^*)^+ \in \boldsymbol{U}^-_{-\nu} \otimes \boldsymbol{U}^+_\nu.$$

Then if M, M' are integrable modules of $\boldsymbol{U}$, then $\Theta = \sum_\nu \Theta_\nu$ is a well-defined operator on $M \otimes M'$ which satisfies $\Delta(u)\Theta = \Theta\overline{\Delta}(u)$ as endomorphisms of $M \otimes M'$, where $\overline{\Delta}(u) = \overline{\Delta(\overline{u})}$. Moreover, Θ is independent of the choice of basis $\boldsymbol{B}$ and is invertible with inverse $\overline{\Theta}$.

Example 2.12 When $n = 1$, the quasi-$\mathcal{R}$–matrix Θ can be explicitly given by the formula

$$\Theta = \sum_{n\ge 0}(-1)^n(\pi q)^{-\binom{n}{2}}[n]^!(\pi q - q^{-1})^n F^{(n)} \otimes E^{(n)} = 1 - (\pi q - q^{-1})F \otimes E + \cdots .$$

(NB there is a typo in the power of πq in [7, Example 3.1.2].)

While $\overline{\Theta}$ can be evaluated easily, it will be more convenient to have the following alternate description of $\overline{\Theta}$ using the properties of the bilinear form on $\boldsymbol{f}$; see [7, Section 1.4].

Lemma 2.13 *With the same notation as in Theorem 2.11, $\overline{\Theta} = \sum_\nu \overline{\Theta}_\nu$ is given by*

$$\overline{\Theta}_\nu = \pi_\nu q^{\nu\cdot\nu/2} \sum_{b\in\boldsymbol{B}_\nu} b^- \otimes \sigma(b^*)^+ \in \boldsymbol{U}^-_{-\nu} \otimes \boldsymbol{U}^+_\nu.$$

Proof Let $\overline{\boldsymbol{B}} = \{\overline{b} \mid b \in \boldsymbol{B}\}$, with dual basis $\overline{\boldsymbol{B}}^*$. Then since Θ is independent of the choice of basis, we see that $\Theta_\nu = (-1)^{\operatorname{ht}\nu}\pi^{\boldsymbol{p}(\nu)}\pi_\nu q_\nu \sum_{b\in\boldsymbol{B}_\nu} \overline{b}^- \otimes (\overline{b}^*)^+$ for $\nu \in \mathbb{N}[I]$. We have

$$\overline{\Theta}_\nu = (-1)^{\operatorname{ht}\nu}\pi^{\boldsymbol{p}(\nu)} q_{-\nu} \sum_{b\in\boldsymbol{B}_\nu} \overline{(\overline{b}^-)} \otimes \overline{(\overline{b}^{*+})},$$

and note that $(\overline{x})^\pm = \overline{(x^\pm)}$, so $\overline{(\overline{b}^-)} = b^-$.

On the other hand, recall from [7, Section 1.4] the variant bilinear form $\{-,-\}$ defined by $\{x,y\} = \overline{(\overline{x},\overline{y})}$. Note that by construction, $(\overline{b}^*, \overline{b'}) = \delta_{b,b'}$. Then for any $b, b' \in \boldsymbol{B}$, we apply Lemma 1.4.3(b) of [loc. cit.] to deduce that

$$\delta_{b,b'} = \overline{(\overline{b}^*,\overline{b'})} = \{\overline{b}^*, b'\} = (-1)^{\operatorname{ht}\nu}\pi^{\boldsymbol{p}(\nu)}\pi_\nu q^{-\nu\cdot\nu/2} q_{-\nu}(\overline{b}^*, \sigma(b')).$$

(We note that while the power of π appears different from that in [loc. cit.], it is equivalent.) Therefore, we have

$$\overline{\overline{b}^*} = (-1)^{\operatorname{ht}\nu}\pi^{\boldsymbol{p}(\nu)}q^{\nu\cdot\nu/2}\pi_\nu q_\nu \sigma(b)^*.$$

Then the lemma follows from the observation that since $(\sigma(x),\sigma(y)) = (x,y)$, we have $\sigma(b)^* = \sigma(b^*)$. □

Now we will proceed to use Θ to define a universal map $\mathcal{R}\colon M\otimes N\to N\otimes M$ for any modules M and N. These constructions will be modified versions of the standard arguments in the nonsuper case; compare [15, Sections 7.3–7.6] or [18, Section 4.2 and Chapter 32].

For $1\le s<t\le 3$, let $\Theta_\nu^{st}\in \boldsymbol{U}\otimes\boldsymbol{U}\otimes\boldsymbol{U}$ be defined by

$$\Theta_\nu^{st} = (-1)^{\operatorname{ht}\nu}\pi^{\boldsymbol{p}(\nu)}\pi_\nu q_\nu \sum_{b\in \boldsymbol{B}_\nu} b_1\otimes b_2\otimes b_3,$$

where $b_s = b^-$,$b_t = (b^*)^+$, and $b_m = 1$ for $m\ne s,t$.

Proposition 2.14 *We have the following identities:*

$$(\Delta\otimes 1)(\Theta_\nu) = \sum_{\nu'+\nu''=\nu}\Theta_{\nu'}^{23}(1\otimes \widetilde{K}_{-\nu''}\otimes 1)\Theta_{\nu''}^{13},$$

$$(\overline{\Delta}\otimes 1)(\Theta_\nu) = \sum_{\nu'+\nu''=\nu}\Theta_{\nu'}^{13}(1\otimes \widetilde{J}_{\nu'}\widetilde{K}_{\nu'}\otimes 1)\Theta_{\nu''}^{23},$$

$$(1\otimes\Delta)(\Theta_\nu) = \sum_{\nu'+\nu''=\nu}\Theta_{\nu'}^{12}(1\otimes \widetilde{J}_{\nu''}\widetilde{K}_{\nu''}\otimes 1)\Theta_{\nu''}^{13},$$

$$(1\otimes\overline{\Delta})(\Theta_\nu) = \sum_{\nu'+\nu''=\nu}\Theta_{\nu'}^{13}(1\otimes \widetilde{K}_{-\nu''}\otimes 1)\Theta_{\nu''}^{12}.$$

Proof These identities are proved exactly as in [18, Section 4.2] using the analogous results in [7]. □

To construct a universal $\boldsymbol{U}$–module homomorphism from Θ, we will need some additional maps. The first is the swap map; that is, the algebra $\boldsymbol{U}\otimes\boldsymbol{U}$ is equipped with an involution $\mathfrak{s}$ defined by $\mathfrak{s}(x\otimes y) = \pi^{p(x)p(y)}y\otimes x$. This induces involutions on $\boldsymbol{U}^{\otimes m}$ by applying $\mathfrak{s}$ to sequential pairs of tensor factors; specifically, these involutions are the maps $\mathfrak{s}_{t,t+1} = 1^{\otimes t-1}\otimes\mathfrak{s}\otimes 1^{m-t-1}$, and it is not hard to see they satisfy the braid relations $\mathfrak{s}_{t-1,t}\mathfrak{s}_{t,t+1}\mathfrak{s}_{t-1,t} = \mathfrak{s}_{t,t+1}\mathfrak{s}_{t-1,t}\mathfrak{s}_{t,t+1}$. In particular, we see that to each element γ of the permutation group $\mathfrak{S}_m$, there is an automorphism $\mathfrak{s}_\gamma$ of $\boldsymbol{U}^m$;

for example, $\mathfrak{s}_{(23)} = \mathfrak{s}_{2,3}$ and $\mathfrak{s}_{(123)} = \mathfrak{s}_{1,2}\mathfrak{s}_{2,3}$. Similarly, to any tensor product of modules $N = \bigotimes_{i=1}^m M_t$ and $\gamma \in \mathfrak{S}_m$, we can define $N_\gamma = \bigotimes_{i=1}^m M_{\gamma(t)}$ and a map $\mathfrak{s}_\gamma\colon N \to N_\gamma$ given by

$$\mathfrak{s}(v) = \pi^{p(\gamma,v)} v_\gamma,$$

where $v = v_1 \otimes \cdots \otimes v_m$, $v_\gamma = v_{\gamma(1)} \otimes \cdots \otimes v_{\gamma(m)}$, and

$$p(\gamma, v) = \sum_{\substack{1 \le s < t \le n \\ \gamma(s) > \gamma(t)}} p(v_t)p(v_s).$$

These maps are compatible in the sense that for $v = \bigotimes_{t=1}^n v_t \in N$ and $u \in \boldsymbol{U}$,

$$\mathfrak{s}_\gamma(\Delta^{m-1}(u)v) = \mathfrak{s}_\gamma(\Delta^{m-1}(u))\mathfrak{s}_\gamma(v).$$

When $m = 2$, we will just write $\mathfrak{s} = \mathfrak{s}_{1,2}$.

The other ingredient is a weight-renormalization operator. This operator is induced by the weight function defined in the following lemma.

Lemma 2.15 *There exists a function* $\mathfrak{f}\colon X \times X \to (\mathbb{Q}(q, \boldsymbol{t})^\tau)^\times$ *satisfying*

$$\mathfrak{f}(\zeta + \mu', \zeta' + \nu')\mathfrak{f}(\zeta, \zeta')^{-1} = (\pi q)^{-\langle \tilde{\mu}, \zeta' \rangle} q^{-\langle \tilde{\nu}, \zeta \rangle - \mu \cdot \nu}$$

for $\zeta, \zeta' \in X$ *and* $\mu, \nu \in \mathbb{Z}[I]$. *Moreover,*

1. *the function* $\mathfrak{r}(\zeta, \zeta') = \mathfrak{f}(\zeta, \zeta')\mathfrak{f}(\zeta, -\zeta')$ *satisfies* $\mathfrak{r}(\zeta + \mu, \zeta' + \nu) = \mathfrak{r}(\zeta, \zeta')$ *for any* $\mu, \nu \in \mathbb{Z}[I]$,
2. *the function* $\mathfrak{l}(\zeta, \zeta') = \mathfrak{f}(\zeta, \zeta')\mathfrak{f}(-\zeta, \zeta')$ *satisfies* $\mathfrak{l}(\zeta + \mu, \zeta' + \nu) = \mathfrak{l}(\zeta, \zeta')$ *for any* $\mu, \nu \in \mathbb{Z}[I]$,
3. *we have* $\mathfrak{f}(\zeta, \zeta')\mathfrak{f}(-\zeta, -\zeta')^{-1} = \pi^{P(\zeta)P(\zeta')}$; *in particular,*

$$\mathfrak{l}(\zeta, \zeta') = \pi^{P(\zeta)P(\zeta')}\mathfrak{r}(\zeta, \zeta').$$

Proof It is easy to verify that such a function $\mathfrak{f}$ exists by choosing a set of coset representatives R for $\mathbb{Z}[I]$ in X. Then (3) follows easily using (2-5) and Lemma 2.2. □

Example 2.16 Let us consider the case $n = 1$. Then the function $\mathfrak{f}$ is determined by the values $\mathfrak{f}(0,0)$, $\mathfrak{f}(0,1)$, $\mathfrak{f}(1,0)$, and $\mathfrak{f}(1,1)$. Then for any $\epsilon_1, \epsilon_2 \in \{0,1\}$,

$$\mathfrak{f}(\epsilon_1 + 2s, \epsilon_2 + 2t) = \mathfrak{f}(\epsilon_1, \epsilon_2)\pi^{s\epsilon_2} q^{-t\epsilon_1 - s\epsilon_2 - 2st}.$$

By direct computation, one finds the corresponding coset functions to be

$$\mathfrak{r}(\epsilon_1 + 2s, \epsilon_2 + 2t) = \mathfrak{f}(\epsilon_1, \epsilon_2)^2 q^{\epsilon_1\epsilon_2},$$
$$\mathfrak{l}(\epsilon_1 + 2s, \epsilon_2 + 2t) = \mathfrak{f}(\epsilon_1, \epsilon_2)^2 \pi^{\epsilon_1\epsilon_2} q^{\epsilon_1\epsilon_2}.$$

Given $\boldsymbol{U}$–modules M and M', define the $\mathbb{Q}(q,\boldsymbol{t})^{\tau}$–linear bijection $\mathfrak{F}\colon M \otimes M' \to M \otimes M'$ by $\mathfrak{F}(m \otimes m') = \mathfrak{f}(|m|,|m'|)m \otimes m'$. For $1 \le s < t \le 3$, we define $\mathfrak{F}^{st}$ on $M_1 \otimes M_2 \otimes M_3$ via $\mathfrak{F}^{st}(m_1 \otimes m_2 \otimes m_3) = \mathfrak{f}(|m_s|,|m_t|)m_1 \otimes m_2 \otimes m_3$. Let ${}^{\mathfrak{F}}\Theta^{st} = \Theta^{st} \circ \mathfrak{F}^{st}$. The following results are then proven entirely analogously to the classical case; see for instance, the arguments in [15, Chapter 7].

Proposition 2.17 (Yang–Baxter equation) *As operators on* $M_1 \otimes M_2 \otimes M_3$,

$$ {}^{\mathfrak{F}}\Theta^{12} \circ {}^{\mathfrak{F}}\Theta^{13} \circ {}^{\mathfrak{F}}\Theta^{23} = {}^{\mathfrak{F}}\Theta^{23} \circ {}^{\mathfrak{F}}\Theta^{13} \circ {}^{\mathfrak{F}}\Theta^{12}. $$

Proposition 2.18 *Define* $\mathcal{R}\colon M \otimes M' \to M' \otimes M$ *by* $\mathcal{R} = \Theta \circ \mathfrak{F} \circ \mathfrak{s}$. *Then* $\mathcal{R}$ *is a* $\boldsymbol{U}$*–module isomorphism.*

We thus obtain the following crucial property of $\mathcal{R}$.

Proposition 2.19 *For any modules* M_1, M_2, *and* M_3, *let* $\mathcal{R}_{st} = {}^{\mathfrak{F}}\Theta^{st} \circ \mathfrak{s}_{(st)}$. *Then*

$$ \mathcal{R}_{12}\mathcal{R}_{23}\mathcal{R}_{12} = \mathcal{R}_{23}\mathcal{R}_{12}\mathcal{R}_{23}\colon M_1 \otimes M_2 \otimes M_3 \to M_3 \otimes M_2 \otimes M_1. $$

Remark 2.20 In [18, Section 32], it is shown that for $\mathfrak{g} = \mathfrak{sl}(2)$, which we can view as the $\pi = 1$ (ie $\tau = \pm 1$) specialization of Example 2.12, we can extend our field $\mathbb{Q}(q)$ to $\mathbb{Q}(\sqrt{q})$ and normalize so that $\mathfrak{f}$ is bimultiplicative; that is $\mathfrak{f}(m+a,n+b) = \mathfrak{f}(m,n)\mathfrak{f}(m,b)\mathfrak{f}(a,n)\mathfrak{f}(a,b)$. This is necessary for the maps $\mathcal{R}$ to satisfy the Hexagon Identities and thus define a braiding on the category of finite dimensional modules.

Note that Example 2.12 shows such a renormalization is impossible in general in the $\pi = -1$ case, so in particular the maps $\mathcal{R}$ can not be normalized to define a braiding on the category of finite dimensional weight modules. It is possible to overcome this difficulty by restricting the class of modules to those of "even" highest weight. One could also expand the definition of $\mathfrak{f}$ to a function on $\widehat{X} \times \widehat{X}$, but as a consequence the analysis in Section 4 becomes much more complex. Since it is not essential to have the braiding, we opt to not do this here.

3 Diagrammatic calculus and knot invariants

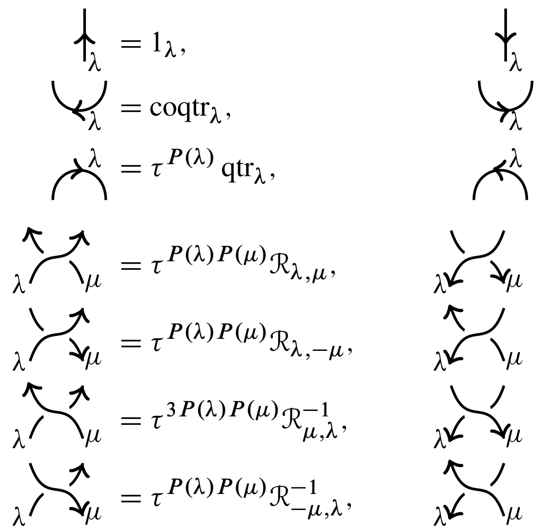
We will now interpret the $\boldsymbol{U}$–module homomorphisms in terms of planar diagrams. At first, these diagrams should be interpreted as slice diagrams; that is, diagrams together with vertical slices at various heights such that between consecutive slices is an elementary diagram corresponding to a $\boldsymbol{U}$–module homomorphism. However, we will ultimately see that diagrams which can be identified by planar isotopies yield the same morphisms.

3.1 Cups, caps, and crossings

Recall that coqtr_λ, coev_λ, qtr_λ, and ev_λ are the maps defined in Lemma 2.7 where $V = V(\lambda)$. Likewise, let $\mathcal{R}_{\pm\lambda,\pm\mu}\colon V(\pm\lambda)\otimes V(\pm\mu) \to V(\pm\mu)\otimes V(\pm\lambda)$ be the map defined in Proposition 2.18. Furthermore, we will use the notation $1_{\pm\lambda} = 1_{V(\pm\lambda)}$.

We will now begin to represent our maps via a graphical calculus in anticipation of constructing tangle invariants. Specifically, we follow [27; 1] and interpret maps between tensor products of the modules $V(\pm\lambda)$ for various $\lambda \in X^+$ as sliced oriented tangle diagrams with X^+–labeled strands; a concise exposition of this approach is lain out in [20, Chapter 3]. The elementary oriented tangle diagrams are interpreted as follows (note that while sideways-oriented crossings aren't considered elementary, we include them here for convenience in later arguments):

(3-1)



$= 1_\lambda$, $\quad = 1_{-\lambda}$,

$= \mathrm{coqtr}_\lambda$, $\quad = \tau^{3P(\lambda)}\,\mathrm{coev}_\lambda$,

$= \tau^{P(\lambda)}\,\mathrm{qtr}_\lambda$, $\quad = \mathrm{ev}_\lambda$,

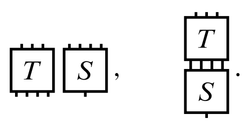
$= \tau^{P(\lambda)P(\mu)}\mathcal{R}_{\lambda,\mu}$, $\quad = \tau^{3P(\lambda)P(\mu)}\mathcal{R}_{-\lambda,-\mu}$,

$= \tau^{P(\lambda)P(\mu)}\mathcal{R}_{\lambda,-\mu}$, $\quad = \tau^{3P(\lambda)P(\mu)}\mathcal{R}_{-\lambda,\mu}$,

$= \tau^{3P(\lambda)P(\mu)}\mathcal{R}^{-1}_{\mu,\lambda}$, $\quad = \tau^{P(\lambda)P(\mu)}\mathcal{R}^{-1}_{-\mu,-\lambda}$,

$= \tau^{P(\lambda)P(\mu)}\mathcal{R}^{-1}_{-\mu,\lambda}$, $\quad = \tau^{3P(\lambda)P(\mu)}\mathcal{R}^{-1}_{\mu,-\lambda}$.

We construct more general diagrams from these elementary ones by the following constructions. If $\boxed{T}$ is some diagram denoting the morphism ϕ and $\boxed{S}$ is some diagram denoting the morphism ψ, then we can combine them in two ways:



- The first is the *horizontal* composition, which denotes the tensor product $\phi\otimes\psi$.
- The second is the *vertical* composition, which denotes the composition $\phi\circ\psi$, or zero if this composition is undefined (which is to say, when the strands on top of S don't match the number and labeling of the strands on the bottom of T).

We will say two diagrams are equal if the corresponding morphisms agree. Note that, by construction and by Proposition 2.19, the following diagrams are equal for any choice of orientation and labeling of strands:

(3-2) 

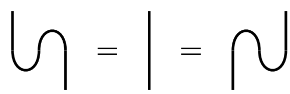
(3-3) 

In (3-2), the symbols $\boxed{T}$ and $\boxed{S}$ stand for arbitrary subdiagrams with an arbitrary number of strands protruding from the top and bottom.

3.2 Graphical identities

Now we shall prove some more substantial diagrammatic identities.

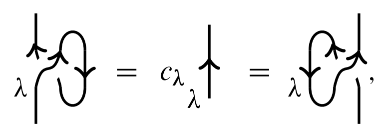
Lemma 3.1 *We have an equality of diagrams*

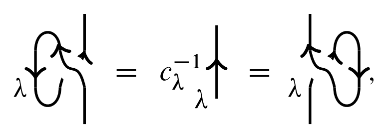


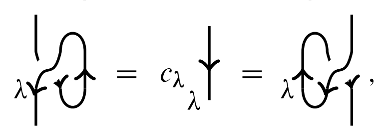
for any choice of orientation or labeling of the strand.

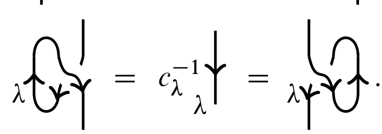
Proof This follows by choosing a homogeneous basis for the module and applying the definitions. □

Lemma 3.2 *Let $\lambda \in X^+$, and define $c_\lambda = \mathfrak{f}(\lambda,\lambda)\pi^{P(\lambda)}q^{-\langle\tilde{\rho},\lambda\rangle}$. Then we have*

(a) 

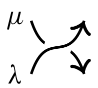
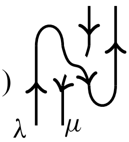
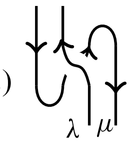
(b) 

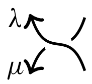
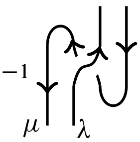
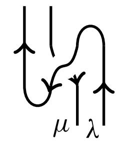
(c) 

(d) 

Proof This is easy to verify directly by evaluating the maps on either a highest-weight or lowest-weight vector. □

Lemma 3.3 *We have equalities of diagrams*

(a)  $=\ \mathfrak{r}(\mu,\lambda)$  $=\ \pi^{P(\mu)P(\lambda)}\mathfrak{r}(\mu,\lambda)$ ,

(b)  $=\ \pi^{P(\mu)P(\lambda)}\mathfrak{r}(\mu,\lambda)^{-1}$  $=\ \mathfrak{r}(\mu,\lambda)^{-1}$ 

for any $\lambda,\mu\in X^+$.

Proof The proof of (a) and (b) being similar, we shall only prove (a) here. First, unpacking the graphical representation, we see that (a) is equivalent to

$$\mathcal{R}_{\lambda,-\mu}=\mathfrak{r}(\mu,\lambda)\phi=\pi^{p(\mu)p(\lambda)}\mathfrak{r}(\mu,\lambda)\psi,$$

where ϕ and ψ are the compositions

$$\phi=(\mathrm{qtr}_\lambda\otimes 1_{-\mu}\otimes 1_\lambda)\circ(1_\lambda\otimes\mathcal{R}^{-1}_{-\lambda,-\mu}\otimes 1_\lambda)\circ(1_\lambda\otimes 1_{-\mu}\otimes\mathrm{coev}_\lambda),$$
$$\psi=(1_{-\mu}\otimes 1_\lambda\otimes\mathrm{qtr}_\mu)\circ(1_{-\mu}\otimes\mathcal{R}^{-1}_{\lambda,\mu}\otimes 1_{-\mu})\circ(\mathrm{coev}_\mu\otimes 1_\lambda\otimes 1_{-\mu}).$$

Let $B(\lambda)$ be a homogeneous basis for $V(\lambda)$. Let $v_0\in B(\lambda)_\kappa$ and $w_0\in V(\mu)_\xi$ for some $\kappa,\xi\in X$. We shall compare the images of our three maps on $v_0\otimes w_0^*$.

First, let $x=\mathcal{R}_{\lambda,-\mu}(v_0\otimes w_0^*)$ and note that

$$x=\pi^{p(w_0)p(v_0)}\mathfrak{f}(-\xi,\kappa)\sum_\nu(-1)^{\mathrm{ht}\,\nu}\pi^{\boldsymbol{p}(\nu)}\pi_\nu q_\nu\pi^{p(\nu)p(w_0)}\times\sum_{b\in\boldsymbol{B}_\nu}b^-w_0^*\otimes(b^*)^+v_0. \tag{3-4}$$

For ϕ, first let us note the effect of each map in the composition separately. The graphical representation tells us which tensor factors are impacted at each step, so we restrict our view to these tensor factors when computing these maps. First, we have the coevaluation which adds two tensor factors on the right:

$$\mathrm{coev}_\lambda(1)=\sum_{v\in\boldsymbol{B}(\lambda)}\pi^{p(v)}q^{\langle\tilde{\rho},|v|\rangle}v^*\otimes v.$$

Next, we apply $\mathcal{R}^{-1}_{-\lambda,-\mu} = \mathfrak{s} \circ \mathfrak{F}^{-1} \circ \overline{\Theta}$ to the middle tensor factors, and thus setting $y = \mathcal{R}^{-1}_{-\lambda,-\mu}(w_0^* \otimes v^*)$, we see that

$$y = \sum_\nu \mathfrak{f}(-\xi - \nu, -|v| + \nu)^{-1} \pi^{p(w_0)p(\nu)+p(\nu)p(\nu)} q^{\nu\cdot\nu/2} \sum_{b \in \boldsymbol{B}_\nu} \sigma(b^*)^+ v^* \otimes b^- w_0^*.$$

Finally, we apply the quantum trace to the two tensor factors on the left; hence we need to compute $\mathrm{qtr}(v_0 \otimes \sigma(b^*)^+ v^*)$. Since $x^*(y) = 0$ unless $\|x\| = \|y\|$ (that is, unless x and y have the same weight and parity), we can assume $|v| = \kappa + \nu$ and $p(v) = p(v_0) + p(\nu)$. Then we have

$$\begin{aligned}
&\mathrm{qtr}_\lambda(v_0 \otimes \sigma(b^*)^+ v^*) \\
&\quad = \pi^{p(v_0)} q^{-\langle \widetilde{\rho}, |v_0| \rangle} (\sigma(b^*)^+ v^*)(v_0) \\
&\quad = (-1)^{\mathrm{ht}\,\nu} \pi^{\boldsymbol{p}(\nu)+p(v_0)+p(\nu)p(v_0)+p(\nu)} q^{-\nu\cdot\nu/2-\langle\widetilde{\rho},\kappa\rangle} (\pi q)^{-\langle \widetilde{\nu},\kappa\rangle} q_{-\nu} v^*((b^*)^+ v_0).
\end{aligned}$$

Putting these computations together, we see that

$$\begin{aligned}
\phi(v_0 \otimes w_0^*) = &\sum_{v \in \boldsymbol{B}(\lambda)} \sum_\nu \sum_{b \in \boldsymbol{B}_\nu} \pi^{p(v_0)+p(\nu)} q^{\langle \widetilde{\rho}, \kappa+\nu\rangle} \\
&\times \mathfrak{f}(-\xi-\nu,-\kappa)^{-1} \pi^{p(w_0)p(v_0)+p(w_0)p(\nu)+p(\nu)p(v_0)+p(\nu)} q^{\nu\cdot\nu/2} \\
&\times (-1)^{\mathrm{ht}\,\nu} \pi^{\boldsymbol{p}(\nu)+p(v_0)+p(\nu)p(v_0)+p(\nu)} q^{-\nu\cdot\nu/2-\langle\widetilde{\rho},\kappa\rangle} (\pi q)^{-\langle\widetilde{\nu},\kappa\rangle} q_{-\nu} \\
&\times v^*((b^*)^+ v_0) b^- w_0^* \otimes v \\
= &\sum_\nu (-1)^{\mathrm{ht}\,\nu} \mathfrak{f}(-\xi-\nu,-\kappa)^{-1} (\pi q)^{-\langle\widetilde{\nu},\kappa\rangle} \pi_\nu q^{\langle\widetilde{\rho},\nu\rangle} q_{-\nu} \pi^{p(v_0)p(w_0)+\boldsymbol{p}(\nu)} \\
&\times \sum_{b\in\boldsymbol{B}_\nu} \pi^{p(w_0)p(\nu)} b^- w_0^* \otimes \Bigl(\sum_{v\in\boldsymbol{B}(\lambda)} v^*((b^*)^+ v_0) v \Bigr).
\end{aligned}$$

But note that we have the following identities:

$$\begin{aligned}
\mathfrak{f}(-\xi-\nu,-\kappa)(\pi q)^{\langle\widetilde{\nu},\kappa\rangle} &= \mathfrak{f}(-\xi,-\kappa), \\
q^{\langle\widetilde{\rho},\nu\rangle} &= q_\nu^2, \\
\sum_{v\in\boldsymbol{B}(\lambda)} v^*((b^*)^+ v_0) v &= (b^*)^+ v_0.
\end{aligned}$$

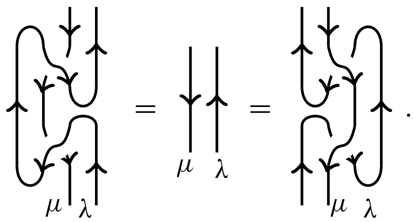
Applying these identities to the computation of $\phi(v_0 \otimes w_0^*)$, we find

$$\phi(v_0 \otimes w_0^*) = \mathfrak{r}(-\xi,\kappa)^{-1} R_{\lambda,-\mu}(v_0 \otimes w_0^*).$$

Finally, since $-\xi \in \mu + \mathbb{Z}[I]$ and $\kappa \in \lambda + \mathbb{Z}[I]$, we can apply Lemma 2.15(1) to conclude that $\phi = \mathfrak{r}(\mu,\lambda)^{-1} R_{\lambda,-\mu}$. A similar computation shows that $\psi = \mathfrak{l}(\mu,\lambda)^{-1}\mathcal{R}_{\lambda,-\mu}$, and the result then follows from Lemma 2.15. □

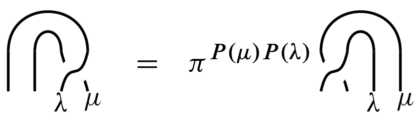
Note that since the maps in Lemma 3.3(a) and (b) are mutually inverse, we obtain the following corollary.

Corollary 3.4 *We have an equality of diagrams*



Finally, we show a somewhat more involved identity, leading us to our final result.

Lemma 3.5 *We have an equality of diagrams*



for any choice of orientation.

Proof In order to prove the identity without referring to a particular orientation, it will be convenient to introduce the following notation. Suppose $m \in V(\zeta)$ and $n \in V(-\zeta)$ for some $\zeta \in X^+$. Let us denote by (n,m) (resp. (m,n)) the evaluation $\mathrm{ev}_\zeta(n \otimes m)$ (resp. the quantum trace $\mathrm{qtr}_\zeta(m \otimes n)$). In particular, one may think of $(-,-)$ as a pairing on $V(\zeta) \oplus V(-\zeta)$ satisfying, for $v, w \in V(\zeta)$,

$$
\begin{aligned}
(v,w) &= (v^*,w^*) = 0, & (v,w^*) &= \pi^{p(v)p(w)} q^{-\langle\widetilde{\rho},|v|\rangle}(w^*,v), \\
(uv,w^*) &= \pi^{p(u)p(v)}(v,S(u)w^*), & (uw^*,v) &= \pi^{p(u)p(w)}(w^*,S(u)v).
\end{aligned}
\tag{3-5}
$$

Indeed, all the statements of (3-5) are obvious except $(uv,w^*) = \pi^{p(u)p(v)}(v,S(u)w^*)$, which follows from a simple calculation on the generators: for example,

$$
(E_i v, w^*) = \pi^{p(v)p(w)} q^{-\langle\widetilde{\rho},|v|\rangle} q_i^{-2}(-E_i \widetilde{J}_i^{-1}\widetilde{K}_i^{-1} w^*)(v) = \pi^{p(v)p(i)}(v, S(E_i)w^*).
$$

In this proof we will use the notation $(-,-)$ as shorthand for ev_ζ and qtr_ζ for both $\zeta = \lambda, \mu$ with the intended map (and highest weight) being clear from context. Using this notation, the diagram equality is equivalent to showing that the maps

$$
\begin{aligned}
\psi &= (-,-) \circ (1_{s\mu} \otimes (-,-) \otimes 1_{-s\mu}) \circ (\mathcal{R}_{s\lambda,t\mu} \otimes 1_{-s\lambda} \otimes 1_{-t\mu}), \\
\phi &= (-,-) \circ (1_{t\lambda} \otimes (-,-) \otimes 1_{-t\lambda}) \circ (1_{s\lambda} \otimes 1_{t\mu} \otimes \mathcal{R}_{-s\lambda,-t\mu})
\end{aligned}
$$

are $\pi^{P(\mu)P(\lambda)}$ multiples of each other for any choice of $s, t \in \{1,-1\}$.

Let $w \in V(s\lambda)$, $x \in V(t\mu)$, $y \in V(-s\lambda)$, and $z \in V(-t\mu)$, where $V(-\xi) = V(\xi)^*$ for $\xi \in X^+$. Then on one hand,

$$\psi(w \otimes x \otimes y \otimes z) = \sum_{\nu} \sum_{b \in \boldsymbol{B}_\nu} \pi^{p(x)p(w)} \mathfrak{f}(|x|,|w|)(-1)^{\operatorname{ht}\nu} \pi^{\boldsymbol{p}(\nu)} \pi_\nu q_\nu \pi^{p(\nu)p(x)} (b^- x, z)((b^*)^+ w, y).$$

On the other hand, using the representation of Θ in the basis $\sigma(\boldsymbol{B})$,

$$\phi(w \otimes x \otimes y \otimes z) = \sum_{\nu} \sum_{b \in \boldsymbol{B}_\nu} \pi^{p(y)p(z)} \mathfrak{f}(|z|,|y|)(-1)^{\operatorname{ht}\nu} \pi^{\boldsymbol{p}(\nu)} \pi_\nu q_\nu \pi^{p(\nu)p(z)} (x, \sigma(b)^- z)(w, \sigma(b^*)^+ y).$$

Thus to see that $\psi(w \otimes x \otimes y \otimes z) = \pi^{P(\mu)P(\lambda)} \phi(w \otimes x \otimes y \otimes z)$, and hence that $\psi = \pi^{p(\mu)p(\lambda)} \phi$ since w, x, y, z are arbitrary, it is enough to show that $l = \pi^{P(\mu)P(\lambda)} r$, where

$$l = \pi^{p(y)p(z)+p(\nu)p(z)} \mathfrak{f}(|z|,|y|)(x, \sigma(b)^- z)(w, \sigma(b^*)^+ y),$$
$$r = \pi^{p(w)p(x)+p(\nu)p(x)} \mathfrak{f}(|x|,|w|)(b^- x, z)((b^*)^+ w, y).$$

Using the properties of $(-,-)$ (see (3-5)) and S (see (2-20)), we see that

$$(x, \sigma(b)^- z)(w, \sigma(b^*)^+ y) = \pi^{p(x)p(\nu)+p(w)p(\nu)} q^{-\nu\cdot\nu+\langle\tilde{\nu},|x|\rangle} (\pi q)^{-\langle\tilde{\nu},|w|\rangle} (b^- x, z)((b^*)^+ w, y).$$

Note that l and r are both zero unless $-\|x\| = \|z\| - \nu$ and $-\|w\| = \|y\| + \nu$. In particular, l and r are both zero unless $p(y) = p(w) + p(\nu)$ and $p(z) = p(x) + p(\nu)$, in which case

$$p(y)p(z) + p(\nu)p(z) + p(x)p(\nu) + p(w)p(\nu) \equiv p(w)p(x) + p(w)p(\nu) \bmod 2.$$

Likewise, l and r are both zero unless $-|y| = |w| + \nu$ and $-|z| = |x| - \nu$, in which case

$$\mathfrak{f}(|z|,|y|) q^{-\nu\cdot\nu+\langle\tilde{\nu},|x|-|w|\rangle} = \mathfrak{f}(-|x|,-|w|).$$

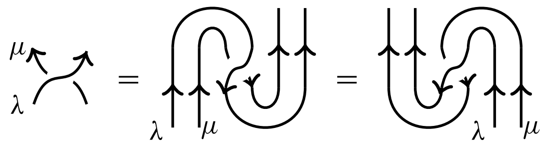
Finally, note that $\mathfrak{f}(-|x|,-|w|) = \pi^{P(-|x|)P(-|w|)} \mathfrak{f}(|x|,|w|)$. Putting these observations together,

$$l = \pi^{p(w)p(x)+p(\nu)p(x)+P(-|x|)P(-|w|)} \mathfrak{f}(|x|,|w|)(b^- x, z)((b^*)^+ w, y)$$
$$= \pi^{P(-|x|)P(-|w|)} r.$$

Since parity in X only depends on the $X/\mathbb{Z}[I]$ cosets and we have $-|x| \in \mu + \mathbb{Z}[I]$ and $-|w| \in \lambda + \mathbb{Z}[I]$, the result follows. □

Lastly, note that Lemmas 3.5 and 3.1 immediately imply the following corollary.

Corollary 3.6 *We have an equality of diagrams*

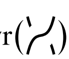
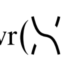


for any $\lambda, \mu \in X^+$.

3.3 Defining the tangle invariant

Recall that the *writhe* $\mathrm{wr}(T)$ of an oriented tangle T is defined by forgetting the orientation and setting

$$\mathrm{wr}(\ \ ) = 1, \quad \mathrm{wr}(\ \ ) = -1, \quad \mathrm{wr}(T) = \sum \mathrm{wr}(X),$$

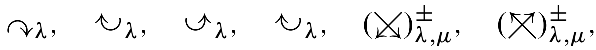



where the sum is over all crossings X in T.

Theorem 3.7 *Let T be an oriented tangle, and $\lambda \in X^+$ be a dominant weight. For any slice diagram $S(T)$ of T, let $S(T)_\lambda$ be the associated map defined by the diagrammatic calculus with strands colored by λ. Then $S(T)_\lambda$ is independent of the choice of slice diagram, and $T_\lambda = S(T)_\lambda$ is an isotopy invariant of oriented framed tangles. Moreover, if $J_T^\lambda = (\pi^{p(\lambda)}\mathfrak{f}(\lambda,\lambda)^{-1}q^{\langle\tilde{\rho},\lambda\rangle})^{\mathrm{wr}(T)}T_\lambda$, then J_T^λ is independent of the framing, hence is an invariant of T.*

Proof To prove the theorem, it suffices to show that the maps $S(T)_\lambda$ (resp. J_T^λ) are invariant under the Turaev moves for framed (resp. unframed) oriented tangles [27, Theorem 3.2; 20, Theorem 3.3, Equations (3.9)–(3.16)]. This is straightforward to check. □

We note that the proof of Theorem 3.7 actually implies a more general result, though we first need to recall some notions. The category of X^+–colored oriented tangles is the strict monoidal category whose objects are finite sequences of pairs (λ, s) where $\lambda \in X^+$ and $s \in \{\pm 1\}$, and whose morphisms from $(\lambda_a, s_a)_{1\le a\le b}$ to $(\mu_c, s_c)_{1\le c\le d}$ are tangle diagrams where the labeling and orientation of the r^{th} strand from the left at the lower (resp. upper) boundary corresponds to (λ_r, s_r) (resp. (μ_c, s_c)); see [27; 1] for more details. In particular, morphisms in this category (and thus colored tangles) are generated from the elementary morphisms



subject to relations which are simply colored versions of the Turaev moves.

We can extend Theorem 3.7 to framed multicolored tangles with the same proof. To obtain the unframed invariant, we must similarly renormalize but now by the factor $\prod_{\lambda\in X^+}(\pi^{p(\lambda)}\mathfrak{f}(\lambda,\lambda)^{-1}q^{\langle\tilde{\rho},\lambda\rangle})^{\mathrm{wr}_\lambda(T)}$, where wr_λ is defined to be the writhe where we exclude from the sum any crossings where there is a strand not labeled by λ. Therefore, we obtain the following corollary.

Corollary 3.8 *There exists a covariant functor J from the category of X^+–colored oriented tangles modulo isotopy to $\mathcal{O}_{\mathrm{fin}}$ satisfying*

$$\curvearrowright_\lambda \mapsto \tau^{P(\lambda)}\,\mathrm{ev}_\lambda,\quad \curvearrowleft_\lambda \mapsto \mathrm{qtr}_\lambda,\quad \circlearrowleft_\lambda \mapsto \mathrm{coqtr}_\lambda,\quad \circlearrowright_\lambda \mapsto \tau^{-P(\lambda)}\,\mathrm{coev}_\lambda,$$
$$(\times)^{\pm}_{\lambda,\mu} \mapsto (\pi^{p(\lambda)}\mathfrak{f}(\lambda,\lambda)^{-1}q^{\langle\tilde{\rho},\lambda\rangle})^{\pm\delta_{\lambda,\mu}}\tau^{\pm P(\lambda)P(\mu)}\mathcal{R}^{\pm}_{\lambda,\mu},$$
$$(\times)^{\pm}_{\lambda,\mu} \mapsto (\pi^{p(\lambda)}\mathfrak{f}(\lambda,\lambda)^{-1}q^{\langle\tilde{\rho},\lambda\rangle})^{\pm\delta_{\lambda,\mu}}\tau^{\mp P(\lambda)P(\mu)}\mathcal{R}^{\pm}_{-\lambda,-\mu}.$$

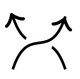
In particular, if L is an oriented colored link, then $J(L)\in\mathbb{Q}(q,\boldsymbol{t})^\tau$ is the associated quantum covering $\mathfrak{osp}(1|2n)$ colored link invariant.

Remark 3.9 Recall from Remark 2.4 that all the module homomorphisms of Section 2 are defined over the subring $\mathbb{Q}(q)^\pi$ of $\mathbb{Q}(q,\boldsymbol{t})^\tau$. We observe that whenever $P(\lambda)=0$, the maps represented by the diagrams are defined over $\mathbb{Q}(q)^\pi$. By Lemma 2.2, this holds whenever λ is an even weight (ie $\langle\underline{n},\lambda\rangle\in 2\mathbb{N}$) or n is even. In particular, this means that the functor J can be defined to the category $\mathcal{O}$ of the quantum group over $\mathbb{Q}(q)^\pi$ provided all weights are even, or n is even.

Example 3.10 Let's take $\mathfrak{g}=\mathfrak{osp}(1|2)$ and $\lambda=1$. Fix $\mathfrak{f}(1,1)=1$, and note that $\langle\tilde{\rho},\lambda\rangle=1$ and $p(\lambda)=1$. We can explicitly compute the maps represented by our diagrams on $V(1)\otimes V(1)$. Let v_1,v_{-1} be the basis of $V(1)$ from Example 2.9. Then with respect to the ordered basis $\{v_1\otimes v_1, v_1\otimes v_{-1}, v_{-1}\otimes v_1, v_{-1}\otimes v_{-1}\}$ of $V(1)\otimes V(1)$, we have

$$\Theta=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&q^{-1}-\pi q&1&0\\0&0&0&1\end{bmatrix},\qquad \mathfrak{F}=\begin{bmatrix}1&0&0&0\\0&q&0&0\\0&0&\pi q&0\\0&0&0&\pi\end{bmatrix},\qquad s=\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&\pi\end{bmatrix},$$

and thus

 $$=\begin{bmatrix}\tau&0&0&0\\0&0&\tau q&0\\0&\tau^3 q&\tau-\tau^3q^2&0\\0&0&0&\tau\end{bmatrix},$$  $$=\begin{bmatrix}\tau^3&0&0&0\\0&\tau^3-\tau q^{-2}&\tau q^{-1}&0\\0&\tau^3q^{-1}&0&0\\0&0&0&\tau^3\end{bmatrix}.$$

Note that $\pi^{p(\lambda)}q^{\langle\widetilde{\rho},\lambda\rangle} = \pi q$. Then it is easy to verify directly that

 $-\,q^2$  $= (\tau - \tau^3 q^2)$  .

Now let T be an unframed oriented link with all strands colored by λ, and fix a subdiagram which consists of two strands with either no crossing or a single crossing. Since $T^\sharp$ is isotopy invariant and independent of framing, we may assume that the strands are directed upward. Let T_+, T_0 and T_- be T with the subdiagram replaced by

,  and ,

respectively. Then using the above relation and the definition in Theorem 3.7,

$$(\pi q^{-1})J^1_{T_+} - (\pi q^3)J^1_{T_-} = (\tau - \tau^3 q^2)J^1_{T_0},$$

hence

$$(\pi q^2)^{-1}J^1_{T_+} - \pi q^2 J^1_{T_-} = (\tau q^{-1} - \tau^3 q)J^1_{T_0}.$$

Moreover, if T is the unknot, then for either orientation we have $J^1_T = \tau^3 q + \tau q^{-1}$. In particular, for any link K, we see that J^1_K is simply a multiple of the Jones polynomial of T in the variable $\tau^3 q = \tau^{-1} q$. In particular, note that using the specialization $\tau = \boldsymbol{t}$, which corresponds to $\pi = -1$, this shows the uncolored $U_q(\mathfrak{osp}(1|2))$ link invariant is equal to the $U_{\boldsymbol{t}^{-1}q}(\mathfrak{sl}_2)$ link invariant.

4 Relating $\mathfrak{so}(2n+1)$ and $\mathfrak{osp}(1|2n)$ invariants

The results of Example 3.10 suggest a connection between the specializations of the tangle invariants in Theorem 3.7. We now make this precise by extending the constructions in [5; 4]. We begin by recalling the definition of the twistor maps.

4.1 Definition of twistors

An *enhancer* ϕ is an function $\phi\colon \mathbb{Z}[I]\times X \to \mathbb{Z}$ satisfying

$$\begin{aligned}
\phi(\nu,\lambda+\mu) &\equiv \phi(\nu,\mu)+\phi(\nu,\lambda) \bmod 4 && \text{for } \nu,\mu\in\mathbb{Z}[I],\\
\phi(\nu+\mu,\lambda) &\equiv \phi(\nu,\lambda)+\phi(\mu,\lambda) \bmod 4 && \text{for } \nu,\mu\in\mathbb{Z}[I],\\
\phi(i,i) = d_i \quad &\text{and}\quad \phi(i,j)\in 2\mathbb{Z} && \text{for } i\neq j\in I,\\
\phi(i,j)-\phi(j,i) &\equiv i\cdot j + 2p(i)p(j) \bmod 4 && \text{for } i,j\in I.
\end{aligned} \tag{4-1}$$

Note that $\phi(i,i)-\phi(i,i)=0\equiv i\cdot i+2p(i)p(i) \bmod 4$ since $i\cdot i=2d_i$ and $2p(i)p(i)=2p(i)=2d_i$. In particular, note that these congruences imply that

$$\begin{aligned}\phi_4\colon \mathbb{Z}[I]\times\mathbb{Z}[I]\to\mathbb{Z}/4\mathbb{Z},\quad (\mu,\nu)\mapsto\phi(\mu,\nu) \bmod 4 \quad \text{is } \mathbb{Z}\text{–bilinear},\\ \phi(\mu,\nu)\equiv\phi(\nu,\mu)+\mu\cdot\nu+2p(\mu)p(\nu) \bmod 4 \quad \text{for } \mu,\nu\in\mathbb{Z}[I].\end{aligned} \tag{4-2}$$

Note that an enhancer can always be defined on $\mathbb{Z}[I]\times\mathbb{Z}[I]$ by defining it for I and extending in $\mathbb{Z}$–bilinearly, and then it can be extended to $\mathbb{Z}[I]\times X$ by translation along a transversal of $X/\mathbb{Z}[I]$.

When I has a unique odd element, as in the present case, the enhancer is closely related to the usual pairing.

Lemma 4.1 *Let ϕ be an enhancer. Then $\phi(\mu,\nu)+\phi(\nu,\mu)\equiv\mu\cdot\nu \bmod 4$.*

Proof It is straightforward using the fact that $|I_1|=1$. □

The *ϕ–enhanced quantum covering group $\widehat{\boldsymbol{U}}$* associated to $\boldsymbol{U}$ and the enhancer ϕ is the semidirect product of $\boldsymbol{U}$ with the algebra $\mathbb{Q}(q,\boldsymbol{t})^\tau[T_\mu,\Upsilon_\mu \mid \mu\in\mathbb{Z}[I]]$ subject to the relations

$$T_\mu T_\nu=T_{\mu+\nu},\quad \Upsilon_\mu\Upsilon_\nu=\Upsilon_{\mu+\nu},\quad T_0=\Upsilon_0=T_\nu^4=\Upsilon_\nu^4=1,\quad T_\mu\Upsilon_\nu=\Upsilon_\nu T_\mu, \tag{4-3}$$

$$T_\mu u=\boldsymbol{t}^{\langle\mu,|u|\rangle}uT_\mu,\qquad u\in\boldsymbol{U},\ \mu\in\mathbb{Z}[I], \tag{4-4}$$

$$\Upsilon_\mu u=\boldsymbol{t}^{\phi(\mu,|u|)}u\Upsilon_\mu,\quad u\in\boldsymbol{U},\ \mu\in\mathbb{Z}[I]. \tag{4-5}$$

See [5; 4] for a more formal definition. The enhanced quantum covering group has a useful $\mathbb{Q}(\boldsymbol{t})$–linear automorphism called a *twistor*. There are several ways to define such a twistor; we will need the following.

Proposition 4.2 [5, Theorems 4.3, 4.12] *Define a product $*$ on $\boldsymbol{f}$ by the following rule: if x and y are homogeneous elements of $\boldsymbol{f}$, let $x*y=\boldsymbol{t}^{\phi(|x|,|y|)}xy$. Let $(\boldsymbol{f},*)$ denote $\boldsymbol{f}$ with this multiplication. Finally, let $\mathfrak{X}\colon\mathbb{Q}(q,\boldsymbol{t})^\tau\to\mathbb{Q}(q,\boldsymbol{t})^\tau$ be the $\mathbb{Q}(\boldsymbol{t})$–linear automorphism satisfying $\mathfrak{X}(q)=\boldsymbol{t}^{-1}q$ and $\mathfrak{X}(\tau)=\boldsymbol{t}^{-1}\tau$.*

(1) *There is a $\mathbb{Q}(\boldsymbol{t})$–linear algebra isomorphism $\mathfrak{X}\colon\boldsymbol{f}\to(\boldsymbol{f},*)$ defined by*

$$\mathfrak{X}(\theta_i)=\theta_i,\quad \mathfrak{X}(q)=\boldsymbol{t}^{-1}q,\quad \mathfrak{X}(\tau)=\boldsymbol{t}^{-1}\tau.$$

(2) *Let $\mathcal{B}$ be the canonical basis of $\boldsymbol{f}$ [8]. Then $\mathfrak{X}$ on $\boldsymbol{f}$ satisfies $\mathfrak{X}(b)=\boldsymbol{t}^{\ell(b)}b$ for all $b\in\mathcal{B}$, where $\ell(b)$ is some integer depending on b.*

(3) *There is a* $\mathbb{Q}(\boldsymbol{t})$*–algebra automorphism* $\mathfrak{X}\colon \hat{\boldsymbol{U}} \to \hat{\boldsymbol{U}}$ *defined by*

$$\mathfrak{X}(E_i) = \boldsymbol{t}_i^{-1}\widetilde{T}_i\Upsilon_i E_i, \quad \mathfrak{X}(F_i) = F_i\Upsilon_{-i}, \quad \mathfrak{X}(K_\mu) = T_{-\mu}K_\mu, \quad \mathfrak{X}(J_\mu) = T_{2\mu}J_\mu,$$
$$\mathfrak{X}(T_\mu) = T_\mu, \qquad \mathfrak{X}(\Upsilon_\mu) = \Upsilon_\mu, \qquad \mathfrak{X}(q) = \boldsymbol{t}^{-1}q, \qquad \mathfrak{X}(\tau) = \boldsymbol{t}^{-1}\tau,$$

where $\widetilde{T}_\mu = \prod_{i\in I} T_{\mu_i d_i i}$ *for* $\mu = \sum_{i\in I}\mu_i i$.

(4) *For* $x \in \boldsymbol{f}[\boldsymbol{t}]$, *we have*
  (a) $\mathfrak{X}(x^+) = t_\nu^2\boldsymbol{t}^{\bullet(|x|)}\mathfrak{X}(x)^+\widetilde{T}_{|x|}\Upsilon_{|x|}$,
  (b) $\mathfrak{X}(x^-) = \mathfrak{X}(x)^-\Upsilon_{-|x|}$.

Later on, we will need some alternate versions of the results in Proposition 4.2 which we shall prove now. First, we note the following analogue of Proposition 4.2(2) for the dual canonical basis.

Lemma 4.3 *Let* $(-,-)$ *be the bilinear form on* $\boldsymbol{f}$ *defined in* (2-13). *Then*

$$\mathfrak{X}^{-1}((\mathfrak{X}(x), \mathfrak{X}(y))) = (-1)^{\boldsymbol{p}(|x|)}(x, y).$$

In particular, $\mathfrak{X}(b^*) = (-1)^{\boldsymbol{p}(b)}\boldsymbol{t}^{-\ell(b)}b^*$ *for any* $b \in \mathcal{B}$.

Proof Let $(x,y)^{\mathfrak{X}} = (-1)^{\boldsymbol{p}(|x|)}\mathfrak{X}^{-1}((\mathfrak{X}(x),\mathfrak{X}(y)))$ and observe that this is a $\mathbb{Q}(q,\boldsymbol{t})^\tau$–bilinear form on $\boldsymbol{f}[\boldsymbol{t}]$. It suffices to show that this bilinear form satisfies the defining properties of $(-,-)$, which is elementary to verify. □

Remark 4.4 Though Lemma 4.3 as stated requires $|I_1| = 1$, a version of it also holds for arbitrary enhanced quantum covering algebras. Indeed, if $|I_1| > 1$, then $\mathfrak{X}^{-1}((\mathfrak{X}(x),\mathfrak{X}(y))) = \boldsymbol{t}^{\binom{\nu}{2}}(x,y)$, where $|x| = \nu = \sum_{i\in I}\nu_i i$ and $\binom{\nu}{2} = \sum_{i\in I}\binom{\nu_i}{2}d_i$.

It will also be more convenient to have the following variant of Proposition 4.2(4b).

Lemma 4.5 *We have*

$$\mathfrak{X}(x^+) = \boldsymbol{t}_{|x|}^{-1}\widetilde{T}_{|x|}\Upsilon_{|x|}\mathfrak{X}(x)^+.$$

Proof This is true if $x = \theta_i$. It suffices to show if it is true for x, then it is true for $\theta_i x$. We have

$$\begin{aligned}\mathfrak{X}(\theta_i x^+) = \mathfrak{X}(\theta_i^+)\mathfrak{X}(x^+) &= \boldsymbol{t}_i^{-1}\widetilde{T}_i\Upsilon_i E_i \boldsymbol{t}_\nu^{-1}T_\nu\Upsilon_\nu\mathfrak{X}(x)^+ \\ &= \boldsymbol{t}_{i+\nu}^{-1}\boldsymbol{t}^{-i\cdot\nu-\phi(\nu,i)}\widetilde{T}_i\Upsilon_i\widetilde{T}_\nu\Upsilon_\nu E_i\mathfrak{X}(x)^+ \\ &= \boldsymbol{t}_{|\theta_i x|}^{-1}\boldsymbol{t}^{-i\cdot\nu-\phi(\nu,i)-\phi(i,\nu)}\widetilde{T}_{i+\nu}\Upsilon_{i+\nu}\mathfrak{X}(\theta_i x)^+.\end{aligned}$$

But then by Lemma 4.1,

$$-i\cdot\nu - \phi(\nu,i) - \phi(i,\nu) \equiv_4 -2i\cdot\nu \equiv_4 0.$$ □

4.2 $\widehat{\boldsymbol{U}}$–modules and Hopf structure

Let M be a $\boldsymbol{U}$–weight module. Then M is naturally a $\widehat{\boldsymbol{U}}$–module by defining

$$T_\mu m = \boldsymbol{t}^{\langle \mu,\lambda\rangle} m, \quad \mu \in \mathbb{Z}[I],\ m \in M_\lambda; \tag{4-6}$$

$$\Upsilon_\mu m = \boldsymbol{t}^{\phi(\mu,\lambda)} m, \quad \mu \in \mathbb{Z}[I],\ m \in M_\lambda. \tag{4-7}$$

To that end, we will call any $\widehat{\boldsymbol{U}}$–module which restricts to a $\boldsymbol{U}$–weight module and satisfies (4-6) a $\widehat{\boldsymbol{U}}$–weight module. If it additionally satisfies (4-7), we shall call it an *untwisted* $\widehat{\boldsymbol{U}}$–weight module.

In particular, any tensor product of $\widehat{\boldsymbol{U}}$–modules can be given an untwisted $\widehat{\boldsymbol{U}}$–weight module structure. However, such a procedure forgets the action of the Υ elements on the factors due to the lack of additivity in the second component of ϕ.

Example 4.6 Consider the case $n=1$. Then $\widehat{\boldsymbol{U}}$ has the untwisted weight module $\widehat{V}(1) = \mathbb{Q}(q,\boldsymbol{t})^\tau v_1 \oplus \mathbb{Q}(q,\boldsymbol{t})^\tau v_{-1}$ which is isomorphic to $V(1)$ as a $\boldsymbol{U}$–module and satisfies $T_i v_1 = \boldsymbol{t} v_1$ and $\Upsilon_{\underline{1}} v_1 = \boldsymbol{t}^{\phi(\underline{1},1)} v_1$. Then $\widehat{V}(1)\otimes\widehat{V}(1)$ is a $\boldsymbol{U}$–weight module hence has a untwisted $\widehat{\boldsymbol{U}}$–module structure, but note that

$$\Upsilon_{\underline{1}} v_1 \otimes v_1 = \boldsymbol{t}^{\phi(\underline{1},2)} v_1 \otimes v_1,$$

and by the definition of ϕ, we have $\phi(\underline{1},2) = \phi(\underline{1},\underline{1}) = 1 \neq \phi(\underline{1},1)+\phi(\underline{1},1)$.

In particular, untwisted module structures will be too naive for our purposes. Instead, we will introduce Hopf structure which will inform our classes of weight modules.

Proposition 4.7 *The algebra $\widehat{\boldsymbol{U}}$ has a Hopf covering algebra structure given by the following:*

(1) *A coassociative coproduct $\Delta\colon \widehat{\boldsymbol{U}} \to \widehat{\boldsymbol{U}} \otimes_{\mathbb{Q}(q,\boldsymbol{t})^\tau} \widehat{\boldsymbol{U}}$ extending $\Delta\colon \boldsymbol{U} \to \boldsymbol{U} \otimes_{\mathbb{Q}(q,\boldsymbol{t})^\tau} \boldsymbol{U}$ such that $\Delta(T_\mu) = T_\mu \otimes T_\mu$ and $\Delta(\Upsilon_\mu) = \Upsilon_\mu \otimes \Upsilon_\mu$ for $\mu \in \mathbb{Z}[I]$. In particular, we inductively define $\Delta^t = (\Delta \otimes 1^{t-1}) \circ \Delta^{t-1}\colon \widehat{\boldsymbol{U}} \to \widehat{\boldsymbol{U}}^{\otimes(t+1)}$ for any integer $t > 1$.*

(2) *An antipode $S\colon \widehat{\boldsymbol{U}} \to \widehat{\boldsymbol{U}}$ extending $S\colon \boldsymbol{U} \to \boldsymbol{U}$ such that $S(T_\mu) = T_{-\mu}$ and $S(\Upsilon_\mu) = \Upsilon_{-\mu}$ for $\mu \in \mathbb{Z}[I]$.*

(3) *A counit map $\epsilon\colon \widehat{\boldsymbol{U}} \to \mathbb{Q}(q,\boldsymbol{t})^\tau$ extending $\epsilon\colon \boldsymbol{U} \to \mathbb{Q}(q,\boldsymbol{t})^\tau$ such that $\epsilon(T_\mu) = \epsilon(\Upsilon_\mu) = 1$ for $\mu \in \mathbb{Z}[I]$.*

Proof To show that these maps define a Hopf structure, we need only check that these morphisms respect (4-3)–(4-5). This is obvious for (4-3), and can be quickly verified for (4-4) and (4-5) by checking it for the generators of $\boldsymbol{U}$. Finally, the coassociativity of Δ on $\widehat{\boldsymbol{U}}$ follows immediately from the coassociativity of Δ on $\boldsymbol{U}$ and the fact that T_μ and Υ_μ are grouplike elements. □

The coproduct gives us another way to define an action of $\widehat{\boldsymbol{U}}$ on tensor products of untwisted $\widehat{\boldsymbol{U}}$–weight modules. Henceforth, given $\widehat{\boldsymbol{U}}$–weight modules M and N, we let $M\widehat{\otimes}N$ denote the space $M\otimes_{\mathbb{Q}(q,\boldsymbol{t})^\pi} N$ with the $\widehat{\boldsymbol{U}}$–weight module structure induced by the coproduct on $\widehat{\boldsymbol{U}}$. (Note that in general, the module $M\widehat{\otimes}N$ is *not* untwisted!)

Example 4.8 Continuing the previous example, the action of $\Upsilon_{\underline{1}}$ on $\widehat{V}(1)\widehat{\otimes}\widehat{V}(1)$ is given by

$$\Delta(\Upsilon_{\underline{1}})v_1\otimes v_1=\boldsymbol{t}^{2\phi(\underline{1},1)}v_1\otimes v_1.$$

Another natural module to consider is the following. Given an untwisted $\widehat{\boldsymbol{U}}$–weight module M, we can construct the restricted linear dual M^*. This space is naturally a $\boldsymbol{U}$–weight module as in Section 2.4, hence has an untwisted $\widehat{\boldsymbol{U}}$ structure. On the other hand, let $M^\natural$ denote the space M^* with the action of $\widehat{\boldsymbol{U}}$ defined by $(uf)(x)=\pi^{p(f)p(u)}f(S(u)x)$. Note that $M^\natural$ is not untwisted: if $f\in(M_\lambda,s)^*$, then $|f|=-\lambda$, but nevertheless,

$$\Upsilon_\mu f=\boldsymbol{t}^{-\phi(\mu,\lambda)}f.$$

Since modules with these unorthodox actions of the Υ_μ will be important, we give the following definitions.

Definition 4.9 We say that a $\widehat{\boldsymbol{U}}$–weight module M is $*$*–twisted* if $\Upsilon_\mu m=\boldsymbol{t}^{-\phi(\mu,-\lambda)}m$ for all $m\in M_\lambda$. More generally, we say that M is a *mixed* weight module if there exists an integer $t\geq 1$ and a sequence $c=(c_1,\dots,c_t)\in\{\pm1\}^t$ such that $M_\lambda=\bigoplus_{(\lambda_s)\in(X^t)_\lambda}M_{(\lambda_s)}$, where

$$\begin{aligned}(X^t)_\lambda&=\{(\lambda_1,\dots,\lambda_t)\in X^t\mid \lambda=\lambda_1+\cdots+\lambda_s\},\\ M_{(\lambda_s)}&=\{m\in M\mid \Upsilon_\mu m=\boldsymbol{t}^{\sum_{1\leq s\leq t}c_s\phi(\mu,c_s\lambda_s)}m \text{ for all } \mu\in\mathbb{Z}[I]\}.\end{aligned}$$

We say c is the *signature* of M, and denote it by $\operatorname{sig}(M)=c$.

Remark 4.10 We note that just as any weight $\boldsymbol{U}$–module can be given a untwisted $\widehat{\boldsymbol{U}}$–module structure, it can also be given an $*$–twisted $\widehat{\boldsymbol{U}}$–module structure. Indeed, suppose M is a weight $\boldsymbol{U}$–module and define $T_i m=\boldsymbol{t}^{\langle i,|m|\rangle}m$ and $\Upsilon_i m=\boldsymbol{t}^{-\phi(i,-|m|)}$. Then this defines an action of $\widehat{\boldsymbol{U}}$, since for any $i\in I$ and $u\in\boldsymbol{U}$, $\Upsilon_i u m=\boldsymbol{t}^{-\phi(i,-(|m|+|u|))}um=\boldsymbol{t}^{\phi(i,|u|)}u\Upsilon_i m$.

In addition to classifying modules by the action of the Υ elements, another property of $\widehat{\boldsymbol{U}}$–weight modules which will be important to us is their interaction with the twistor map $\mathfrak{X}\colon \widehat{\boldsymbol{U}} \to \widehat{\boldsymbol{U}}$.

Definition 4.11 Let M be a $\widehat{\boldsymbol{U}}$–weight module. We say M carries a twistor $\mathfrak{X}$ (or $\mathfrak{X}$ is a twistor on M) if there exists a homogeneous $\mathbb{Q}(t)$–linear bijection $\mathfrak{X}\colon M \to M$ such that $\mathfrak{X}(um) = \mathfrak{X}(u)\mathfrak{X}(m)$.

Modules which carry twistors are not hard to find. Indeed, the simple $\boldsymbol{U}$–modules $V(\lambda)$ are themselves examples when given untwisted (or $*$–twisted) actions of $\widehat{\boldsymbol{U}}$.

Lemma 4.12 [4, Lemma 6.9] *Let $\lambda \in X^+$, and let $\widehat{V}(\lambda)$ be the space $V(\lambda)$ with the untwisted action of $\widehat{\boldsymbol{U}}$. There is a $\mathbb{Q}(\boldsymbol{t})$–linear map $\mathfrak{X}\colon \widehat{V}(\lambda) \to \widehat{V}(\lambda)$ which satisfies $\mathfrak{X}(v_\lambda) = v_\lambda$ and $\mathfrak{X}_\lambda(um) = \mathfrak{X}(u)\mathfrak{X}(m)$ for all $u \in \widehat{\boldsymbol{U}}$ and $m \in \widehat{V}(\lambda)$.*

In light of Lemma 2.8, it follows that the $\boldsymbol{U}$–module $V(\lambda)^*$, viewed as an untwisted $\widehat{\boldsymbol{U}}$–module, also carries a twistor. A similar argument to [4, Lemma 6.9] can be used to construct a twistor on $V(\lambda)$ with a $*$–twisted action of $\boldsymbol{U}$; hence the $\widehat{\boldsymbol{U}}$–module $\widehat{V}(\lambda)^\natural$ carries a twistor. However, this construction is not very compatible with the dual basis, since it relies on an isomorphism $V(\lambda) \to \Pi^{P(\lambda)} V(\lambda)$ and is defined by descent from the highest weight vector. To obtain a convenient definition of a twistor on the dual modules, we will define a map directly on $\widehat{V}(\lambda)^\natural$.

Define the *dual twistor* on $\widehat{\boldsymbol{U}}$ to be the map $\mathfrak{X}^\natural(u) = S \circ \mathfrak{X} \circ S^{-1}(u)$. This map is clearly a bijection, and for any $u, v \in \widehat{\boldsymbol{U}}$ we have

$$\begin{aligned}\mathfrak{X}^\natural(uv) &= S(\mathfrak{X}(S^{-1}(uv)))\\ &= \boldsymbol{t}^{2p(u)p(v)} S(\mathfrak{X}(S^{-1}(u)))S(\mathfrak{X}(S^{-1}(v)))\\ &= \boldsymbol{t}^{2p(u)p(v)} \mathfrak{X}^\natural(u)\mathfrak{X}^\natural(v).\end{aligned}$$

Therefore, it is determined by the images of the generators, which are

$$\mathfrak{X}^\natural(E_i) = t_i E_i \Upsilon_{-i}, \quad \mathfrak{X}^\natural(F_i) = \Upsilon_i F_i \widetilde{T}_i, \quad \mathfrak{X}^\natural(K_\mu) = T_{-\mu} K_\mu, \quad \mathfrak{X}^\natural(J_\mu) = T_{2\mu} J_\mu,$$
$$\mathfrak{X}^\natural(q) = \boldsymbol{t}^{-1} q, \qquad \mathfrak{X}^\natural(\tau) = \boldsymbol{t}\tau.$$

In particular, note that

$$\mathfrak{X}^\natural(x^-) = \Upsilon_\nu \mathfrak{X}(x)^- \widetilde{T}_\nu. \tag{4-8}$$

While $\mathfrak{X}^\natural$ is not an algebra automorphism of $\widehat{\boldsymbol{U}}$, it shares many properties with $\mathfrak{X}$. In particular, we have a version of Lemma 4.12.

Lemma 4.13 *Let $\lambda \in X^+$. There is a $\mathbb{Q}(\boldsymbol{t})$–linear map $\mathfrak{X}^\natural\colon \widehat{V}(\lambda) \to \widehat{V}(\lambda)$ which satisfies $\mathfrak{X}^\natural(v_\lambda) = v_\lambda$ and $\mathfrak{X}^\natural(um) = \boldsymbol{t}^{2p(u)p(m)}\mathfrak{X}^\natural(u)\mathfrak{X}^\natural(m)$ for all $u \in \widehat{\boldsymbol{U}}$ and $m \in \widehat{V}(\lambda)$.*

Proof This follows from more or less the same proof as [4, Lemmas 6.8 and 6.9]. To wit, we can identify the Verma module of highest weight λ for $\boldsymbol{U}$ with $\boldsymbol{f}$ (see [loc. cit.] for details), and in particular this is naturally an untwisted $\widehat{\boldsymbol{U}}$–module. Then we define a map $\mathfrak{X}^\natural_\lambda\colon \boldsymbol{f} \to \boldsymbol{f}$ via $\mathfrak{X}^\natural_\lambda(x) = \boldsymbol{t}^{\langle\widetilde{\nu},\lambda\rangle+\phi(\nu,\lambda-\nu)}\mathfrak{X}(x)$. Then it is straightforward to verify that $\mathfrak{X}^\natural_\lambda(ux) = \boldsymbol{t}^{2p(u)p(x)}\mathfrak{X}^\natural_\lambda(u)\mathfrak{X}^\natural_\lambda(x)$ for $x \in \boldsymbol{f}$ and $u = F_i, T_\mu, J_\mu, K_\mu, \Upsilon_\mu$. From the calculations in [loc. cit.] and the definition, we see that

$$\mathfrak{X}^\natural_\lambda(E_i x) = \boldsymbol{t}^\star E_i \mathfrak{X}^\natural_\lambda(x),$$

where

$$\star = \langle\widetilde{\nu}-\widetilde{i},\lambda\rangle - \langle\widetilde{\nu},\lambda\rangle + \phi(\nu-i,\lambda-\nu+i) - \phi(\nu,\lambda-\nu) - d_i + \langle\widetilde{i},\lambda-\nu+i\rangle - \phi(i,\nu-i).$$

Now we can simplify $\star$ and apply (4-2) to see that

$$\star \equiv \phi(\nu,i) - \phi(i,\nu) - \phi(i,\lambda-\nu) + i\cdot\nu + d_i = 2p(\nu)p(i) - \phi(i,\lambda-\nu) + d_i \mod 4,$$

and thus

$$\mathfrak{X}^\natural_\lambda(E_i x) = \boldsymbol{t}^{p(\nu)p(i)}\boldsymbol{t}_i E_i \Upsilon_{-i}\mathfrak{X}^\natural_\lambda(x) = \boldsymbol{t}^{p(\nu)p(i)}\mathfrak{X}^\natural(E_i)\mathfrak{X}^\natural_\lambda(x).$$

Finally, we note that the kernel of the projection $\boldsymbol{f} \to \widehat{V}(\lambda)$ is trivially preserved by $\mathfrak{X}^\natural_\lambda$; hence it descends to a map on $\widehat{V}(\lambda)$. □

The dual twistor $\mathfrak{X}^\natural$ is what will allow us to define a convenient twistor map on dual modules, as follows. Recall that $V(-\lambda)$ denotes the $\boldsymbol{U}$–module $V(\lambda)^*$. We will adapt this notation to $\widehat{V}(\lambda)^\natural$.

Lemma 4.14 *For $\lambda \in X^+$, let $\widehat{V}(-\lambda) = \widehat{V}(\lambda)^\natural$; that is, the space $V(\lambda)^*$ with the action of $\widehat{\boldsymbol{U}}$ induced by the antipode $S\colon \widehat{\boldsymbol{U}} \to \widehat{\boldsymbol{U}}$. Define a map $\mathfrak{X}$ on $\widehat{V}(-\lambda)$ by $\mathfrak{X}(f)(x) = \boldsymbol{t}^{2p(f)p(x)}\mathfrak{X}(f((\mathfrak{X}^\natural)^{-1}(x)))$ for homogeneous $x \in \widehat{V}(\lambda)$ and $f \in \widehat{V}(-\lambda)$. Then $\mathfrak{X}(uf) = \mathfrak{X}(u)\mathfrak{X}(f)$ for all $u \in \widehat{\boldsymbol{U}}$ and $f \in \widehat{V}(-\lambda)$.*

Proof Let $f \in \widehat{V}(-\lambda)$ and $x \in \widehat{V}(\lambda)$ be homogeneous. First, observe that since $\mathfrak{X}^\natural$ preserves the $\widehat{X}$–grading, $\mathfrak{X}(f)(x) = 0$ unless $\|x\| = \|f\|$. Moreover, if $a \in \mathbb{Q}(q,\boldsymbol{t})^\tau$,

$$\begin{aligned}\mathfrak{X}(f)(ax) &= \boldsymbol{t}^{2p(f)p(x)}\mathfrak{X}(f((\mathfrak{X}^\natural)^{-1}(ax))) \\ &= \boldsymbol{t}^{2p(f)p(x)}\mathfrak{X}(\mathfrak{X}^{-1}(a)f((\mathfrak{X}^\natural)^{-1}(x))) = a\mathfrak{X}(f)(x),\end{aligned}$$

so $\mathfrak{X}(f)$ is indeed an element of $\widehat{V}(-\lambda)$.

Now suppose $u \in \widehat{\boldsymbol{U}}$. We compute that

$$\begin{aligned}\mathfrak{X}(uf)(x) &= \boldsymbol{t}^{2p(uf)p(x)}\mathfrak{X}((uf)((\mathfrak{X}^{\natural})^{-1}(x)))\\ &= \boldsymbol{t}^{2p(u)p(x)+2p(f)p(x)}\mathfrak{X}(\pi^{p(u)p(f)}f(S(u)(\mathfrak{X}^{\natural})^{-1}(x)))\end{aligned}$$

and

$$\begin{aligned}&\mathfrak{X}(u)\mathfrak{X}(f)(x)\\ &\quad= \boldsymbol{t}^{2p(f)p(ux)}\pi^{p(u)p(f)}\mathfrak{X}(f((\mathfrak{X}^{\natural})^{-1}(S(\mathfrak{X}(u))x)))\\ &\quad= \boldsymbol{t}^{2p(f)p(u)+2p(f)p(x)+2p(u)p(x)}\pi^{p(u)p(f)}\mathfrak{X}(f((\mathfrak{X}^{\natural})^{-1}(S(\mathfrak{X}(u)))\mathfrak{X}'^{-1}(x)))\\ &\quad= \boldsymbol{t}^{2p(f)p(x)+2p(u)p(x)}\mathfrak{X}(\pi^{p(u)p(f)}f(S(u)(\mathfrak{X}^{\natural})^{-1}(x))).\end{aligned}$$

Therefore, $\mathfrak{X}(uf) = \mathfrak{X}(u)\mathfrak{X}(f)$. □

4.3 Twistor on tensor products

Now let us return to the question of relating the $\mathfrak{osp}(1|2)$ and $\mathfrak{sl}(2)$ link invariants. Since the invariants arise from maps between tensor products of simple modules and their duals, we shall also need variants of the twistor maps on the corresponding $\widehat{\boldsymbol{U}}$–modules. In the following, we shall define a number of versions of $\mathfrak{X}$ in different settings. However, they will all be compatible in natural ways, so rather than label these maps differently, we shall treat them en suite as an operator on $\widehat{\boldsymbol{U}}$ and its modules.

The following proposition takes the first step in this direction by showing that there is a natural extension of the twistor maps to tensor powers of $\boldsymbol{U}$.

Proposition 4.15 *For each positive integer t, there exists a $\mathbb{Q}(\boldsymbol{t})$–algebra isomorphism $\mathfrak{X}\colon \widehat{\boldsymbol{U}}^{\otimes t+1} \to \widehat{\boldsymbol{U}}^{\otimes t+1}$ which satisfies*

$$\mathfrak{X}(x \otimes y) = \mathfrak{X}(x)\Delta^{s}(\Upsilon_{|y|}) \otimes \Delta^{s'}(\widetilde{T}_{|x|}\Upsilon_{|x|})\mathfrak{X}(y)$$

for any positive integers s, s' satisfying $s + s' = t + 1$, $x \in \widehat{\boldsymbol{U}}^{\otimes s}$, and $y \in \widehat{\boldsymbol{U}}^{\otimes s'}$. Moreover, $\Delta^t(\mathfrak{X}(x)) = \mathfrak{X}(\Delta^t(x))$ for any $x \in \boldsymbol{U}$.

Proof Define $\mathfrak{X}'\colon \widehat{\boldsymbol{U}}^{\otimes t+1} \to \widehat{\boldsymbol{U}}^{\otimes t+1}$ as follows: for $x = \bigotimes_{s=1}^{t+1} x_s \in \widehat{\boldsymbol{U}}^{\otimes t+1}$, let $\mathfrak{X}(x) = \bigotimes_{s=1}^{t+1} \mathfrak{X}(x)_s$, where

$$\mathfrak{X}(x)_s = \widetilde{T}_{|x_1|+\cdots+|x_{s-1}|}\Upsilon_{|x_1|+\cdots+|x_{s-1}|}\mathfrak{X}(x_s)\Upsilon_{|x_{s+1}|+\cdots+|x_{t+1}|}. \tag{4-9}$$

It is elementary to check that

$$\mathfrak{X}(x \otimes y) = \mathfrak{X}(x)\Delta^{s}(\Upsilon_{|y|}) \otimes \Delta^{s'}\widetilde{T}_{|x|}\Upsilon_{|x|}\mathfrak{X}(y)$$

for any positive integers s, s' satisfying $s + s' = t + 1$, $x \in \widehat{\boldsymbol{U}}^{\otimes s}$, and $y \in \widehat{\boldsymbol{U}}^{\otimes s'}$. Moreover, since $\mathfrak{X}$ on $\widehat{\boldsymbol{U}}$ is a bijection, it is easy to see that so is $\mathfrak{X}$ on $\widehat{\boldsymbol{U}}^{\otimes t+1}$.

We will prove that $\mathfrak{X}$ is an algebra homomorphism, and hence isomorphism, by induction. Since $\mathfrak{X}$ on $\widehat{\boldsymbol{U}}$ is an isomorphism, let us assume $\mathfrak{X}$ on $\widehat{\boldsymbol{U}}^{\otimes t}$ is an isomorphism. Then for $x, w \in \widehat{\boldsymbol{U}}^{\otimes t}$ and $y, z \in \widehat{\boldsymbol{U}}$,

$$\begin{aligned}
&\mathfrak{X}(x \otimes y)\mathfrak{X}(w \otimes z) \\
&\quad = (\mathfrak{X}(x)\Upsilon_{|y|} \otimes \widetilde{T}_{|x|}\Upsilon_{|x|}\mathfrak{X}(y))(\mathfrak{X}(w)\Upsilon_{|z|} \otimes \widetilde{T}_{|w|}\Upsilon_{|w|}\mathfrak{X}(z)) \\
&\quad = \pi^{p(y)p(w)}\mathfrak{X}(x)\Upsilon_{|y|}\mathfrak{X}(w)\Upsilon_{|z|} \otimes \widetilde{T}_{|x|}\Upsilon_{|x|}\mathfrak{X}(y)\widetilde{T}_{|w|}\Upsilon_{|w|}\mathfrak{X}(z) \\
&\quad = \pi^{p(y)p(w)}\boldsymbol{t}^{\phi(|y|,|w|)-\phi(|w|,|y|)-|w|\cdot|y|}\mathfrak{X}(xw)\Upsilon_{|yz|} \otimes \widetilde{T}_{|xw|}\Upsilon_{|xw|}\mathfrak{X}(yz) \\
&\quad = \pi^{p(y)p(w)}\boldsymbol{t}^{2p(y)p(z)}\mathfrak{X}(xw \otimes yz) = \mathfrak{X}(\pi^{p(y)p(w)}xw \otimes yz) \\
&\quad = \mathfrak{X}((x \otimes y)(w \otimes z)).
\end{aligned}$$

This completes the induction showing $\mathfrak{X}$ on $\widehat{\boldsymbol{U}}^{\otimes t+1}$ is an isomorphism as claimed. Finally, showing that $\mathfrak{X}$ commutes with Δ^t is straightforward using (4-9) and checking on the generators. □

Now that we have a viable twistor map on tensor powers of $\widehat{\boldsymbol{U}}$, we need an analogue on the tensor powers of modules. In particular, suppose we have a collection of $\widehat{\boldsymbol{U}}$ modules which are untwisted or $*$–twisted, and which carry twistors. We will produce a twistor on the tensor product of these modules.

As might be suggested by (4-9), this is not as simple as taking the tensor power of the twistors. A version of such a twistor is produced in [4, Proposition 6.11] by rescaling the tensor product of twistors by a power of $\boldsymbol{t}$ given by a function of the weights of the tensor factors. We will do something similar, but it turns out that we will need functions which depend not only on the weights of tensor factors but also their parities, as well as the signature of the tensor product.

Lemma 4.16 *Let $c = (c_1, c_2)$ where $c_1, c_2 \in \{1, -1\}$. There exists a function $\kappa_c\colon \widehat{X}^2 \to \mathbb{Z}$ satisfying $\kappa((0,0),\zeta) \equiv \kappa(\zeta,(0,0)) \equiv 0 \bmod 4$ and*

$$\begin{aligned}
&\kappa_c(\zeta + \mu, \zeta' + \nu) - \kappa_c(\zeta, \zeta') \\
&\quad \equiv \langle \widetilde{\mu}, |\zeta'| \rangle + c_2\phi(\mu, c_2|\zeta'|) + 2p(\zeta)p(\nu) + c_1\phi(\nu, c_1|\zeta|) + \mu \cdot \nu + \phi(\mu, \nu) \bmod 4
\end{aligned}$$

for all $\zeta, \zeta' \in \widehat{X}$ and $\mu, \nu \in \mathbb{Z}[I]$.

Proof Fix $c = (c_1, c_2)$ where $c_1, c_2 \in \{1, -1\}$. Note that it suffices to show such a function $\kappa = \kappa_c$ exists on each coset of $\mathbb{Z}[I] \times \mathbb{Z}[I]$ (where as in (2-6), we view $\mathbb{Z}[I]$

as a subset of $\widehat{X}$), so fix a set of representatives C of $\widehat{X}/\mathbb{Z}[I]$. For $\zeta_0, \zeta_1 \in C$, set

$$\begin{aligned}\kappa(\zeta_0+\mu, \zeta_1+\nu) \\ &= \langle \widetilde{\mu}, |\zeta_1| \rangle + c_2\phi(\mu, c_2|\zeta_1|) + 2p(\zeta_0)p(\nu) + c_1\phi(\nu, c_1|\zeta_0|) + \mu\cdot\nu + \phi(\mu,\nu).\end{aligned}$$

It is elementary to verify that this has the desired properties. □

We henceforth suppose we have fixed choices of κ_c for each $c \in \{1,-1\}^2$. We can extend κ naturally to larger powers of $\widehat{X}$. Let $t > 1$ be a positive integer and fix a sequence $c = (c_s) \in \{\pm 1\}^t$. Let $\kappa_c\colon \widehat{X}^t \to \mathbb{Z}$ be the function defined by

$$\kappa_c(\zeta) = \sum_{1\le r<s\le t} \kappa_{(c_r,c_s)}(\zeta_r,\zeta_s), \quad \zeta = (\zeta_s) \in \widehat{X}^t.$$

Then if we have $\zeta = (\zeta_s)$ and $\zeta' = (\zeta'_s)$ in $\widehat{X}^t$ with $\zeta'_s = \zeta_s + \delta_{r,s} i$ for some $1 \le r \le t$, then

$$\kappa(\zeta') - \kappa(\zeta) \equiv \sum_{r<s\le t} (\langle \widetilde{i}, |\zeta_s| \rangle + c_s\phi(i, c_s|\zeta_s|)) + \sum_{1\le s'<r} (2p(\zeta_{s'})p(i) + c_{s'}\phi(i, c_{s'}|\zeta_{s'}|))$$

modulo 4.

We can observe some convenient properties of the maps κ_c.

Lemma 4.17 *Let $c = (c_s) \in \{\pm 1\}^t$, and let $\zeta = (\zeta_s)$ and $\zeta' = (\zeta'_s)$ in $\widehat{X}^t$.*

(1) *Let $1 \le r \le t$, and define $c_{\le r} = (c_1, \dots, c_r)$ and $c_{>r} = (c_{r+1}, \dots, c_t)$. Likewise, define $\zeta''_{\le r} = (\zeta''_1, \dots, \zeta''_r)$ and $\zeta''_{>r} = (\zeta''_{r+1}, \dots, \zeta''_t)$ for any $\zeta'' = (\zeta''_s) \in \widehat{X}^t$. Then*

$$\kappa_c(\zeta,\zeta') = \kappa_{c_{\le r}}(\zeta_{\le r}, \zeta'_{\le r}) + \kappa_{c_{>r}}(\zeta_{>r}, \zeta'_{>r}) + \sum_{1\le s<r<s'\le t} \kappa_{(c_s,c_{s'})}(\zeta_s, \zeta'_{s'}).$$

(2) *Suppose that there exists $1 \le r < t$ such that $\zeta_r = \zeta'_r + \nu$, $\zeta_{r+1} = \zeta'_{r+1} - \nu$, and $\zeta_s = \zeta'_s$ for $s \ne r, r+1$ and some $\nu \in \mathbb{Z}[I]$. Then*

$$\begin{aligned}\kappa_c(\zeta) - \kappa_c(\zeta') \\ &= \langle \widetilde{\nu}, \zeta_{r+1} \rangle + c_{r+1}\phi(\nu, c_{r+1}\zeta_{r+1}) + 2p(\nu)p(\zeta_r) - c_r\phi(\nu, c_r\zeta_r) - \nu\cdot\nu - \phi(\nu,\nu).\end{aligned}$$

(3) *For any $\zeta \in \widehat{X}$ and $c_1 = \pm 1$, we have*

$$\kappa_{c_1,\pm1,\mp1}(\zeta+\widehat{\nu}, (\pm\lambda,0), (\mp\lambda,0)) = \kappa_{c_1,\pm1,\mp1}(\zeta, (\pm\lambda,0), (\mp\lambda,0)),$$

$$\kappa_{\pm1,\mp1,c_1}((\pm\lambda,0), (\mp\lambda,0), \zeta+\widehat{\nu}) = \kappa_{\pm1,\mp1,c_1}((\pm\lambda,0), (\mp\lambda,0), \zeta).$$

Proof We note that (1) is an immediate consequence of the definition of κ_c. On the other hand, (2) and (3) both follow from direct computations and the definition. □

The functions κ_c allows us to define a twistor on tensor product modules as follows.

Proposition 4.18 *Let $M_1, M_2, \dots, M_t$ be untwisted or $*$–twisted $\widehat{\boldsymbol{U}}$–modules carrying twistors and let $M = M_1 \widehat{\otimes} M_2 \otimes \cdots \widehat{\otimes} M_t$ be the $\widehat{\boldsymbol{U}}^{\otimes t}$–module (and hence a mixed $\widehat{\boldsymbol{U}}$–module via Δ^{t-1}) with the natural action. Set $c = \mathrm{sig}(M) = (c_1, \dots, c_t)$. Then the automorphism*

$$\mathfrak{X}(m_1 \otimes \cdots \otimes m_t) = \boldsymbol{t}^{\kappa_c((\|m\|_i))} \mathfrak{X}(m_1) \otimes \cdots \otimes \mathfrak{X}(m_t)$$

satisfies

$$\mathfrak{X}((x_1 \otimes \cdots \otimes x_t)(m_1 \otimes \cdots \otimes m_t)) = \mathfrak{X}(x_1 \otimes \cdots \otimes x_t)\mathfrak{X}(m_1 \otimes \cdots \otimes m_t).$$

In particular, $\mathfrak{X}(um) = \mathfrak{X}(u)\mathfrak{X}(m)$ for $u \in \widehat{\boldsymbol{U}}$ and $m \in M$.

Proof First, observe it is enough to show

$$\mathfrak{X}((1^{s-1} \otimes x_s \otimes 1^{t-s})(m_1 \otimes \cdots \otimes m_t)) = \mathfrak{X}(1^{s-1} \otimes x_s \otimes 1^{t-s})\mathfrak{X}(m_1 \otimes \cdots \otimes m_t),$$

where $1 \le s \le t$ and x_s is a generator of $\widehat{\boldsymbol{U}}$. This is trivial when x_s is K_μ, J_μ, T_μ and Υ_μ for some $\mu \in \mathbb{Z}[I]$ so it suffices to check the case $x_s = E_i, F_i$ for $i \in I$. To do this, let us make our equations more compact with the following notation: for $m_1 \otimes \cdots \otimes m_t \in M$, let

$$\begin{aligned} m_{<s} &= m_1 \otimes \cdots \otimes m_{s-1}, & m_{>s} &= m_{s+1} \otimes \cdots \otimes m_t, \\ \mathfrak{X}(m)_{<s} &= \mathfrak{X}(m_1) \otimes \cdots \otimes \mathfrak{X}(m_{s-1}), & \mathfrak{X}(m)_{>s} &= \mathfrak{X}(m_{s+1}) \otimes \cdots \otimes \mathfrak{X}(m_t), \\ \|m\|_{<s} &= (\|m_1\|, \dots, \|m_{s-1}\|), & \|m\|_{>s} &= (\|m_{s+1}\|, \dots, \|m_t\|), \\ \phi'(i, m_{<s}) &= \sum_{1 \le r < s} c_r \phi(i, c_r |m_r|), & \phi''(i, m_{>s}) &= \sum_{s < r \le t} c_r \phi(i, c_r |m_r|). \end{aligned}$$

Using this notation, we compute that

$$\begin{aligned} &\mathfrak{X}((1^{s-1} \otimes E_i \otimes 1^{t-s})(m_{<s} \otimes m_s \otimes m_{>s})) \\ &\quad = \mathfrak{X}(\pi_i^{p(m_{<s})} m_{<s} \otimes E_i m_s \otimes m_{>s}) \\ &\quad = \boldsymbol{t}^{2p(i)p(m_{<s}) + \kappa_c(\|m\|_{<s}, \|m_s\| + \widehat{i}, \|m\|_{>s})} \pi^{p(i)p(m_{<s})} \mathfrak{X}(m)_{<s} \otimes \mathfrak{X}(E_i m_s) \otimes \mathfrak{X}(m_{>s}) \\ &\quad = \boldsymbol{t}^{\kappa(\|m_{<s}\|, \|m_s\|, \|m_{>s}\|) + \phi'(i, m_{<s}) + \phi''(i, m_{>s}) + \langle \widetilde{i}, |m_{>s}| \rangle} \\ &\qquad \times \pi^{p(i)p(m_{<s})} \mathfrak{X}(m)_{<s} \otimes \mathfrak{X}(E_i)\mathfrak{X}(m_s) \otimes \mathfrak{X}(m_{>s}) \\ &\quad = \boldsymbol{t}^{\kappa(\|m_{<s}\|, \|m_s\|, \|m_{>s}\|)} \pi^{p(i)p(m_{<s})} \\ &\qquad \times (\Upsilon_i^{\otimes(s-1)} \mathfrak{X}(m)_{<s}) \otimes (\mathfrak{X}(E_i)\mathfrak{X}(m_s)) \otimes ((\Upsilon_i \widetilde{T}_i)^{\otimes(t-s)} \mathfrak{X}(m_{>s})) \\ &\quad = \mathfrak{X}(1^{s-1} \otimes E_i \otimes 1^{t-s}) \mathfrak{X}(m_{<s} \otimes m_s \otimes m_{>s}). \end{aligned}$$

The case $x_s = F_i$ proceeds similarly. □

We now have defined a family of compatible twistor maps on untwisted modules, $*$–twisted modules and their tensor products. Moreover, the twistor maps on tensor products of modules are compatible with one another in the following sense. Let $M_1, \dots, M_t$, $c_1, \dots, c_s$ and M be as in Proposition 4.18. Fix $1 \le r \le t$ and set $m_{\le r} = m_1 \otimes \cdots \otimes m_r$ and $m_{>r} = m_{r+1} \otimes \cdots \otimes m_t$. Then by Lemma 4.17(1),

$$\mathfrak{X}(m_{\le r} \otimes m_{>r}) = \Bigg(\prod_{1 \le s \le r < s' \le t} \boldsymbol{t}^{\kappa_{c_s, c_{s'}}(\|m_s\|, \|m_{s'}\|)} \Bigg) \mathfrak{X}(m_{\le r}) \otimes \mathfrak{X}(m_{>r}). \tag{4-10}$$

4.4 Twisting the crossings, caps, and cups

We have now lain the groundwork for studying the atomic maps in our graphical calculus from Section 3 under the twistor functor. Specifically, we will show that the twistor almost commutes with cups, caps, and crossings up to a factor of an integral power of $\boldsymbol{t}$, where the power depends on the map. We begin by considering the cups and caps on their domains of definition.

Proposition 4.19 *Let $\lambda \in X^+$. Then the map ev_λ (resp. qtr_λ, coev_λ, and coqtr_λ) viewed as a function $\widehat{V}(-\lambda) \widehat{\otimes} \widehat{V}(\lambda) \to \mathbb{Q}(q, \boldsymbol{t})^\tau$ (resp. $\widehat{V}(\lambda) \widehat{\otimes} \widehat{V}(-\lambda) \to \mathbb{Q}(q, \boldsymbol{t})^\tau$, $\mathbb{Q}(q, \boldsymbol{t})^\tau \to \widehat{V}(-\lambda) \widehat{\otimes} \widehat{V}(\lambda)$, and $\mathbb{Q}(q, \boldsymbol{t})^\tau \to \widehat{V}(\lambda) \widehat{\otimes} \widehat{V}(-\lambda)$) is a $\widehat{\boldsymbol{U}}$–module homomorphism. Moreover, we have*

1. $\mathrm{ev}_\lambda \mathfrak{X} = \boldsymbol{t}^{\kappa_{(-1,1)}((-\lambda,0),(\lambda,0))} \mathfrak{X}\, \mathrm{ev}_\lambda$;
2. $\mathrm{qtr}_\lambda \mathfrak{X} = \boldsymbol{t}^{\kappa_{(1,-1)}((\lambda,0),(-\lambda,0)) - \langle \widetilde{\rho}, \lambda \rangle} \mathfrak{X}\, \mathrm{qtr}_\lambda$;
3. $\mathrm{coev}_\lambda \mathfrak{X} = \boldsymbol{t}^{-\kappa_{(-1,1)}((-\lambda,0),(\lambda,0)) + \langle \widetilde{\rho}, \lambda \rangle} \mathfrak{X}\, \mathrm{coev}_\lambda$;
4. $\mathrm{coqtr}_\lambda \mathfrak{X} = \boldsymbol{t}^{-\kappa_{(1,-1)}((\lambda,0),(-\lambda,0))} \mathfrak{X}\, \mathrm{coqtr}_\lambda$.

Proof First, observe that since these maps are $\boldsymbol{U}$–module homomorphisms, they preserve weight spaces hence preserve the action of T_i for $i \in I$. Therefore, it only remains to check that they commute with the action of Υ_i for $i \in I$, As the arguments are all similar, let us show this for ev_λ. Let $f \in \widehat{V}(-\lambda)$ and $x \in V(\lambda)$. Then

$$\Upsilon_i\, \mathrm{ev}_\lambda(f \otimes x) = \boldsymbol{t}^{\phi(i,0)}\, \mathrm{ev}_\lambda(f \otimes x) = f(x).$$

On the other hand, $\Upsilon_i(f \otimes x) = (\Upsilon_i f) \otimes (\Upsilon_i x) = \boldsymbol{t}^{-\phi(i,-|f|)+\phi(i,|x|)} f \otimes x$, hence

$$\mathrm{ev}_\lambda(\Upsilon_i(f \otimes x)) = \boldsymbol{t}^{\phi(i,|x|)-\phi(i,-|f|)} f(x).$$

However, since $f(x) = 0$ if $|f| \neq -|x|$, we see that

$$\mathrm{ev}_\lambda(\Upsilon_i(f \otimes x)) = \boldsymbol{t}^{\phi(i,|x|)-\phi(i,|x|)} f(x) = f(x) = \Upsilon_i\, \mathrm{ev}_\lambda(f \otimes x).$$

In order to verify (1)–(4), it suffices to compute the images $\mathfrak{X}(b^- v_\lambda \otimes (b^- v_\lambda)^*)$ and $\mathfrak{X}((b^- v_\lambda)^* \otimes b^- v_\lambda)$ for $b \in \mathcal{B}_\nu = \mathcal{B} \cap \boldsymbol{f}_\nu$. We compute directly that

$$\mathfrak{X}(b^- v_\lambda) = \boldsymbol{t}^{\ell(b) - \phi(\nu,\lambda)} b^- v_\lambda,$$
$$\mathfrak{X}^\natural(b^- v_\lambda) = \boldsymbol{t}^{\ell(b) + \langle \nu,\lambda\rangle + \phi(\nu,\lambda-\nu)} b^- v_\lambda.$$

This implies that for any $b, b' \in \mathcal{B}_\nu$,

$$\begin{aligned}\mathfrak{X}((b^- v_\lambda)^*)(b'^- v_\lambda) &= \boldsymbol{t}^{2p(\nu)} \mathfrak{X}((b^- v_\lambda)^*((\mathfrak{X}^\natural)^{-1}(b'^- v_\lambda))) \\ &= \boldsymbol{t}^{2p(\nu) - \ell(b) - \langle \nu,\lambda\rangle - \phi(\nu,\lambda-\nu)} \delta_{b,b'},\end{aligned}$$

and hence $\mathfrak{X}((b^- v_\lambda)^*) = \boldsymbol{t}^{2p(\nu) - \ell(b) - \langle \nu,\lambda\rangle - \phi(\nu,\lambda-\nu)} (b^- v_\lambda)^*$. In particular, for $c = (1,-1)$, observe that

$$\begin{aligned}&\mathfrak{X}((b^- v_\lambda) \otimes (b^- v_\lambda)^*) \\ &\quad = \boldsymbol{t}^{\kappa_c((\lambda,0)-\hat{\nu},(-\lambda,0)+\hat{\nu}) + \ell(b) - \phi(\nu,\lambda) + 2p(\nu) - \ell(b) - \langle \nu,\lambda\rangle - \phi(\nu,\lambda-\nu)} (b^- v_\lambda) \otimes (b^- v_\lambda)^* \\ &\quad = \boldsymbol{t}^{\kappa_c((\lambda,0),(-\lambda,0)) + 2p(\nu) - \nu\cdot\nu} (b^- v_\lambda) \otimes (b^- v_\lambda)^*.\end{aligned}$$

It is easy to verify that $\frac{1}{2}\nu\cdot\nu = p(\nu) \bmod 2$ by induction; hence we see that

$$\mathfrak{X}((b^- v_\lambda) \otimes (b^- v_\lambda)^*) = \boldsymbol{t}^{\kappa_{(1,-1)}((\lambda,0),(-\lambda,0))} (b^- v_\lambda) \otimes (b^- v_\lambda)^*.$$

A similar computation shows that

$$\mathfrak{X}((b^- v_\lambda) * \otimes (b^- v_\lambda)) = \boldsymbol{t}^{\kappa_{(-1,1)}((-\lambda,0),(\lambda,0)} (b^- v_\lambda)^* \otimes (b^- v_\lambda).$$

Note that in either case, the power of $\boldsymbol{t}$ is independent of $b \in \mathcal{B}$, and applying this to the definition of the maps proves (1) and (4). For (2) and (3), also note that $\pi^{p(\nu)} q^{\pm\langle \tilde{\rho},\lambda-\nu\rangle} = \pi_\nu q_\nu^{\mp 2} q^{\pm\langle \tilde{\rho},\lambda\rangle}$, and we compute that $\mathfrak{X}(\pi_\nu q_\nu^{\mp 2} q^{\pm\langle \tilde{\rho},\lambda\rangle}) = \boldsymbol{t}^{\mp\langle \tilde{\rho},\lambda\rangle} \pi_\nu q_\nu^{\mp 2} q^{\pm\langle \tilde{\rho},\lambda\rangle}$; the result follows. □

Example 4.20 Consider the case $n = 1$ and $\lambda = m$. As noted in Example 2.10, $\langle \tilde{\rho},\lambda\rangle = m$ and $\mathrm{ev}_m \circ \mathrm{coev}_m = \pi^m [m+1]$. Then we have $\mathrm{ev}_m \circ \mathrm{coev}_m \circ \mathfrak{X}(1) = \pi^m [m+1]$, and

$$\mathfrak{X} \circ \mathrm{ev}_m \circ \mathrm{coev}_m(1) = \mathfrak{X}(\pi^m [m+1]) = \boldsymbol{t}^{-m} \pi^m [m+1] = \boldsymbol{t}^{-m}\, \mathrm{ev}_m \circ \mathrm{coev}_m \circ \mathfrak{X}(1).$$

Note that this is consistent with Proposition 4.19, as we see that

$$\mathrm{ev}_m \circ \mathrm{coev}_m \circ \mathfrak{X} = \boldsymbol{t}^{-\kappa_{(-1,1)}((-\lambda,0),(\lambda,0)) + \langle \tilde{\rho},\lambda\rangle}\, \mathrm{ev}_m \circ \mathfrak{X} \circ \mathrm{coev}_m = \boldsymbol{t}^m \mathfrak{X} \circ \mathrm{ev}_m \circ \mathrm{coev}_m .$$

The last elementary diagram to consider is the crossing, which is to say the automorphism $R = \Theta \mathfrak{F}\mathfrak{s}$ of a tensor product of two modules. In order to have a concrete comparison of $R\mathfrak{X}$ and $\mathfrak{X}R$ on tensor products carrying twistors, it will be necessary to have a precise description of $\mathfrak{X}(\mathfrak{f}(\zeta,\eta))$ for any $\zeta,\eta\in X$. To that end, let us once and for all fix a transversal T of $X/\mathbb{Z}[I]$ and note that $\widehat{T}=\{(\zeta,0),(\zeta,1)\mid\zeta\in T\}$ is a transversal of $\widehat{X}/\mathbb{Z}[I]$. Then for $\zeta_0,\zeta_1\in T$, we shall henceforth require that

$$\mathfrak{f}(|\zeta_0|,|\zeta_1|)=1. \tag{4-11}$$

Then we have the following proposition.

Proposition 4.21 *Let $\lambda,\lambda'\in X^+\cup -X^+$. Let $\hat{\xi},\hat{\xi}'\in\widehat{T}$ be the corresponding coset representatives of $(\lambda,0)$ and $(\lambda',0)$ in $\widehat{X}/\mathbb{Z}[I]$, and let $(c_1,c_2)=\mathrm{sig}(\widehat{V}(\lambda)\widehat{\otimes}\widehat{V}(\lambda'))$. Let $\mathcal{R}\colon \widehat{V}(\lambda)\widehat{\otimes}\widehat{V}(\lambda')\to\widehat{V}(\lambda')\widehat{\otimes}\widehat{V}(\lambda)$ be the map described in Proposition 2.18. Then $\mathcal{R}$ is a $\widehat{\boldsymbol{U}}$–module homomorphism. Moreover, as maps on $\widehat{V}(\lambda)\widehat{\otimes}\widehat{V}(\lambda')$, we have*

$$\mathfrak{X}\mathcal{R}=\boldsymbol{t}^{\kappa_{(c_2,c_1)}(\hat{\xi}',\hat{\xi})-\kappa_{(c_1,c_2)}(\hat{\xi},\hat{\xi}')+2p(\hat{\xi})p(\hat{\xi}')}\mathcal{R}\mathfrak{X}.$$

Proof Recall that $\mathcal{R}=\Theta\mathfrak{f}\mathfrak{s}$ by definition. It is easy to see that $\mathcal{R}$ is a $\widehat{\boldsymbol{U}}$–module homomorphism: indeed, since $\mathcal{R}$ preserve weight spaces, it commutes with the action of the T_i for $i\in I$; moreover, $\mathfrak{f}\mathfrak{s}$ obviously commutes with the diagonal action of Υ_i, and it is easy to check directly that $\Theta_\nu\Delta(\Upsilon_i)=\Delta(\Upsilon_i)\Theta_\nu$. We will prove the remainder of the proposition in two steps.

First we shall show that $\mathfrak{X}(\Theta_\nu)=\Theta_\nu$ for any $\nu\in\mathbb{Z}_{\geq 0}[I]$, and thus $\mathfrak{X}\Theta=\Theta\mathfrak{X}$ as maps on $V(\lambda)\otimes V(\lambda')$. This is straightforward: applying Lemmas 4.5, 4.3 and Proposition 4.15, we compute that

$$\begin{aligned}
&\mathfrak{X}(\Theta_\nu)\\
&\quad=(-1)^{\mathrm{ht}\,\nu}\boldsymbol{t}^{2\boldsymbol{p}(\nu)}\pi^{\boldsymbol{p}(\nu)}\boldsymbol{t}_\nu^2\pi_\nu\boldsymbol{t}_\nu^{-1}q_\nu\sum_{b\in\mathcal{B}_\nu}\mathfrak{X}(b^-)\Upsilon_\nu\otimes\widetilde{T}_{-\nu}\Upsilon_{-\nu}\mathfrak{X}((b^*)^+)\\
&\quad=(-1)^{\mathrm{ht}\,\nu+\boldsymbol{p}(\nu)}\pi^{\boldsymbol{p}(\nu)}\boldsymbol{t}_\nu\pi_\nu q_\nu\sum_{b\in\mathcal{B}_\nu}(\mathfrak{X}(b)^-\Upsilon_{-\nu})\Upsilon_\nu\otimes\widetilde{T}_{-\nu}\Upsilon_{-\nu}(\boldsymbol{t}_\nu^{-1}\widetilde{T}_\nu\Upsilon_\nu\mathfrak{X}(b^*)^+)\\
&\quad=(-1)^{\mathrm{ht}\,\nu+\boldsymbol{p}(\nu)}\pi^{\boldsymbol{p}(\nu)}\boldsymbol{t}_\nu\pi_\nu q_\nu\sum_{b\in\mathcal{B}_\nu}(\boldsymbol{t}^{\ell(b)}b^-)\otimes(\boldsymbol{t}_\nu^{-1}\boldsymbol{t}^{-\ell(b)}(-1)^{\boldsymbol{p}(\nu)}(b^*)^+)\\
&\quad=(-1)^{\mathrm{ht}\,\nu}\pi^{\boldsymbol{p}(\nu)}\pi_\nu q_\nu\sum_{b\in\mathcal{B}_\nu}b^-\otimes(b^*)^+=\Theta_\nu.
\end{aligned}$$

Now it remains to show that we have $\mathfrak{X}\mathfrak{F}\mathfrak{s}=\boldsymbol{t}^{\kappa_{(c_2,c_1)}(\hat{\xi}',\hat{\xi})-\kappa_{(c_1,c_2)}(\hat{\xi},\hat{\xi}')+2p(\hat{\xi})p(\hat{\xi}')}\mathfrak{F}\mathfrak{s}\mathfrak{X}$ as maps on $V(\lambda)\otimes V(\lambda')$. Set $c=(c_1,c_2)$, and $\tilde{c}=(c_2,c_1)$. Let $m\in V(\lambda)$ and $n\in V(\lambda')$. Then we see directly that

$$\mathfrak{X}\mathfrak{F}\mathfrak{s}(m\otimes n)=\boldsymbol{t}^{2p(m)p(n)+\kappa_{\tilde{c}}(\|n\|,\|m\|)}\mathfrak{X}(\mathfrak{f}(|n|,|m|))\pi^{p(m)p(n)}\mathfrak{X}(n)\otimes\mathfrak{X}(m),$$

$$\mathfrak{F}\mathfrak{s}\mathfrak{X}(m\otimes n)=\boldsymbol{t}^{\kappa_c(\|m\|,\|n\|)}\mathfrak{f}(|n|,|m|)\pi^{p(m)p(n)}\mathfrak{X}(n)\otimes\mathfrak{X}(m).$$

The proposition then follows by verifying that

$$t^{2p(m)p(n)+\kappa_{\tilde{c}}(\|n\|,\|m\|)}\mathfrak{X}(\mathfrak{f}(|n|,|m|)) = t^{\kappa_{\tilde{c}}(\hat{\zeta}',\hat{\zeta})-\kappa_c(\hat{\zeta},\hat{\zeta}')+2p(\hat{\zeta})p(\hat{\zeta}')+\kappa_c(\|m\|,\|n\|)}\mathfrak{f}(|n|,|m|).$$

Note that $\hat{\zeta} = \|m\| + \mu$ and $\hat{\zeta}' = \|n\| + \nu$ for some $\mu, \nu \in \mathbb{Z}[I]$. Let $\zeta = |\hat{\zeta}| \in X$ and $\zeta' = |\hat{\zeta}'| \in X$. Then in particular, (4-11) implies

$$\mathfrak{f}(|n|,|m|) = (\pi q)^{\langle\tilde{\nu},\zeta\rangle} q^{\langle\tilde{\mu},\zeta'\rangle-\mu\cdot\nu},$$

so $\mathfrak{X}(\mathfrak{f}(|n|,|m|)) = t^{\langle\tilde{\nu},\zeta\rangle-\langle\tilde{\mu},\zeta'\rangle+\mu\cdot\nu}\mathfrak{f}(|n|,|m|)$. Therefore, we are reduced to showing that $\ell \equiv r \mod 4$, where

$$\ell = 2p(m)p(n) + \langle\tilde{\nu},\zeta\rangle - \langle\tilde{\mu},\zeta'\rangle + \mu\cdot\nu + \kappa_{\tilde{c}}(\|n\|,\|m\|),$$
$$r = 2p(\hat{\zeta})p(\hat{\zeta}') + \kappa_{\tilde{c}}(\hat{\zeta}',\hat{\zeta}) - \kappa_c(\hat{\zeta},\hat{\zeta}') + \kappa_c(\|m\|,\|n\|) \mod 4.$$

We compute directly that

$$\begin{aligned}
&\kappa_{\tilde{c}}(\|n\|,\|m\|) - \kappa_{\tilde{c}}(\hat{\zeta}',\hat{\zeta}) + \kappa_c(\hat{\zeta},\hat{\zeta}') - \kappa_c(\|m\|,\|n\|)\\
&\quad= \kappa_{\tilde{c}}(\hat{\zeta}'-\nu,\hat{\zeta}-\mu) - \kappa_{\tilde{c}}(\hat{\zeta}',\hat{\zeta}) + \kappa_c(\hat{\zeta},\hat{\zeta}') - \kappa_c(\hat{\zeta}-\mu,\hat{\zeta}'-\nu)\\
&\quad\equiv_4 -\langle\tilde{\nu},\zeta\rangle - c_1\phi(\nu,c_1\zeta) + 2p(\hat{\zeta})p(\mu) - c_2\phi(\mu,c_2\zeta) + \mu\cdot\nu + \phi(\mu,\nu)\\
&\qquad\quad + \langle\tilde{\mu},\zeta\rangle + c_1\phi(\mu,c_1\zeta) - 2p(\hat{\lambda})p(\nu) + \phi(\nu,\lambda) - \mu\cdot\nu - \phi(\nu,\mu)\\
&\quad\equiv_4 2p(\hat{\lambda})p(\nu) + 2p(\hat{\zeta})p(\mu) + 2p(\mu)p(\nu) - \langle\tilde{\nu},\lambda\rangle + \langle\tilde{\mu},\zeta\rangle + \mu\cdot\nu\\
&\quad\equiv_4 2p(m)p(n) - 2p(\hat{\lambda})p(\hat{\zeta}) - \langle\tilde{\nu},\lambda\rangle + \langle\tilde{\mu},\zeta\rangle + \mu\cdot\nu,
\end{aligned}$$

where here $\equiv_4$ denotes equivalence modulo 4. This finishes the proof. □

We have seen that the twistor map commutes (up to an integral power of $\boldsymbol{t}$) with the elementary functions in our graphical calculus. However, note that in Theorem 3.7, the typical composition factor of a tangle invariant is not just one of these maps, but in fact is a tensor product of these maps with various identities. It is important to note that a consequence of Proposition 4.18 is that the twistor maps on tensor products are not local, since the power of $\boldsymbol{t}$ in the construction depends on the weight and signature of each tensor factor. Nevertheless, we can extend Propositions 4.19 and 4.21 to this more general setting.

Proposition 4.22 *Let $M_1,\dots,M_t$ be $\hat{\boldsymbol{U}}$–modules such that for each $1 \le s \le t$, $M_s = \hat{V}(\mu_s)$ for some $\mu_s \in X^+ \cup -X^+$. Let $M = M_1\hat{\otimes}\cdots\hat{\otimes}M_t$ and let $c = (c_1,\dots,c_t) = \mathrm{sig}(M)$. For any $\lambda \in X^+$ and $0 \le r \le t$, we define $M_{\le r} = M_1\hat{\otimes}\cdots\hat{\otimes}M_r$, $M_{>r} = M_{r+1}\hat{\otimes}\cdots\hat{\otimes}M_t$, and*

$$M(r,\pm\lambda) = M_{\le r}\hat{\otimes}\hat{V}(\pm\lambda)\hat{\otimes}\hat{V}(\mp\lambda)\hat{\otimes}M_{>r}.$$

(1) *Let* $\mathcal{R}_s = 1_{M_{\leq s-1}} \otimes \mathcal{R} \otimes 1_{M_{>s+1}} \colon M \to M$ *for some* $1 \leq s \leq t-1$. *Then as maps on* M, *we have that* $\mathfrak{X} R_s$ *and* $R_s \mathfrak{X}$ *are proportional up to an integral power of* $\boldsymbol{t}$.

(2) *Let* $\mathrm{ev}(M, r, \lambda) = 1_{M_{\leq r}} \otimes \mathrm{ev}_\lambda \otimes 1_{M_{>r}}$ *for some* $1 \leq r \leq t$. *Then as maps on* $M(r, -\lambda)$, *we have that* $\mathfrak{X}\,\mathrm{ev}(M, r, \lambda)$ *and* $\mathrm{ev}(M, r, \lambda)\mathfrak{X}$ *are proportional up to an integral power of* $\boldsymbol{t}$.

(3) *Let* $\mathrm{qtr}(M, r, \lambda) = 1_{M_{\leq r}} \otimes \mathrm{qtr}_\lambda \otimes 1_{M_{>r}}$ *for some* $1 \leq r \leq t$. *Then as maps on* $M(r, \lambda)$, *we have that* $\mathfrak{X}\,\mathrm{qtr}(M, r, \lambda)$ *and* $\mathrm{qtr}(M, r, \lambda)\mathfrak{X}$ *are proportional up to an integral power of* $\boldsymbol{t}$.

(4) *Let* $\mathrm{coev}(M, r, \lambda) = 1_{M_{\leq r}} \otimes \mathrm{coev}_\lambda \otimes 1_{M_{>r}}$ *for some* $1 \leq r \leq t$. *Then as maps on* M, *we have that* $\mathfrak{X}\,\mathrm{coev}(M, r, \lambda)$ *and* $\mathrm{coev}(M, r, \lambda)\mathfrak{X}$ *are proportional up to an integral power of* $\boldsymbol{t}$.

(5) *Let* $\mathrm{coqtr}(M, r, \lambda) = 1_{M_{\leq r}} \otimes \mathrm{coqtr}_\lambda \otimes 1_{M_{>r}}$ *for some* $1 \leq r \leq t$. *Then as maps on* M, *we have that* $\mathfrak{X}\,\mathrm{coqtr}(M, r, \lambda)$ *and* $\mathrm{coqtr}(M, r, \lambda)\mathfrak{X}$ *are proportional up to an integral power of* $\boldsymbol{t}$.

Remark 4.23 The precise constants of proportionality can be determined directly as in Propositions 4.19 and 4.21 (and can be worked out from the following proof), but we leave them out of the statement of Proposition 4.22 because they are not particularly illuminating, and are not necessary for Theorem 4.24

Proof As the proofs of (2)–(5) are similar, we shall only prove (1) and (2) here in detail.

We will begin with the proof of (1), which is essentially the same as the proof of Proposition 4.21. To wit, we first observe that for any $a, b \geq 0$ and $\nu \in \mathbb{N}[I]$,

$$\mathfrak{X}(1^{\otimes a} \otimes \Theta_\nu \otimes 1^{\otimes b}) = (\Upsilon_{|\Theta_\nu|})^{\otimes a} \otimes \mathfrak{X}(\Theta_\nu) \otimes (\Upsilon_{|\Theta_\nu|}\widetilde{T}_{|\Theta_\nu|})^{\otimes b},$$

and the result follows from the observation that $|\Theta_\nu| = \nu - \nu = 0$. Then $\mathfrak{X} R_s = (1^{\otimes s-1} \otimes \Theta \otimes 1^{t-s-1})\mathfrak{X}\mathfrak{F}_s\mathfrak{s}_s$. Then we verify directly that

$$\mathfrak{X}\mathfrak{F}_s\mathfrak{s}_s = \boldsymbol{t}^{\kappa(c_{s+1}, c_s)(\widehat{\zeta}', \widehat{\zeta}) - \kappa(c_s, c_{s+1})(\widehat{\zeta}, \widehat{\zeta}') + 2p(\widehat{\zeta})p(\widehat{\zeta}')}\mathfrak{F}_s\mathfrak{s}_s\mathfrak{X},$$

where $\widehat{\zeta}$ (resp. $\widehat{\zeta}'$) is the coset representative for $(\mu_s, 0)$ (resp. $(\mu_{s+1}, 0)$).

Now, we shall prove (2). Note that an arbitrary element of $M(r, -\lambda)$ is a linear combination of simple tensors of the form $x = m_{\leq r} \otimes (b^- v_\lambda)^* \otimes (b'^- v_\lambda) \otimes m_{>r}$, where $b, b' \in \mathcal{B}$, $m_{\leq r} = m_1 \otimes \cdots \otimes m_r \in M_{\leq r}$ and $m_{>r} = m_{r+1} \otimes \cdots \otimes m_t \in M_{>r}$; hence we need only prove (1) holds when evaluating both sides at such elements. Since

$\mathrm{ev}_\lambda((b^- v_\lambda)^* \otimes (b'^- v_\lambda)) = \delta_{b,b'}$, note that (1) is trivially true when $b \neq b'$, so let's assume $b = b' \in \mathcal{B}_\nu$. Then

$$\mathrm{ev}(M,r,\lambda)\mathfrak{X}(x) = \boldsymbol{t}^{\diamondsuit(m_1,\dots,m_t)+\clubsuit}\mathfrak{X}(m_{\leq r}) \otimes \mathrm{ev}_\lambda\, \mathfrak{X}((b^- v_\lambda)^* \otimes (b^- v_\lambda)) \otimes \mathfrak{X}(m_{>r}),$$

where we set

$$\diamondsuit(m_1,\dots,m_t) = \sum_{s<r<s'} \kappa_{(c_s,c'_s)}(\|m_s\|, \|m'_s\|),$$

$$\begin{aligned}\clubsuit &= \sum_{s<r}(\kappa_{(c_s,-1)}(\|m_s\|, (-\lambda,0)+\hat{\nu}) + \kappa_{(c_s,1)}(\|m_s\|, (\lambda,0)-\hat{\nu})) \\ &\quad + \sum_{s>r}(\kappa_{(-1,c_s)}((-\lambda,0)+\hat{\nu}, \|m_s\|) + \kappa_{(1,c_s)}((\lambda,0)-\hat{\nu}, \|m_s\|)).\end{aligned}$$

Now $\clubsuit$ can be simplified to

$$\begin{aligned}\clubsuit = \sum_{s<r}&(\kappa_{(c_s,-1)}((\mu_s,0),(-\lambda,0)) + \kappa_{(c_s,1)}((\mu_s,0),(\lambda,0))) \\ &+ \sum_{s>r}(\kappa_{(-1,c_s)}((-\lambda,0),(\mu_s,0)) + \kappa_{(1,c_s)}((\lambda,0),(\mu_s,0))).\end{aligned}$$

Note that $\clubsuit$ is independent of x. Then

$$\begin{aligned}\mathrm{ev}(M,r,\lambda)\mathfrak{X}(x) &= \boldsymbol{t}^{\diamondsuit(m_1,\dots,m_t)+\clubsuit}\mathfrak{X}(m_{\leq r}) \otimes \mathrm{ev}_\lambda\, \mathfrak{X}((b^- v_\lambda)^* \otimes (b^- v_\lambda)) \otimes \mathfrak{X}(m_{>r}) \\ &= \boldsymbol{t}^{\diamondsuit(m_1,\dots,m_t)+\clubsuit+\kappa_{(-1,1)}((-\lambda,0),(\lambda,0))}\mathfrak{X}(m_{\leq r}) \otimes \mathfrak{X}(m_{>r}) \\ &= \boldsymbol{t}^{\clubsuit+\kappa_{(-1,1)}((-\lambda,0),(\lambda,0))}\mathfrak{X}(m_{\leq r} \otimes m_{>r}).\end{aligned}$$

Since $\mathfrak{X}(m_{\leq r} \otimes m_{>r}) = \mathfrak{X}(\mathrm{ev}_\lambda(x))$ and the exponent of $\boldsymbol{t}$ is independent of x, this completes the proof of (2). □

Theorem 4.24 *Let K be any oriented knot, and let $J_K^\lambda(q,\tau) \in \mathbb{Q}(q,\boldsymbol{t})^\tau$ be the λ–colored knot invariant defined in Theorem 3.7. Let ${}_{\mathfrak{so}}J_K^\lambda(q) = J_K^\lambda(q,1)$ and ${}_{\mathfrak{osp}}J_K^\lambda(q) = J_K^\lambda(q,\boldsymbol{t})$. Then*

$${}_{\mathfrak{osp}}J_K^\lambda(q) = \boldsymbol{t}^{\star(K,\lambda)}\,{}_{\mathfrak{so}}J_K^\lambda(\boldsymbol{t}^{-1}q)$$

for some $\star(K,\lambda) \in \mathbb{Z}$.

Proof Let $J = J_K^\lambda(q,\tau)$. First, observe that J can be thought of as a function $\mathbb{Q}(q,\boldsymbol{t})^\tau \to \mathbb{Q}(q,\boldsymbol{t})^\tau$, and in that spirit $\mathfrak{X}(J)$ is $\mathfrak{X}\circ J(1)$. On the other hand, $J = W_K \circ S$, where $W_K = (\mathfrak{f}(\lambda,\lambda)^{-1}\pi^{P(\lambda)}q^{\langle\rho,\lambda\rangle})^{\mathrm{wr}(K)}$ (interpreted as a function $\mathbb{Q}(q,\boldsymbol{t})^\tau \to \mathbb{Q}(q,\boldsymbol{t})^\tau$) and S is a slice diagram of K interpreted as a composition of morphisms as described in Section 3 (with strands colored by λ). In particular, observe that $\mathfrak{X}(\mathfrak{f}(\lambda,\lambda)) = \boldsymbol{t}^x\mathfrak{f}(\lambda,\lambda)$ for some $x \in \mathbb{Z}$ depending on the coset representative of λ in $X/\mathbb{Z}[I]$, and that $\mathfrak{X}(\pi^{P(\lambda)}q^{\langle\rho,\lambda\rangle}) = \pi^{P(\lambda)}q^{\langle\rho,\lambda\rangle)}$. Then $\mathfrak{X}W = \boldsymbol{t}^{-x\,\mathrm{wr}(K)}W\mathfrak{X}$.

Likewise, note that S can be written as a composition of maps of the form $\mathrm{ev}(M,r,\lambda)$, $\mathrm{coev}(M,r,\lambda)$, $\mathrm{qtr}(M,r,\lambda)$, $\mathrm{coqtr}(M,r,\lambda)$, and $R_s\colon M\to M$ for various $r,s\in\mathbb{N}$ with the notation being the same as in Proposition 4.22. In particular, we see that $\mathfrak{X}\circ S=\boldsymbol{t}^{y}S\circ\mathfrak{X}$ for some $y\in\mathbb{Z}$, and thus

$$\mathfrak{X}(J)=\mathfrak{X}\circ W_K\circ S(1)=\boldsymbol{t}^{-x\mathrm{wr}(K)+y}W_K\circ S\circ\mathfrak{X}(1)=\boldsymbol{t}^{-x\mathrm{wr}(K)+y}J.$$

On the other hand, observe that $\mathfrak{X}(J_K^\lambda(q,\tau))=J_K^\lambda(\boldsymbol{t}^{-1}q,\boldsymbol{t}^{-1}\tau)$, and so

$$\boldsymbol{t}^{y-x\mathrm{wr}(K)}J_K^\lambda(\boldsymbol{t}^{-1}q,\boldsymbol{t}^{-1}\tau)=J_K^\lambda(q,\tau).$$

The theorem follows from specializing $\tau=\boldsymbol{t}$. □

Remark 4.25 Note that since ${}_{\mathfrak{so}}J_K^\lambda(q)\in\mathbb{Z}[q,q^{-1}]$, Theorem 4.24 implies that (after a renormalization) ${}_{\mathfrak{osp}}J_K^\lambda(q)={}_{\mathfrak{so}}J_K^\lambda(v)\in\mathbb{Z}[v,v^{-1}]$ where $v=q\boldsymbol{t}^{-1}$. Furthermore, note that when n or $\langle\underline{n},\lambda\rangle$ is even, ${}_{\mathfrak{osp}}J_K^\lambda(q)\in\mathbb{Q}(q)$ (see Remark 2.4); thus in this case, ${}_{\mathfrak{osp}}J_K^\lambda(q)\equiv{}_{\mathfrak{so}}J_K^\lambda(q)\bmod 2$.

References

[1] **Y Akutsu**, **T Deguchi**, **T Ohtsuki**, *Invariants of colored links*, J. Knot Theory Ramifications 1 (1992) 161–184 MR

[2] **S C Blumen**, *The Birman–Wenzl–Murakami algebra, Hecke algebra and representations of $U_q(osp(1|2n))$*, preprint (2006) arXiv

[3] **S C Blumen**, *On the $U_q(\mathrm{osp}(1|2n))$ and $U_{-q}(\mathrm{so}(2n+1))$ uncolored quantum link invariants*, J. Knot Theory Ramifications 19 (2010) 335–353 MR

[4] **S Clark**, *Quantum supergroups, IV: The modified form*, Math. Z. 278 (2014) 493–528 MR

[5] **S Clark**, **Z Fan**, **Y Li**, **W Wang**, *Quantum supergroups, III: Twistors*, Comm. Math. Phys. 332 (2014) 415–436 MR

[6] **S Clark**, **D Hill**, *Quantum supergroups, V: Braid group action*, Comm. Math. Phys. 344 (2016) 25–65 MR

[7] **S Clark**, **D Hill**, **W Wang**, *Quantum supergroups, I: Foundations*, Transform. Groups 18 (2013) 1019–1053 MR

[8] **S Clark**, **D Hill**, **W Wang**, *Quantum supergroups, II: Canonical basis*, Represent. Theory 18 (2014) 278–309 MR

[9] **S Clark**, **W Wang**, *Canonical basis for quantum $\mathfrak{osp}(1|2)$*, Lett. Math. Phys. 103 (2013) 207–231 MR

[10] **A P Ellis**, **M Khovanov**, **A D Lauda**, *The odd nilHecke algebra and its diagrammatics*, Int. Math. Res. Not. 2014 (2014) 991–1062 MR

[11] **A P Ellis**, **A D Lauda**, *An odd categorification of $U_q(\mathfrak{sl}_2)$*, Quantum Topol. 7 (2016) 329–433 MR

[12] **N Geer**, **B Patureau-Mirand**, *Multivariable link invariants arising from Lie superalgebras of type I*, J. Knot Theory Ramifications 19 (2010) 93–115 MR

[13] **M D Gould**, **J R Links**, **Y-Z Zhang**, *Type-I quantum superalgebras, q–supertrace, and two-variable link polynomials*, J. Math. Phys. 37 (1996) 987–1003 MR

[14] **D Hill**, **W Wang**, *Categorification of quantum Kac–Moody superalgebras*, Trans. Amer. Math. Soc. 367 (2015) 1183–1216 MR

[15] **J C Jantzen**, *Lectures on quantum groups*, Graduate Studies in Mathematics 6, Amer. Math. Soc., Providence, RI (1996) MR

[16] **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426 MR

[17] **M Khovanov**, **A D Lauda**, *A categorification of quantum* $\mathrm{sl}(n)$, Quantum Topol. 1 (2010) 1–92 MR

[18] **G Lusztig**, *Introduction to quantum groups*, Progress in Mathematics 110, Birkhäuser, Boston (1993) MR

[19] **V Mikhaylov**, **E Witten**, *Branes and supergroups*, Comm. Math. Phys. 340 (2015) 699–832 MR

[20] **T Ohtsuki**, *Quantum invariants: a study of knots, 3–manifolds, and their sets*, Series on Knots and Everything 29, World Scientific, River Edge, NJ (2002) MR

[21] **P S Ozsváth**, **J Rasmussen**, **Z Szabó**, *Odd Khovanov homology*, Algebr. Geom. Topol. 13 (2013) 1465–1488 MR

[22] **H Queffelec**, **A Sartori**, *Mixed quantum skew Howe duality and link invariants of type A*, preprint (2015) arXiv

[23] **N Y Reshetikhin**, **V G Turaev**, *Ribbon graphs and their invariants derived from quantum groups*, Comm. Math. Phys. 127 (1990) 1–26 MR

[24] **R Rouquier**, *2–Kac–Moody algebras*, preprint (2008) arXiv

[25] **A Sartori**, *The Alexander polynomial as quantum invariant of links*, Ark. Mat. 53 (2015) 177–202 MR

[26] **P Tingley**, *A minus sign that used to annoy me but now I know why it is there*, preprint (2010) arXiv

[27] **V G Turaev**, *Operator invariants of tangles, and R–matrices*, Izv. Akad. Nauk SSSR Ser. Mat. 53 (1989) 1073–1107 MR In Russian; translated in Math. of the USSR-Izvestiya 35 (1990) 411–444

[28] **B Webster**, *Knot invariants and higher representation theory*, preprint (2013) arXiv To appear in Mem. Amer. Math. Soc.

[29] **R B Zhang**, *Finite-dimensional representations of* $U_q(\mathrm{osp}(1/2n))$ *and its connection with quantum* $\mathrm{so}(2n+1)$, Lett. Math. Phys. 25 (1992) 317–325 MR

[30] **R B Zhang**, *Quantum supergroups and topological invariants of three-manifolds*, Rev. Math. Phys. 7 (1995) 809–831 MR

Department of Mathematics, Northeastern University
Boston, MA, United States

se.clark@northeastern.edu

https://web.northeastern.edu/sclark/

Received: 18 October 2016 Revised: 28 February 2017

msp

Algebraic & Geometric Topology 17 (2017) 3007–3020

Localization of cofibration categories and groupoid C^*–algebras

MARKUS LAND
THOMAS NIKOLAUS
KAROL SZUMIŁO

We prove that relative functors out of a cofibration category are essentially the same as relative functors which are only defined on the subcategory of cofibrations. As an application we give a new construction of the functor that assigns to a groupoid its groupoid C^*–algebra and thereby its topological K–theory spectrum.

55U35; 46L80

Let $(\mathcal{C}, w\mathcal{C}, c\mathcal{C})$ be a *cofibration category*, ie a structure dual to a category of fibrant objects in the sense of Brown [1]. Here, $w\mathcal{C}$ and $c\mathcal{C}$ are the subcategories of weak equivalences and cofibrations, ie they have the same objects as $\mathcal{C}$ but morphisms are the weak equivalences or the cofibrations, respectively. Similarly, $wc\mathcal{C}$ will denote the subcategory of acyclic cofibrations. In addition to Brown's axioms, we will assume that $\mathcal{C}$ has *good cylinders*, which is a mild technical condition explained in Definition 9.

In this paper we will prove the following theorem. It will be formulated using the language of ∞–categories, following the notation of Lurie [11; 12]. In particular, an ordinary category $\mathcal{C}$ can be considered as an ∞–category by taking its nerve $N\mathcal{C}$.

Theorem 1 *If a cofibration category $\mathcal{C}$ has good cylinders, then the map induced by the inclusion*

$$Nc\mathcal{C}[wc^{-1}] \xrightarrow{\simeq} N\mathcal{C}[w^{-1}]$$

is an equivalence of ∞–categories.

By $N\mathcal{C}[w^{-1}]$ we denote the universal ∞–category obtained from $N\mathcal{C}$ by inverting the weak equivalences; see [12, Definition 1.3.4.1 and Remark 1.3.4.2]. By passing to opposite categories, the dual statement of Theorem 1 for fibration categories also holds.

The proof of Theorem 1 will be given at the end of the paper, but let us first establish a consequence and the application to C^*–algebras associated to groupoids.

Let $\mathcal{C}$ be a cofibration category with good cylinders and $\mathcal{M}$ a model category which is Quillen equivalent to a combinatorial model category and has functorial fibrant and cofibrant replacements, eg any of the model categories of spectra.

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3007

Proposition 2 *For any functor $F\colon c\mathcal{C} \to \mathcal{M}$ that sends acyclic cofibrations in $c\mathcal{C}$ to weak equivalences in $\mathcal{M}$ there exists a functor $\widehat{F}\colon \mathcal{C} \to \mathcal{M}$ with the following properties:*

(1) *$\widehat{F}$ sends weak equivalences in $\mathcal{C}$ to weak equivalences in $\mathcal{M}$.*

(2) *$\widehat{F}$ extends F in the sense that there exists a zigzag of natural weak equivalences between F and $\widehat{F}\,|_{c\mathcal{C}}$.*

Moreover, $\widehat{F}$ is unique in the following sense: for any other functor $\widehat{F}'\colon \mathcal{C} \to \mathcal{M}$ that satisfies (1) and (2) there exists a zigzag of natural weak equivalences between $\widehat{F}$ and $\widehat{F}'$.

Proof We denote the ∞–category $N\mathcal{M}[w^{-1}]$ associated to the model category $\mathcal{M}$ by $\mathcal{M}_\infty$. We claim that for any ordinary category $\mathcal{A}$ the canonical map

$$N\mathrm{Fun}(\mathcal{A}, \mathcal{M})[\ell^{-1}] \to \mathrm{Fun}(N\mathcal{A}, \mathcal{M}_\infty)$$

is an equivalence of ∞–categories, where ℓ is the class of levelwise weak equivalences. Here $\mathrm{Fun}(-,-)$ is used both for the ordinary category of functors between ordinary categories and the ∞–category of functors between ∞–categories; we hope that it is clear from the context which of the two is meant. If $\mathcal{M}$ is a simplicial, combinatorial model category, this is a special case of [11, Proposition 4.2.4.4], using that for a simplicial model category $\mathcal{M}$, the ∞–category $\mathcal{M}_\infty$ is equivalent to the homotopy coherent nerve of the simplicial subcategory of $\mathcal{M}$ on the fibrant and cofibrant objects; see [12, Theorem 1.3.4.20]. From the existence of functorial (co)fibrant replacements and Hovey [8, Proposition 1.3.13] it follows that a Quillen equivalence $\mathcal{M} \simeq \mathcal{M}'$ induces a Quillen equivalence $\mathrm{Fun}(\mathcal{A}, \mathcal{M}) \simeq \mathrm{Fun}(\mathcal{A}, \mathcal{M}')$. Thus the domain of the map in question is invariant under Quillen equivalences in $\mathcal{M}$. The same is true for the codomain, thus the statement that this map is an equivalence is invariant under Quillen equivalences in $\mathcal{M}$. Hence it is also true for all model categories $\mathcal{M}$ with functorial (co)fibrant replacements that are Quillen equivalent to a combinatorial, simplicial model category. Since every combinatorial model category is equivalent to a combinatorial, simplicial model category by a result of Dugger [6, Corollary 1.2], the claim holds in our generality. If $\mathcal{A}$ is a relative category, it also follows that the induced functor

$$N\mathrm{Fun}^w(\mathcal{A}, \mathcal{M})[\ell^{-1}] \to \mathrm{Fun}^w(N\mathcal{A}, \mathcal{M}_\infty)$$

is an equivalence, where the superscript w refers to functors that send weak equivalences in $\mathcal{A}$ to weak equivalences or equivalences in the target. This follows immediately from the nonrelative case, noting that both sides are just full subcategories of $N\mathrm{Fun}(\mathcal{A}, \mathcal{M})[\ell^{-1}]$ and $\mathrm{Fun}(N\mathcal{A}, \mathcal{M}_\infty)$. Thus in the canonical commuting square

$$\begin{array}{ccc} N\mathrm{Fun}^w(\mathcal{C},\mathcal{M})[\ell^{-1}] & \longrightarrow & N\mathrm{Fun}^w(c\mathcal{C},\mathcal{M})[\ell^{-1}] \\ \downarrow & & \downarrow \\ \mathrm{Fun}^w(N\mathcal{C},\mathcal{M}_\infty) & \longrightarrow & \mathrm{Fun}^w(Nc\mathcal{C},\mathcal{M}_\infty) \end{array}$$

the vertical maps are equivalences of ∞–categories. By Theorem 1 the lower map is also an equivalence, therefore also the upper one is. Passing to homotopy categories we obtain the desired result, using that isomorphisms in homotopy categories of functor categories are represented by zigzags of natural weak equivalences. □

Applications

Groupoids

We denote by Gpd the 1–category of small groupoids and by Gpd_2 the ∞–category associated to the $(2,1)$–category of groupoids in which the 2–morphisms are natural isomorphisms. The category Gpd admits a simplicial model structure in which the equivalences are equivalences of categories and the cofibrations are functors that are injective on the set of objects. In this model structure all objects are cofibrant and fibrant, compare Casacuberta, Golasiński and Tonks [2]. Furthermore, if we denote by Gpd^ω the full subcategory on groupoids with at most countable many morphisms then Gpd^ω inherits the structure of a cofibration category.

The following lemma is a well-known fact, but we had difficulties finding a clear reference for this so we state it as an extra lemma.

Lemma 3 *The canonical map $N\mathrm{Gpd}[w^{-1}] \to \mathrm{Gpd}_2$ is an equivalence of ∞–categories.*

Proof This follows from the description of the ∞–category associated to a simplicial model category — see [12, Theorem 1.3.4.20] — as being the homotopy coherent nerve of the simplicial category of cofibrant and fibrant objects. □

Corollary 4 *Let $\mathcal{C}$ be an ∞–category. Then the canonical map $Nc\mathrm{Gpd} \to \mathrm{Grp}_2$ induces an equivalence*

$$\mathrm{Fun}(\mathrm{Gpd}_2,\mathcal{C}) \xrightarrow{\simeq} \mathrm{Fun}^w(Nc\mathrm{Gpd},\mathcal{C}),$$

where the superscript w refers to functors that send equivalences of groupoids to equivalences in $\mathcal{C}$.

Proof Since the canonical map $N\mathrm{Gpd}[w^{-1}] \to \mathrm{Gpd}_2$ is an equivalence by Lemma 3, this is a direct application of Theorem 1. □

The following corollary of Proposition 2 implies that in the approach to assembly maps discussed by Davis and Lück [4, Section 2], one can directly restrict to functors from groupoids to spectra that are only defined for maps of groupoids that are injective on objects. This resolves the issues illustrated in [4, Remark 2.3].

Corollary 5 *Let* Sp *be any of the standard model categories of spectra. Then every functor* $F\colon c\mathrm{Gpd} \to \mathrm{Sp}$ *which sends equivalences of groupoids to weak equivalences in* Sp *extends uniquely* (in the sense of Proposition 2) *to a functor* $\widehat{F}\colon \mathrm{Gpd} \to \mathrm{Sp}$ *which also sends weak equivalences of groupoids to weak equivalences of spectra.*

Remark The statements of Corollaries 4 and 5 remain true if we replace Gpd by Gpd^{ω}. Furthermore, Corollary 5 does not depend on the exact choice of model category of spectra as long as it is Quillen equivalent to a combinatorial model category. Notice that this is automatically fulfilled if the model category is stable, due to the rigidity result of Schwede; see [17].

Next we want to demonstrate how to apply these results by *functorially* constructing C^*–algebras and topological K–theory spectra associated to groupoids. This discussion is similar to the one given by Joachim [9, Section 3] but we use our main theorem to obtain full functoriality instead of an explicit construction.

Definition 6 Let $\mathcal{G}$ be a groupoid. We let $\mathbb{C}\mathcal{G}$ be the $\mathbb{C}$–linearization of the set of morphisms of $\mathcal{G}$. This is a $\mathbb{C}$–algebra by linearization of the multiplication on morphisms given by

$$f \cdot g = \begin{cases} f \circ g & \text{if } f \text{ and } g \text{ are composable,} \\ 0 & \text{otherwise.} \end{cases}$$

We remark that $\mathbb{C}\mathcal{G}$ is unital if and only if the set of objects of $\mathcal{G}$ is finite. Then we complete $\mathbb{C}\mathcal{G}$ in a universal way, like for the full group C^*–algebra, to obtain a C^*–algebra $C^*\mathcal{G}$. More precisely, the norm is given by the supremum over all norms of representations of $\mathbb{C}\mathcal{G}$ on a separable Hilbert space. This is isomorphic to the C^*–algebra associated to the maximal groupoid C^*–category of Dell'Ambrogio [5, Definition 3.16] using the construction $\mathcal{C} \mapsto A_{\mathcal{C}}$ of Joachim [9, Section 3].

The association $\mathcal{G} \mapsto C^*\mathcal{G}$ is functorial for cofibrations of groupoids but not for general morphisms, since it can happen that morphisms are not composable in a groupoid, but become composable after applying a functor; compare the remark in Davis and Lück [4, page 214]. We observe that the C^*–algebra $C^*\mathcal{G}$ is separable provided $\mathcal{G} \in \mathrm{Gpd}^{\omega}$.

Lemma 7 *Let $F\colon \mathcal{G}_1 \to \mathcal{G}_2$ be an acyclic cofibration of groupoids. Then the induced morphism*
$$C^*F\colon C^*\mathcal{G}_1 \to C^*\mathcal{G}_2$$
is a KK*–equivalence.*

Proof The C^*–algebra associated to a groupoid with finitely many connected components is the product of the C^*–algebras associated to each connected component. For an infinite number of components, the associated C^*–algebra is the filtered colimit of the C^*–algebras associated to finitely many connected components. Since finite products of KK–equivalences are again KK–equivalences, and the filtered colimit of these KK–equivalences is again a KK–equivalence, we may assume that $\mathcal{G}_1$ (and thus $\mathcal{G}_2$) is connected. Let $x \in \mathcal{G}$ be an object. We let $G_1 = \mathrm{End}(x)$ and $G_2 = \mathrm{End}(Fx)$ be the endomorphism groups and notice the fact that F is an equivalence implies that F induces an isomorphism $G_1 \cong G_2$. Then we consider the diagram

$$\begin{array}{ccc} C^*\mathcal{G}_1 & \xrightarrow{C^*F} & C^*\mathcal{G}_2 \\ \uparrow & & \uparrow \\ C^*G_1 & \xrightarrow{\cong} & C^*G_2 \end{array}$$

in which the lower horizontal arrow is an isomorphism. Thus to show the lemma it suffices to prove the lemma in the special case where F is the inclusion of the endomorphisms of an object x of a connected groupoid $\mathcal{G}$.

This can be done in the abstract setting of corner algebras. For this suppose A is a C^*–algebra and $p \in A$ is a projection. It is called full if ApA is dense in A. The algebra pAp is called the corner algebra of p in A. It is called a full corner if p is a full projection. We write i_p for the inclusion $pAp \subset A$. Given a projection p the module pA is an imprimitivity $pAp - \overline{ApA}$ bimodule; see eg [15, Example 3.6]. Thus if p is full, then pA gives rise to an invertible element $[pA, i_p, 0] = \mathcal{F}(p) \in \mathrm{KK}(pAp, A)$. In this KK–group we have an equality

$$\mathcal{F}(p) = [\,pA, i_p, 0] + [(1-p)A, 0, 0] = [\,pA \oplus (1-p)A, i_p, 0] = [A, i_p, 0] = [i_p];$$

in other words, the inclusion $pAp \to A$ of a corner algebra associated to a full projection is a KK–equivalence.

To come back to our situation let us suppose $\mathcal{G}$ is a groupoid, $x \in \mathcal{G}$ is an object and let us denote its endomorphism group by $G = \mathrm{End}(x)$. We can consider the element $p = \mathrm{id}_x \in C^*\mathcal{G}$, which is clearly a projection. Its corner algebra is given by

$$p \cdot C^*\mathcal{G} \cdot p \cong C^*G.$$

If $\mathcal{G}$ is connected, it follows that every morphism in $\mathcal{G}$ may be factored through id_x and thus p is full. Hence it follows that the inclusion $C^*G \to C^*\mathcal{G}$ is an embedding of a full corner algebra. Thus, by the general theory, this inclusion is a KK–equivalence, which proves the lemma. □

Let us denote by KK_∞ the ∞–category given by the localization of the category $C^*\mathrm{Alg}$ of separable C^*–algebras at the KK–equivalences; see eg [10, Definition 3.2]. In formulas we have $\mathrm{KK}_\infty := NC^*\mathrm{Alg}[w^{-1}]$, where w denotes the class of KK–equivalences. The homotopy category of KK_∞ is Kasparov's KK–category of C^*–algebras.

Corollary 8 *There exists a functor*

$$\mathrm{Gpd}_2^\omega \to \mathrm{KK}_\infty$$

which on objects sends a groupoid $\mathcal{G}$ to the full groupoid C^*–algebra $C^*\mathcal{G}$.

Remark The $(2,1)$–category Orb^ω consisting of (countable) groups, group homomorphisms and conjugations is the full subcategory of the $(2,1)$–category of (countable) groupoids on connected groupoids and hence along this inclusion we also obtain a functor

$$\mathrm{Orb}^\omega \to \mathrm{KK}_\infty$$

which on objects sends a group to its full group C^*–algebra. This will be used by the first two authors in [10] to compare the L–theoretic Farrell–Jones conjecture and the Baum–Connes conjecture.

Proof of Corollary 8 By Corollary 4 and the remark after Corollary 5, we have an equivalence

$$\mathrm{Fun}^w(Nc\mathrm{Gpd}^\omega, \mathrm{KK}_\infty) \simeq \mathrm{Fun}(\mathrm{Gpd}_2^\omega, \mathrm{KK}_\infty),$$

and thus it suffices to construct a functor

$$c\mathrm{Gpd}^\omega \to C^*\mathrm{Alg}$$

which has the *property* that it sends equivalences of groupoids to KK–equivalences. We have established in Lemma 7 that the functor of Definition 6 satisfies this property. □

Remark In [10, Proposition 3.7] it is shown that the topological K–theory functor

$$K\colon NC^*\mathrm{Alg} \to \mathrm{Sp}$$

factors over KK_∞, in fact becomes corepresentable there. It thus follows from Corollary 8 that there is a functor sending a groupoid to the topological K–theory spectrum of its C^*–algebra.

The proof of Theorem 1

In this section we will prove Theorem 1. Recall that we consider a cofibration category $(\mathcal{C}, w\mathcal{C}, c\mathcal{C})$ and aim to compare the ∞–categories associated to the relative categories $(\mathcal{C}, w\mathcal{C})$ and $(c\mathcal{C}, wc\mathcal{C})$. As our model of the homotopy theory of $(\infty, 1)$–categories we will use the *complete Segal spaces* of Rezk; see [16]. This homotopy theory is modelled by the Rezk model structure on the category of bisimplicial sets in which fibrant objects are the complete Segal spaces. The model structure is constructed as a Bousfield localization of the Reedy model structure and hence every levelwise weak equivalence of bisimplicial sets is a Rezk equivalence, ie an equivalence of ∞–categories.

The ∞–category associated to a relative category $(\mathcal{D}, w\mathcal{D})$ is modelled by the *classification diagram* $N^{\mathrm{R}}\mathcal{D}$ of Rezk, which is given by

$$(N^{\mathrm{R}}\mathcal{D})_k \mapsto Nw(\mathcal{D}^{[k]}),$$

where the weak equivalences in $\mathcal{D}^{[k]}$ are levelwise weak equivalences; compare [16, Section 3.3; 13, Theorem 3.8]. See also Cisinski's response in [3]. Here, again, the notation N refers to the nerve of a category, which is a simplicial set, and here it should be thought of as a homotopy type as opposed to an ∞–category. The classification diagram is not fibrant in the Rezk model structure, but it is levelwise equivalent to a fibrant object if $\mathcal{D}$ is a cofibration category.

Let X be an object of a cofibration category $\mathcal{C}$. Recall that a *cylinder* on X is a factorization of the canonical morphism $X \sqcup X \to X$ via a cofibration $X \sqcup X \to IX$ and a weak equivalence $IX \to X$. A *cylinder functor* on $\mathcal{C}$ is a functor $I\colon \mathcal{C} \to \mathcal{C}$ equipped with natural transformations that provide such factorizations for all objects of $\mathcal{C}$. In the introduction we stated Theorem 1 under the following assumption on $\mathcal{C}$.

Definition 9 A cofibration category $\mathcal{C}$ has *good cylinders* if it has a cylinder functor I such that for every cofibration $X \rightarrowtail Y$ the induced morphism $IX \sqcup_{X \sqcup X} (Y \sqcup Y) \to IY$ is a cofibration.

For example, any cofibration category arising from a monoidal model category (or a model category enriched over a monoidal model category) has good cylinders, since they are given by tensoring with a chosen interval object. In particular the cofibration category underlying the model category of groupoids we discussed has good cylinders.

Theorem 10 *If $\mathcal{C}$ has good cylinders, then the inclusion $c\mathcal{C} \to \mathcal{C}$ induces a levelwise weak equivalence of the classification diagrams $N^{\mathrm{R}}c\mathcal{C} \to N^{\mathrm{R}}\mathcal{C}$.*

For the proof we will need a series of auxiliary definitions and lemmas. Let us first fix some notation. If J is a category, then $\widehat{J}$ denotes J considered as a relative category with all morphisms as weak equivalences. If J is any relative category, then $\mathcal{C}^J$ stands for the cofibration category of all relative diagrams $J \to \mathcal{C}$ with levelwise weak equivalences and cofibrations. If J is any relative direct category, then $\mathcal{C}^J_{\mathrm{R}}$ stands for the cofibration category of all relative Reedy cofibrant diagrams $J \to \mathcal{C}$ with levelwise weak equivalences and Reedy cofibrations. See [14, Theorem 9.3.8] for the construction of these cofibration categories and [14, Sections 9.1 and 9.2] for definitions of (relative) direct categories and Reedy cofibrations. (Note that Radulescu-Banu [14] uses the word "restricted" instead of "relative".) For our purposes we only need the direct category $J = [k]$, so we will recall the definitions just in this case. A diagram over $[k]$ is Reedy cofibrant if all its structure maps are cofibrations. A morphism $X \to Y$ of such diagrams is a Reedy cofibration if all the induced morphisms $X_{i+1} \sqcup_{X_i} Y_i \to Y_{i+1}$ are cofibrations. In [14] cofibration categories are assumed to have certain infinite colimits that are necessary for these results to hold for arbitrary J. However, as mentioned above, we will only use finite categories $J = [k]$, in which case the cited theorem is valid with Brown's original definition, which asserts only existence of an initial object and pushouts along cofibrations.

Definition 11 A subcategory $g\mathcal{C}$ of a cofibration category $\mathcal{C}$ is said to be *good* if

- all cofibrations are in $g\mathcal{C}$;
- the morphisms of $g\mathcal{C}$ are stable under pushouts along cofibrations;
- $\mathcal{C}$ has functorial factorizations that preserve $g\mathcal{C}$, in the sense that if

$$\begin{array}{ccc} A_0 & \longrightarrow & B_0 \\ \downarrow & & \downarrow \\ A_1 & \longrightarrow & B_1 \end{array}$$

is a square in $\mathcal{C}$ such that both vertical morphisms are in $g\mathcal{C}$ and

$$\begin{array}{ccccc} A_0 & \rightarrowtail & \widetilde{B}_0 & \xrightarrow{\sim} & B_0 \\ \downarrow & & \downarrow & & \downarrow \\ A_1 & \rightarrowtail & \widetilde{B}_1 & \xrightarrow{\sim} & B_1 \end{array}$$

is the resulting factorization, then the induced morphism $A_1 \sqcup_{A_0} \widetilde{B}_0 \to \widetilde{B}_1$ is also in $g\mathcal{C}$. In particular, so is $\widetilde{B}_0 \to \widetilde{B}_1$ by the second condition and most of the time only this conclusion will be used. However, the stronger property that $A_1 \sqcup_{A_0} \widetilde{B}_0 \to \widetilde{B}_1$ is in $g\mathcal{C}$ is necessary for the inductive argument in the proof of Lemma 15(3).

Now suppose that $\mathcal{C}$ is a cofibration category with a good subcategory $g\mathcal{C}$. We let $W\mathcal{C}$ be the bisimplicial set whose (m,n)–bisimplices are all diagrams in $\mathcal{C}$ of the form

$$\begin{array}{ccccccc}
X_{0,0} & \overset{\sim}{\rightarrowtail} & X_{0,1} & \overset{\sim}{\rightarrowtail} & \cdots & \overset{\sim}{\rightarrowtail} & X_{0,n} \\
{\scriptstyle\sim}\downarrow{\scriptstyle g} & & {\scriptstyle\sim}\downarrow{\scriptstyle g} & & & & \downarrow{\scriptstyle g}{\scriptstyle\sim} \\
X_{1,0} & \overset{\sim}{\rightarrowtail} & X_{1,1} & \overset{\sim}{\rightarrowtail} & \cdots & \overset{\sim}{\rightarrowtail} & X_{1,n} \\
{\scriptstyle\sim}\downarrow{\scriptstyle g} & & {\scriptstyle\sim}\downarrow{\scriptstyle g} & & & & \downarrow{\scriptstyle g}{\scriptstyle\sim} \\
\vdots & & \vdots & & & & \vdots \\
{\scriptstyle\sim}\downarrow{\scriptstyle g} & & {\scriptstyle\sim}\downarrow{\scriptstyle g} & & & & \downarrow{\scriptstyle g}{\scriptstyle\sim} \\
X_{m,0} & \overset{\sim}{\rightarrowtail} & X_{m,1} & \overset{\sim}{\rightarrowtail} & \cdots & \overset{\sim}{\rightarrowtail} & X_{m,n}
\end{array}$$

ie relative diagrams $\widehat{[m]} \times \widehat{[n]} \to \mathcal{C}$ where all horizontal morphisms are cofibrations and all vertical morphisms are in $g\mathcal{C}$. In other words, $W\mathcal{C}$ is the nerve of a double category with the same objects as $\mathcal{C}$, whose horizontal morphisms are acyclic cofibrations, vertical morphisms are weak equivalences in $g\mathcal{C}$, and double morphisms are just commutative squares.

Lemma 12 *The bisimplicial set $W\mathcal{C}$ is vertically homotopically constant, ie every simplicial operator $[n] \to [n']$ induces a weak homotopy equivalence $(W\mathcal{C})_{*,n'} \to (W\mathcal{C})_{*,n}$.*

Proof Note that $(W\mathcal{C})_{*,n} = N\widetilde{\mathcal{C}}_n$, where $\widetilde{\mathcal{C}}_n$ is a category whose objects are diagrams $\widehat{[n]} \to c\mathcal{C}$ and whose morphisms are weak equivalences with all components in $g\mathcal{C}$. It is enough to consider the case $n' = 0$, ie to show that the constant functor const: $\widetilde{\mathcal{C}}_0 \to \widetilde{\mathcal{C}}_n$ is a homotopy equivalence. The evaluation at n functor $\mathrm{ev}_n\colon \widetilde{\mathcal{C}}_n \to \widetilde{\mathcal{C}}_0$ satisfies $\mathrm{ev}_n \,\mathrm{const} = \mathrm{id}_{\widetilde{\mathcal{C}}_0}$. Moreover, the structure maps of every diagram $X \in \widetilde{\mathcal{C}}_n$ form a natural weak equivalence $X \to \mathrm{const}\,\mathrm{ev}_n X$ since every cofibration is in $g\mathcal{C}$. □

Lemma 13 *The bisimplicial set $W\mathcal{C}$ is horizontally homotopically constant, ie every simplicial operator $[m] \to [m']$ induces a weak homotopy equivalence $(W\mathcal{C})_{m',*} \to (W\mathcal{C})_{m,*}$.*

Proof Note that $(W\mathcal{C})_{m,*} = N\bar{\mathcal{C}}_m$, where $\bar{\mathcal{C}}_m$ is a category whose objects are diagrams $\widehat{[m]} \to g\mathcal{C}$ and whose morphisms are acyclic levelwise cofibrations. Again, it is enough to consider the case $m' = 0$ and to show that the constant functor const: $\bar{\mathcal{C}}_0 \to \bar{\mathcal{C}}_m$ and the evaluation at m functor $\mathrm{ev}_m\colon \bar{\mathcal{C}}_n \to \bar{\mathcal{C}}_0$ form a homotopy equivalence.

We have $\mathrm{ev}_m \,\mathrm{const} = \mathrm{id}_{\bar{\mathcal{C}}_0}$. Moreover, given any object $X \in \bar{\mathcal{C}}_m$ and $i \in [m]$ we consider the composite weak equivalence $X_i \xrightarrow{\sim} X_m$. We combine it with the identity $X_m \to X_m$ and factor functorially the resulting morphism $X_i \sqcup X_m \to X_m$ as $X_i \sqcup X_m \rightarrowtail \widetilde{X}_i \xrightarrow{\sim} X_m$. In the square

$$\begin{array}{ccc} X_m \sqcup X_i & \longrightarrow & X_m \\ \downarrow & & \downarrow \\ X_m \sqcup X_{i+1} & \longrightarrow & X_m \end{array}$$

both vertical morphisms are in $g\mathcal{C}$ (since $g\mathcal{C}$ is closed under pushouts). Thus the induced morphism $\widetilde{X}_i \to \widetilde{X}_{i+1}$ is in $g\mathcal{C}$. Moreover, we obtain acyclic cofibrations $X_i \xrightarrow{\sim} \widetilde{X}_i$ and $X_m \xrightarrow{\sim} \widetilde{X}_i$ that constitute a zigzag of natural weak equivalences connecting $\mathrm{const}\,\mathrm{ev}_m$ and $\mathrm{id}_{\bar{\mathcal{C}}_m}$. □

Lemma 14 *The inclusion $Nwc\mathcal{C} \to Nwg\mathcal{C}$ is a weak homotopy equivalence.*

Proof Observe that the 0^{th} row and the 0^{th} column of $W\mathcal{C}$ are $Nwg\mathcal{C}$ and $Nwc\mathcal{C}$, respectively. Since $W\mathcal{C}$ is homotopically constant in both directions, it follows from [7, Proposition IV.1.7] that we have weak equivalences

$$Nwg\mathcal{C} \xrightarrow{\sim} \operatorname{diag} W\mathcal{C} \xleftarrow{\sim} Nwc\mathcal{C}.$$

Moreover, the restrictions along the diagonal inclusions $[m] \to [m] \times [m]$ induce a simplicial map $\operatorname{diag} W\mathcal{C} \to Nwg\mathcal{C}$ whose composites with the two maps above are the identity on $Nwg\mathcal{C}$ and the inclusion $Nwc\mathcal{C} \to Nwg\mathcal{C}$. Hence the latter is a weak equivalence by 2-out-of-3. □

Next we establish that under specific circumstances certain subcategories of $\mathcal{C}$ are good.

Lemma 15 *Let $\mathcal{C}$ be a cofibration category.*

1. *If $\mathcal{C}$ has functorial factorizations, then $\mathcal{C}$ itself is a good subcategory.*
2. *If $\mathcal{C}$ has good cylinders, then $c\mathcal{C}$ is a good subcategory of $\mathcal{C}$.*
3. *If $c\mathcal{C}$ is a good subcategory of $\mathcal{C}$, then the subcategory of levelwise cofibrations is a good subcategory of $\mathcal{C}^{[k]}_{\mathrm{R}}$ for all k.*

Proof (1) This is vacuously true.

(2) We will show that the standard mapping cylinder factorization makes $c\mathcal{C}$ into a good subcategory. Let

$$\begin{array}{ccc} A_0 & \longrightarrow & B_0 \\ \downarrow\!\!\!\!\!\!\rightarrowtail & & \downarrow\!\!\!\!\!\!\rightarrowtail \\ A_1 & \longrightarrow & B_1 \end{array}$$

be a square where both vertical morphisms are cofibrations. The mapping cylinder of $A_i \to B_i$ is constructed as $IA_i \sqcup_{A_i \sqcup A_i} (A_i \sqcup B_i)$. We need to show that the morphism induced by the square

$$\begin{array}{ccc} A_0 & \longrightarrow & IA_0 \sqcup_{A_0 \sqcup A_0} (A_0 \sqcup B_0) \\ \downarrow & & \downarrow \\ A_1 & \longrightarrow & IA_1 \sqcup_{A_1 \sqcup A_1} (A_1 \sqcup B_1) \end{array}$$

is a cofibration. This morphism coincides with

$$IA_0 \sqcup_{A_0 \sqcup A_0} (A_1 \sqcup B_0) \to IA_1 \sqcup_{A_1 \sqcup A_1} (A_1 \sqcup B_1),$$

which factors as

$$IA_0 \sqcup_{A_0 \sqcup A_0} (A_1 \sqcup B_0) \to IA_0 \sqcup_{A_0 \sqcup A_0} (A_1 \sqcup B_1) \to IA_1 \sqcup_{A_1 \sqcup A_1} (A_1 \sqcup B_1).$$

The first morphism is a pushout of $A_1 \sqcup B_0 \to A_1 \sqcup B_1$, which is a cofibration since $B_0 \to B_1$ is. Comparing the pushouts of rows and columns in the diagram

$$\begin{array}{ccccc} IA_0 & \longleftarrow & IA_0 & \longrightarrow & IA_1 \\ \uparrow & & \uparrow & & \uparrow \\ A_0 \sqcup A_0 & \longleftarrow & A_0 \sqcup A_0 & \longrightarrow & A_1 \sqcup A_1 \\ \downarrow & & \downarrow & & \downarrow \\ A_1 \sqcup B_1 & \longleftarrow & A_1 \sqcup A_1 & \longrightarrow & A_1 \sqcup A_1 \end{array}$$

shows that the second morphism above is a pushout of $IA_0 \sqcup_{A_0 \sqcup A_0} (A_1 \sqcup A_1) \to IA_1$, which is a cofibration since $A_0 \to A_1$ is and $\mathcal{C}$ has good cylinders.

(3) Clearly, every Reedy cofibration is a levelwise cofibration and levelwise cofibrations are stable under pushouts. Consider a diagram

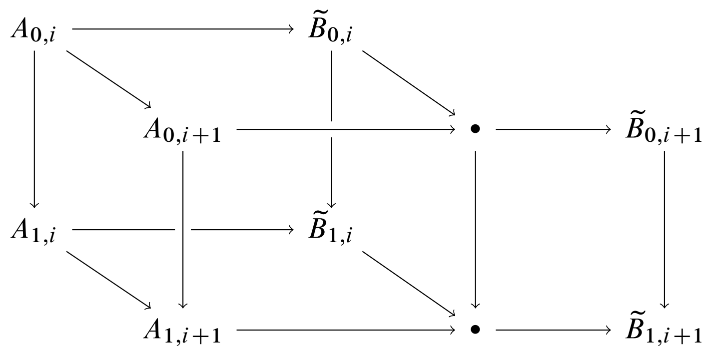
$$\begin{array}{ccccc} A_0 & \rightarrowtail & \widetilde{B}_0 & \xrightarrow{\sim} & B_0 \\ \downarrow & & \downarrow & & \downarrow \\ A_1 & \rightarrowtail & \widetilde{B}_1 & \xrightarrow{\sim} & B_1 \end{array}$$

in $\mathcal{C}^J_{\mathrm{R}}$, where $\widetilde{B}_0$ and $\widetilde{B}_1$ are obtained by the standard Reedy factorization (see [14, Proof of Theorem 9.2.4(v)]) induced by the given functorial factorization in $\mathcal{C}$. Assuming that $A_0 \to A_1$ and $B_0 \to B_1$ are levelwise cofibrations, we need to check that $A_{1,i} \sqcup_{A_{0,i}} \widetilde{B}_{0,i} \to \widetilde{B}_{1,i}$ is a cofibration for every $i \in [m]$.

For $i = 0$, this follows directly from the assumption that $c\mathcal{C}$ is a good subcategory of $\mathcal{C}$. The Reedy factorization is constructed by induction over $[m]$, so assume that the conclusion is already known for $i < m$. The factorization at level $i+1$ arises as

$$\begin{array}{ccccc} A_{0,i+1} \sqcup_{A_{0,i}} \widetilde{B}_{0,i} & \rightarrowtail & \widetilde{B}_{0,i+1} & \xrightarrow{\sim} & B_{0,i+1} \\ \downarrow & & \downarrow & & \downarrow \\ A_{1,i+1} \sqcup_{A_{1,i}} \widetilde{B}_{1,i} & \rightarrowtail & \widetilde{B}_{1,i+1} & \xrightarrow{\sim} & B_{1,i+1} \end{array}$$

where the left square comes from the diagram



where the bullets stand for the pushouts above. The conclusion we need to obtain amounts to the composite of the two squares in the front being a Reedy cofibration when seen as a morphism from left to right. The right square is a Reedy cofibration since $c\mathcal{C}$ is a good subcategory of $\mathcal{C}$ and so is the left one since it is a pushout of the back square, which is a Reedy cofibration by the inductive hypothesis. □

Lemma 16 *The inclusion $Nw(\mathcal{C}^{[k]}_{\mathrm{R}}) \to Nw(\mathcal{C}^{[k]})$ is a weak homotopy equivalence.*

Proof Functorial factorization induces a functor in the opposite direction as well as natural weak equivalences connecting both composites with identities. □

Proof of Theorem 10 Recall that we want to show that $Nw((c\mathcal{C})^{[k]}) \to Nw(\mathcal{C}^{[k]})$ is a weak equivalence for all k. In the diagram

$$
\begin{array}{ccccc}
 & & Nwc(\mathcal{C}_{\mathrm{R}}^{[k]}) & \xrightarrow{(1)} & Nw(\mathcal{C}_{\mathrm{R}}^{[k]}) \\
 & \overset{(3)}{\swarrow} & \downarrow & & \downarrow{\scriptstyle (4)} \\
Nw((c\mathcal{C})^{[k]}) & \longrightarrow & Nwc(\mathcal{C}^{[k]}) & \xrightarrow{(2)} & Nw(\mathcal{C}^{[k]})
\end{array}
$$

the labelled maps are weak equivalences. The map (1) is a weak equivalence by Lemma 14 applied to $\mathcal{C}_{\mathrm{R}}^{[k]}$ with itself as a good subcategory and so is the map (2) by the same argument applied to $\mathcal{C}^{[k]}$. The map (3) is a weak equivalence by Lemma 14 applied to $\mathcal{C}_{\mathrm{R}}^{[k]}$ with the good subcategory of levelwise cofibrations, which is indeed good by Lemma 15. Finally, the map (4) is a weak equivalence by Lemma 16. Hence by 2-out-of-3, the bottom composite is also a weak equivalence as required. □

Acknowledgements We would like to thank Ulrich Bunke for comments on an earlier version of this article. M Land was supported by Wolfgang Lück's ERC Advanced Grant "KL2MG-interactions" (no. 662400) granted by the European Research Council.

References

[1] **K S Brown**, *Abstract homotopy theory and generalized sheaf cohomology*, Trans. Amer. Math. Soc. 186 (1973) 419–458 MR

[2] **C Casacuberta**, **M Golasiński**, **A Tonks**, *Homotopy localization of groupoids*, Forum Math. 18 (2006) 967–982 MR

[3] **D-C Cisinski**, *Does the classification diagram localize a category weak equivalences?*, MathOveflow post (2012) Available at `https://mathoverflow.net/q/92916`

[4] **J F Davis**, **W Lück**, *Spaces over a category and assembly maps in isomorphism conjectures in K– and L–theory*, K–Theory 15 (1998) 201–252 MR

[5] **I Dell'Ambrogio**, *The unitary symmetric monoidal model category of small* C*–*categories*, Homology Homotopy Appl. 14 (2012) 101–127 MR

[6] **D Dugger**, *Combinatorial model categories have presentations*, Adv. Math. 164 (2001) 177–201 MR

[7] **P G Goerss**, **J F Jardine**, *Simplicial homotopy theory*, Progress in Mathematics 174, Birkhäuser, Basel (1999) MR

[8] **M Hovey**, *Model categories*, Mathematical Surveys and Monographs 63, Amer. Math. Soc., Providence, RI (1999) MR

[9] **M Joachim**, *K–homology of C^*–categories and symmetric spectra representing K–homology*, Math. Ann. 327 (2003) 641–670 MR

[10] **M Land**, **T Nikolaus**, *On the relation between K– and L–theory of C^*–algebras*, preprint (2016) arXiv

[11] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR

[12] **J Lurie**, *Higher algebra*, book in progress, Harvard (2017) Available at `http://math.harvard.edu/~lurie/papers/HA.pdf`

[13] **A Mazel-Gee**, *The universality of the Rezk nerve*, preprint (2015) arXiv

[14] **A Radulescu-Banu**, *Cofibrations in homotopy theory*, preprint (2009) arXiv

[15] **I Raeburn**, **D P Williams**, *Morita equivalence and continuous-trace C^*–algebras*, Mathematical Surveys and Monographs 60, Amer. Math. Soc., Providence, RI (1998) MR

[16] **C Rezk**, *A model for the homotopy theory of homotopy theory*, Trans. Amer. Math. Soc. 353 (2001) 973–1007 MR

[17] **S Schwede**, *The stable homotopy category is rigid*, Ann. of Math. 166 (2007) 837–863 MR

Mathematical Institute, University of Regensburg
Regensburg, Germany

Max Planck Institut für Mathematik
Bonn, Germany

Department of Mathematics, University of Western Ontario
London, ON, Canada

`markus.land@mathematik.uni-regensburg.de`, `thoni@mpim-bonn.mpg.de`, `kszumilo@uwo.ca`

`http://www.markus-land.de`, `http://people.mpim-bonn.mpg.de/thoni/`, `http://www-home.math.uwo.ca/~kszumilo/`

Received: 28 October 2016 Revised: 2 May 2017

msp

Algebraic & Geometric Topology 17 (2017) 3021–3056

HOMFLY-PT homology for general link diagrams and braidlike isotopy

MICHAEL ABEL

Khovanov and Rozansky's categorification of the HOMFLY-PT polynomial is invariant under braidlike isotopies for any general link diagram and Markov moves for braid closures. To define HOMFLY-PT homology, they required a link to be presented as a braid closure, because they did not prove invariance under the other oriented Reidemeister moves. In this text we prove that the Reidemeister IIb move fails in HOMFLY-PT homology by using virtual crossing filtrations of the author and Rozansky. The decategorification of HOMFLY-PT homology for general link diagrams gives a deformed version of the HOMFLY-PT polynomial, $P^b(D)$, which can be used to detect nonbraidlike isotopies. Finally, we will use $P^b(D)$ to prove that HOMFLY-PT homology is not an invariant of virtual links, even when virtual links are presented as virtual braid closures.

57M25, 57M27

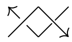
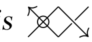
1 Introduction

Khovanov and Rozansky in [12] introduced a triply graded link homology theory categorifying the HOMFLY-PT polynomial. The construction given in [12] of Khovanov–Rozansky HOMFLY-PT homology, or briefly HOMFLY-PT homology, is an invariant of link diagrams up to braidlike isotopy (isotopies which locally resemble isotopies of a braid) and Markov moves for closed braid diagrams. However, Khovanov and Rozansky were not able to prove invariance under all oriented Reidemeister moves. In particular, they could not prove the Reidemeister IIb move, and in fact expected that it would fail in general. Because of this, they required that a link be presented as a braid closure so that HOMFLY-PT homology would be an invariant of links. In this text we will directly address this issue by proving the failure of the Reidemeister IIb move in HOMFLY-PT homology, and explore the consequences of this failure.

The framework of HOMFLY-PT homology can be extended to include the use of "virtual crossings", degree-4 vertices which are not actually positive or negative crossings. Virtual links (links with virtual crossings) were first introduced by Kauffman in [9]. The author and Rozansky in [1] proved that a filtration can be placed on the chain complex whose homology is HOMFLY-PT homology. The associated graded complex of this

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3021

filtration is described using diagrams containing only virtual crossings. The filtration allows us to rewrite the chain complexes in an illuminating manner, allowing us to see new isomorphisms which would be difficult to see otherwise. Using the framework of virtual crossing filtrations we prove the following theorem.

Theorem 1.1 (see Theorem 4.11) *Let $\mathcal{H}(D)$ denote the* HOMFLY-PT *homology of the virtual link diagram D. Suppose D_1, D_2, and D_3 are oriented virtual link diagrams which are identical except in the neighborhood of a single point. Suppose in the neighborhood of that point, D_1 is* *, D_2 is* *, and D_3 is* *. Then $\mathcal{H}(D_1) \simeq \mathcal{H}(D_3)$ up to a grading shift, while $\mathcal{H}(D_1) \not\simeq \mathcal{H}(D_2)$ in general.*

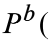
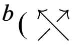
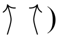
In Section 4 we prove the above theorem and give an explicit example of a diagram of the unknot that does not have the HOMFLY-PT homology of the unknot (Example 4.12). Recall $\mathcal{H}(D)$ is a triply graded vector space. Suppose $d_{ijk} = \dim(\mathcal{H}(D)_{i,j,k})$. Then we can define the Poincaré series of $\mathcal{H}(D)$ as

$$\mathcal{P}(D) = \sum_{i,j,k \in \mathbb{Z}} d_{ijk} q^i a^j t^k. \tag{1-1}$$

Let $P(D)$ denote the HOMFLY-PT polynomial of the link diagram D. In [13], Murakami, Ohtsuki and Yamada introduced a state-sum formulation of the HOMFLY-PT polynomial commonly called the MOY construction. Their approach resolves a link diagram into a $\mathbb{Z}(q,a)$–linear combination of oriented planar 4–regular graphs. They give relations which evaluate each such planar graph as an element of $\mathbb{Z}(q,a)$. The resulting rational function from this process for any link diagram D is its HOMFLY-PT polynomial $P(D)$.

We now define a deformed HOMFLY-PT polynomial $P^b(D) = \mathcal{P}(D)|_{t=-1}$. In the case that D is presented as a braid closure then $P^b(D) = P(D)$. However, this is not true for general link diagrams. We collect known relations and properties of $P^b(D)$ into the following theorem.

Theorem 1.2 (see Theorem 5.1) *Let D be a link diagram. $P^b(D)$ is an invariant of link diagrams up to braidlike isotopy satisfying the skein relation*

$q P^b($  $) - q^{-1} P^b($  $) = (q - q^{-1}) P^b($  $).$

Furthermore, $P^b(D)$ satisfies the relations in Figure 1 in addition to the virtual MOY/Reidemeister moves and Z-moves (see Figures 14 and 15).

In Section 5 we use $P^b(D)$ to show that $\mathcal{H}(D)$ is not an invariant of virtual links, even when presented as a virtual braid closure, by showing it violates the virtual exchange move (see Kamada [8]).

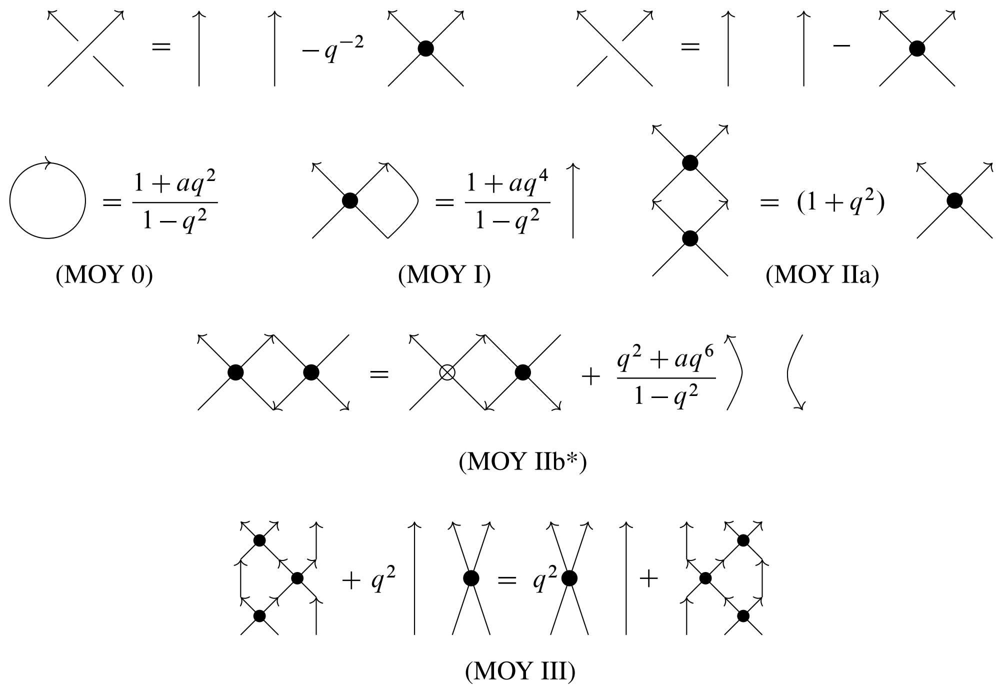


Figure 1: Relations for $P^b(D)$ (the notation $P^b(\cdot)$ omitted for readability)

We note that the relations in Figure 1 may not always be enough to determine $P^b(D)$, though in many examples the relations do suffice. We expect that in fact these relations will not compute $P^b(D)$ in general. The "nonbraidlike" MOY III diagram in Figure 2 does not split in a tractable manner and applying the MOYIIb relation introduces virtual crossings. Kauffman and Manturov in [10] construct an $\mathfrak{sl}_3$ specialization of the HOMFLY-PT polynomial for virtual links as formal $\mathbb{Z}[q, q^{-1}]$–linear combinations of directed graphs which are irreducible under MOY I, MOY IIa, MOY III, virtual MOY moves and Z-moves. Since our $P^b(D)$ for virtual braid closures specializes to their invariant, we expect that certain virtual MOY graphs will be irreducible with respect to the relations in Figure 1.

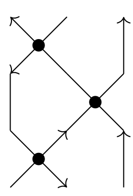
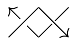
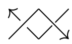
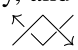
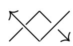


Figure 2: A "nonbraidlike" MOY III configuration

Recent research involving annular link homology by Auroux, Grigsby and Wehrli in the $\mathfrak{sl}_2$ case [4] and Queffelec and Rose in the $\mathfrak{sl}_n$ case [15] give some insight into why to expect this issue. We consider links in the thickened annulus as closed braids in a 3–ball D^3 with the braid axis ℓ removed from D^3. Annular link homology theories, that is homology theories of closed braids in the thickened annulus $D^3 - \ell$, are normally constructed via the use of Hochschild homology on chain complexes of bimodules associated to braids. Hochschild homology $\mathrm{HH}(C)$ acts as a (horizontal) trace on the homotopy category of bimodules, but in general does not act like a Markov trace. In particular, if β_1 and β_2 are two braids which are Markov equivalent and $C(\beta_1)$ and $C(\beta_2)$ are their associated chain complexes of bimodules, then $\mathrm{HH}(C(\beta_1))$ is not necessarily homotopy equivalent to $\mathrm{HH}(C(\beta_2))$. This corresponds to the fact that even though the braid closures of β_1 and β_2 are isotopic as links in S^3, they may not be isotopic in $D^3 - \ell$.

From this viewpoint, $\mathcal{H}(D)$ is an annular link invariant that happens to satisfy the Markov moves (that is, $\mathcal{H}(D)$ is a categorified Markov trace). This is why $\mathcal{H}(D)$ gives invariants of links in S^3 when D is presented as the closure of a braid. The Reidemeister IIb configuration  can only appear in a braid closure when the braid axis is between the two strands. In the case of annular invariants the braid axis is an obstruction to isotopy, and disallows the isotopy  $\sim$ . In an annular invariant the exchange move  $\sim$  is disallowed; however, this move preserves the isomorphism type of $\mathcal{H}(D)$.

Outline of the paper In Section 2 we review the definition of the HOMFLY-PT polynomial and the MOY construction of the HOMFLY-PT polynomial. We use nonstandard conventions in this text to illuminate the connections with HOMFLY-PT homology. In Section 3 we review the construction of HOMFLY-PT homology of links using closed braid diagrams. We also review some homological algebra, in particular properties of Koszul complexes. In Section 4 we explore the properties of HOMFLY-PT homology for general link diagrams. We introduce the role of virtual crossings in this framework and use virtual crossings as a tool to prove that HOMFLY-PT homology is not invariant under the Reidemeister IIb move. Finally, in Section 5 we explore the decategorification (Euler characteristic) of HOMFLY-PT homology and use it to prove that HOMFLY-PT homology cannot be extended to an invariant of virtual links.

Acknowledgements The author would like to thank Mikhail Khovanov, Lev Rozansky and Lenny Ng for many helpful conversations and their feedback. The author would like to also thank Matt Hogancamp for encouraging him to further explore an observation which eventually became the text here. The author was supported in part by NSF grant DMS-1406371.

2 MOY construction of the HOMFLY-PT polynomial

We begin by recalling two constructions of the HOMFLY-PT polynomial for oriented links. Most of this material is well-known, but we introduce it with the purpose of setting our conventions for the sequel. The first construction is given by a skein relation and first appeared in [6]. The second construction, first introduced by Murakami, Ohtsuki and Yamada in [13], constructs the HOMFLY-PT polynomial in terms of a state-sum formula. It is this second construction which is categorified in the construction of Khovanov and Rozansky's HOMFLY-PT homology.

2.1 The HOMFLY-PT polynomial of an oriented link diagram

Let L denote a link in $\mathbb{R}^3$. In this text we will assume all links are oriented. Let D denote a link diagram of L, that is a regular projection of L onto a copy of $\mathbb{R}^2$. The HOMFLY-PT polynomial is an invariant of links which takes (oriented) link diagrams to elements of $\mathbb{Z}(q,a)$.

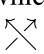
Definition 2.1 Let D be a link diagram and let O be a simple closed curve in the plane of the link diagram. We define the HOMFLY-PT *polynomial,* $P(D) \in \mathbb{Z}(q,a)$, via the following relations:

(1) $P(\varnothing) = 1$ and $P(O) = (1+aq^2)/(1-q^2)$.

(2) $P(D \sqcup O) = P(D)P(O)$.

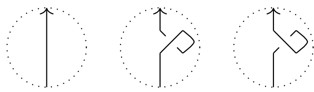
(3) $qP(D_+) - q^{-1}P(D_-) = (q-q^{-1})P(D_0)$, where D_+, D_-, and D_0 are link diagrams which are the same except in the neighborhood of a single point where $D_+ =$ , $D_- =$ , and $D_0 =$ .

(4) If D and D' differ by a sequence of Reidemeister II and III moves (with any orientation), then $P(D) = P(D')$.

We will call a crossing which locally looks like  a *positive* crossing, and a crossing that locally looks like  a *negative crossing.* Let $f, g \in \mathbb{Z}(q,a)$ be nonzero. We will write $f \doteq g$ if $f = (-1)^i a^j q^k g$ for some $i, j, k \in \mathbb{Z}$. In other words, we write $f \doteq g$ if f/g is a unit in $\mathbb{Z}[q^{\pm 1}, a^{\pm 1}]$.

Theorem 2.2 (HOMFLY [6]; PT [14]) *Let D and D' be two link diagrams of a link L. Then $P(D) \doteq P(D')$. Furthermore, $P(D_2) = -q^{-2}P(D_1)$ and $P(D_3) = aq^2P(D_1)$, where D_1, D_2, D_3 are link diagrams which are the same except in the neighborhood of a single point where they are as in Figure 3.*

We will often denote the HOMFLY-PT polynomial of a link by $P(L)$, suppressing the choice of link diagram. In this case $P(L)$ is well defined up to a unit in $\mathbb{Z}[q^{\pm 1}, a^{\pm 1}]$.



D_1 D_2 D_3

Figure 3: The diagrams D_1, D_2, and D_3 from Theorem 2.2

Remark 2.3 We are using nonstandard conventions for the HOMFLY-PT polynomial in this text. The HOMFLY-PT polynomial as defined here is not a polynomial, but rather is a rational function. One may choose a different normalization where both $P(D) \in \mathbb{Z}[a^{\pm 1}, q^{\pm 1}]$ is honestly a (Laurent) polynomial and $P(D_1) = P(D_2) = P(D_3)$ where D_1, D_2, and D_3 are as in Figure 3. The choice of normalization here coincides with our conventions for HOMFLY-PT homology in the sequel.

2.2 The MOY construction of the HOMFLY-PT polynomial

Murakami, Ohtsuki and Yamada in [13] give a construction of the $\mathfrak{sl}_n$ polynomial, $P_n(L) \in \mathbb{Z}[q, q^{-1}]$, of a link L using evaluations of oriented colored trivalent plane graphs. These trivalent plane graphs correspond to the intertwiners between tensor powers of fundamental representations of $\mathcal{U}_q(\mathfrak{sl}_n)$. The $\mathfrak{sl}_n$ polynomial is actually a specialization of the HOMFLY-PT polynomial, that is $P_n(L)(q) = P(L)(q, a = q^{2-2n})$ in our conventions. We may adjust the MOY construction of the $\mathfrak{sl}_n$ polynomial to compute the HOMFLY-PT polynomial. We will replace the "wide edge" graph of [13] with a single degree-4 vertex (see Figure 4) which we will call a *MOY vertex.*

Recall an oriented graph is 4*–regular* if every vertex has degree 4, that is if each vertex has a total of 4 outgoing/incoming edges. The MOY state model of the HOMFLY-PT polynomial writes a link diagram as a formal $\mathbb{Z}(q, a)$–linear combination of planar, oriented, 4–regular graphs. The orientation locally at each vertex is the same as the orientation of the MOY vertex in Figure 4. We call such planar, oriented, 4–regular graphs *MOY graphs.*

We now define the MOY construction of the HOMFLY-PT polynomial. Let D be a link diagram. We can resolve any crossing c into either an oriented smoothing  or a MOY vertex  (with consistent orientation). To each resolution of c we associate a *weight*. If we smooth the crossing then the resolution has weight 0. If we replace the crossing with a MOY vertex, then the weight is -2 if the crossing was positive and 0

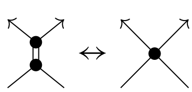


Figure 4: The MOY wide edge graph and our MOY vertex

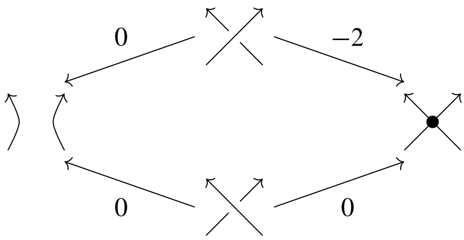


Figure 5: Resolution chart

if the crossing was negative. A resolution chart is given in Figure 5 for reference. We define a *state* σ of D as a choice of resolution for every crossing c in D. If D has n crossings, then it has 2^n possible states. We define the *weight of a state* $\mu(\sigma)$ to be the sum of the weights of the chosen resolutions of σ. Finally we will set $\nu(\sigma)$ to be the number of MOY vertices in σ.

Theorem 2.4 (Murakami, Ohtsuki and Yamada [13]) *The relations given in Figure 6 are sufficient to compute $\overline{P}(D_\sigma)$ as an element of $\mathbb{Z}(q,a)$ for any link diagram D and any state σ.*

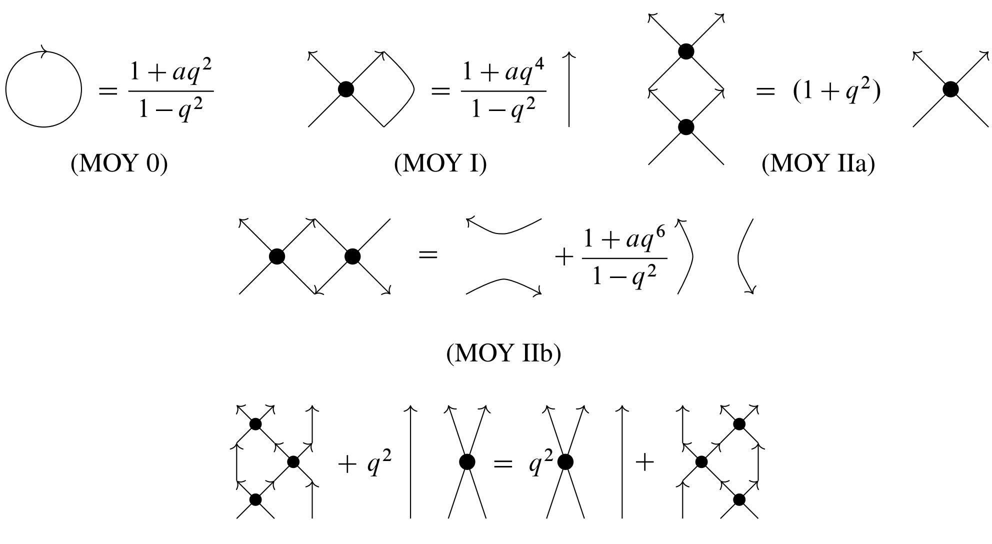


Figure 6: MOY relations (the notation $\overline{P}(\cdot)$ omitted for readability)

Definition 2.5 The *MOY polynomial,* $\overline{P}(D)$, is given by

$$\overline{P}(D) = \sum_{\sigma} (-1)^{\nu(\sigma)} q^{\mu(\sigma)}\, \overline{P}(D_\sigma). \tag{2-1}$$

Theorem 2.6 (Murakami, Ohtsuki and Yamada [13]) *Let D be an oriented link diagram. Then $\overline{P}(D) = P(D)$.*

Example 2.7 Using the relations in Figure 6, we compute $\overline{P}(D)$ for the diagram of the left-handed trefoil knot given in Figure 7. We leave it as an exercise to the reader to show

$$\overline{P}(D) = \left(\frac{1+aq^2}{1-q^2}\right)(q^2+aq^2+aq^6). \tag{2-2}$$

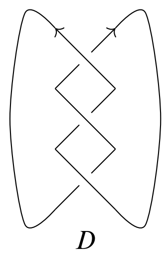


Figure 7: Diagram of the left-handed trefoil knot

3 HOMFLY-PT homology for closed braid diagrams

In this section we introduce the construction of Khovanov and Rozansky's HOMFLY-PT homology. The approach of this construction is to associate a chain complex of modules to every MOY graph and a *bicomplex* of modules to every link diagram. Our approach in this section is most similar to the approach of Rasmussen in [16] where we ignore his "$\mathfrak{sl}_n$" differential, as it is not needed in the construction of HOMFLY-PT homology.

3.1 Koszul complexes

Before introducing HOMFLY-PT homology, we recall some terminology and notation involving Koszul complexes. Let $R = \bigoplus_{i\in\mathbb{Z}} R_i$ be a $\mathbb{Z}$–graded commutative $\mathbb{Q}$–algebra and $M = \bigoplus_{i\in\mathbb{Z}} M_i$ be a $\mathbb{Z}$–graded R–module. It will be instructive to keep the example of $R = \mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]$ in mind, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are finite lists of variables (not necessarily of the same length). We define the grading shift functor $\bullet(k)$ by $M(k)_j = M_{j-k}$ for all $j\in\mathbb{Z}$. We will commonly use a nonstandard notation for grading shifts. In particular, we will set $q^kM := M(k)$ and say $\deg_q(x) = q^j$ if $x\in M_j$.

Definition 3.1 Let $p \in R$ be an element of degree k. The *Koszul complex* of p is defined as the chain complex

$$[p]_R = q^k R_1 \xrightarrow{p} R_0,$$

where p is used to denote the algebra endomorphism of R given by multiplication by p. Here $R_0 = R_1 = R$ and the subscript is simply used to denote the homological degree of the module. We will often write $[p] = [p]_R$ when there can be no confusion. Now let $\boldsymbol{p} = p_1, \dots, p_k$ be a sequence of elements in R. Then we define the *Koszul complex* of $\boldsymbol{p}$ as the complex

$$\begin{bmatrix} p_1 \\ \vdots \\ p_k \end{bmatrix} = [p_1] \otimes_R \cdots \otimes_R [p_k],$$

where $\otimes_R$ denotes the ordinary tensor product of chain complexes.

As a convention, we will call the homological grading in Koszul complexes the *Hochschild grading* and denote it by $\deg_a$. We write $\deg_a(x) = a^k$ to say that x is in Hochschild degree k and similarly write $a^k M$ to denote that M is being shifted k in Hochschild degree.

We say a sequence of elements $\boldsymbol{p} = p_1, \dots, p_k$ in R is a *regular sequence* if p_m is not a zero divisor in $R/(p_1, \dots, p_{m-1})$ for all $m = 1, \dots, k$. The following proposition is a standard fact in homological algebra and is proven in many introductory texts such as [18].

Proposition 3.2 *Let $\boldsymbol{p} = p_1, \dots, p_n$ be a regular sequence in R. Then the Koszul complex of $\boldsymbol{p}$ is a graded free R–module resolution of $R/(p_1, \dots, p_n)$.*

The notation we use for Koszul complexes is reminiscent of the notation for a column vector in $R^{\oplus n}$. Note that we will always use square brackets for Koszul complexes and round brackets for row vectors in $R^{\oplus n}$ to eliminate any confusion. Along these lines, we can look at "row operations" on Koszul complexes.

Proposition 3.3 *Let $\boldsymbol{p} = p_1, \dots, p_k$ be a sequence of elements in R, and let $\lambda \in \mathbb{Q}$. Then*

$$\begin{bmatrix} \vdots \\ p_i \\ \vdots \\ p_j \\ \vdots \end{bmatrix} \simeq \begin{bmatrix} \vdots \\ p_i + \lambda p_j \\ \vdots \\ p_j \\ \vdots \end{bmatrix}.$$

A homotopy equivalence of this form will be called a change of basis.

Proof We will omit grading shifts in the proof for clarity. We consider the map $\Phi\colon [p_i]\otimes_R[p_j] \to [p_i+\lambda p_j]\otimes_R[p_j]$ given by:

$$\begin{array}{ccccccc}
[p_i]\otimes_R[p_j] & = & R & \xrightarrow{\begin{pmatrix} p_i \\ p_j \end{pmatrix}} & R\oplus R & \xrightarrow{(-p_j \;\; p_i)} & R \\
\Big\downarrow{\scriptstyle \Phi} & & \Big\downarrow{\scriptstyle 1} & & \Big\downarrow{\scriptstyle \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}} & & \Big\downarrow{\scriptstyle 1} \\
[p_i+\lambda p_j]\otimes_R[p_j] & = & R & \xrightarrow[\begin{pmatrix} p_i+\lambda p_j \\ p_j \end{pmatrix}]{} & R\oplus R & \xrightarrow[(-p_j \;\; p_i+\lambda p_j)]{} & R
\end{array}$$

This map is clearly invertible. □

3.2 Marked MOY graphs

A *marked MOY graph* is a MOY graph Γ (possibly with boundary) with markings such that the marks partition the graph into some combination of *elementary MOY graphs* as shown in Figure 8. We label the marks and the endpoints of the graph (if any) with variables. Typically, though not necessarily, we will label outgoing edges by variables y_i, incoming edges by variables x_i, and internal marks by variables t_i. An example of this process is given in Figure 8.

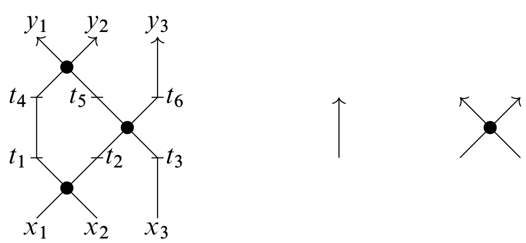


Figure 8: An example of a marked MOY graph and the elementary MOY graphs

To a marked MOY graph Γ, we will associate a collection of rings. Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{t}$ denote the lists of incoming, outgoing, and internal variables respectively. We first define the *total ring* $E^t(\Gamma)$ of Γ as the polynomial ring $\mathbb{Q}[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{t}]$ containing all variables. We make this ring into a graded ring by setting $\deg_q(x_i) = \deg_q(y_i) = \deg_q(t_i) = q^2$. We call this grading the *internal* or *quantum grading*. We also suppose that all elements in $E^t(\Gamma)$ have Hochschild degree a^0. The other rings we will define will be subrings of $E^t(\Gamma)$. The *edge ring*, $E(\Gamma)$, is the polynomial ring of incoming and outgoing ("edge") variables $\mathbb{Q}[\boldsymbol{x}, \boldsymbol{y}]$. The total ring $E^t(\Gamma)$ has a natural free $E(\Gamma)$–module structure. We also define the *incoming ring* (resp. *outgoing ring*) by $E^i(\Gamma) = \mathbb{Q}[\boldsymbol{x}]$ (resp. $E^o(\Gamma) = \mathbb{Q}[\boldsymbol{y}]$). Since $E(\Gamma) \cong E^i(\Gamma) \otimes_{\mathbb{Q}} E^o(\Gamma)$ as $\mathbb{Q}$–algebras, any $E(\Gamma)$–module can be considered as an $E^i(\Gamma)$–$E^o(\Gamma)$–bimodule. Note that if Γ does not have any boundary (eg if it is a resolution of a link diagram), then $E(\Gamma) \cong E^i(\Gamma) \cong E^o(\Gamma) \cong \mathbb{Q}$.

We will now define chain complexes $C(\Gamma)$ of free $E(\Gamma)$–modules associated to a marked MOY graph Γ. The chain modules of $C(\Gamma)$ will be direct sums of shifted copies of $E^t(\Gamma)$. We do this by first defining Koszul complexes associated to the elementary MOY graphs and then give rules for how gluing the graphs together affects the complexes associated to them. We will use the symbols $\uparrow$ and $\times$ to denote the elementary arc and vertex MOY graphs. To the arc, we associate the Koszul complex of modules over $E^t(\uparrow) = E(\uparrow) = \mathbb{Q}[x, y]$,

$$C(\uparrow) = [y - x]_{E(\uparrow)} = q^2 a E(\uparrow) \xrightarrow{y-x} E(\uparrow), \tag{3-1}$$

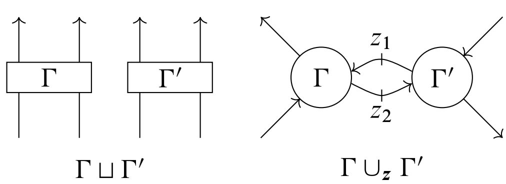
and to the vertex graph, we associate the Koszul complex of modules over $E^t(\times) = E(\times) = \mathbb{Q}[x_1, x_2, y_1, y_2]$,

$$\begin{aligned} C(\times) &= \begin{bmatrix} y_1 + y_2 - x_1 - x_2 \\ (y_1 - x_1)(y_1 - x_2) \end{bmatrix}_{E(\times)} \\ &= q^6 a^2 E(\times) \xrightarrow{A} q^4 a E(\times) \oplus q^2 a E(\times) \xrightarrow{B} E(\times), \end{aligned} \tag{3-2}$$

where

$$A = \begin{pmatrix} y_1 + y_2 - x_1 - x_2 \\ (y_1 - x_1)(y_1 - x_2) \end{pmatrix}, \quad B = \begin{pmatrix} -(y_1 - x_1)(y_1 - x_2) & y_1 + y_2 - x_1 - x_2 \end{pmatrix}.$$

Now suppose Γ is a marked MOY graph with edge ring E and total ring E^t. Also let Γ' be another marked MOY graph with edge ring E' and total ring E'^t. The disjoint union of these graphs $\Gamma \sqcup \Gamma'$ has edge ring $E'' \cong E \otimes_{\mathbb{Q}} E'$ and total ring $E''^t \cong E^t \otimes_{\mathbb{Q}} E'^t$. To the marked MOY graph $\Gamma \sqcup \Gamma'$ we will associate the complex of E''–modules $C(\Gamma \sqcup \Gamma') := C(\Gamma) \otimes_{\mathbb{Q}} C(\Gamma')$. A picture of the corresponding diagram is shown in Figure 9.



$\Gamma \sqcup \Gamma'$ $\Gamma \cup_{\boldsymbol{z}} \Gamma'$

Figure 9: Examples of disjoint union and gluing of marked MOY graphs

Finally, we define a complex for when we glue two marked MOY graphs together. Let Γ and Γ' be two marked MOY graphs. We can glue outgoing edges of Γ to incoming edges of Γ' (or vice versa) to get a new marked MOY graph. First suppose only one pair of endpoints, one from each graph, are being glued together. Suppose that in both Γ and Γ' the endpoint being glued is labeled by the variable z (that is, $z \in E^i(\Gamma) \cap E^o(\Gamma')$ or $z \in E^i(\Gamma') \cap E^o(\Gamma)$). Then we define the new graph $\Gamma \cup_z \Gamma'$ by identifying the endpoints labeled by z and associate to $\Gamma \cup_z \Gamma'$ the complex

$$C(\Gamma \cup_z \Gamma') := C(\Gamma) \otimes_{\mathbb{Q}[z]} C(\Gamma'). \tag{3-3}$$

The edge ring of $\Gamma \cup_z \Gamma'$ is $E(\Gamma \cup_z \Gamma') = (E(\Gamma) \otimes_{\mathbb{Q}[z]} E(\Gamma'))/(z)$ and the total ring is $E^t(\Gamma \cup_z \Gamma') = E^t(\Gamma) \otimes_{\mathbb{Q}[z]} E^t(\Gamma')$. Note that after gluing, z is no longer in the edge ring as it is an internal variable. We may glue multiple edges at once in a similar manner. If $\boldsymbol{z} = z_1, \dots, z_n$ are the variables at the marked endpoints being identified, then we define $C(\Gamma \cup_{\boldsymbol{z}} \Gamma') := C(\Gamma) \otimes_{\mathbb{Q}[\boldsymbol{z}]} C(\Gamma')$. Similar to the case where we only identified one pair of edges, the edge ring of $\Gamma \cup_{\boldsymbol{z}} \Gamma'$ is given by $E(\Gamma \cup_{\boldsymbol{z}} \Gamma') = (E(\Gamma) \otimes_{\mathbb{Q}[\boldsymbol{z}]} E(\Gamma'))/(z_1, \dots, z_n)$ and the total ring is $E^t(\Gamma \cup_{\boldsymbol{z}} \Gamma') = E^t(\Gamma) \otimes_{\mathbb{Q}[\boldsymbol{z}]} E^t(\Gamma')$.

We can also describe disjoint union and gluing of marked MOY graphs in terms of Koszul complexes. Suppose $C(\Gamma)$ and $C(\Gamma')$ are given by the Koszul complexes

$$C(\Gamma) = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix}_{E^t(\Gamma)} \quad \text{and} \quad C(\Gamma') = \begin{bmatrix} p'_1 \\ \vdots \\ p'_n \end{bmatrix}_{E^t(\Gamma')}.$$

We can present $C(\Gamma \sqcup \Gamma')$ and $C(\Gamma \cup_{\boldsymbol{z}} \Gamma')$ as the Koszul complexes

$$C(\Gamma \sqcup \Gamma') = \begin{bmatrix} p_1 \\ \vdots \\ p_m \\ p_{1'} \\ \vdots \\ p'_n \end{bmatrix}_{E^t(\Gamma \sqcup \Gamma')} \quad \text{and} \quad C(\Gamma \cup_{\boldsymbol{z}} \Gamma') = \begin{bmatrix} p_1 \\ \vdots \\ p_m \\ p_{1'} \\ \vdots \\ p'_n \end{bmatrix}_{E^t(\Gamma \cup_{\boldsymbol{z}} \Gamma')}. \tag{3-4}$$

Here the distinction comes from the difference in total and edge rings. $C(\Gamma \sqcup \Gamma')$ is a chain complex of free $E(\Gamma \sqcup \Gamma')$–modules with the chain modules as direct sums of shifted copies of $E^t(\Gamma \sqcup \Gamma')$. However $C(\Gamma \cup_{\boldsymbol{z}} \Gamma')$ is a chain complex of free $E(\Gamma \cup_{\boldsymbol{z}} \Gamma')$–modules with the chain modules as direct sums of shifted copies of $E^t(\Gamma \cup_{\boldsymbol{z}} \Gamma')$. We now give another useful technique for simplifying the complexes associated to marked MOY graphs, called *mark removal.*

Lemma 3.4 *Suppose that z is an internal variable of a marked MOY graph Γ and $C(\Gamma)$ is the Koszul complex of the sequence $\boldsymbol{p} = p_1, \dots, z - p_i, \dots, p_k$, where $p_1, \dots, p_k \in E(\Gamma)$. Let $\psi\colon E^t(\Gamma) \to E^t(\Gamma)/(z-p_i)$ be the quotient map identifying z with p_i. Then we have*

$$C(\Gamma) \simeq \psi\left(\begin{bmatrix} p_1 \\ \vdots \\ \widehat{z-p_i} \\ \vdots \\ p_k \end{bmatrix}\right) \simeq \begin{bmatrix} p_1 \\ \vdots \\ \widehat{z-p_i} \\ \vdots \\ p_k \end{bmatrix}_{E^t(\Gamma)/(z-p_i)}$$

as complexes of $E(\Gamma)$–modules, omitting the term $z-p_i$ from the sequence.

Various forms of this lemma are proven in other texts on HOMFLY-PT homology, such as the original work of Khovanov and Rozansky [12] or work of Rasmussen [16]. We refer the reader to Lemma 3.8 in [16] for this exact form, omitting the "backward" differentials of the matrix factorizations. Lemma 3.4 allows us to freely add or remove marks without changing the homotopy type of the complex (as a complex of $E(\Gamma)$–modules). This implies the following very useful statement.

Corollary 3.5 *Let Γ and Γ' be two marked MOY graphs whose underlying (unmarked) MOY graphs are the same (isomorphic as oriented graphs). Then $C(\Gamma) \simeq C(\Gamma')$ as complexes of modules over $E(\Gamma) = E(\Gamma')$.*

Example 3.6 Consider the marked MOY graph from Figure 10. The marks partition the MOY graphs into six elementary MOY graphs (three MOY vertices and three arcs) which are drawn in Figure 10.

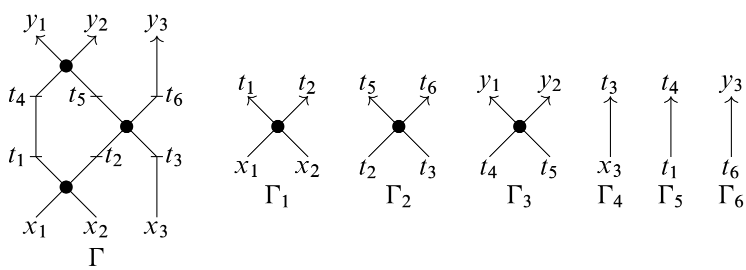


Figure 10: The marked MOY graph in Example 3.6 and its elementary MOY graphs

We can write Γ as $(\Gamma_1 \sqcup \Gamma_4) \cup_{t_1,t_2,t_3} (\Gamma_2 \sqcup \Gamma_5) \cup_{t_4,t_5,t_6} (\Gamma_3 \sqcup \Gamma_6)$, and therefore

$$C(\Gamma) = C(\Gamma_1 \sqcup \Gamma_4) \otimes_{\mathbb{Q}[t_1,t_2,t_3]} C(\Gamma_2 \sqcup \Gamma_5) \otimes_{\mathbb{Q}[t_4,t_5,t_6]} C(\Gamma_3 \sqcup \Gamma_6).$$

We can write $C(\Gamma)$, after some applications of mark removal to remove t_3, t_4, and t_6, as:

$$C(\Gamma) \simeq \begin{bmatrix} y_1 + y_2 - t_1 - t_5 \\ y_1 y_2 - t_1 t_5 \\ t_5 + y_3 - t_2 - x_3 \\ t_5 y_3 - t_2 x_3 \\ t_1 + t_2 - x_1 - x_2 \\ t_1 t_2 - x_1 x_2 \end{bmatrix}_{\mathbb{Q}[x_1,x_2,x_3,y_1,y_2,y_3,t_1,t_2,t_5]}$$

We invite the reader to finish the process of removing the internal variables t_1, t_2, and t_5 to get a *finite-rank* complex of $\mathbb{Q}[x_1, x_2, x_3, y_1, y_2, y_3]$–modules.

3.3 MOY braid graphs

A *MOY braid graph* is a graph formed by taking a braid and replacing every crossing with a MOY vertex, whose incoming and outgoing edges are consistent with the orientation of the braid. The complexes associated to MOY braid graphs and their "braid closures" satisfy the following local relations (as proven in [12; 16]):

Proposition 3.7 *Let Γ_0, Γ_{1a}, Γ_{1b}, Γ_{2a}, Γ_{2b}, Γ_{3a}, Γ_{3b}, Γ_{3c}, and Γ_{3d} be MOY graphs as in Figure 11. Then*

$$C(\Gamma_0) \simeq \bigoplus_{i=0}^{\infty} q^{2i}(\mathbb{Q} \oplus aq^2\mathbb{Q}), \tag{3-5}$$

$$C(\Gamma_{1a}) \simeq \bigoplus_{i=0}^{\infty} q^{2i}(C(\Gamma_{1b}) \oplus aq^4 C(\Gamma_{1b})), \tag{3-6}$$

$$C(\Gamma_{2a}) \simeq C(\Gamma_{2b}) \oplus q^2 C(\Gamma_{2b}), \tag{3-7}$$

$$C(\Gamma_{3a}) \oplus q^2 C(\Gamma_{3b}) \simeq q^2 C(\Gamma_{3c}) \oplus C(\Gamma_{3d}), \tag{3-8}$$

where $\simeq$ denotes homotopy equivalence over the corresponding edge rings.

To compare the isomorphisms in Proposition 3.7 to the relations in Figure 6 we introduce the notation of a "Laurent series shift functor". Suppose $F(q,a) \in \mathbb{N}[\![q^{\pm 1}, a^{\pm 1}]\!]$, that is

$$F(q,a) = \sum_{i,j \in \mathbb{Z}} c_{ij} q^i a^j, \quad c_{ij} \in \mathbb{N} \cup \{0\}.$$

Suppose M is a $\mathbb{Z}\times\mathbb{Z}$–graded R–module with grading shifts denoted by $q^i a^j$. Then

$$F(q,a)M := \bigoplus_{i,j \in \mathbb{Z}} q^i a^j M^{\oplus c_{ij}}. \tag{3-9}$$

We can write similar expressions for chain complexes C of $\mathbb{Z}\times\mathbb{Z}$–graded R–modules.

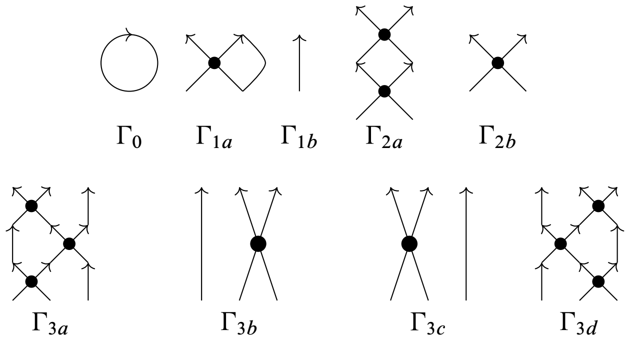


Figure 11: MOY graphs for Proposition 3.7

Let

$$F(q,a,t) = \sum_{i,j,k\in\mathbb{Z}} c_{ijk} q^i a^j t^k \in \mathbb{N}[\![q^{\pm 1}, a^{\pm 1}, t^{\pm 1}]\!],$$

and let the homological grading shift on C be denoted by t^k. Then

$$F(q,a,t)C := \bigoplus_{i,j,k\in\mathbb{Z}} q^i a^j t^k C^{\oplus c_{ijk}}. \tag{3-10}$$

The coefficients in MOY 0 and MOY I are not actually Laurent series, but rather rational functions. However, considering the rational function with Laurent polynomial numerator $F(q,a)$ as a geometric series, we can write the rational functions as a Laurent series

$$\frac{F(q,a)}{1-q^2} = F(q,a)\sum_{i=0}^{\infty} q^{2i}.$$

With this notation in mind, we can rewrite the isomorphism (3-5) as

$$\begin{aligned} C(\Gamma_0) &\simeq \bigoplus_{i=0}^{\infty} q^{2i}(\mathbb{Q}\oplus aq^2\mathbb{Q}) = \bigoplus_{i=0}^{\infty} q^{2i}(1+aq^2)\mathbb{Q} \\ &= (1+aq^2)\sum_{i=0}^{\infty} q^{2i}\mathbb{Q} = \frac{1+aq^2}{1-q^2}\mathbb{Q} \end{aligned} \tag{3-11}$$

and the isomorphism (3-6) as

$$\begin{aligned} C(\Gamma_{1a}) &\simeq \bigoplus_{i=0}^{\infty} q^{2i}(C(\Gamma_{1b})\oplus aq^4 C(\Gamma_{1b})) \\ &= (1+aq^4)\sum_{i=0}^{\infty} q^{2i} C(\Gamma_{1b}) = \frac{1+aq^4}{1-q^2}C(\Gamma_{1b}). \end{aligned} \tag{3-12}$$

We invite the reader to compare the rewritten relation (3-11) to (MOY 0) from Figure 6 and (3-12) to (MOY I). This comparison can be made for (3-7) to (MOY 2a) and (3-8) to (MOY 3) as well.

3.4 Khovanov–Rozansky HOMFLY-PT homology

We now have recalled the necessary tools to define Khovanov–Rozansky HOMFLY-PT homology, or briefly HOMFLY-PT homology. We first define two q–degree 0 maps $\chi_i\colon \uparrow\uparrow \to q^{-2}\overset{\bullet}{\times}$ and $\chi_o\colon \overset{\bullet}{\times} \to \uparrow\uparrow$. Set $E = \mathbb{Q}[x_1, x_2, y_1, y_2]$ to be the edge ring of both $\uparrow\uparrow$ and $\overset{\bullet}{\times}$. Then we define χ_i by

$$
(3\text{-}13)\qquad
\begin{array}{ccccccc}
\uparrow\uparrow = & a^2q^4E & \xrightarrow{\begin{pmatrix} y_1-x_1 \\ y_2-x_2 \end{pmatrix}} & aq^2E \oplus aq^2E & \xrightarrow{(x_2-y_2\ \ y_1-x_1)} & E \\
\chi_i \downarrow & \downarrow 1 & & \downarrow \begin{pmatrix} y_1-x_2 & 0 \\ 1 & 1 \end{pmatrix} & & \downarrow y_1-x_2 \\
q^{-2}\overset{\bullet}{\times} = & a^2q^4E & \xrightarrow[\begin{pmatrix} P_2 \\ P_1 \end{pmatrix}]{} & aE \oplus aq^2E & \xrightarrow[(-P_1\ \ P_2)]{} & q^{-2}E
\end{array}
$$

and χ_o by

$$
(3\text{-}14)\qquad
\begin{array}{ccccccc}
\overset{\bullet}{\times} = & a^2q^6E & \xrightarrow{\begin{pmatrix} P_2 \\ P_1 \end{pmatrix}} & aq^2E \oplus aq^4E & \xrightarrow{(-P_1\ \ P_2)} & E \\
\chi_o \downarrow & \downarrow y_1-x_2 & & \downarrow \begin{pmatrix} 1 & 0 \\ -1 & y_1-x_2 \end{pmatrix} & & \downarrow 1 \\
a\uparrow\uparrow = & a^2q^4E & \xrightarrow[\begin{pmatrix} y_1-x_1 \\ y_2-x_2 \end{pmatrix}]{} & aq^2E \oplus aq^2E & \xrightarrow[(x_2-y_2\ \ y_1-x_1)]{} & E
\end{array}
$$

Above, we set $P_1 = y_1 + y_2 - x_1 - x_2$ and $P_2 = (y_1 - x_1)(y_1 - x_2)$ for the sake of legibility. We now define two bicomplexes of free E–modules for the positive crossing $\times^{+}$ and the negative crossing $\times^{-}$:

$$
(3\text{-}15)\qquad C(\times^{+}) := C(\uparrow\uparrow) \xrightarrow{\chi_i} tq^{-2}C(\overset{\bullet}{\times}),
$$

$$
(3\text{-}16)\qquad C(\times^{-}) := t^{-1}C(\overset{\bullet}{\times}) \xrightarrow{\chi_o} C(\uparrow\uparrow).
$$

Note that we use the notation $t^kC(\Gamma)$ to mean that the complex for Γ sits in homological degree k. This is a different homological degree than our Hochschild degree we introduced earlier. We will simply denote this degree by $\deg_t$ and call it the *homological degree.* We will say that x has (total) degree $\deg(x) = q^i a^j t^k$ if it has quantum degree i, Hochschild degree k, and homological degree j. We will denote the differential in the complexes for the MOY graphs as d_g and the differentials in the complexes (built from χ_i and χ_o) associated to crossings as d_c. Both $C($  $)$ and $C($  $)$ are bicomplexes with commuting differentials d_g and d_c.

A *marked tangle diagram* is a tangle diagram with markings such that the marks partition the tangle diagram into arcs, positive crossings, and negative crossings. We label the marks and the endpoints (if any) by variables in a similar fashion to marked MOY graphs. We define rings associated to each marked tangle diagram τ in a similar manner to our constructions for marked MOY graphs.

Before defining a bicomplex for a tangle diagram, we recall a definition from homological algebra.

Definition 3.8 Let $C = (C_{\bullet\bullet}, d_h, d_v)$ and $C' = (C'_{\bullet\bullet}, d'_h, d'_v)$ be two bicomplexes. We define the *tensor product bicomplex* $C \otimes C' = ((C \otimes C')_{\bullet\bullet}, d_h^{\otimes}, d_v^{\otimes})$ as follows:

$$(C \otimes C')_{mn} = \bigoplus_{i+k=m,\, j+\ell=n} (C_{ij} \otimes C'_{k\ell}),$$

$$d_h^{\otimes}(x \otimes y) = d_h(x) \otimes y + (-1)^i x \otimes d'_h(y) \quad \text{for } x \in C_{ij} \text{ and } y \in C'_{k\ell},$$

$$d_v^{\otimes}(x \otimes y) = d_v(x) \otimes y + (-1)^k x \otimes d'_v(y) \quad \text{for } x \in C_{ij} \text{ and } y \in C'_{k\ell}.$$

We can now build a bicomplex for any tangle diagram (and link diagram) in a similar manner to what we did in Section 3.3 for MOY graphs. To a disjoint union of (marked) tangles $\tau = \tau_1 \sqcup \tau_2$ we associate the bicomplex of $E(\tau)$–modules

$$C(\tau) := C(\tau_1) \otimes_{\mathbb{Q}} C(\tau_2).$$

Similarly if we are gluing two tangles τ_1 and τ_2 at the marked points $\boldsymbol{z} = z_1, \dots, z_k$ in such a way that the orientations are consistent, then we define a bicomplex of $E(\tau_1 \cup_{\boldsymbol{z}} \tau_2)$–modules

$$C(\tau_1 \cup_{\boldsymbol{z}} \tau_2) = C(\tau_1) \otimes_{\mathbb{Q}[\boldsymbol{z}]} C(\tau_2).$$

We omit the rest of the details in this case, and leave it to the reader to compare with the analogous conventions for marked MOY graphs. Now let $\beta \in \mathrm{Br}_n$ be a braid with n strands. We can mark β in such a way that we partition it into arcs and crossings of the form  or  and we label the endpoints and markings in a similar manner to our conventions for marked MOY graphs. Therefore we can use the rules of disjoint unions and gluing of tangles to write a bicomplex $C(\beta)$ of $E(\beta)$–modules. In this case, $E(\beta)=\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]$ where $|\boldsymbol{x}|=|\boldsymbol{y}|=n$.

We now describe the construction of the HOMFLY-PT homology of a link L. Suppose that $\beta\in \mathrm{Br}_n$ is a braid representative of L, that is L is the circular closure of β in $\mathbb{R}^3$. We will often use the notation L_β for the link diagram of the closure of β. Then we can describe the bicomplex $C(L_\beta)=C(\beta)\otimes_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]}C(1_n)$, where 1_n denotes the identity braid (oriented downwards) with the top endpoints labeled by $\boldsymbol{y}$ and the bottom endpoints labeled by $\boldsymbol{x}$. We refer the reader to Figure 12 for an example of this decomposition of a braid closure.

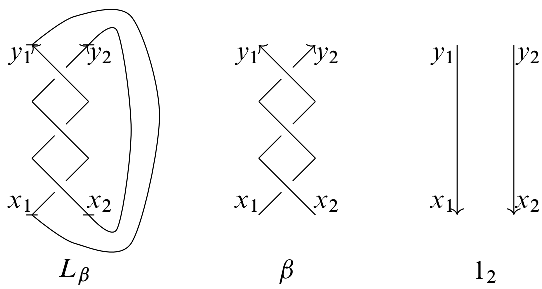


Figure 12: A link presented as a braid closure and the constituent tangles

Definition 3.9 Suppose L is a link with braid representative $\beta\in\mathrm{Br}_n$. The HOMFLY-PT homology of L_β is $\mathcal{H}(L_\beta)=H_{d_{c*}}(H_{d_g}(C(L_\beta)))$.

Remark 3.10 $\mathcal{H}(L_\beta)$, as defined above, arises as the E^2–page of a spectral sequence. Let L be an n–component link. It is easily shown that the E^∞–page of that spectral sequence is the homology of the n–component unlink (up to a grading shift). In particular, $H_{d_c}(C(L))$ is isomorphic to the E^∞–page.

Theorem 3.11 (Khovanov and Rozansky [12]) *Suppose $\beta\in\mathrm{Br}_n$ and $\beta'\in\mathrm{Br}_{n'}$ are two braid representatives of a link L. Then $\mathcal{H}(L_\beta)\cong\mathcal{H}(L_{\beta'})$ up to a grading shift. Furthermore, suppose the Poincaré series* (see (1-1)) *of $\mathcal{H}(L_\beta)$ is given by*

$$\mathcal{P}(L_\beta)=\sum_{i,j,k\in\mathbb{Z}} d_{i,j,k}q^ia^jt^k,\quad \text{where } d_{i,j,k}=\dim_{\mathbb{Q}}(\mathcal{H}(L_\beta))_{i,j,k}.$$

Then

$$\mathcal{P}(L_\beta)|_{t=-1}=\sum_{i,j,k\in\mathbb{Z}} d_{i,j,k}q^ia^j(-1)^k=P(L_\beta).$$

4 HOMFLY-PT homology for general link diagrams

In this section we study what happens when we consider general link diagrams in the construction of HOMFLY-PT homology. We will see that not all Reidemeister moves are respected, and that in general HOMFLY-PT homology is only an invariant up to braidlike isotopy.

4.1 Virtual crossings and marked MOY graphs

We start by introducing virtual crossings into the framework of (marked) MOY graphs. We will not fully discuss virtual knot theory here, but rather refer the reader to Kauffman [9]. Virtual crossings were first considered as a tool in HOMFLY-PT and $\mathfrak{sl}_n$ homologies by Khovanov and Rozansky in [11], and studied further by the author and Rozansky in [1].

A *virtual MOY graph* is a MOY graph where we allow the underlying graph to be nonplanar. Such a graph can always be drawn where the intersections forced by the projection onto the plane are transverse double points. An example of this is given in Figure 13. To the marked virtual crossing graph we associate the following complex of free $E(\otimes) = \mathbb{Q}[x_1, x_2, y_1, y_2]$–modules:

$$C(\otimes) = \begin{bmatrix} y_1 - x_2 \\ y_2 - x_1 \end{bmatrix}_{E(\otimes)} = q^4 E(\otimes) \xrightarrow{A} q^2 E(\otimes) \oplus q^2 E(\otimes) \xrightarrow{B} E(\otimes), \tag{4-1}$$

where

$$A = \begin{pmatrix} y_1 - x_2 \\ y_2 - x_1 \end{pmatrix}, \quad B = \begin{pmatrix} x_1 - y_2 & y_1 - x_2 \end{pmatrix}.$$

Note that $C(\otimes)$ resembles $C(\uparrow\ \uparrow)$ except for a transposition of x_1 and x_2 in the definition of the complexes. In this sense, we can think of a virtual crossing as being a permutation of strands with no additional crossing data or vertex at the intersection.

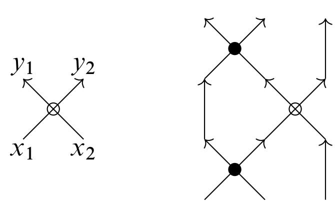


Figure 13: A (marked) virtual crossing and an example of a virtual MOY graph

Proposition 4.1 *The moves in Figure 14 preserve the homotopy equivalence type of* $C(\Gamma)$.

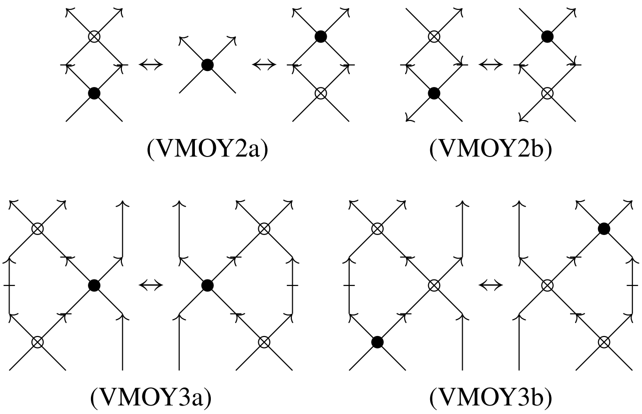


Figure 14: Virtual MOY moves (marks included for reference for Proposition 4.1)

Proof We begin with (VMOY2a). Let t_1 and t_2 be the variables associated to the marks (from left to right) in the diagram on the left for (VMOY2a). The left-hand side of (VMOY2a) is presented as the Koszul complex of the sequence

$$(y_2 - t_1, y_1 - t_2, t_1 + t_2 - x_1 - x_2, (t_1 - x_1)(t_1 - x_2)).$$

Let $w = y_1 + y_2 - x_1 - x_2$. Then

$$\begin{aligned} \begin{bmatrix} y_2 - t_1 \\ y_1 - t_2 \\ t_1 + t_2 - x_1 - x_2 \\ (t_1 - x_1)(t_1 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} &\simeq \begin{bmatrix} y_2 - t_1 \\ y_1 - t_2 \\ w \\ (t_1 - x_1)(t_1 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \\ &\simeq \begin{bmatrix} y_1 - t_2 \\ w \\ (y_1 - x_1)(y_1 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},t_2]} \\ &\simeq \begin{bmatrix} w \\ (y_1 - x_1)(y_1 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]}. \end{aligned}$$

The first isomorphism is a change of basis and the other two isomorphisms are mark removals. The last term is the Koszul complex for $C(\;)$. The second isomorphism in (VMOY2a) is proven by a similar argument. Next we prove (VMOY2b). For consistency, we label the bottom endpoints x_1 and x_2, the top endpoints y_1 and y_2 and the marks by t_1 and t_2 (reading from left to right). The left-hand side of (VMOY2b) can be written as

$$\begin{bmatrix} y_2-t_1 \\ t_2-y_1 \\ t_1+x_1-t_2-y_2 \\ (t_1-x_2)(t_1-t_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \simeq \begin{bmatrix} y_2+x_1-y_1-y_2 \\ (y_2-x_2)(y_2-t_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]},$$

where the isomorphism is given by removing the internal variables t_1 and t_2. Likewise, the right-hand side of (VMOY2b) can be written as

$$\begin{bmatrix} y_2+t_2-y_1-t_1 \\ (y_2-y_1)(y_2-t_1) \\ t_1-x_2 \\ x_1-t_2 \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \simeq \begin{bmatrix} y_2+x_1-y_1-y_2 \\ (y_2-x_2)(y_2-t_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]},$$

where again the isomorphism is given by removing the internal variables t_1 and t_2. This proves (VMOY2b).

Now we approach (VMOY3a). For both diagrams in (VMOY3a), label the top variables as y_1, y_2, y_3 and bottom variables as x_1, x_2, x_3 from left to right. Also label the marks as t_1, t_2, t_3 from bottom to top. The associated Koszul complex to the left-hand side is

$$\begin{bmatrix} y_1-t_3 \\ y_2-t_2 \\ t_2-x_2 \\ t_1-x_1 \\ y_3+t_3-t_1-x_3 \\ (t_3-t_1)(t_3-x_3) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \simeq \begin{bmatrix} y_2-x_2 \\ y_1+y_3-x_1-x_3 \\ (y_1-x_1)(y_3-x_3) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]},$$

where the isomorphism is given by removing the internal variables t_1, t_2, t_3. Likewise, the associated Koszul complex to the right hand-side is given by

$$\begin{bmatrix} y_3-t_3 \\ y_2-t_2 \\ t_1-x_3 \\ t_2-x_2 \\ 2_1+t_3-t_1-x_1 \\ (y_1-x_1)(y_1-t_1) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \simeq \begin{bmatrix} y_2-x_2 \\ y_1+y_3-x_1-x_3 \\ (y_1-x_1)(y_3-x_3) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]},$$

where the isomorphism is given by removing the internal variables t_1, t_2, t_3. This proves (VMOY3a).

The approach to proving (VMOY3b) is almost identical, but we include it for completeness. For both diagrams in (VMOY3b), label the top variables as y_1, y_2, y_3 and bottom variables as x_1, x_2, x_3 from left to right. Also label the marks as t_1, t_2, t_3 from bottom to top. The associated Koszul complex to the left-hand side is

$$\begin{bmatrix} y_1 - t_3 \\ y_2 - t_2 \\ y_3 - t_1 \\ t_3 - x_3 \\ t_2 + t_1 - x_1 - x_2 \\ (t_2 - x_1)(t_2 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \simeq \begin{bmatrix} y_1 - x_3 \\ y_2 + y_3 - x_1 - x_2 \\ (y_2 - x_1)(y_2 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]},$$

where the isomorphism is given by removing the internal variables t_1, t_2, t_3. Likewise, the associated Koszul complex to the right hand-side is given by

$$\begin{bmatrix} y_1 - t_1 \\ y_2 + y_3 - t_3 - t_2 \\ (y_2 - t_3)(y_2 - t_2) \\ t_3 - x_1 \\ t_2 - x_2 \\ t_1 - x_3 \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}]} \simeq \begin{bmatrix} y_1 - x_3 \\ y_2 + y_3 - x_1 - x_2 \\ (y_2 - x_1)(y_2 - x_2) \end{bmatrix}_{\mathbb{Q}[\boldsymbol{x},\boldsymbol{y}]},$$

where the isomorphism is given by removing the internal variables t_1, t_2, t_3. Therefore, (VMOY3b) is proven. □

For any *virtual link diagram* D, that is a link diagram with virtual crossings, we can repeat the procedure from Section 3.4 to build a bicomplex of $E(D)$–modules. We now record the additional "virtual" Reidemeister moves.

Proposition 4.2 *The moves in Figure 15 preserve the homotopy equivalence type of $C(D)$. The isomorphisms (VR1), (VR2a), (VR2b), (VR3), and (SVR) are called* virtual Reidemeister moves, *and the isomorphisms (Z1±) and (Z2±) are called* Z-moves.

The proofs of (VR1), (VR2a), (VR2b), and (VR3) follow the same outline (write Koszul complexes for both sides, and compare after mark removal) as the proof of Proposition 4.1. (SVR) and the Z-moves follow from resolving the single crossing and applying the moves (VR2a), (VR2b), (VMOY2a), (VMOY2b), and (VMOY3a). Note that our Koszul complexes for many of our diagrams are free resolutions of certain bimodules. On the level of these bimodules these moves are proven in other sources. (VR1) is proven in Lemma 6.5 of [1], (VR2a), (VR3), (SVR), (Z1+), and (Z1−) are proven in [17, Lemma 3.1, Lemma 3.2 and Theorem 3.4] and [1, Theorem 2.2]. Strictly speaking, a different move from the Z-moves, called "virtualization moves" are proven in these texts. However, the Z-moves follow via tensoring with a virtual crossing

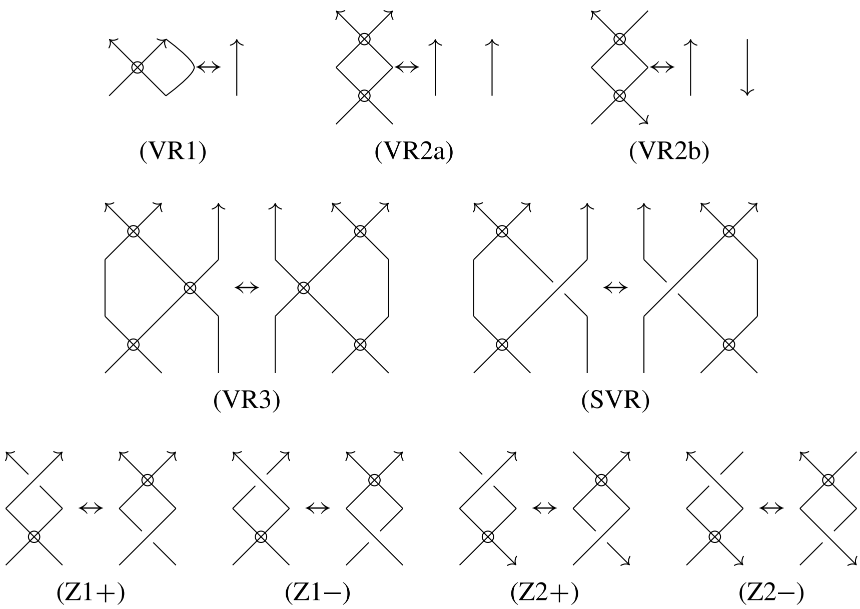


Figure 15: Virtual Reidemeister moves and Z-moves

and applying (VR2a). A little more needs to be said about the moves (Z2±). After resolving the crossing we apply (VMOY2b) and note, up to a relabeling of variables, the differential does not change.

4.2 Virtual filtrations of signed MOY graphs and a key lemma

Virtual crossing filtrations were first introduced by the author and Rozansky in [1]. We now introduce these filtrations in our current setting as an eventual tool in proving the failure of certain Reidemeister moves. In this text, we use virtual filtrations to prove a key lemma in the process of finding an explicit example of failure of Reidemeister IIb.

Definition 4.3 Let R be a commutative ring and let C and D be chain complexes of objects in an additive category. Let $\bullet[i]$ denote the homological shift functor given by $C_j[i] = C_{j-i}$. Define $\mathrm{Hom}^k(C, D)$ to be the $\mathbb{Z}$–module of chain maps $f\colon C \to D[-k]$ quotiented by the submodule of chain maps homotopic to the zero map.

In [11], Khovanov and Rozansky make the following observation.

Proposition 4.4 *There exists a unique map* $F \in \mathrm{Hom}^1(C(\uparrow\,\uparrow), q^2 C(\otimes))$*, up to rescaling, such that* $\mathrm{Cone}(F)$ *is homotopy equivalent to* $C(\bullet)$*. Likewise there exists a unique map, up to rescaling,* $G \in \mathrm{Hom}^1(C(\otimes), q^2 C(\uparrow\,\uparrow))$ *such that* $\mathrm{Cone}(G)$ *is homotopy equivalent to* $C(\bullet)$*.*

We will call the maps F and G *virtual saddle maps.* In our presentation of $C(\uparrow\,\uparrow)$ and $C(\otimes)$ we can write the virtual saddle maps explicitly. We give the following explicit presentation of the virtual saddle map G and leave it to the reader to do the same for the analogous map F:

$$\begin{array}{ccccccc} C(\otimes) & = & a^2q^4E & \xrightarrow{\begin{pmatrix}x_1-y_2\\x_2-y_1\end{pmatrix}} & aq^2E\oplus aq^2E & \xrightarrow{(x_2-y_1\ \ y_2-x_1)} & E \\ \Big\downarrow G & & & \searrow^{\begin{pmatrix}1\\-1\end{pmatrix}} & & \searrow^{(1\ 1)} & \\ q^2C(\uparrow\,\uparrow) & = & a^2q^6E & \xrightarrow[\begin{pmatrix}x_1-y_1\\x_2-y_2\end{pmatrix}]{} & aq^4E\oplus aq^4E & \xrightarrow[(x_2-y_2\ \ y_1-x_1)]{} & q^2E \end{array}$$

The mapping cone presentations give rise to filtrations. In particular, $\mathrm{Cone}(F)$ has $q^2C(\otimes)$ as a submodule and $C(\uparrow\,\uparrow)$ as the quotient $\mathrm{Cone}(F)/q^2C(\otimes)$. We call this filtration the *negative filtration* and denote it as $C_-(\bullet)$. Likewise $\mathrm{Cone}(G)$ has $q^2C(\uparrow\,\uparrow)$ as a subcomplex and $C(\otimes)$ as the quotient complex $\mathrm{Cone}(F)/q^2C(\uparrow\,\uparrow)$. We call this filtration the *positive filtration* and denote it as $C_+(\bullet)$.

We will often identify $C_-(\bullet)$ with $\mathrm{Cone}(F)$ (and $C_+(\bullet)$ with $\mathrm{Cone}(G)$) and consider the filtered complexes as mapping cones. This process simplifies the differential d_c so that it can be presented in the following manner (as proven in [1]).

Proposition 4.5 *The bicomplex $C(\times)$ is homotopy equivalent to the bicomplex $C(\uparrow\,\uparrow)\xrightarrow{\phi_i} tq^{-2}C_+(\bullet)$, where ϕ_i denotes the canonical inclusion of $C(\uparrow\,\uparrow)$ into* $\mathrm{Cone}(G)$*. Suppose $C(\uparrow\,\uparrow)$ has the trivial filtration. Then ϕ_i is a filtered map with respect to the filtration on $C_+(\bullet)$ and thus $C(\times)$ is a filtered bicomplex.*

In addition, $C(\times)$ is homotopy equivalent to the bicomplex $t^{-1}C(\bullet)\xrightarrow{\phi_o} C(\uparrow\,\uparrow)$, where ϕ_0 is the canonical projection of $C(\uparrow\,\uparrow)$ from $\mathrm{Cone}(F)$*. The projection ϕ_o is a filtered map with respect to the filtration on $C_-(\bullet)$ and thus $C(\times)$ is a filtered bicomplex.*

We can extend this filtration to any tangle or link diagram via the tensor product filtration. We will also refer to the given filtration on the bicomplex associated to a tangle as a *virtual filtration.* The following theorem was the main focus of [1].

Theorem 4.6 *Let β be a braid on n strands, and L_β denote its circular closure. Then the virtual filtration on $C(\beta)$ is invariant under Reidemeister IIa and is violated by at most two levels by Reidemeister III. Furthermore, the virtual filtration on $\mathcal{H}(L_\beta)$ is invariant under the Markov moves (up to a possible shift in filtration).*

We now focus on describing the filtrations on MOY graphs, which will be useful in proving Lemma 4.10. A *signed MOY graph* is a MOY graph where each vertex is marked by a sign $+$ or $-$. Suppose Γ is a signed MOY graph marked so that it is partitioned into graphs of the form ⟨MOY vertex⟩ and ⟨strand⟩. Then we can define a filtration on $C(\Gamma)$. To each MOY vertex marked with a $+$ we associate $C_+(\text{MOY vertex})$, and to each MOY vertex marked with a $-$ we associate $C_-(\text{MOY vertex})$. We give the trivial filtration to $C(\text{strand})$. Then the filtration on $C(\Gamma)$ is given by the tensor product filtration.

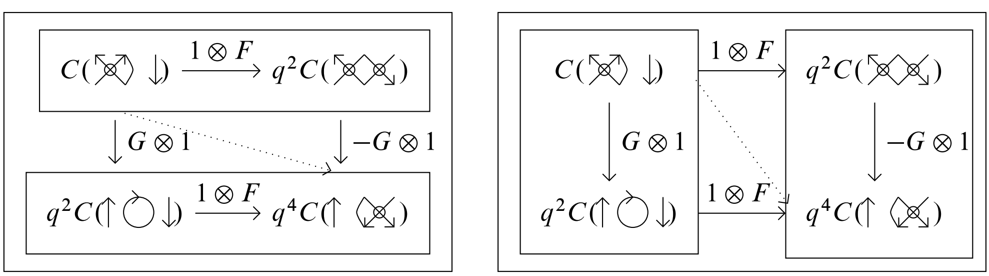
If we choose different sign assignments, then we receive homotopy equivalent complexes for $C(\Gamma)$, but not necessarily *filtered* homotopy equivalent complexes (eg $C_-(\text{MOY vertex})$ and $C_+(\text{MOY vertex})$ are not filtered homotopy equivalent). For this reason, if $\underline{\varepsilon}$ is a assignment of signs to each MOY vertex of Γ, then we will write $C_{\underline{\varepsilon}}(\Gamma)$ for the filtered complex we get from the above construction.

With this construction in mind, we may present every MOY graph as an *iterated mapping cone* of graphs with only virtual crossings and no MOY vertices. We will commonly use the alternate notation for mapping cones shown in Figure 16.

$$\boxed{A \xrightarrow{\ f\ } B} := \mathrm{Cone}(f\colon A \to B)$$

Figure 16: Alternate notation for mapping cones

Example 4.7 We now consider the signed MOY graph $\Gamma =$ ⟨graph⟩ whose left vertex is labeled by $+$ and right vertex is labeled by $-$. We can present $C_{+-}(\Gamma)$ as one of the two equal iterated mapping cones:



Equality of the above iterated mapping cones follows from the associativity of the mapping cone operation. Note that the dotted arrow is the zero map, but is drawn in the diagram above as a reminder that such a map may be needed after simplifying the above complex.

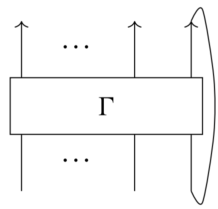


Figure 17: Partial braid closure of a virtual braidlike MOY graph

Next we state a few simplifications and tools which will be necessary in proving our key lemma. The following lemma is proven in [1].

Lemma 4.8 *Let* $\widehat{\boxtimes}$ *and* $\widehat{\uparrow\uparrow}$ *denote the partial braid closure of* $\boxtimes$ *and* $\uparrow\uparrow$ *respectively* (see Figure 17). *We write* $\widehat{G}\colon C(\widehat{\boxtimes}) \to q^2 C(\widehat{\uparrow\uparrow})$ *for the map induced by* G *under partial braid closure, and similarly for* $\widehat{F}$. *Then*

(1) $\operatorname{Cone}(\widehat{G}) \simeq \operatorname{Cone}(0) \simeq C(\uparrow) \oplus q^2 C(\uparrow \bigcirc)$,

(2) $\operatorname{Cone}(\widehat{F}) \simeq \operatorname{Cone}\left(aq^2 C(\uparrow) \oplus \dfrac{1+aq^4}{1-q^2} C(\uparrow) \xrightarrow{(1\ 0)} q^2 C(\uparrow) \right)$,

(3) $\operatorname{Cone}(\widehat{F}) \simeq \operatorname{Cone}(\widehat{G}) \simeq \dfrac{1+aq^4}{1-q^2} C(\uparrow)$.

Furthermore, the homotopy equivalences in (1) *and* (2) *are filtered.*

A proof of the following result can be found in other texts on link homology such as [5].

Proposition 4.9 *Consider the complex*

$$A \xrightarrow{\binom{\bullet}{\alpha}} B \oplus C \xrightarrow{\left(\begin{smallmatrix}\varphi & \lambda \\ \mu & \nu\end{smallmatrix}\right)} D \oplus E \xrightarrow{\binom{\bullet}{\varepsilon}} F,$$

where $\varphi\colon B \to D$ *is an isomorphism and all other maps are arbitrary up to the condition that* $d^2 = 0$. *Then there exists a homotopy equivalence:*

$$\begin{array}{ccccccc}
A & \xrightarrow{\binom{\bullet}{\alpha}} & B \oplus C & \xrightarrow{\left(\begin{smallmatrix}\varphi & \lambda \\ \mu & \nu\end{smallmatrix}\right)} & D \oplus E & \xrightarrow{(\bullet\ \varepsilon)} & F \\
{\scriptstyle (1)}\downarrow\ \uparrow{\scriptstyle (1)} & & {\scriptstyle (0\ 1)}\downarrow\ \uparrow{\scriptstyle \left(\begin{smallmatrix}-\varphi^{-1}\lambda \\ 1\end{smallmatrix}\right)} & & {\scriptstyle (-\mu\varphi^{-1}\ 1)}\downarrow\ \uparrow{\scriptstyle \left(\begin{smallmatrix}0 \\ 1\end{smallmatrix}\right)} & & {\scriptstyle (1)}\downarrow\ \uparrow{\scriptstyle (1)} \\
A & \xrightarrow{(\alpha)} & C & \xrightarrow{(\nu-\mu\varphi^{-1}\lambda)} & E & \xrightarrow{(\varepsilon)} & F
\end{array}$$

We call this homotopy equivalence Gaussian elimination.

We now look at the analogue of (MOYIIb) from Figure 6. This will be our key lemma in simplifying the Reidemeister IIb complex.

Lemma 4.10 *Let*

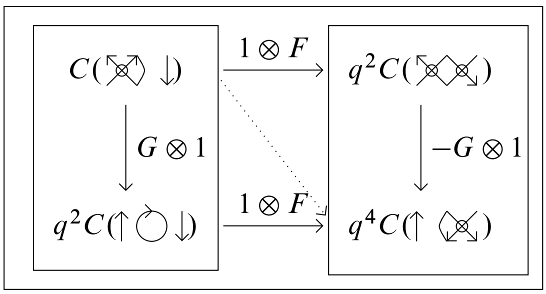
$$\widetilde{C}(\uparrow\ \downarrow) = \frac{q^2 + aq^6}{1-q^2} C(\uparrow\ \downarrow).$$

Then

$$C(\ ) \simeq \widetilde{C}(\uparrow\ \downarrow) \oplus C(\ ).$$

Proof Consider $C(\times_\bullet \times_\bullet)$ with the filtration described in Example 4.7. We then consider $C(\times_\bullet \times_\bullet)$ as the iterated mapping cone:

(4-2)

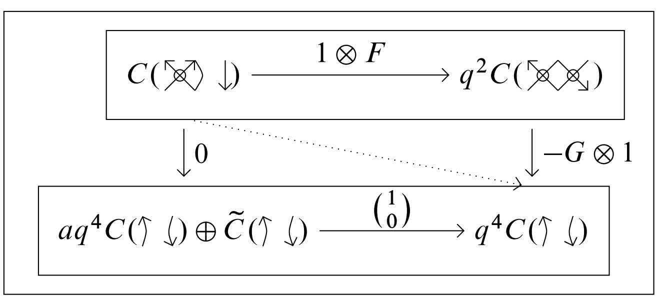


The maps

$$G \otimes 1\colon C(\ \downarrow) \to q^2 C(\uparrow \circlearrowright \downarrow) \quad \text{and} \quad 1 \otimes F\colon q^2(C(\uparrow \circlearrowright \downarrow)) \to q^4 C(\uparrow\ \times_\otimes)$$

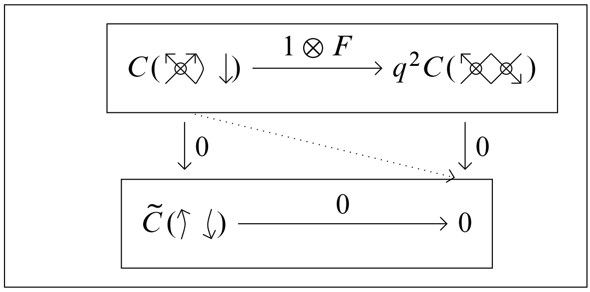
can be viewed as maps between partial braid closures $\widehat{G} \otimes 1_{\downarrow}$ and $1_{\uparrow} \otimes \widehat{F}$ respectively. We now apply isomorphisms (1) and (2) of Lemma 4.8 and associativity of the mapping cone operation to simplify the complex (4-2) to:

(4-3)



Note the dotted arrow may no longer be the zero map, but knowledge of the exact map will ultimately not be necessary. Now we apply Gaussian elimination to the bottom mapping cone in (4-3) to get:

(4-4)



The top mapping cone is isomorphic to $C(\;)$ by Proposition 4.5. Therefore $C(\;) \simeq \widetilde{C}(\uparrow\ \downarrow) \oplus C(\;)$ as desired. □

4.3 Failure of Reidemeister IIb for $\mathcal{H}(D)$

Now with the introduction of virtual crossings, we can study the exact outcome of allowing general link diagrams in the computation of $\mathcal{H}(D)$.

We now state the main result of this section.

Theorem 4.11 *The isomorphism* $C(\;) \simeq tq^{-2}C(\;)$ *holds.*

Proof We first write $C(\;)$ using (3-15) and (3-16):

$$C(\;) = t^{-1}C(\uparrow\ ) \xrightarrow{A} C(\uparrow \bigcirc \downarrow) \oplus q^{-2}C(\;) \xrightarrow{B} tq^{-2}C(\ \downarrow),$$

where the tensor product above is over the appropriate ring of internal variables and

$$A = \begin{pmatrix} 1 \otimes \chi_0 \\ \chi_i \otimes 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \chi_i \otimes 1 & -1 \otimes \chi_o \end{pmatrix}.$$

We use the isomorphisms from Proposition 3.7 and Lemma 4.10 to rewrite the complex as

$$q^{-2}\widetilde{C}(\uparrow\ \downarrow) \xrightarrow{D} \underline{q^{-2}\widetilde{C}(\uparrow\ \downarrow) \oplus C(\uparrow \bigcirc \downarrow) \oplus q^{-2}C(\;)} \xrightarrow{F} C(\uparrow \bigcirc \downarrow) \oplus q^{-2}C(\uparrow\ \downarrow),$$

where the underlined term is in homological degree 0 (omitting homological shifts for compactness of notation) and

$$D = \begin{pmatrix} 1 \\ * \\ * \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} * & 1 & 0 \\ * & * & -1 \otimes \chi_o \end{pmatrix}.$$

Note the maps $*$ are irrelevant to our discussion. Note above, we use the map $-1 \otimes \chi_o$ to denote the map $-1 \otimes \chi_o : q^{-2}C(\;) \to C(\ \downarrow) \simeq C(\uparrow\ \downarrow)$ (the isomorphism following from (VR1)). After performing Gaussian elimination on the complex above, we are left with

$$\begin{aligned} C(\ \ ) &\simeq q^{-2}C(\ \ ) \xrightarrow{1\otimes\chi_o} tq^{-2}C(\uparrow\ \downarrow) \\ &\simeq tq^{-2}C(\ \ ) \otimes (t^{-1}C(\ \ ) \xrightarrow{\chi_o} C(\downarrow\ \downarrow)) \\ &\simeq tq^{-2}C(\ \ ). \end{aligned}$$

□

Example 4.12 We now give an example where the local failure of Reidemeister IIb gives a failure of isotopy invariance for a certain link diagram. Let D be the diagram for the unknot given in Figure 18, and let O denote the standard diagram of the unknot as a circle bounding a disc in $\mathbb{R}^2$. Then we have the following chain of isomorphisms

$$\mathcal{H}(D) \simeq t^2q^{-4}\mathcal{H}(D') \simeq t^2q^{-4}\mathcal{H}(D'') \simeq t^2q^{-4}\mathcal{H}(D''').$$

The first isomorphism is given by applying Theorem 4.11 twice. The second isomorphism following from applying (Z2−) from Proposition 4.1. The last isomorphism follows from applying (VR2b) from Proposition 4.1. D''' is the diagram of the left-handed trefoil knot, and we computed $P(D''')$ in Example 2.7. This is enough to show that $\mathcal{H}(D) \not\simeq \mathcal{H}(O)$; however, it is an easy exercise to show that

$$\mathcal{H}(D''') = (aq^2 + t^2q^2 + t^2aq^4)\frac{1+aq^2}{1-q^2}\mathbb{Q} \not\simeq \frac{1+aq^2}{1-q^2}\mathbb{Q} \simeq \mathcal{H}(O).$$

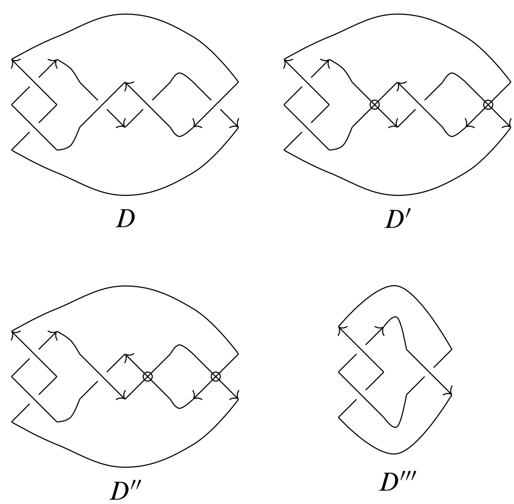


Figure 18: The failure of Reidemeister IIb for an unknot diagram. The above diagrams all have the same HOMFLY-PT homology up to a grading shift.

4.4 Braidlike isotopy and $\mathcal{H}(L)$

As we saw in Example 4.12, HOMFLY-PT homology for general link diagrams is *not* an isotopy invariant. However, it was proven in [12] that it was a *braidlike isotopy* invariant. We now carefully recall the definition of braidlike isotopy.

Definition 4.13 Two oriented link diagrams D and D' are said to represent *braidlike isotopic* links if they differ by a sequence of planar isotopies and the following moves in Figure 19. Such a sequence of moves will be called a *braidlike isotopy*.

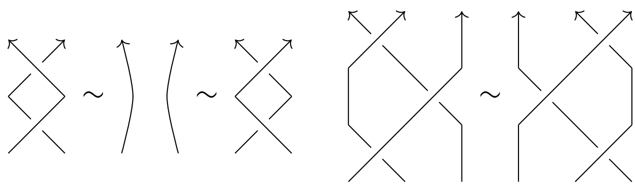


Figure 19: Braidlike Reidemeister moves

Theorem 4.14 (Khovanov and Rozansky [12]) *Let D and D' be two braidlike isotopic link diagrams. Then $\mathcal{H}(D) \simeq \mathcal{H}(D')$.*

Braidlike isotopy is an important notion in studying links in the solid torus. It is a well-known fact that two braid closures in the solid torus give isotopic links if and only if the braids are equivalent, or rather if they are braidlike isotopic. Audoux and Fiedler in [3] give a deformation of Khovanov homology which detects braidlike isotopy of links in $\mathbb{R}^3$. Their invariant decategorifies to a deformation of the Jones polynomial which can be computed using a Kauffman bracket-like relation. In the case of closed braid diagrams, their invariant corresponds with the homology theory studied in [2] by Asaeda, Przytycki and Sikora and the decategorification corresponds with the polynomial invariant studied by Hoste and Przytycki in [7].

We can now interpret Example 4.12 in the following manner: The unknot diagram D shown in Figure 18 is isotopic, but not braidlike isotopic to the standard unknot diagram O. In particular, there is no sequence of Reidemeister moves transforming D to O which does not contain the Reidemeister IIb move. In this sense, we see that $\mathcal{H}(D)$ can detect nonbraidlike isotopy. Note that $\mathcal{H}(D)$ is isomorphic to the homology of the left-handed trefoil knot (after two negative Reidemeister I moves), though clearly D is not a diagram for the left-handed trefoil knot. Therefore, this viewpoint of $\mathcal{H}(D)$ may not be useful in determining isotopy type of general diagrams, but can be very useful when we know the two diagrams are of the same isotopy type and we wish to determine if they are of the same braidlike isotopy type.

We can also see easily that the Poincaré series of $\mathcal{H}(D)$ and $\mathcal{H}(O)$ differ. Direct calculation shows that $\mathcal{P}(D) = (aq^2 + t^2q^2 + t^2aq^4)\mathcal{P}(O)$. This implies that we may be able to detect nonbraidlike isotopies on the level of the MOY calculus. As we will see in the next section, after a deformation of the MOY theory, this is indeed the case.

5 Decategorification of $\mathcal{H}(D)$ for general link diagrams

In this section we study the decategorification of $\mathcal{H}(D)$, which we will denote by $P^b(D)$. As we saw in Example 4.12, $P^b(D) \neq P(D)$ in general. In particular, when this occurs, this implies that D is not braidlike isotopic to a closed braid presentation of a link. We will end this section with a note on virtual links and give an explicit example of where $P^b(D)$ is not invariant under the virtual exchange move, which implies that $\mathcal{H}(D)$ cannot be extended to a virtual link invariant.

5.1 A deformation of the HOMFLY-PT polynomial

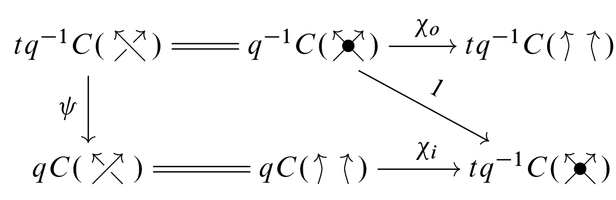
Let D be a link diagram and recall $\mathcal{P}(D)$ is defined as the Poincaré series of $\mathcal{H}(D)$. We define our deformed HOMFLY-PT polynomial as

(5-1) $$P^b(D) = \mathcal{P}(D)|_{t=-1} \in \mathbb{Z}(q,a).$$

Theorem 5.1 *Let D and D' be two link diagrams which are braidlike isotopic. Then $P^b(D) = P^b(D')$. Furthermore, P^b satisfies the following skein relation:*

(5-2) $$qP^b(\nwarrow\!\!\nearrow) - q^{-1}P^b(\nwarrow\!\!\nearrow) = (q - q^{-1})P^b(\uparrow\,\uparrow).$$

Proof The first statement is an immediate corollary of Theorem 4.14. For the second part of the statement, note that we have a map of homological degree 1 $\psi\colon tq^{-1}C(\nwarrow\!\!\nearrow) \to qC(\nwarrow\!\!\nearrow)$ given by:



The mapping cone of ψ, after Gaussian elimination, is homotopy equivalent to

$$qC(\uparrow\,\uparrow) \xrightarrow{-\chi_o\chi_i} tq^{-1}C(\uparrow\,\uparrow),$$

and therefore

(5-3) $$\mathrm{Cone}\big(tq^{-1}C(\nwarrow\!\!\nearrow) \xrightarrow{\psi} C(\nwarrow\!\!\nearrow)\big) \simeq \big(qC(\uparrow\,\uparrow) \xrightarrow{-\chi_o\chi_i} tq^{-1}C(\uparrow\,\uparrow)\big).$$

Properties of the Euler characteristic of a chain complex and (5-3) gives us the relation (5-2) as desired. □

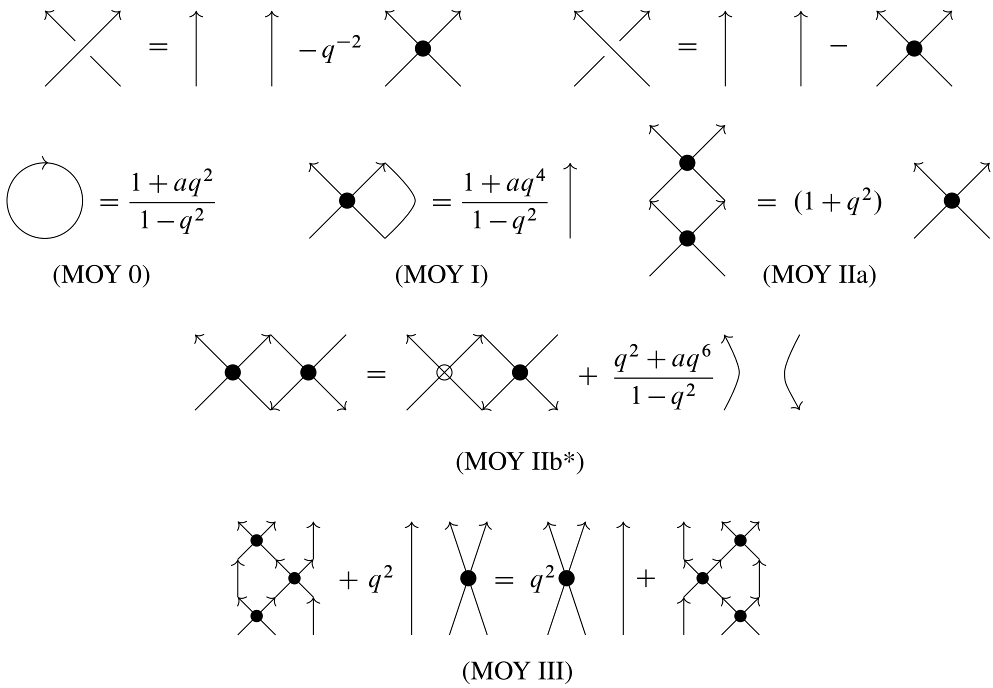
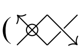


Figure 20: MOY relations for $P^b(D)$ and the deformed MOY IIb relation (the notation $P^b(\cdot)$ omitted for readability)

Corollary 5.2 *If D is a link diagram presented as a braid closure, then $P^b(D) = P(D)$. Equivalently if D is a link diagram for some link L and $P^b(D) \neq P(D)$, then D is not braidlike isotopic to a braid presentation of L.*

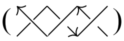
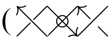
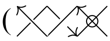
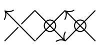
Corollary 5.3 *The relation $P^b($$) = -q^{-2} P^b($$)$ holds. We call this relation the deformed Reidemeister II relation and denote it by (RIIb*).*

We can also give a MOY-style construction for $P^b(D)$. In particular, we address the relation (MOY IIb) from Figure 6. As we saw in Lemma 4.10, the categorified MOY IIb relation does not hold as we would expect. However, we can decategorify the results in Proposition 3.7, Proposition 4.1, Theorem 4.11 and Lemma 4.10 in a natural way.

Proposition 5.4 *Let D be a link diagram. The braidlike Reidemeister moves, virtual MOY moves, virtual Reidemeister moves, and the relations in Figure 20 hold for $P^b(D)$.*

Example 5.5 Let D be the diagram of the $(2, 2k+1)$–torus knot given in Figure 21. First note that we can transform D to D' using (RIIb*) so that $P^b(D) = q^{-4} P^b(D')$.

More precisely, we use the fact that

$P^b($  $) = -q^{-2} P^b($  $) = -q^{-2} P^b($  $)$
$= q^{-4} P^b($  $) = q^{-4} P^b($  $).$

The equalities above follow (from left to right) by (R2b*), (Z2+), (MOYIIb*), and (R2b*). D' is isotopic to the $(2, 2k-1)$–torus knot via a planar isotopy and a (braidlike) Reidemeister IIa move. Therefore via a straightforward calculation similar to that in Example 2.7,

$$P^b(D) = \frac{1 + aq^2}{1 - q^2}\left((a+1)\sum_{i=1}^{k-1} q^{-i} + q^{-4k}\right).$$

However, if $\widetilde{D}$ is a braid presentation of the $(2, 2k+1)$–torus knot, then

$$P^b(\widetilde{D}) = \frac{1 + aq^2}{1 - q^2}\left((a+1)\sum_{i=1}^{k} q^{-i-2} + q^{-4k-2}\right).$$

Therefore D is not braidlike isotopic to a braid presentation of the $(2, 2k+1)$–torus knot for all $k \geq 1$.

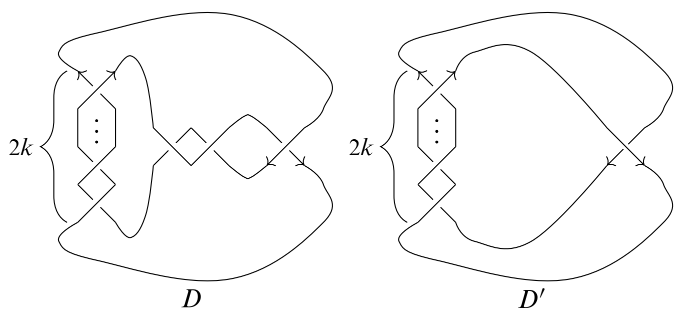


Figure 21: A nonbraidlike diagram D for the $(2, 2k+1)$–torus knot and a braidlike link diagram D' for the $(2, 2k-1)$–torus knot such that $P^b(D) = q^{-4} P^b(D')$.

5.2 An obstruction to extending HOMFLY-PT homology to virtual links

Finally we wish to present an argument showing that the current definition of $\mathcal{H}(D)$ cannot be extended to virtual links, even when they are presented as closures of virtual braids. We will ultimately use $P^b(D)$ to justify this statement.

A *virtual braid* is a braid in which we allow virtual crossings alongside positive and negative crossings. Two virtual braids are said to be equivalent if they differ by the moves from Figure 19 and the moves (VR2a), (VR3), and (SVR) from Figure 15. Kauffman in [9] proves that every virtual link can be presented as the closure of a virtual braid. There is also an analogue of the Markov theorem for virtual links.

Theorem 5.6 (Kamada [8]) *Let β and β' be two virtual braids. Their braid closures are equivalent virtual links if and only if they differ by a sequence of virtual braid equivalence moves (the braidlike Reidemeister moves, (VR2a), (VR3), and (SVR)), the Markov moves, and the virtual exchange move. The Markov moves and virtual exchange more are pictured in Figure 22.*

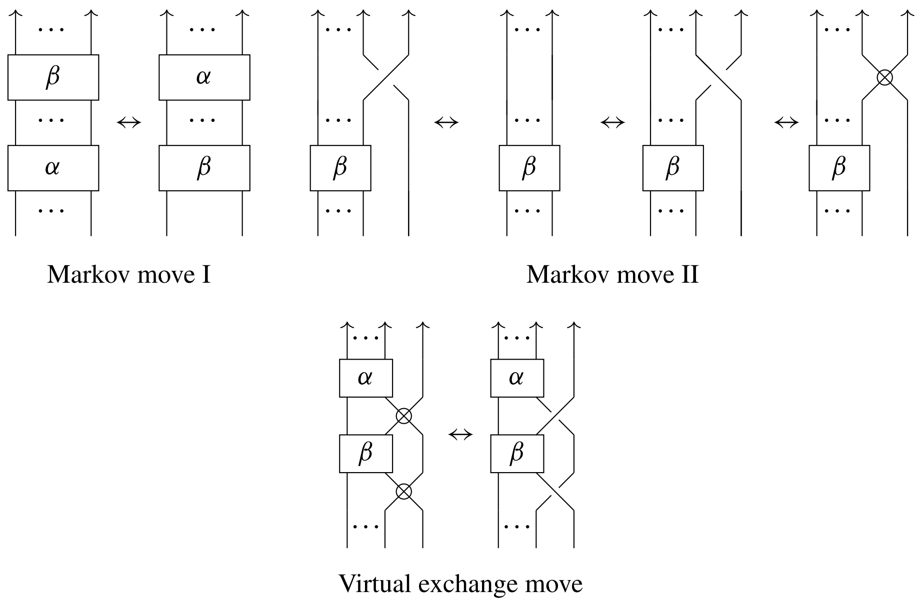


Figure 22: Markov moves for virtual links and the virtual exchange move. α and β are virtual braids.

Now we show by example that $P^b(D)$ is *not* invariant under the virtual exchange move. Therefore $\mathcal{H}(D)$ is not an invariant of virtual links, even when the links are presented as virtual braid closures.

Example 5.7 Let L be a connected sum of two virtual Hopf links as shown in Figure 23. β_1 and β_2 are two virtual braids whose closures are equivalent as virtual links to L. In particular, β_1 and β_2 are related by Markov move I and the virtual exchange move shown in Figure 22. Let D_1 be the braid closure of β_1 and D_2 be the

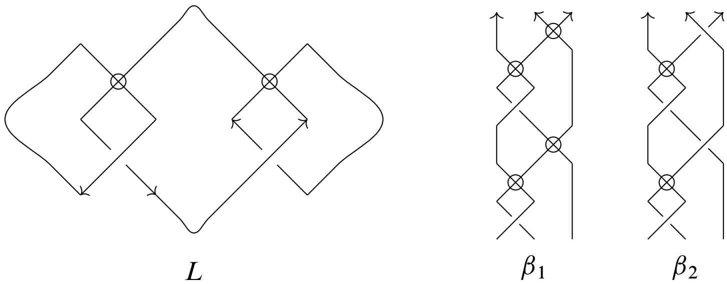


Figure 23: A connected sum of two virtual Hopf links, L, and two braid presentations of L

braid closure of β_2. Using the relations from Figure 20 we can directly compute that

$$P^b(D_1) = \frac{1+aq^2}{1-q^2}\left(1 - q^{-2}\left(\frac{1+aq^4}{1-q^2}\right)\right)^2,$$
$$P^b(D_2) = \frac{1+aq^2}{1-q^2}\left(aq^2 + 2\left(\frac{1+aq^4}{1-q^2}\right) - q^{-2}\left(\frac{1+aq^4}{1-q^2}\right)^2\right).$$

It is easy to see that $P^b(D_2) \neq P^b(D_1)$. In particular,

$$P^b(D_2) - P^b(D_1) = q^2(1+a)\frac{1+aq^2}{1-q^2}.$$

Therefore, $P^b(D)$ is not an invariant of virtual links and thus neither is $\mathcal{H}(D)$.

References

[1] **M Abel**, **L Rozansky**, *Virtual crossings and a filtration of the triply graded link homology of a link diagram*, J. Knot Theory Ramifications 26 (2017) art. id. 1750052, 38 pages MR

[2] **M M Asaeda**, **J H Przytycki**, **A S Sikora**, *Categorification of the Kauffman bracket skein module of I–bundles over surfaces*, Algebr. Geom. Topol. 4 (2004) 1177–1210 MR

[3] **B Audoux**, **T Fiedler**, *A Jones polynomial for braid-like isotopies of oriented links and its categorification*, Algebr. Geom. Topol. 5 (2005) 1535–1553 MR

[4] **D Auroux**, **J E Grigsby**, **S M Wehrli**, *Sutured Khovanov homology, Hochschild homology, and the Ozsváth–Szabó spectral sequence*, Trans. Amer. Math. Soc. 367 (2015) 7103–7131 MR

[5] **D Bar-Natan**, *Khovanov's homology for tangles and cobordisms*, Geom. Topol. 9 (2005) 1443–1499 MR

[6] **P Freyd**, **D Yetter**, **J Hoste**, **W B R Lickorish**, **K Millett**, **A Ocneanu**, *A new polynomial invariant of knots and links*, Bull. Amer. Math. Soc. 12 (1985) 239–246 MR

[7] **J Hoste**, **J H Przytycki**, *An invariant of dichromatic links*, Proc. Amer. Math. Soc. 105 (1989) 1003–1007 MR

[8] **S Kamada**, *Braid presentation of virtual knots and welded knots*, Osaka J. Math. 44 (2007) 441–458 MR

[9] **L H Kauffman**, *Virtual knot theory*, European J. Combin. 20 (1999) 663–690 MR

[10] **L H Kauffman**, **V O Manturov**, *Graphical constructions for the* $\mathrm{sl}(3)$*,* C_2 *and* G_2 *invariants for virtual knots, virtual braids and free knots*, J. Knot Theory Ramifications 24 (2015) art. id. 1550031, 47 pages MR

[11] **M Khovanov**, **L Rozansky**, *Virtual crossings, convolutions and a categorification of the* $\mathrm{SO}(2N)$ *Kauffman polynomial*, J. Gökova Geom. Topol. 1 (2007) 116–214 MR

[12] **M Khovanov**, **L Rozansky**, *Matrix factorizations and link homology, II*, Geom. Topol. 12 (2008) 1387–1425 MR

[13] **H Murakami**, **T Ohtsuki**, **S Yamada**, *Homfly polynomial via an invariant of colored plane graphs*, Enseign. Math. 44 (1998) 325–360 MR

[14] **J H Przytycki**, **P Traczyk**, *Conway algebras and skein equivalence of links*, Proc. Amer. Math. Soc. 100 (1987) 744–748 MR

[15] **H Queffelec**, **D E V Rose**, *Sutured annular Khovanov–Rozansky homology*, preprint (2015) arXiv To appear in Trans. Amer. Math. Soc.

[16] **J Rasmussen**, *Some differentials on Khovanov–Rozansky homology*, Geom. Topol. 19 (2015) 3031–3104 MR

[17] **A-L Thiel**, *Categorification of the virtual braid groups*, Ann. Math. Blaise Pascal 18 (2011) 231–243 MR

[18] **C A Weibel**, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics 38, Cambridge Univ. Press (1994) MR

Department of Mathematics, Duke University
Durham, NC, United States

maabel@math.duke.edu

http://services.math.duke.edu/~maabel/

Received: 30 October 2016 Revised: 7 March 2017

msp

Algebraic & Geometric Topology 17 (2017) 3057–3079

The topological sliceness of 3–strand pretzel knots

ALLISON N MILLER

We give a complete characterization of the topological slice status of odd 3–strand pretzel knots, proving that an odd 3–strand pretzel knot is topologically slice if and only if it either is ribbon or has trivial Alexander polynomial. We also show that topologically slice even 3–strand pretzel knots, except perhaps for members of Lecuona's exceptional family, must be ribbon. These results follow from computations of the Casson–Gordon 3–manifold signature invariants associated to the double branched covers of these knots.

57M25; 57N70

1 Introduction

In the years since Fox first posed the slice-ribbon conjecture (Problem 1.33 on Kirby's list [14]), its validity has been established for several families of knots. The usual strategy is to give an explicit list of ribbon knots in the family and then to provide an obstruction to the smooth sliceness of all others in the family. An early example of this is the following classification of the smoothly slice rational knots due to Lisca.

Theorem 1.1 (Lisca [16]) *A rational knot is smoothly slice if and only if it is ribbon if and only if it is in* $\mathcal{R}$.

$\mathcal{R}$ is an explicit family of rational knots known to be ribbon at least since the work of Casson and Gordon [4]. Lisca argues that if K is not in $\mathcal{R}$, then Donaldson's diagonalization theorem obstructs $\Sigma_2(K)$ from smoothly bounding a rational homology ball, and hence obstructs K from being smoothly slice.

In a similar spirit, though with entirely different methods, we give an almost complete characterization of the topological sliceness of 3–strand pretzels via the computation of Casson–Gordon signatures corresponding to the double branched cover. In particular, we have the following complete characterization of topologically slice odd 3–strand pretzel knots. (Note that we call a pretzel knot $P(p_1, \dots, p_n)$ *odd* if all of its parameters p_i are odd and *even* if one parameter is even.)

Theorem 1.2 (Main Theorem A) *Let K be an odd 3–strand pretzel knot with nontrivial Alexander polynomial. Then K is topologically slice if and only if K is of the form $\pm P(p, q, -q)$ or $\pm P(1, q, -q-4)$ for some odd $p, q \in \mathbb{N}$, in which case it is obviously ribbon.*

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3057

By work of Freedman in [9], every knot with trivial Alexander polynomial is topologically slice. The following result, originally proved by Fintushel and Stern, illustrates that this is far from true for 3–strand pretzel knots in the smooth category.

Theorem 1.3 (Fintushel and Stern [8]) *Let K be a nontrivial odd 3–strand pretzel knot with $\Delta_K(t) = 1$. Then K is not smoothly slice.*

Theorems 1.2 and 1.3 therefore together give an alternate proof of the following complete characterization of smoothly slice 3–strand pretzel knots given by Greene and Jabuka in [11]. Their arguments, like Lisca's, are smooth in nature and rely on Donaldson's theorem along with additional obstructions coming from Heegaard Floer homology.

Theorem 1.4 (Greene and Jabuka [11]) *Let K be an odd 3–strand pretzel knot. Then K is smoothly slice if and only if it is ribbon if and only if K is of the form $\pm P(p,q,-q)$ or $\pm P(1,q,-q-4)$ for odd $p,q \in \mathbb{N}$.*

Note that both Lisca and Greene and Jabuka actually prove stronger results that completely characterize the order of rational knots and odd 3–strand pretzel knots in the smooth concordance group. Theorem 1.2 has the following nice corollary.

Corollary 1.5 *Let K be a genus-one alternating knot. Then K is topologically slice if and only if K is ribbon.*

Proof Let K be a genus-one alternating knot. Then by work of Stoimenow in [18], K either is an odd 3–strand pretzel knot with all parameters of the same sign (and hence has nonzero signature and is not even algebraically slice) or is rational. Therefore we may assume that K is a genus-one rational knot and hence (up to reflection) corresponds to the fraction $(4ab+1)/(2a)$ for some $a,b>0$; see for example Burde and Zieschang [2, Proposition 12.26]. Note that K has determinant $4ab+1>1$ and hence does not have trivial Alexander polynomial. Therefore, since such knots can also be described as the 3–strand pretzel knot $P(1,2a-1,-(2b+1))$, Theorem 1.2 implies that K is topologically slice if and only if it is ribbon. □

We also consider the topological slice status of even 3–strand pretzel knots, and are able to use Casson–Gordon signatures to prove the following theorem, where for odd $a>0$ we define P_a to be the even 3–strand pretzel knot $P(a,-a-2,-(a+1)^2/2)$.

Theorem 1.6 (Main Theorem B) *Let K be an even 3–strand pretzel knot that is not of the form $\pm P_a$ for $a \equiv 1, 11, 37, 47, 59 \bmod 60$. Then K is topologically slice if and only if K is of the form $P(p,q,-q)$ for some even p and odd q, in which case it is obviously ribbon.*

The family $\{\pm P_a\}$ was first considered by Lecuona in [15]. Lecuona uses techniques analogous to those of Greene and Jabuka to describe the smooth sliceness of even 3–strand pretzel knots, except for this exceptional family $\{\pm P_a\}$. In fact, Lecuona's results are much broader, essentially characterizing the smooth sliceness up to mutation of all even pretzel knots not in this exceptional family. It follows from work of Jabuka in [13] that the knots $\{\pm P_a\}$ are exactly the even 3–strand pretzel knots with trivial rational Witt class and determinant one.

Theorem 1.7 (Lecuona [15]) *Let K be an even 3–strand pretzel knot that is not of the form $\pm P_a$ for any $a \equiv 1, 11, 37, 47, 49, 59 \bmod 60$. Then K is smoothly slice if and only if it is ribbon if and only if it is of the form $P(p, q, -q)$ for some even p and odd q.*

Lecuona conjectures that the (non)existence of a Fox–Milnor factorization for the Alexander polynomial obstructs even the algebraic sliceness of the $\{\pm P_a\}$ family. When combined with Theorem 1.6, this would imply an affirmative answer to the following conjecture.

Conjecture 1.8 *Let K be an even 3–strand pretzel knot. Then K is topologically slice if and only if K is ribbon.*

We conveniently summarize Theorems 1.2 and 1.6 in this (slightly weaker) statement:

Theorem 1.9 *Let K be a 3–strand pretzel knot with nontrivial determinant. Then K is topologically slice if and only if K is ribbon.*

Note that despite our almost complete understanding of topological sliceness for 3–strand pretzel knots, it remains open whether smoothly slice equals topologically slice for rational knots. Recent work of Feller and McCoy [7] shows that there are rational knots with distinct smooth and topological 4–genera.

A natural next question is the extent to which double branched cover Casson–Gordon signatures obstruct the topological sliceness of pretzel knots with more than three strands. However, several difficulties arise. First, pretzel knots with more than three strands have nontrivial mutations which often persist in concordance. (See the work of Herald, Kirk and Livingston [12] for examples.) However, even if we are willing to consider knots only up to mutation we cannot expect a complete answer from these techniques. In particular, there exist algebraically slice odd 5–strand pretzel knots with nontrivial Alexander polynomial but trivial determinant. (For example, consider $P(7, 11, 53, -5, -19)$.) There is no reason to believe that these knots are topologically slice, but there are also no double branched cover Casson–Gordon signatures to serve as sliceness obstructions.

Outline of the paper In Section 2, we provide background and basic results on Casson–Gordon signatures. In Section 3, we provide necessary results concerning the colored signatures of links. In Section 4, we prove Main Theorem A (Theorem 1.2), completely characterizing which odd 3–strand pretzel knots are topologically slice. Finally, in Section 5 we briefly outline the arguments used to prove Main Theorem B (Theorem 1.6), our result for even 3–strand pretzel knots.

2 Casson–Gordon signature invariants

Casson and Gordon associate to a knot K and a map $\chi\colon H_1(\Sigma_n(K)) \to \mathbb{Z}_d$ the invariant $\tau(K,n,\chi) \in L_0(\mathbb{Q}(\omega)(t)) \otimes \mathbb{Q}$. Note that $L_0(\mathbb{Q}(\omega)(t))$ is the Witt group of nonsingular Hermitian forms on finite-dimensional $\mathbb{Q}(\omega)(t)$–modules, where $\omega = e^{2\pi i/d}$. These invariants obstruct the topological sliceness of K as follows.

Theorem 2.1 (Casson and Gordon [4]) *Let K be a topologically slice knot and n a prime power. Then there exists a square-root order subgroup $M \leq H_1(\Sigma_n(K))$, invariant under the action of the covering transformations, with the linking form of $\Sigma_n(K)$ vanishing on $M \times M$ (ie M is a* ***metabolizer*** *for the linking form) such that if χ is a prime-power order character with $\chi|_M = 0$, then $\tau(K,n,\chi) = 0$.*

While this is a powerful sliceness obstruction, $\tau(K,n,\chi)$ cannot generally be directly computed. Instead, as originated in [4], one relates the Witt class signature $\overline{\sigma}_1(\tau(K,n,\chi))$ to a simpler signature associated to any 3–manifold Y and character from $H_1(Y)$ to a cyclic group. We give the definition of this signature, following [3].

First, whenever $X_\chi \to X$ is a cyclic d–fold cover, perhaps branched, we let $\omega = e^{2\pi i/d}$ and define the χ–twisted homology of X to be the $\mathbb{Q}(\omega)$ vector space

$$H_*^\chi(X) := H_*(C_*(X_\chi) \otimes_{\mathbb{Z}[\mathbb{Z}_d]} \mathbb{Q}(\omega)) \cong H_*(X_\chi) \otimes_{\mathbb{Z}[\mathbb{Z}_d]} \mathbb{Q}(\omega).$$

We now let Y be a closed 3–manifold and $\chi\colon H_1(Y) \to \mathbb{Z}_d$ an onto homomorphism. The map χ induces a d–fold cyclic cover $Y_\chi \to Y$ with a canonical generator τ for the group of covering transformations. Suppose that there is some d–fold branched cyclic cover of 4–manifolds $W_\chi \to W$ with branch set a closed surface $F \subset \operatorname{int}(W)$ such that $\partial(W_\chi \to W) = r(Y_\chi \to Y)$ for some $r \in \mathbb{N}$. Suppose also that the covering transformation $\tilde{\tau}$ of W_χ that induces rotation by $2\pi/d$ on the fibers of the normal bundle of the preimage of F in W_χ induces the canonical covering transformation τ on Y_χ. We can always choose either $F = \varnothing$ or $r = 1$ by bordism group considerations and an explicit description in [3], respectively, and all of our work will be in one of these cases. The action of $\widetilde{\tau}$ on $H := H_2(W_\chi, \mathbb{C})$ allows us to decompose H as the direct sum of eigenspaces $H_2^k(W_\chi)$ corresponding to eigenvalues ω^k for $k = 0, \dots, d-1$. For $k > 0$, define $\epsilon_k(W_\chi)$ to be the signature of the intersection form of W_χ when restricted to $H_2^k(W_\chi)$. Note that $\epsilon_1(W_\chi)$ can be equivalently be defined as the signature of the twisted intersection form on $H_2^\chi(W) = H_2(W_\chi) \otimes_{\mathbb{Z}[\mathbb{Z}_d]} \mathbb{Q}(\omega)$.

Definition 2.2 With the above setup, the k^{th} Casson–Gordon signature of (Y, χ) is

$$\sigma_k(Y, \chi) = \frac{1}{r}\Big(\sigma(W) - \epsilon_k(W_\chi) - \frac{2k(d-k)}{d^2}([F]\cdot[F])\Big).$$

Those familiar with the definition of $\tau(K, n, \chi)$ should note that we generally have $\sigma_1(\Sigma_n(K), \chi) \neq \overline{\sigma}_1(\tau(K, n, \chi))$. However, we can bound the difference between $\sigma_1(\Sigma_n(K), \chi)$ and $\overline{\sigma}_1(\tau(K, n, \chi))$, in a straightforward extension of [4, Theorem 3].

Theorem 2.3 (Casson and Gordon [4]) *Let χ: $H_1(\Sigma_n(K)) \to \mathbb{Z}_d$ be an onto homomorphism. Then*

$$|\sigma_1(\Sigma_n(K), \chi) - \overline{\sigma}_1(\tau(K, n, \chi))| \leq \dim H_1^\chi(\Sigma_n(K)) + 1.$$

Proof We follow the proof of [4, Theorem 3]. Let M_n denote the n–fold cyclic cover of the 3–manifold $S_0^3(K)$ obtained by doing 0–surgery along K. For convenience we let $\Sigma_n = \Sigma_n(K)$. Note that χ determines a map $H_1(M_n) \to \mathbb{Z}_d$, which by an abuse of notation we also refer to as χ. By the usual bordism group considerations, for some $r \in \mathbb{N}$ there is a compact 4–manifold W_n with boundary $r\Sigma_n$ such that χ extends over $H_1(W_n)$. Note that M_n can be obtained from Σ_n by a single 0–framed surgery along $\widetilde{K}$, the preimage of K under the branched covering map. Therefore rM_n bounds a 4–manifold V_n obtained by attaching r 0–framed 2–handles to W_n. Let ν denote the nullity of the twisted intersection form on $H_2^\chi(V_n)$. The arguments of the proof of [4, Theorem 3] carry over verbatim to establish the inequality

$$|\sigma_1(M_n, \chi) - \overline{\sigma}_1(\tau(K, n, \chi))| \leq \frac{\nu}{r}.$$

Since our covers are unbranched, Definition 2.2 gives us

$$\sigma_1(\Sigma_n, \chi) = \frac{1}{r}\big(\sigma(W_n) - \sigma(H_2^\chi(W_n))\big),$$
$$\sigma_1(M_n, \chi) = \frac{1}{r}\big(\sigma(V_n) - \sigma(H_2^\chi(V_n))\big).$$

By our construction of V_n from W_n, it is straightforward to verify that $\sigma(V_n) = \sigma(W_n)$ and that $H_2^\chi(V_n)$ has a codimension-r subspace which is isometric to $H_2^\chi(W_n)$. Note that by duality the intersection form on $H_2^\chi(V_n)$ has nullity equal to $r \dim H_1^\chi(\Sigma_n)$, whereas by definition the intersection form on $H_2^\chi(W_n)$ has nullity ν. We thus have the following, which when combined with our previous inequality gives the desired result:

$$\begin{aligned}|\sigma_1(\Sigma_n,\chi)-\sigma_1(M_n,\chi)| &= \left|\frac{1}{r}\big[\sigma(W_n)-\sigma(H_2^\chi(W_n))\big]-\frac{1}{r}\big[\sigma(V_n)-\sigma(H_2^\chi(V_n))\big]\right| \\ &= \frac{1}{r}\big|\sigma(H_2^\chi(W_n))-\sigma(H_2^\chi(V_n))\big| \\ &\leq \frac{1}{r}\big[r-(\nu-r\dim H_1^\chi(\Sigma_n))\big] \\ &= \dim H_1^\chi(\Sigma_n)+1-\frac{\nu}{r}. \end{aligned}$$ □

The following corollary will be our main obstruction to topological sliceness.

Corollary 2.4 [4] *Suppose that K is a topologically slice knot and that $n = p^r$ is a prime power. Then there exists a metabolizer M for the linking form on $H_1(\Sigma_n(K))$ such that if χ is a character of prime-power order d vanishing on M, then for any $k = 1, \dots, d-1$,*

$$|\sigma_k(\Sigma_n(K), \chi)| \leq \dim H_1^\chi(\Sigma_n(K)) + 1.$$

Proof Replacing χ with a nonzero multiple of itself permutes $\{\sigma_k(\Sigma_n(K), \chi)\}_{k=1}^{d-1}$ while preserving the property of vanishing on M, so Theorems 2.1 and 2.3 combine to give the desired result. □

If the obstruction of Corollary 2.4 vanishes for characters from $H_1(\Sigma_2(K))$ to $\mathbb{Z}_d$, then we will refer to K as *CG-slice* at d. The following proposition is often convenient in recognizing that $\Sigma_n(K)_\chi$ is a rational homology sphere, and hence that the bound of Corollary 2.4 reduces to $|\sigma_1(\Sigma_n(K), \chi)| \leq 1$.

Proposition 2.5 (Casson and Gordon [3]) *Suppose that Y is a rational homology sphere with $H_1(Y, \mathbb{Z}_p)$ cyclic for some prime p. Then any cyclic p^n–fold cover of Y is also a rational homology sphere.*

In order to effectively apply this obstruction, we would like to be able to compute $\sigma_k(Y, \chi)$ from an arbitrary integral surgery description of Y.

Definition 2.6 Let K be an oriented knot, and A an embedded annulus such that $\partial A = K \sqcup -K'$ and $\operatorname{lk}(K, K') = \lambda$. An *$\lambda$–twisted a–cable* of K is any oriented link L obtained as the union of $n = n_+ + n_-$ parallel copies of K in A such that n_+ are oriented with K, n_- opposite to K, and $n_+ - n_- = a$.

Let $L = \bigcup_{i=1}^n L_i$ be an oriented link in S^3 such that surgery along L with integer framings $\{\lambda_i\}_{i=1}^n$ gives Y. We refer to the meridian of component L_i as μ_i and let

$A = [a_{ij}]$ be the linking matrix of L. The following proposition is a generalization of [3, Lemma 3.1].

Proposition 2.7 (Gilmer [10]) *Let Y be obtained by integer surgery on L as above and $\chi\colon H_1(Y) \to \mathbb{Z}_d$ be an onto homomorphism. Let L_χ be a satellite of L obtained by replacing each L_i by a nonempty λ_i–twisted m_i–cable of L_i, such that $\chi(\mu_i) \equiv m_i \bmod d$. Then for any $0 < k < d$,*

$$\sigma_k(Y,\chi) = \sigma(A) - \sigma_{L_\chi}(\omega^k) - \frac{2k(d-k)}{d^2}\left(\sum_{i,j=1}^{n} m_i m_j a_{ij}\right).$$

In order to effectively apply Proposition 2.7 we will need to compute the Tristram–Levine signatures of cables of links. The techniques of colored signatures prove useful for this, as well as providing an independent means of computation for $\sigma_1(Y,\chi)$.

3 Colored signatures of colored links

A *n–colored link* is an oriented link L together with a surjective map assigning to each component of L a color in $\{1, 2, \ldots, n\}$. We let L_i denote the sublink of L consisting of i–colored components, and call each L_i a *colored component*. A *C-complex* for a colored link L consists of a union of Seifert surfaces for the colored components of L which intersect only in a prescribed way (in "clasps"; see [5] for the precise definition).

The *colored signature* of L is a map $\sigma_L\colon (S^1)^n \to \mathbb{Z}$ that is defined via the C-complex in a way exactly analogous to the definition of the Tristram–Levine signatures in terms of a Seifert surface for a link. The colored signature shares many properties, including a 4–dimensional interpretation, with the ordinary signatures. We need the following results, due primarily to Cimasoni and Florens [5]:

Recovery of Tristram–Levine signatures Let L be a n–component, n–colored link, and call the underlying ordinary link L'. Then for any $\omega \in S^1 - \{1\}$, we have $\sigma_L(\omega, \ldots, \omega) = \sigma_{L'}(\omega) + \sum_{i<j} \operatorname{lk}(L_i, L_j)$.

Additivity Let $L' = L'_1 \cup \cdots \cup L'_m$ and $L'' = L''_{m+1} \cup \cdots \cup L''_{m+n}$ be colored links and L be the $(m+n-1)$–colored link obtained by connected summing any component of L'_m with any component of L''_{m+1}. Then $\sigma_L(\omega_1, \ldots, \omega_m, \ldots, \omega_{m+n-1}) = \sigma_L(\omega_1, \ldots, \omega_m) + \sigma_{L''}(\omega_m, \ldots, \omega_{m+n-1})$.

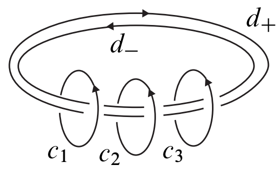
Behavior under reversal and mirroring The colored signature is invariant under global reversal of orientations. Also, letting $\overline{L}$ denote the mirror of L we have $\sigma_{\overline{L}}(\omega_1, \ldots, \omega_n) = -\sigma_L(\omega_1, \ldots, \omega_n)$.

Behavior at 1 (Degtyarev, Florens and Lecuona [6]) Let L be an n–colored link and L' be the $(n-1)$–colored link obtained by deleting the n^{th} colored component of L. Then $\sigma_L(\omega_1, \dots, \omega_{n-1}, 1) = \sigma_{L'}(\omega_1, \dots, \omega_{n-1})$.

Hopf link computation Let L be either Hopf link, considered as a 2–colored link. Then the colored signature function of L is identically 0.

We also need the following consequence of Degtyarev, Florens and Lecuona's general description of the signature of a splice in [6].

Example 3.1 Let L be the following 5–colored link:



Let $\Phi(L)$ be the satellite of L obtained by replacing each component c_i with a coherently oriented torus link $T(a_i, p_i a_i)$ for $i = 1, 2, 3$. Observe that as an ordinary oriented link, L is isotopic to its mirror image in a way that swaps components d_+ and d_- and preserves all other components. It follows that $\sigma_L(\omega_0, \omega_0, \vec{\omega}) = 0$ for all $\omega_0 \in S^1$ and $\vec{\omega} \in (S^1)^3$. Let $\theta \in S^1$ be such that $\theta^{a_i} \neq 1$ for $i = 1, 2, 3$. Then [6, Theorem 2.2] and the above results imply that $\sigma_{\Phi(L)}(\theta) = \sum_{i=1}^{3} \sigma_{T(a_i, p_i a_i)}(\theta)$.

Finally, in some cases colored signatures give us an alternate computational method for Casson–Gordon signatures.

Theorem 3.2 (Cimasoni and Florens [5]) *Let Y be a 3–manifold obtained by surgery on a framed n–component link L with linking matrix $A = [a_{ij}]$. Let $\chi\colon H_1(Y) \to \mathbb{Z}_d$ be a character of prime-power order that takes the meridian of each component of L to a unit in $\mathbb{Z}_d$. Denote the lift of the image of the i^{th} meridian of L to $\{1, \dots, d-1\}$ by m_i. Consider L as a n–colored link, and let $\omega_\chi = (\omega^{m_1}, \dots, \omega^{m_n})$. Then*

$$\sigma_1(Y, \chi) = \sigma(A) - \Big(\sigma_L(\omega_\chi) - \sum_{i<j} a_{ij}\Big) - \frac{2}{d^2}\Big(\sum_{i,j}(d - m_i)m_j a_{ij}\Big).$$

Note that in the case that every meridian is sent to 1 and $k = 1$, Theorems 2.7 and 3.2 both reduce to the original [3, Lemma 3.1].

4 Casson–Gordon signatures of 3–strand pretzels

We now give the outline of the proof of Theorem 1.2, deferring computations to later propositions.

Proof of Theorem 1.2 Suppose that K is an algebraically slice odd 3–strand pretzel knot with nontrivial Alexander polynomial. We will argue that either the Casson–Gordon signatures of $\Sigma_2(K)$ obstruct the topological sliceness of K or the knot K is in fact ribbon. Since K is algebraically slice, the ordinary signature of K vanishes, and so an easy computation from the standard genus-one Seifert surface for K shows that $pq+qr+pr<0$; see also [13]. Also, $|H_1(\Sigma_2(K))| = -pq-qr-pr = D^2$ for some odd $D \in \mathbb{N}$. Note that since K is a genus-one algebraically slice knot with nontrivial Alexander polynomial, $D^2 \neq 1$ and hence D has prime divisors. Since $pq+pr+qr<0$, the parameters p, q and r are not all of the same sign and so via reflection and the symmetries of 3–strand pretzel knots we can assume that $p,q>0$ and $r<0$.

In the following cases, the existence of a prime d that divides D and satisfies the given conditions implies that the Casson–Gordon signatures of $\Sigma_2(K)$ corresponding to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K:

Case 1 (Proposition 4.1) d divides p and q but not r.

Case 2 (Proposition 4.3) d divides r and exactly one of p and q.

Case 3 (Proposition 4.6) d divides all of p, q and r.

Case 4 (Proposition 4.10) d divides D but none of p, q and r; $p \not\equiv q \bmod d$; and $r \neq -(4p+q)$ (assuming without loss of generality that $q>p$).

Case 5 (Proposition 4.11) d divides D but none of p, q and $r=-(4p+q)$.

Case 6 (Proposition 4.12) d divides D but none of p, q and r; $p \equiv q \bmod d$; and $d \neq 3$.

Now suppose that there is no prime satisfying any of the above. It follows that p, q and r are relatively prime, $p \equiv q \bmod 3$, and D is a power of three. We show that in this case the Casson–Gordon signatures corresponding to characters of order 3 and 9 obstruct topological sliceness in Proposition 4.13. □

We now set up for our various computations. Note that if r equals one of $-p$ and $-q$, there is a single band move taking K to a 2–component unlink, and hence K is ribbon. So we suppose $r \neq -p,-q$. We start with the surgery diagram for $\Sigma_2(K)$ in Figure 1, with linking matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & p & 0 & 0 \\ 1 & 0 & q & 0 \\ 1 & 0 & 0 & r \end{bmatrix}$$

and $\sigma(A)=0$. We refer to the meridians of each component by μ_0, μ_p, μ_q and μ_r according to their framings.

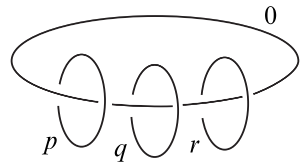


Figure 1: A surgery diagram L_0 for $\Sigma_2(P(p,q,r))$

Note that A is a presentation matrix for $H_1(\Sigma_2(K)$, and that it is straightforward to use row and column moves and obtain the smaller presentation matrix $A' = \left[\begin{smallmatrix} p+q & p \\ p & p+r \end{smallmatrix}\right]$. Let d be any prime dividing D and suppose that d does not divide all of p, q and r. Observe that this implies that some entry of A' is a unit in $\mathbb{Z}_d$, and hence by choosing this as our pivot entry and working over $\mathbb{Z}_d$ we can use row and column moves to obtain $A'' = \left[\begin{smallmatrix} 1 & 0 \\ 0 & * \end{smallmatrix}\right]$. Observe that A'' is a presentation matrix for $H_1(\Sigma_2(K), \mathbb{Z}_d)$, and so we see that $H_1(\Sigma_2(K), \mathbb{Z}_d)$ is cyclic and hence every regular d^n–fold cyclic cover of $\Sigma_2(K)$ is a rational homology sphere (Proposition 2.5). In addition, when $H_1(\Sigma_2(K), \mathbb{Z}_d)$ is cyclic any character $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ will vanish on any metabolizer for the linking form; see [12, Lemma 8.2]. So we have the following:

Useful fact Suppose that $K = P(p,q,r)$ is topologically slice, d is a prime dividing $pq + qr + pr$ that does not divide all of p, q and r, and χ is any character $H_1(\Sigma_2(K)) \to \mathbb{Z}_d$. Then $|\sigma_1(\Sigma_2(K), \chi)| \leq 1$.

4.1 Cases 1 and 2: d divides some but not all of p, q and r

Proposition 4.1 (**Case 1**) *Let $K = K(p,q,r)$, where*

$$p, q > 0, \quad r < 0 \quad \text{and} \quad pq + pr + qr = -D^2.$$

Suppose that d is a prime that divides p and q but not r. Then the Casson–Gordon signatures of $\Sigma_2(K)$ associated to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K.

Proof We start by manipulating our surgery description for $\Sigma_2(K)$. Slide the curves with framing p and q over the curve with framing r. Then convert the 0–framed 2–handle to a 1–handle, and cancel the 1–handle with the r–framed 2–handle. We end with a new surgery description for $\Sigma_2(K)$ with underlying link $L = T(2, 2r)$ and framings $p + r$ and $q + r$. The linking matrix of L is $A = \left[\begin{smallmatrix} p+r & r \\ r & q+r \end{smallmatrix}\right]$ and has $\sigma(A) = 0$. Note that if we consider the entries of A mod d we get a presentation matrix for $H_1(\Sigma_2(K), \mathbb{Z}_d)$ with respect to basis $\{\mu_p, \mu_q\}$, which immediately implies that $H_1(\Sigma_2(K), \mathbb{Z}_d) \cong \mathbb{Z}_d$, with generator $\mu_p = -\mu_q$.

By our useful fact, it suffices to show that for some $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ we have that $|\sigma_1(\Sigma_2(K), \chi)| > 1$. Define χ on $H_1(\Sigma_2(K))$ by $\chi(\mu_p) = \chi(-\mu_q) = 1$. So L_χ is the torus link $T(2, 2r)$ with strands oppositely oriented. Note that $\sigma_{L_\chi}(\omega^k) = -1$ for $0 < k < d$ and so we have by Proposition 2.7 that

$$\sigma_k(\Sigma_2(K), \chi) = 1 - 2((p+r) - 2r + (q+r))\frac{k(d-k)}{d^2} = 1 - 2\Big(\frac{p+q}{d}\Big)\Big(\frac{k(d-k)}{d}\Big).$$

Note that d divides p and q, so $p+q \geq 2d$. Note that $k(d-k) \geq (d-1)$ for all choices of $k = 1, \dots, d-1$. Since $d \geq 3$, we have

$$|\sigma_k(\Sigma_2(K), \chi)| \geq 2 \cdot 2 \cdot \big(1 - \tfrac{1}{3}\big) - 1 = \tfrac{8}{3} - 1 > 1. \qquad \square$$

The above proof shows $\sigma_k(\Sigma_2(K), \chi) < -1$ for all choices of $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ and $k = 1, \dots, d-1$, giving the following easy corollary.

Corollary 4.2 *For each odd prime s, let $K_s = P(p_s, q_s, r_s)$ be an odd 3–strand pretzel knot such that $p_s, q_s > 0$ are divisible by s; $r_s < 0$ is not divisible by s; and $p_s q_s + p_s r_s + q_s r_s = -s^2$. Then $\{K_s\}$ is a basis of algebraically slice knots for a $\mathbb{Z}^\infty$ subgroup of the topological concordance group.*

Note that such K_s exist; for example, we can take $K_s = (s^2, s^2, -(s^2+1)/2)$. (Note since s is odd s^2+1 is equivalent to $2 \bmod 4$ and so this is an odd pretzel as desired.)

Proof Suppose that $K = \sum_{i=1}^n a_i K_{s_i}$ is topologically slice, where each a_i is nonzero. By reflecting K, we can assume without loss of generality that $a_1 > 0$. Since K is topologically slice and $H_1(\Sigma_2(K), \mathbb{Z}_{s_i})$ is nonzero, it follows from Theorem 2.1 that there is some nontrivial character $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_{s_1}$ such that $\overline{\sigma}_1(\tau(K, 2, \chi)) = 0$. Observe that

$$H_1(\Sigma_2(K)) = \bigoplus_{i=1}^n \big(H_1(\Sigma_2(K_{s_i}))^{\oplus |a_i|}\big) = \bigoplus_{i=1}^n \big(\mathbb{Z}_{s_i}[t]/\langle t+1\rangle\big)^{\oplus |a_i|}.$$

Note that χ is trivial on each $H_1(\Sigma_2(K_{s_i}))$ factor for $i \neq 1$, and that χ can be decomposed as $\chi = \bigoplus_{j=1}^{|a_1|} \chi_j$, where each $\chi_j\colon H_1(\Sigma_2(K_{s_1})) \to \mathbb{Z}_{s_1}$ and at least one χ_j is nontrivial. By the additivity of Casson–Gordon signatures,

$$\overline{\sigma}_1(\tau(K, 2, \chi)) = \sum_{j=1}^{|a_1|} \overline{\sigma}_1(\tau(K_{s_1}, 2, \chi_j)).$$

However, the proof of Proposition 4.1 shows that $\sigma_1(\Sigma_2(K_{s_1}), \chi_j) < -1$ whenever χ_j is nontrivial, and that

$$\big|\overline{\sigma}_1(\tau(K_{s_1}, 2, \chi_j) - \sigma_1(\Sigma_2(K_{s_1}), \chi_j)\big| \leq 1.$$

It follows that $\overline{\sigma}_1(\tau(K,2,\chi_j))$ is strictly negative whenever χ_j is nontrivial (and zero when χ_j is trivial), and so $\overline{\sigma}_1(\tau(K,2,\chi)) < 0$, which is our desired contradiction. □

Now we continue to the next case.

Proposition 4.3 (**Case 2**) *Let $K = K(p,q,r)$. Suppose that there exists a prime d that divides r and exactly one of p and q, but that $r \neq -p,-q$. Then the Casson–Gordon signatures of $\Sigma_2(K)$ associated to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K.*

Proof The argument is exactly analogous to that of the proof of Proposition 4.1, except that we choose k to be $(d-1)/2$; the details are left to the reader. □

4.2 Case 3: d divides all of p, q and r

In this case, we have that $H_1(\Sigma_2(K),\mathbb{Z}_d) \cong \mathbb{Z}_d \oplus \mathbb{Z}_d$, and so there may be metabolizers $M \leq H_1(\Sigma_2(K))$ with nontrivial image in $H_1(\Sigma_2(K),\mathbb{Z}_d)$. For each such metabolizer we provide a character χ to $\mathbb{Z}_d$ vanishing on M such that the corresponding Casson–Gordon signature has sufficiently large absolute value. We first determine what "sufficiently large" is in the context of Corollary 2.4.

Lemma 4.4 *Let $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$. Then $\dim H_1^{\chi}(\Sigma_2(K))$ is 1 if $\chi(\mu_p)$, $\chi(\mu_q)$ and $\chi(\mu_r)$ are all nonzero and 0 otherwise.*

Proof By slight simplifications of the Wirtinger presentation, we obtain

$$\pi_1(S^3 - L_0) = \langle \mu_0,\mu_p,\mu_q,\mu_r : \mu_0\mu_p = \mu_p\mu_0,\ \mu_0\mu_q = \mu_q\mu_0,\ \mu_0\mu_r = \mu_r\mu_0\rangle,$$

where μ_* is any meridian of the $*$–framed curve, for $* = 0,p,q,r$. Note that the 0–framed longitudes of the surgery curves are given with respect to this generating set by $\lambda_0 = \mu_r\mu_q\mu_p$ and $\lambda_p = \lambda_q = \lambda_r = \mu_0$. Gluing in solid tori according to the surgery framings gives new relations

$$\lambda_0 = \mu_r\mu_q\mu_p = 1, \quad \mu_p^p\lambda_p = \mu_p^p\mu_0 = 1, \quad \mu_q^q\lambda_q = \mu_q^q\mu_0 = 1, \quad \mu_r^r\lambda_r = \mu_r^r\mu_0 = 1.$$

We therefore have the following presentation for $\pi_1(\Sigma_2(K))$, in which generators and relators correspond respectively to the 1– and 2–cells of a cell-complex structure (with a single 0–cell) on a space homotopy equivalent to $\Sigma_2(K)$:

$$\begin{aligned}\pi_1(\Sigma_2(K)) &= \left\langle \mu_0,\mu_p,\mu_q,\mu_r : \begin{array}{l}[\mu_0,\mu_p] = [\mu_0,\mu_q] = [\mu_0,\mu_r] = 1,\\ \mu_r\mu_q\mu_p = \mu_p^p\mu_0 = \mu_q^q\mu_0 = \mu_r^r\mu_0 = 1\end{array}\right\rangle \\ &= \left\langle \mu_p,\mu_q,\mu_r : \mu_r\mu_q\mu_p = \mu_p^p\mu_q^{-q} = \mu_p^p\mu_r^{-r} = 1\right\rangle.\end{aligned}$$

Any choice of $x, y, z \in \mathbb{Z}_d$ such that $x + y + z \equiv 0 \bmod d$ will define a character χ via $\mu_p \mapsto x$, $\mu_q \mapsto y$ and $\mu_r \mapsto z$. First suppose that none of x, y and z are equivalent to 0. Then by replacing χ with a nonzero multiple, which does not change the underlying cover, we may assume that $x = 1$.

We now follow the Reidemeister–Schreier algorithm to lift these 0–, 1–, and 2–cells to obtain a 2–complex with the same fundamental group as $\Sigma_2(K)_\chi$. Note that all subscripts are considered mod d. First, lift the single 0–cell to d 0–cells $o_1, \dots, o_d$. Note that μ_p has d lifts $\alpha_1, \dots, \alpha_d$, where α_i is a 1–cell from o_i to o_{i+1}; μ_q has d lifts $\beta_1, \dots, \beta_d$, where β_i is a 1–cell from o_i to o_{i+y}; and μ_r has d lifts $\gamma_1, \dots, \gamma_d$, where γ_i is a 1–cell from o_i to o_{i+z}. We similarly compute the attaching maps of the d lifts of each of the 2–cells. For example, the lifts of the 2–cell corresponding to the relator $\mu_r\mu_q\mu_p$ have attaching maps of the form $\gamma_i\beta_{z+i}\alpha_{y+z+i}$ for $i = 1, \dots, d$. Now contract along $\alpha_2, \dots, \alpha_d$ to obtain a complex with a single 0–cell, $(2d+1)$ 1–cells, and $(3d)$ 2–cells, with a corresponding presentation for $\pi_1(\Sigma_2(K)_\chi)$. Abelianizing gives a presentation for $H_1(\Sigma_2(K)_\chi)$ with generators $a, b_1, \dots, b_d, c_1, \dots, c_d$ and relations $a + b_1 + c_x = 0$; $b_k + c_{x+k-1} = 0$ for $k = 2, \dots, d$; and $(p/d)a = (q/d)(b_1 + \cdots + b_d) = (r/d)(c_1 + \cdots + c_d)$. This simplifies to

$$H_1(\Sigma_2(K)_\chi) = \left\langle a, b_1, \dots, b_d : \frac{p}{d}a = \frac{q}{d}(b_1 + \cdots + b_d) = -\frac{r}{d}(b_1 + \cdots + b_d + a)\right\rangle.$$

So as a $\mathbb{Q}$–module, $H_1(\Sigma_2(K)_\chi, \mathbb{Q})$ has generators $b_1, \dots, b_d$ and single relation $(pq + pr + qr)(b_1 + \cdots + b_d) = 0$. Note that the covering transformation of $\Sigma_2(K)_\chi$ sends b_i onto b_{i+1} for $i = 1, \dots, d-1$, and we have that $H_1(\Sigma_2(K)_\chi, \mathbb{Q})$ is a cyclic $\mathbb{Q}[\mathbb{Z}_d]$–module with generator b_1 and relator $(pq+pr+qr)(1+t+t^2+\cdots+t^{d-1})b_1$. Since $1 + \xi_d + \xi_d^2 + \cdots + \xi_d^{d-1} = 0$, we have

$$H_1^\chi(\Sigma_2(K)) = H_1(\Sigma_2(K)_\chi, \mathbb{Q}) \otimes_{\mathbb{Q}[\mathbb{Z}_d]} \mathbb{Q}(\xi_d) \cong \mathbb{Q}(\xi_d).$$

When one of x, y and z is 0, an extremely similar argument shows that $\Sigma_2(K)_\chi$ is a rational homology sphere and so $\dim H_1^\chi(\Sigma_2(K)) = 0$. □

By considering the linking matrix A for L_0 with its entries taken mod d, we see that $H_1(\Sigma_2(K), \mathbb{Z}_d)$ is generated as a $\mathbb{Z}_d$–module by the images of μ_p, μ_q and μ_r (which we continue to refer to as μ_p, μ_q and μ_r by a mild abuse of notation) and has single relation $\mu_p + \mu_q + \mu_r = 0$. Suppose that $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ sends μ_p to a, μ_q to b and μ_r to c, where $0 < a, b, c < d$. We must have $\chi(\mu_0) \equiv 0$ and $a + b + c \equiv 0 \bmod d$. We will use Proposition 2.7 to compute $\sigma_1(\Sigma_2(K), \chi)$, letting L_χ be the distant union of $T(a, pa)$, $T(b, qb)$ and $T(c, rc)$, each with all strands coherently oriented, along with two incoherently oriented linking 0 strands parallel to λ_0, as in Figure 2.

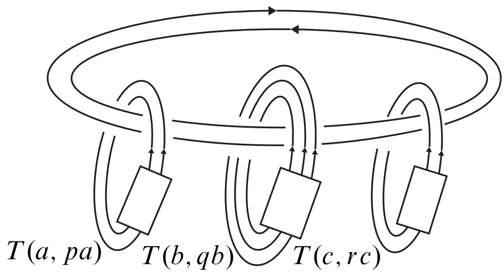


Figure 2: The link L_χ, pictured with $a = 2$, $b = 3$ and $c = 2$

Note, as computed in Example 3.1, $\sigma_{L_\chi}(\omega) = \sigma_{T(a,pa)}(\omega) + \sigma_{T(b,qb)}(\omega) + \sigma_{T(c,rc)}(\omega)$. Also, Litherland's formula in [17] for the Tristram–Levine signature of a torus link implies that $\sigma_{T(j,jkn)}(e^{2\pi i/n}) = -2j(j-1)k$ for $0 < j < n$. While Litherland's result is stated only for torus knots, it holds for torus links as well. In particular, the underlying computation in [1] of the signature of the Brieskorn manifold $V(p,q,r)_\delta = \{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1^p + z_2^q + z_3^r = \delta\} \cap \mathbb{D}^6$ does not depend on any relative primeness of the parameters p, q and r.

Therefore, we have that

$$\begin{aligned}
&\sigma_1(\Sigma_2(K), \chi) \\
&\qquad = 0 - \sigma_{L_\chi}(\omega) - 2(a^2 p + b^2 q + c^2 r)\Big(\frac{d-1}{d^2}\Big) \\
&\qquad = -\sigma_{T(a,pa)}(\omega) - \sigma_{T(b,qb)}(\omega) - \sigma_{T(c,rc)}(\omega) - 2(a^2 p + b^2 q + c^2 r)\Big(\frac{d-1}{d^2}\Big) \\
&\qquad = 2a(a-1)\frac{p}{d} + 2b(b-1)\frac{q}{d} + 2c(c-1)\frac{r}{d} - 2(a^2 p + b^2 q + c^2 r)\Big(\frac{d-1}{d^2}\Big) \\
&\qquad = \frac{2}{d^2}(a(d-a)p + b(d-b)q + c(d-c)r).
\end{aligned}$$

Unfortunately, we cannot conclude that $|\sigma_1(\Sigma_2(K), \chi)| > 1$ for all such choices of χ. For example, when $K = P(3\cdot 7, 5\cdot 7, -17\cdot 7)$, $d = 7$, and χ sends μ_p to 2, μ_q to 4 and μ_r to 1, we have $|\sigma_1(\Sigma_2(K), \chi)| = \frac{8}{11}$. However, this choice of χ does not vanish on any metabolizer for the linking form $\lambda\colon H_1(\Sigma_2(K)) \times H_1(\Sigma_2(K)) \to \mathbb{Q}/\mathbb{Z}$, and so there is still some hope to obstruct the sliceness of K via double branched cover Casson–Gordon signatures.

Lemma 4.5 *Suppose M is a metabolizer for the linking form on $H_1(\Sigma_2(K))$ with nonzero image in $H_1(\Sigma_2(K), \mathbb{Z}_d)$. If $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ vanishes on M and takes μ_p, μ_q and μ_r to nonzero elements of $\mathbb{Z}_d$, then $\sigma_1(\Sigma_2(K), \chi)$ is an integer that is divisible by 4.*

Proof For convenience, we write $p = dp'$, $q = dq'$ and $r = dr'$. Note we have assumed that M has nontrivial image in $H_1(\Sigma_2(K), \mathbb{Z}_d)$, and hence we can assume there is $\alpha = x\mu_p + y\mu_q \in M$ such that not both of x and y are equivalent to 0 mod d.

The linking form is given with respect to our μ_0, μ_p, μ_q, μ_r generating set for $H_1(\Sigma_2(K))$ by $-A^{-1}$ (Gordon and Litherland). Direct computation shows that $\lambda(x\mu_p + y\mu_q, x\mu_p + y\mu_q) = (1/D^2)((q+r)x^2 - 2rxy + (p+r)y^2)$. Since $\alpha \in M$, we know D^2 and hence d^2 divides $(q+r)x^2 - 2rxy + (p+r)y^2$, and so we have

$$(*) \qquad (q' + r')x^2 - 2r'xy + (p' + r')y^2 \equiv 0 \bmod d.$$

Now, let $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ be a character vanishing on M. As usual, we write $a = \chi(\mu_p)$, $b = \chi(\mu_q)$ and $c = \chi(\mu_r)$, with $a + b + c \equiv 0 \bmod d$. Since $\chi(\alpha) = ax + by \equiv 0 \bmod d$, we can write $y = -a\overline{b}x$, and so neither x nor y is equivalent to 0 mod d. Substituting into $(*)$, we obtain

$$\begin{aligned} 0 &\equiv (q' + r')x^2 - 2r'xy + (p' + r')y^2 \\ &\equiv (q' + r')x^2 + 2r'a\overline{b}x^2 + (p' + r')a^2\overline{b}^2x^2 \\ &\equiv [a^2\overline{b}^2p' + q' + (a\overline{b} + 1)^2r']x^2 \quad \bmod d. \end{aligned}$$

Multiplying through by (b^2/x^2) and recalling that $c^2 \equiv (a+b)^2 \bmod d$ gives us that $a^2p' + b^2q' + c^2r' \equiv 0 \bmod d$. Finally, we can write

$$\begin{aligned} \frac{d^2}{2}\sigma_1(\Sigma_2(K), \chi) &= a(d-a)p + b(d-b)q + c(d-c)r \\ &= d(a(d-a)p' + b(d-b)q' + c(d-c)r') \\ &= d^2(p' + q' + r') - d(a^2p' + b^2q' + c^2r'). \end{aligned}$$

Observe that the right side is divisible by d^2, and hence $\sigma_1(\Sigma_2(K))$ is an integer. Also, since d is odd, $a(d-a)p + b(d-b)q + c(d-c)r$ is even for any choice of a, b and c and $\sigma_1(\Sigma_2(K), \chi)$ is divisible by 4. □

Proposition 4.6 (**Case 3**) *Let $K = P(p, q, r)$, with $p, q \neq -r$ and suppose that d is a prime dividing all of p, q and r. Then the Casson–Gordon signatures of $\Sigma_2(K)$ associated to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K.*

Proof Suppose that K is CG-slice at d, for a contradiction. So there exists a metabolizer $M \leq H_1(\Sigma_2(K))$ such that any character χ_0 of prime-power order that vanishes on M has $|\sigma_1(\Sigma_2(K), k\chi_0)| \leq \dim H_1^{\chi}(\Sigma_2(K)) + 1$ for all $0 < k < d$. If there exists χ to $\mathbb{Z}_d$ vanishing on M that takes any of μ_p, μ_q and μ_r to 0, then $\Sigma_2(K)_\chi$ is a rational homology sphere and arguments as in Cases 1 and 2 show that there is some k such that $|\sigma_1(\Sigma_2(K), k\chi)| > 1$.

So we can now assume that no such χ exists. In particular, this implies that the image of M in $H_1(\Sigma_2(K), \mathbb{Z}_d)$ is nontrivial. So let $\chi_0\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ be a nontrivial character vanishing on M and taking none of μ_p, μ_q and μ_r to 0. Since K is CG-slice, Corollary 2.4 and Lemma 4.4 combine to give us that $|\sigma_1(\Sigma_2(K), k\chi_0)| \leq 2$ for all k. Lemma 4.5 gives us that $\sigma_1(K, k\chi_0)$ is an integer divisible by 4 and so $\sigma_1(\Sigma_2(K), k\chi_0) = 0$.

Now, let χ be a multiple of χ_0 such that $\chi(\mu_p) = 1$ and $\chi(\mu_q) = b$, and so $\chi(\mu_r) = d - b - 1$. We therefore have

$$0 = \frac{d^2}{2}\sigma_1(K, \chi) = (d-1)p + b(d-b)q + (b+1)(d-b-1)r. \tag{1}$$

We split into cases depending on the value of b.

Case I $(\mathbf{0 < b < (d-1)/2})$ In this case, we have $(2\chi)(\mu_p) = 2$, $(2\chi)(\mu_q) = 2b$, and $(2\chi)(\mu_r) = d - 2b - 2$, so

$$0 = \frac{d^2}{2}(\sigma_1(K, 2\chi)) = 2(d-2)p + 2b(d-2b)q + (2b+2)(d-2b-2)r. \tag{2}$$

We then have that

$$\begin{aligned} \frac{1}{2}(2\,\mathrm{eq}(1) - \mathrm{eq}(2)) &= p + b^2 q + (b+1)^2 r = 0, \\ \frac{1}{2d}(4\,\mathrm{eq}(1) - \mathrm{eq}(2)) &= p + bq + (b+1)r = 0. \end{aligned}$$

It follows that $(b+1)r = -(b-1)q$ and finally that $p + q = 0$, which is our desired contradiction.

Case II $(\mathbf{b = (d-1)/2})$ In this case, (1) simplifies to show that $q + r = -4p/(d+1)$. Also, $(2\chi)(\mu_p) = 2$ and $(2\chi)(\mu_q) = (2\chi)(\mu_r) = d - 1$, so

$$0 = 2(d-2)p + (d-1)q + (d-1)r. \tag{3}$$

Substituting our expression for $q + r$ into (3), we obtain that $(d^2 - 3d)p = 0$, and so $d = 3$. But this implies that $q + r = -p$, and hence that p is even, which is our desired contradiction.

Case III $(\mathbf{d/2 < b < d})$ In this case, we have $(2\chi)(\mu_p) = 2$, $(2\chi)(\mu_q) = 2b - d$ and $(2\chi)(\mu_r) = 2d - 2b - 2$. Therefore

$$\begin{aligned} 0 &= \frac{d^2}{2}(\sigma_1(K, 2\chi)) \\ &= 2(d-2)p + (2b-d)(2d-2b)q + (2b-d+2)(2d-2b-2)r. \end{aligned} \tag{4}$$

We then have that

$$\frac{1}{2}(2\,\mathrm{eq}(1)-\mathrm{eq}(4)) = p+(d-b)^2q+(d-b-1)^2r = 0,$$
$$\frac{1}{2d}(4\,\mathrm{eq}(1)-\mathrm{eq}(4)) = p+(d-b)q+(d-b-1)r = 0$$

It follows that $(d-b)q = -(d-b-2)r$, and finally that $p+r=0$, which is our desired contradiction. □

4.3 Cases 4, 5 and 6: d divides $pq+pr+qr$ but not any of p, q, r

The link L_0 considered as a 4–colored link has identically 0 colored signature, since it is a connected sum of 2–colored Hopf links. Note that since d divides none of p, q and r, every nontrivial character χ to $\mathbb{Z}_d$ has all of $\chi(\mu_p)$, $\chi(\mu_q)$, $\chi(\mu_r)$ and $\chi(\mu_0)$ nonzero, and so Theorem 3.2 applies and we have the following simple formula for $\sigma_1(\Sigma_2(K),\chi)$.

Proposition 4.7 *Let $K=P(p,q,r)$ and suppose $\chi\colon H_1(\Sigma_2(K))\to\mathbb{Z}_d$ has $\chi(\mu_p)$, $\chi(\mu_q)$, $\chi(\mu_r)$ and $\chi(\mu_0)$ all nonzero. Let a, b, c and ϵ be the unique lifts of $\chi(\mu_p)$, $\chi(\mu_q)$, $\chi(\mu_r)$ and $\chi(\mu_0)$ to $\{1,\dots,d-1\}$. Then*

$$\sigma_1(\Sigma_2(K),\chi) = 3-\frac{2}{d^2}f(\chi),$$

where $f(\chi) := (d-\epsilon)(a+b+c)+(d-a)(ap+\epsilon)+(d-b)(bq+\epsilon)+(d-c)(cr+\epsilon)$.

Remark 4.8 Note that the parity of $a+b+c$ and of ϵ together determine the parity of $f(\chi)$; in particular, if $a+b+c$ is odd then ϵ and $f(\chi)$ have opposite parities. Also, when $a+b+c=d$ we have that

$$f(\chi) = d^2+d\epsilon+a(d-a)p+b(d-b)q+(a+b)(d-(a+b))r.$$

Lemma 4.9 *Let $\chi\colon H_1(\Sigma_2(K))\to\mathbb{Z}_d$, where d divides none of p, q and r. Then $f(\chi)$ is divisible by d^2.*

Proof First, recall that $H_1(\Sigma_2(K))$ is presented by linking matrix A, and so our a, b, c and ϵ values must satisfy

$$a+b+c \equiv ap+\epsilon \equiv bq+\epsilon \equiv cr+\epsilon \equiv 0 \mod d.$$

We can rewrite $f(\chi)$ as

$$f(\chi) = d\big[(a+b+c)+(ap+\epsilon)+(bq+\epsilon)+(cr+\epsilon)\big] - \big[\epsilon(a+b+c)+a(ap+\epsilon)+b(bq+\epsilon)+c(cr+\epsilon)\big].$$

The first term can immediately be seen to be divisible by d^2, and so it suffices to show that $g(\chi)=\epsilon(a+b+c)+a(ap+\epsilon)+b(bq+\epsilon)+c(cr+\epsilon)$ is also divisible by d^2. Writing $ap+\epsilon=k_1d$, $bq+\epsilon=k_2d$ and $cr+\epsilon=k_3d$ for $k_1,k_2,k_3\in\mathbb{Z}$, we have

$$\begin{aligned}g(\chi)&=a(ap+\epsilon+\epsilon)+b(bq+\epsilon+\epsilon)+c(cr+\epsilon+\epsilon)\\&=\frac{k_1d-\epsilon}{p}(k_1d+\epsilon)+\frac{k_2d-\epsilon}{q}(k_2d+\epsilon)+\frac{k_3d-\epsilon}{r}(k_3d+\epsilon)\\&=\frac{k_1^2d^2-\epsilon^2}{p}+\frac{k_2^2d^2-\epsilon^2}{q}+\frac{k_3^2d^2-\epsilon^2}{r}.\end{aligned}$$

Note that since d is relatively prime to all of p, q and r, we can multiply through by pqr without changing the divisibility of $g(\chi)$ by d^2. We therefore have the desired result, since

$$\begin{aligned}g(\chi)pqr&=(k_1^2d^2-\epsilon^2)qr+(k_2^2d^2-\epsilon^2)pr+(k_3^2d^2-\epsilon^2)pq\\&=d^2(k_1^2qr+k_2^2qr+k_3^2pr)-(pq+qr+pr)\epsilon^2.\end{aligned}$$ □

Proposition 4.10 (**Case 4**) *Let $K=P(p,q,r)$ with p, q and r odd, $q\geq p>0$, and $r<0$, and let d be some prime dividing $pq+pr+qr$ which divides none of p, q and r. Suppose also that $r\neq-(4p+q)$ and that $p\not\equiv q \bmod d$. Then the Casson–Gordon signatures of $\Sigma_2(K)$ associated to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K.*

Proof For the sake of contradiction, assume K is CG-slice at d. Since $H_1(\Sigma_2(K),\mathbb{Z}_d)$ is cyclic, for any $\chi\colon H_1(\Sigma_2(K))\to\mathbb{Z}_d$ we must have

$$|\sigma_1(\Sigma_2(K),\chi)|=\left|3-\frac{2}{d^2}f(\chi)\right|\leq 1.$$

Note that the first equality comes from Proposition 4.7 in the above equation. Therefore, by Lemma 4.9 we have $f(\chi)=d^2$ or $2d^2$.

We will work with two characters. Note that our formula for $f(\chi)$ uses the unique integer lifts of $\chi(\mu_i)$ to $\{1,\dots,d-1\}$, so we will be careful to only write $\chi(\mu_i)=x$ if $0<x<d$. We define χ_1 to have $\chi_1(\mu_r)=1$, and $\chi_2=2\chi_1$. It follows that $\chi_1(\mu_0)$ is the unique integer ϵ in $(0,d)$ such that $\epsilon+r\equiv 0 \bmod d$, $\chi_1(\mu_p)$ is the unique integer a in $(0,d)$ such that $\epsilon+ap\equiv 0 \bmod d$, and $\chi_1(\mu_q)=d-a-1$. Note that $\chi_i(\mu_p)+\chi_i(\mu_q)+\chi_i(\mu_r)=d$, so by Remark 4.8, $f(\chi_i)$ has the opposite parity as $\chi_i(\mu_0)$ for $i=1,2$. We now define some convenient notation:

$$\begin{bmatrix}x_1\\x_2\end{bmatrix}_y=\begin{cases}x_1 & \text{if } 0<y<d/2,\\ x_2 & \text{if } d/2<y<d\end{cases}\quad\text{and}\quad\begin{bmatrix}x_1\\x_2\end{bmatrix}_{p(y)}=\begin{cases}x_1 & \text{if } y \text{ is even},\\ x_2 & \text{if } y \text{ is odd}.\end{cases}$$

We therefore have

$$\chi_2(\mu_p)=\begin{bmatrix}2a\\2a-d\end{bmatrix}_a,\quad \chi_2(\mu_q)=\begin{bmatrix}d-2a-2\\2d-2a-2\end{bmatrix}_a,\quad \chi_2(\mu_0)=\begin{bmatrix}2\epsilon\\2\epsilon-d\end{bmatrix}_\epsilon,$$
$$f(\chi_1)=\begin{bmatrix}d^2\\2d^2\end{bmatrix}_{p(\epsilon)}\quad\text{and}\quad f(\chi_2)=\begin{bmatrix}d^2\\2d^2\end{bmatrix}_\epsilon.$$

(Note that if $a=(d-1)/2$, then χ_1 sends both μ_p and μ_q to $(d-1)/2$. But this implies that $p\equiv q \bmod d$, which we have assumed is not the case.)

We thus have the following two equations from our formulas for $f(\chi_1)$ and $f(\chi_2)$:

$$\text{(5)}\quad \begin{bmatrix}0\\d^2\end{bmatrix}_{p(\epsilon)}=d\epsilon+a(d-a)p+(a+1)(d-a-1)q+(d-1)r,$$

$$\text{(6)}\quad \begin{bmatrix}0\\d^2\end{bmatrix}_{\epsilon}=d\epsilon+\begin{bmatrix}a(d-2a)p+(a+1)(d-2a-2)q\\(2a-d)(d-a)p+(2+2a-d)(d-a-1)q\end{bmatrix}_a+(d-2)r.$$

Consider eq(7) $=$ eq(5) $-$ eq(6) and eq(7) $=(1/d)(2\,\text{eq(5)}-\text{eq(6)})$:

$$\text{(7)}\quad \begin{bmatrix}0\\d^2\end{bmatrix}_{p(\epsilon)}-\begin{bmatrix}0\\d^2\end{bmatrix}_{\epsilon}=\begin{bmatrix}a^2p+(a+1)^2q\\(d-a)^2p+(d-a-1)^2q\end{bmatrix}_a+r,$$

$$\text{(8)}\quad \begin{bmatrix}0\\2d\end{bmatrix}_{p(\epsilon)}-\begin{bmatrix}0\\d\end{bmatrix}_{\epsilon}=\epsilon+\begin{bmatrix}ap+(a+1)q\\(d-a)p+(d-a-1)q\end{bmatrix}_a+r.$$

Note that the left side of (8) is even exactly when $\epsilon<d/2$, while the right side has the same parity as ϵ. So we can assume $\epsilon<d/2$ if and only if ϵ is even, and (7) and (8) simplify to the following:

$$\text{(9)}\quad 0=\begin{bmatrix}a^2p+(a+1)^2q\\(d-a)^2p+(d-a-1)^2q\end{bmatrix}_a+r,$$

$$\text{(10)}\quad \begin{bmatrix}0\\d\end{bmatrix}_{\epsilon}=\epsilon+\begin{bmatrix}ap+(a+1)q\\(d-a)p+(d-a-1)q\end{bmatrix}_a+r.$$

We can use (9) to see that if $a<d/2$ then $D=ap+(a+1)q$ and if $a>d/2$ then $D=(d-a)p+(d-a-1)q$. We will now split into cases, and show that each leads to a contradiction by using (9) to write r in terms of a, d, p and q and substituting this expression into (10). Note that since d divides D, we certainly have that $d\le D$.

Case I $(a,\epsilon<d/2)$ By combining (9) and (10) in this case, we see that we have $\epsilon=a^2(p+q)+a(q-p)$, and so

$$2a^2(p+q)<2a^2(p+q)+2a(q-p)=2\epsilon<d\le D=ap+(a+1)q.$$

It follows that $(2a^2-a)p+(2a^2-a-1)q<0$, which gives the desired contradiction.

Case II $(a<d/2<\epsilon)$ In this case we have

$$0<d-\epsilon=-a(a-1)p-a(a+1)q<0.$$

Case III $(\epsilon<d/2<a)$ First, suppose $a=d-2$. Then (9) implies that $r=-(4p+q)$, which we have assumed is not the case. So we can assume that $a<d-2$, and so

$$D=(d-a)p+(d-a-1)q<(d-a)(d-a-1)p+(d-a-1)(d-a-2)q=\epsilon<d.$$

Case IV $(d/2<a,\epsilon)$ As in Case III, we can assume that $a<d-2$, and so

$$0<d-\epsilon=-(d-a)(d-a-1)p-(d-a-1)(d-a-2)q<0. \qquad \square$$

Proposition 4.11 (**Case 5**) *Suppose $K=P(p,q,r)$ for $r=-(4p+q)$. Suppose d is a prime that divides $pq+pr+qr$ but none of p, q and r. Then either $K=P(1,q,-(q+4))$, in which case K is ribbon, or the Casson–Gordon signatures of $\Sigma_2(K)$ corresponding to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K.*

Note that $K=P(1,q,-(q+4))$ is a 2–bridge knot. If we write $q=2k+1$, then K is a generalized twist knot corresponding to the fraction $-(4(k+1)(k+2)+1)/(2(k+1))$ and has been known to be ribbon at least since [3].

Proof Let χ be the character sending μ_p to $d-2$, μ_q and μ_r to 1 and μ_0 to ϵ. Then $\chi'=\frac{1}{2}(d-1)\chi$ sends μ_p to 1, μ_q and μ_r to $\frac{1}{2}(d-1)$ and μ_0 to ϵ'. Arguments as in the proof of Proposition 4.10 show that if $p>1$ then at least one of $|\sigma_1(\Sigma_2(K),\chi)|$ and $|\sigma_1(\Sigma_2(K),\chi')|$ is strictly larger than 1, and hence that K is not CG-slice at d. $\square$

Proposition 4.12 (**Case 6**) *Suppose d divides $pq+pr+qr$ but none of p, q and r, $p\equiv q \bmod d$ and $d\neq 3$. Then the Casson–Gordon signatures of $\Sigma_2(K)$ associated to characters to $\mathbb{Z}_d$ obstruct the topological sliceness of K.*

Proof For $i=1,2$, consider the characters $\chi_i\colon H_1(\Sigma_2(K))\to\mathbb{Z}_d$ defined by $\chi_i(\mu_p)=\chi_i(\mu_q)=i$, $\chi_i(\mu_r)=d-2i$ and $\chi_i(\mu_0)=\epsilon_i$. (Note that since $d\neq 3$ we have that $d-2i>0$ for $i=1,2$.) Arguments as in the proof of Proposition 4.10 show that at least one of $|\sigma_1(\Sigma_2(K),\chi_i)|$ is strictly larger than 1, and hence that K is not CG-slice at d. $\square$

Proposition 4.13 *Suppose that $K=P(p,q,r)$ has p, q and r relatively prime, $|H_1(\Sigma_2(K)|=|pq+pr+qr|=3^{2n}$ for some $n\in\mathbb{N}$, and $p\equiv q \bmod 3$. Then either K is ribbon or the Casson–Gordon signatures associated to characters of order 3 and 9 obstruct the topological sliceness of K.*

Proof First, suppose that $n \geq 2$. Since p, q and r are pairwise relatively prime, $H_1(\Sigma_2(K))$ is cyclic, and any character to $\mathbb{Z}_{3^n}$ will vanish on the unique metabolizer for the linking form. Proposition 2.5 implies that the associated covers are rational homology spheres, and so it suffices to find such a character χ with $|\sigma_1(\Sigma_2(K), \chi)| > 1$. The arguments of Propositions 4.10, 4.11 and 4.12 applied to $d = 9$ (according to whether $r = -(4p+q)$ and whether $p \equiv q \bmod 9$) show that this is the case.

Now suppose that $n = 1$ and so $pq + pr + pq = -9$ and $r = -(pq+9)/(p+q)$. A slight variation on our usual arithmetic arguments then implies that $\sigma_1(\Sigma_2(K), \chi) < -1$ for some $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_3$, and hence that K is not CG-slice at $d = 3$. □

5 Topological sliceness of even 3–strand pretzel knots.

We now outline the proof of our argument that all topologically slice even 3–strand pretzel knots are either ribbon or in Lecuona's family $\{\pm P_a\}$, leaving the details of arithmetic to the reader.

Theorem 5.1 *Let K be an even 3–strand pretzel knot. Suppose that K is topologically slice. Then, up to reflection, either $K = P(p, -p, q)$ for some $p, q \in \mathbb{N}$ (and K is ribbon) or $K = P_a = P(a, -a-2, -(a+1)^2/2)$ for some $a \equiv 1, 11, 37, 47, 59 \bmod 60$.*

Proof Suppose that K is an algebraically slice even 3–strand pretzel. First, note that by Jabuka's computation of the rational Witt classes of pretzel knots, we can assume that either $K = P(p, -p, q)$ for some odd p and even q or $K = P(-p, p \pm 2, q)$ for some odd p and even q such that $\det(K) = \pm 2q - p^2 \mp 2p = m^2 > 0$ [13, Theorem 1.11]. In the first case K is ribbon, and so we assume that we are in the second case. By the symmetries of 3–strand pretzel knots, we can also assume that up to reflection $K = P(-p, p+2, q)$ for $p \in \mathbb{N}$. Then our condition that $\det(K) = 2q - p^2 - 2p > 0$ implies that $q > 0$ as well.

First, observe that if $\det(K) = 1$ then $q = (p+1)^2/2$ and up to reflection K is an element of Lecuona's family $\{P_a\}$. For $a \not\equiv 1, 11, 37, 47, 49, 59 \bmod 60$, Theorem 4.5 of [15] states that K is not algebraically slice. When $a \equiv 49 \bmod 60$, an argument analogous to the proof of [15, Theorem 4.5] shows that $\Delta_K(t)$ does not have a Fox–Milnor factorization and hence that K is not algebraically slice. (In particular, note that since $a \equiv 49 \bmod 60$ we have that 5 divides $(a+1)^2/4$ and 3 divides $a+2$. Working mod 5, we have $\Delta_{P_a}(t) \equiv \Pi_{1 \neq d | a} \Phi_d(t) \Pi_{1 \neq d | a+2} \Phi_d(t)$, where $\Phi_d(t)$ denotes the d^{th} cyclotomic polynomial. Since $\Phi_3(t)$ is symmetric, irreducible mod 5, and relatively prime to each $\Phi_d(t)$ for $d \neq 3$ dividing a or $a+2$, the desired result follows.)

So we can assume that $\det(K) = m^2 > 1$, and in particular that there is an (odd) prime d dividing $\det(K)$. Arguments as in the proof of Proposition 4.1 show that $\Sigma_2(K)$ has a surgery presentation with underlying link the coherently oriented torus link $-T(2,2p)$ and linking matrix $\left[\begin{smallmatrix} 2 & -p \\ -p & q-p \end{smallmatrix}\right]$. It follows that $H_1(\Sigma_2(K))$ is cyclic, and hence that $H_1(\Sigma_2(K), \mathbb{Z}_d)$ is certainly cyclic as well. It therefore suffices to show that there is a single $\chi\colon H_1(\Sigma_2(K)) \to \mathbb{Z}_d$ with $|\sigma_k(\Sigma_2(K),\chi)| > 1$ for some $1 \le k < d$.

The construction of χ and computation of the corresponding Casson–Gordon signatures is extremely similar to the arguments of Section 4, and therefore we only list the cases one must consider and leave the verification of the details to the interested reader. It is convenient to consider six cases, according to the values mod d of the parameters of K: $-p \equiv q \equiv 0$; $p+2 \equiv q \equiv 0$; $-p \equiv 2q \not\equiv 0$; $p+2 \equiv 2q \not\equiv 0$; $-p \equiv p+2 \not\equiv 0$; and $-p$, $p+2$ and q are mutually distinct and nonzero. □

Acknowledgements I would like to thank Tye Lidman for suggesting Corollary 4.2, Lisa Piccirillo for her help with manipulating surgery descriptions, and my advisor Cameron Gordon for his advice, support and helpful comments. I would also like to thank the referee for thoughtful suggestions.

References

[1] **E Brieskorn**, *Beispiele zur Differentialtopologie von Singularitäten*, Invent. Math. 2 (1966) 1–14 MR

[2] **G Burde**, **H Zieschang**, *Knots*, 2nd edition, de Gruyter Studies in Math. 5, de Gruyter, Berlin (2003) MR

[3] **A J Casson**, **C M Gordon**, *On slice knots in dimension three*, from "Algebraic and geometric topology, 2" (R J Milgram, editor), Proc. Sympos. Pure Math. XXXII, Amer. Math. Soc., Providence, RI (1978) 39–53 MR

[4] **A J Casson**, **C M Gordon**, *Cobordism of classical knots*, from "À la recherche de la topologie perdue" (L Guillou, A Marin, editors), Progr. Math. 62, Birkhäuser, Boston (1986) 181–199 MR

[5] **D Cimasoni**, **V Florens**, *Generalized Seifert surfaces and signatures of colored links*, Trans. Amer. Math. Soc. 360 (2008) 1223–1264 MR

[6] **A Degtyarev**, **V Florens**, **A G Lecuona**, *The signature of a splice*, Int. Math. Res. Not. 2017 (2017) 2249–2283 MR

[7] **P Feller**, **D McCoy**, *On 2–bridge knots with differing smooth and topological slice genera*, Proc. Amer. Math. Soc. 144 (2016) 5435–5442 MR

[8] **R Fintushel**, **R J Stern**, *Pseudofree orbifolds*, Ann. of Math. 122 (1985) 335–364 MR

[9] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR

[10] **P M Gilmer**, *Configurations of surfaces in 4–manifolds*, Trans. Amer. Math. Soc. 264 (1981) 353–380 MR

[11] **J Greene**, **S Jabuka**, *The slice-ribbon conjecture for 3–stranded pretzel knots*, Amer. J. Math. 133 (2011) 555–580 MR

[12] **C Herald**, **P Kirk**, **C Livingston**, *Metabelian representations, twisted Alexander polynomials, knot slicing, and mutation*, Math. Z. 265 (2010) 925–949 MR

[13] **S Jabuka**, *Rational Witt classes of pretzel knots*, Osaka J. Math. 47 (2010) 977–1027 MR

[14] **R Kirby**, *Problems in low dimensional manifold theory*, from "Algebraic and geometric topology, 2" (R J Milgram, editor), Proc. Sympos. Pure Math. XXXII, Amer. Math. Soc., Providence, RI (1978) 273–312 MR

[15] **A G Lecuona**, *On the slice-ribbon conjecture for pretzel knots*, Algebr. Geom. Topol. 15 (2015) 2133–2173 MR

[16] **P Lisca**, *Lens spaces, rational balls and the ribbon conjecture*, Geom. Topol. 11 (2007) 429–472 MR

[17] **R A Litherland**, *Signatures of iterated torus knots*, from "Topology of low-dimensional manifolds" (R A Fenn, editor), Lecture Notes in Math. 722, Springer (1979) 71–84 MR

[18] **A Stoimenow**, *Knots of genus one or on the number of alternating knots of given genus*, Proc. Amer. Math. Soc. 129 (2001) 2141–2156 MR

Department of Mathematics, University of Texas
Austin, TX, United States

amiller@math.utexas.edu

http://www.ma.utexas.edu/users/amiller/

Received: 1 November 2016 Revised: 15 April 2017

msp

Algebraic & Geometric Topology 17 (2017) 3081–3094

An index obstruction to positive scalar curvature on fiber bundles over aspherical manifolds

RUDOLF ZEIDLER

We exhibit geometric situations where higher indices of the spinor Dirac operator on a spin manifold N are obstructions to positive scalar curvature on an ambient manifold M that contains N as a submanifold. In the main result of this note, we show that the Rosenberg index of N is an obstruction to positive scalar curvature on M if $N \hookrightarrow M \twoheadrightarrow B$ is a fiber bundle of spin manifolds with B aspherical and $\pi_1(B)$ of finite asymptotic dimension. The proof is based on a new variant of the multipartitioned manifold index theorem which might be of independent interest. Moreover, we present an analogous statement for codimension-one submanifolds. We also discuss some elementary obstructions using the $\hat{\mathrm{A}}$-genus of certain submanifolds.

58J22; 46L80, 53C23

1 Introduction

We consider the following setup:

Geometric Setup 1.1 Let M be a closed connected spin m–manifold and $N \subseteq M$ a closed connected submanifold of codimension q with trivial normal bundle. Moreover, we denote the fundamental groups of M and N by Γ and Λ, respectively, and let $j\colon \Lambda \to \Gamma$ be the map induced by the inclusion $\iota\colon N \hookrightarrow M$.

Hanke, Pape and Schick [6] have found that if the codimension q is 2 and some assumptions on homotopy groups hold, then the Rosenberg index of N is an obstruction to positive scalar curvature on M. Motivated by this result, it is an interesting endeavor to find further situations where the Rosenberg index of N is an obstruction to positive scalar curvature on the ambient manifold M. In this note, we exhibit certain cases where it is possible to relax the restrictions on the codimension.

Recall the *Rosenberg index* $\alpha^\Gamma(M) \in \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$ of a closed spin manifold M, where $\Gamma = \pi_1(M)$ and $\mathrm{C}^*_\epsilon\Gamma$ denotes the maximal ($\epsilon = \max$) or reduced ($\epsilon = \mathrm{red}$) group C^*-algebra. Abstractly, it is obtained by applying the Baum–Connes assembly map

$$\mu\colon \mathrm{K}_*(\mathrm{B}\Gamma) \to \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma),$$

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3081

to the image of the K-homological fundamental class of M under the map $u\colon M \to \mathrm{B}\Gamma$ that classifies the universal covering. The (*maximal, if* $\epsilon = \max$) *strong Novikov conjecture* predicts that $\mu \otimes \mathbb{Q}$ is injective.

All statements in the introduction are made under implicit assumption of Geometric Setup 1.1. We start by recalling the precise statement of Hanke, Pape and Schick's codimension-two obstruction.

Theorem 1.2 [6, Theorem 1.1] *Let $\epsilon \in \{\mathrm{red}, \max\}$. Let N have codimension $q = 2$ and suppose that $j\colon \Lambda \to \Gamma$ is injective and that $\pi_2(M) = 0$. If $\alpha^\Lambda(N) \neq 0 \in \mathrm{K}_{m-2}(\mathrm{C}^*_\epsilon \Lambda)$, then M does not admit a metric of positive scalar curvature.*

Remark 1.3 The theorem was proved by applying methods from Roe's coarse index theory to a manifold that arises out of a modification of a certain covering of M. As discussed in [6, Section 3], this proof only shows that M does not admit positive scalar curvature and does not give $\alpha^\Gamma(M) \neq 0$. However, the theorem actually implies that M stably does not admit positive scalar curvature and hence nonvanishing of $\alpha^\Gamma(M)$ would be a consequence of the stable Gromov–Lawson–Rosenberg conjecture (at least if we worked with real K-theory throughout). It is an open question whether it is possible to prove nonvanishing of $\alpha^\Gamma(M)$ directly under the hypotheses of Theorem 1.2.

1.1 Obstructions on fiber bundles and codimension one

Hanke, Pape and Schick state the following application of Theorem 1.2 to fiber bundles:

Corollary 1.4 [6, Corollary 4.5] *Let $\epsilon \in \{\mathrm{red}, \max\}$. Suppose that $N \hookrightarrow M \twoheadrightarrow \Sigma$ is a fiber bundle, where $\pi_2(N) = 0$ and Σ is a closed surface different from S^2, $\mathbb{RP}^2$. If $\alpha^\Lambda(N) \neq 0 \in \mathrm{K}_{m-2}(\mathrm{C}^*_\epsilon \Lambda)$, then M does not admit a metric of positive scalar curvature.*

In this special case it turns out that we can settle the question from Remark 1.3 by a more direct argument. Indeed, in the following main result of this note, we generalize Corollary 1.4 to base manifolds of arbitrary dimension and obtain the stronger conclusion that $\alpha^\Gamma(M)$ is nonvanishing:

Theorem 1.5 *Let $\epsilon \in \{\mathrm{red}, \max\}$. Suppose that $N \overset{\iota}{\hookrightarrow} M \overset{\pi}{\twoheadrightarrow} B$ is a fiber bundle, where B is aspherical and $\pi_1(B) = \Gamma/\Lambda$ has finite asymptotic dimension. If $\alpha^\Lambda(N) \neq 0 \in \mathrm{K}_{m-q}(\mathrm{C}^*_\epsilon \Lambda)$, then $\alpha^\Gamma(M) \neq 0 \in \mathrm{K}_m(\mathrm{C}^*_\epsilon \Gamma)$. In particular, M does not admit positive scalar curvature in this case.*

In the proof we also employ coarse index theory. More specifically, we apply the multipartitioned manifold index theorem. Although variants of it have been established previously by Siegel [13] and Schick and Zadeh [11], neither of these references provides the theorem in sufficient generality for our purposes. Thus, in Section 2, we have included a concise proof of the required result, which might be of independent interest (see Theorem 2.7).

Unlike Theorem 1.2, in the proof of Theorem 1.5 we apply the q–multipartitioned manifold index theorem directly to a covering of M (without modifying it further) and thereby obtain the stronger conclusion that $\alpha^{\Gamma}(M) \neq 0$. If B is a surface or, more generally, admits nonpositive sectional curvature, then the fact that a suitable covering of M is q–partitioned follows from the Cartan–Hadamard theorem applied to B. To obtain the level of generality as stated, we apply a result of Dranishnikov [1, Theorem 3.5] which says that an aspherical manifold with a fundamental group of finite asymptotic dimension has a stably hypereuclidean universal covering.

Remark 1.6 Unlike Corollary 1.4, the condition $\pi_2(N) = 0$ is not required by Theorem 1.5. This, however, is not just a feature of our method: in fact, a careful reading of the proof from [6] reveals that in Theorem 1.2 the hypothesis $\pi_2(M) = 0$ can be weakened to surjectivity of the map $\pi_2(N) \to \pi_2(M)$.

Moreover, the idea for Theorem 1.5 works even in full generality in codimension one (without assumptions on higher homotopy groups or on being a fiber bundle):

Theorem 1.7 *Let $\epsilon \in \{\mathrm{red}, \max\}$. Let N have codimension $q = 1$ and suppose that $j\colon \Lambda \to \Gamma$ is injective. If $\alpha^{\Lambda}(N) \neq 0 \in \mathrm{K}_{m-1}(\mathrm{C}^*_{\epsilon}\Lambda)$, then $\alpha^{\Gamma}(M) \neq 0 \in \mathrm{K}_m(\mathrm{C}^*_{\epsilon}\Gamma)$. In particular, M does not admit positive scalar curvature in this case.*

Remark 1.8 In the proofs of Theorems 1.5 and 1.7, a homomorphism $\Psi\colon \mathrm{K}_*(\mathrm{C}^*_{\epsilon}\Gamma) \to \mathrm{K}_{*-q}(\mathrm{C}^*_{\epsilon}\Lambda)$ with $\Psi(\alpha^{\Gamma}(M)) = \alpha^{\Lambda}(N)$ is constructed, which might be of independent interest.

1.2 Higher $\widehat{\mathrm{A}}$ obstructions via submanifolds

In addition to our result on fiber bundles, we have some obstructions via the $\widehat{\mathrm{A}}$-genus of submanifolds of arbitrary codimension under some restriction on the homotopy groups. In contrast to the results above, the proofs of the results below do not employ coarse index theory and essentially only rely on elementary techniques from (co)homology theory.

First we state a result that applies to intersections of codimension-two submanifolds. We continue to work in Geometric Setup 1.1.

Theorem 1.9 *Let $N = N_1 \cap \cdots \cap N_k$, where $N_1, \ldots, N_k \subseteq M$ are closed submanifolds that intersect mutually transversely and have trivial normal bundles. Suppose that the codimension of N_i is at most two for all $i \in \{1, \ldots, k\}$ and that $\pi_2(N) \to \pi_2(M)$ is surjective.*

If $\widehat{\mathrm{A}}(N) \neq 0$, then $\alpha^\Gamma(M) \neq 0 \in \mathrm{K}_*(\mathrm{C}^*_{\max}\Gamma)$. In particular, M does not admit a metric of positive scalar curvature in this case.

In particular, specializing to a single codimension-two submanifold, this settles the question of Remark 1.3 in the case when $\widehat{\mathrm{A}}(N) \neq 0$ (which implies $\alpha^\Lambda(N) \neq 0$).

The proof of this theorem (see Section 3) proceeds as follows: First we show that the surjectivity of $\pi_2(N) \to \pi_2(M)$ allows us to rewrite $\widehat{\mathrm{A}}(N)$ as a higher $\widehat{\mathrm{A}}$-genus of M. Afterwards we appeal to a result of Hanke and Schick [7, Theorem 1.2] about the maximal strong Novikov conjecture in low cohomological degrees and conclude that $\alpha^\Gamma(M) \neq 0 \in \mathrm{K}_*(\mathrm{C}^*_{\max}\Gamma)$. If we allow higher codimensions for the submanifolds N_i, our method still works but we are no longer in a position to apply [7, Theorem 1.2] and hence need to assume the strong Novikov conjecture:

Theorem 1.10 *Let $\epsilon \in \{\mathrm{red}, \max\}$. Let $N = N_1 \cap \cdots \cap N_k$, where $N_1, \ldots, N_k \subseteq M$ are closed submanifolds that intersect mutually transversely and have trivial normal bundles. Let d be the maximum of the codimensions of N_i over all $i \in \{1, \ldots, k\}$ and suppose that $\pi_j(M) = 0$ for $2 \leqslant j \leqslant d$.*

If $\widehat{\mathrm{A}}(N) \neq 0$ and Γ satisfies the (maximal, if $\epsilon = \max$) strong Novikov conjecture, then $\alpha^\Gamma(M) \neq 0 \in \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$.

Note that the conditions on the homotopy groups are also slightly different than in Theorem 1.9. In fact, in Proposition 3.2, we prove our results under a more general homological condition which includes the restrictions on the homotopy groups from Theorems 1.9 and 1.10 as a special case (see Lemma 3.3).

If we restrict Theorem 1.10 to a single submanifold, we obtain:

Corollary 1.11 *Let $\epsilon \in \{\mathrm{red}, \max\}$. Suppose N has codimension q and $\pi_j(M) = 0$ for $2 \leqslant j \leqslant q$. If $\widehat{\mathrm{A}}(N) \neq 0$ and Γ satisfies the (maximal, if $\epsilon = \max$) strong Novikov conjecture, then $\alpha^\Gamma(M) \neq 0 \in \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$.*

In the special case that Γ is virtually nilpotent (which implies the strong Novikov conjecture), the consequence of Corollary 1.11 that M cannot admit positive scalar curvature was proved by Engel [2, Theorem 4.10] using a different method.

Moreover, under the assumptions of Corollary 1.11, even higher $\widehat{\mathrm{A}}$-genera of N are obstructions to positive scalar curvature on M. This was also discovered by Engel using yet a different method; see [3, Application A].

2 The multipartitioned manifold index theorem

2.1 Coarse index theory

Here we briefly review the relevant aspects of coarse index theory; see [10; 8]. Let $\epsilon \in \{\mathrm{red}, \max\}$ be fixed in this section. Let X be a proper metric space endowed with an isometric, free and proper action of a discrete group Γ. We denote the *Γ–equivariant Roe algebra* of X by $\mathrm{C}^*(X)^\Gamma$. It is defined to be the (spacial if $\epsilon = \mathrm{red}$ or maximal if $\epsilon = \max$) completion of the $*$-algebra of all Γ–equivariant locally compact operators of finite propagation defined over a fixed suitable Hilbert space representation of $\mathcal{C}_0(X)$. Recall the *index map* (or *assembly map*) from locally finite K-homology of the quotient $\Gamma\backslash X$ to the K-theory of the equivariant Roe algebra:

$$\mathrm{Ind}^\Gamma\colon \mathrm{K}_*^{\mathrm{lf}}(\Gamma\backslash X) \to \mathrm{K}_*(\mathrm{C}^*(X)^\Gamma). \tag{1}$$

For an explicit definition of the assembly in the nonequivariant case (also pertaining to $\epsilon = \max$), see for instance [4, Subsection 4.6]. A straightforward generalization of the same formulas to the equivariant case then yields the equivariant assembly map $\mathrm{K}_*^{\mathrm{lf},\Gamma}(X) \to \mathrm{K}_*(\mathrm{C}^*(X)^\Gamma)$. To obtain the map as displayed in (1), we precompose with the induction isomorphism $\mathrm{K}_*^{\mathrm{lf}}(\Gamma\backslash X) \cong \mathrm{K}_*^{\mathrm{lf},\Gamma}(X)$ in analytic K-homology as it is exhibited via the Paschke duality picture in [8, Lemma 12.5.4; 12, Theorem 4.3.25].

If X is a complete spin m–manifold, we may apply the index map to the class $[\not{D}_{\Gamma\backslash X}] \in \mathrm{K}_m^{\mathrm{lf}}(\Gamma\backslash X)$ of the spinor Dirac operator on $\Gamma\backslash X$. We will use the notation $\mathrm{Ind}^\Gamma(\not{D}_X) := \mathrm{Ind}^\Gamma([\not{D}_{\Gamma\backslash X}])$. If $X = \widetilde{M}$ is the universal covering of a closed spin manifold M and $\Gamma = \pi_1(M)$, then there is a canonical isomorphism $\mathrm{K}_*(\mathrm{C}^*(\widetilde{M})^\Gamma) \cong \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$ and $\mathrm{Ind}^\Gamma(\not{D}_{\widetilde{M}})$ recovers the Rosenberg index $\alpha^\Gamma(M)$.

In the following we introduce some notation which will feature in our formulation of the multipartitioned manifold index theorem. Let Γ be a countable discrete group and fix a model for the classifying space $\mathrm{B}\Gamma$ as a locally finite simplicial complex. As usual, we denote its universal covering by $\mathrm{E}\Gamma$.

Definition 2.1 Let Y be a proper metric space and define

$$\Gamma\mathrm{K}_i^{\mathrm{lf}}(Y) := \operatorname*{colim}_Z \mathrm{K}_i^{\mathrm{lf}}(Z),$$
$$\Gamma\mathcal{C}_i(Y) := \operatorname*{colim}_Z \mathrm{K}_i(\mathrm{C}^*(\widetilde{Z})^\Gamma),$$

where the colimits range over *admissible subsets* $Z \subseteq \mathrm{B}\Gamma \times Y$ and Z is called admissible if it is closed and $\mathrm{pr}_2|_Z\colon Z \to Y$ is proper. Moreover, $\widetilde{Z}$ denotes the lift of Z to $\mathrm{E}\Gamma \times Y$.

Roughly speaking, $\Gamma\mathrm{K}_i^{\mathrm{lf}}(Y)$ behaves like locally finite K-homology in Y and like ordinary K-homology in the $\mathrm{B}\Gamma$–slot.

Recall that a map $f\colon (Y,d) \to (Y',d')$ between metric spaces is called *coarse* if $f^{-1}(B')$ is bounded for each bounded set $B' \subseteq Y'$ and there exists a function $\rho\colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ such that $d'(f(x),f(y)) \leqslant \rho(d(x,y))$ for all $x,y \in Y$. Since the K-theory of the (equivariant) Roe algebra is functorial with respect to (equivariant) coarse maps [8, Definition 6.3.15], the group $\Gamma\mathcal{C}_i(Y)$ is functorial in Y with respect to coarse maps.

The index map (1) induces an index map in the limit $\mathrm{Ind}^\Gamma\colon \Gamma\mathrm{K}_*^{\mathrm{lf}}(Y) \to \Gamma\mathcal{C}_*(Y)$ which is natural in Y with respect to continuous coarse maps.

Example 2.2 Taking $Y = \mathrm{pt}$ to be a point, we have $\Gamma\mathrm{K}_*(\mathrm{pt}) = \mathrm{K}_*(\mathrm{B}\Gamma)$ as defined via the K-theory spectrum and $\Gamma\mathcal{C}_*(\mathrm{pt}) \cong \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$. Moreover, the index map $\mathrm{Ind}^\Gamma\colon \Gamma\mathrm{K}_*^{\mathrm{lf}}(\mathrm{pt}) \to \Gamma\mathcal{C}_*(\mathrm{pt})$ recovers the assembly map $\mu\colon \mathrm{K}_*(\mathrm{B}\Gamma) \to \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$ featuring in the strong Novikov conjecture.

The external product in K-homology also induces an external product,

$$\Gamma\mathrm{K}_n^{\mathrm{lf}}(X) \otimes \mathrm{K}_d^{\mathrm{lf}}(Y) \xrightarrow{\times} \Gamma\mathrm{K}_{n+d}^{\mathrm{lf}}(X \times Y).$$

Proposition 2.3 (suspension isomorphism) *Let Y be a proper metric space. There are isomorphisms s and σ which make the diagram*

$$\begin{array}{ccc} \Gamma\mathrm{K}_{*+1}^{\mathrm{lf}}(Y \times \mathbb{R}) & \xrightarrow{\mathrm{Ind}^\Gamma} & \Gamma\mathcal{C}_{*+1}(Y \times \mathbb{R}) \\ {\scriptstyle s}\downarrow\cong & & \cong\downarrow{\scriptstyle\sigma} \\ \Gamma\mathrm{K}_*^{\mathrm{lf}}(Y) & \xrightarrow{\mathrm{Ind}^\Gamma} & \Gamma\mathcal{C}_*(Y) \end{array}$$

commute, and such that $s(x \times [\not{D}_\mathbb{R}]) = x$ for all $x \in \Gamma\mathrm{K}_*^{\mathrm{lf}}(Y)$.

Proof To construct s and σ we use the Mayer–Vietoris boundary maps associated to the cover $Y \times \mathbb{R} = Y \times \mathbb{R}_{\geqslant 0} \cup Y \times \mathbb{R}_{\leqslant 0}$ for K-homology and for the K-theory of the Roe algebra, respectively. Indeed, take an admissible subset $Z \subseteq \mathrm{B}\Gamma \times Y \times \mathbb{R}$ such that the cover

$$(*) \qquad Z = \big(Z \cap (\mathrm{B}\Gamma \times Y \times \mathbb{R}_{\geqslant 0})\big) \cup \big(Z \cap (\mathrm{B}\Gamma \times Y \times \mathbb{R}_{\leqslant 0})\big)$$

is coarsely excisive, so that we have a Mayer–Vietoris sequence both in K-homology and for the K-theory of the Roe algebra; see for example [9]. Let

$$s_Z\colon \mathrm{K}_{*+1}^{\mathrm{lf}}(Z) \xrightarrow{\partial_{\mathrm{MV}}} \mathrm{K}_*^{\mathrm{lf}}(Z \cap (\mathrm{B}\Gamma \times Y \times \{0\})) \to \Gamma\mathrm{K}_*^{\mathrm{lf}}(Y),$$
$$\sigma_Z\colon \mathrm{K}_{*+1}(\mathrm{C}^*(\widetilde{Z})^\Gamma) \xrightarrow{\partial_{\mathrm{MV}}} \mathrm{K}_*(\mathrm{C}^*(\widetilde{Z} \cap (\mathrm{E}\Gamma \times Y \times \{0\}))^\Gamma) \to \Gamma\mathcal{C}_*(Y).$$

The family of those admissible subsets where the cover $(*)$ is coarsely excisive is cofinal in the directed set of all admissible subsets, hence the maps s_Z and σ_Z induce the required maps s and σ in the limit. Moreover, one can verify that the family of admissible Z where s_Z and σ_Z are both defined and an isomorphism is also cofinal in the family of all admissible sets. The isomorphism statement relies on showing that we have a cofinal collection of admissible Z such that $Z \cap (\mathrm{B}\Gamma \times Y \times \mathbb{R}_{\geqslant 0})$ and $Z \cap (\mathrm{B}\Gamma \times Y \times \mathbb{R}_{\leqslant 0})$ are *flasque*. A more detailed version of this argument can be found in [14, Proposition 4.2.3].

Thus s and σ are isomorphisms. Finally, the claim $s(x \times [\not{D}_{\mathbb{R}}]) = x$ for all $x \in \Gamma\mathrm{K}^{\mathrm{lf}}_*(Y)$ is a standard fact in K-homology which follows from $\partial_{\mathrm{MV}}([\not{D}_{\mathbb{R}}]) = 1$ for the Mayer–Vietoris boundary map associated to $\mathbb{R} = \mathbb{R}_{\geqslant 0} \cup \mathbb{R}_{\leqslant 0}$. □

Corollary 2.4 *For every $\varepsilon > 0$, we have*

$$\Gamma\mathrm{K}^{\mathrm{lf}}_*(\mathbb{R}^q) \cong \operatorname*{colim}_{K \subset \mathrm{B}\Gamma} \mathrm{K}^{\mathrm{lf}}_*(K \times \mathbb{R}^q) \overset{\iota^!}{\cong} \operatorname*{colim}_{K \subset \mathrm{B}\Gamma} \mathrm{K}^{\mathrm{lf}}_*(K \times B_\varepsilon(0)),$$

where the colimit ranges over compact subsets $K \subseteq \mathrm{B}\Gamma$ and the second isomorphism is induced by the inclusion of the open ball $\iota\colon B_\varepsilon(0) \hookrightarrow \mathbb{R}^q$.

Proof Since for a compact subset $K \subseteq \mathrm{B}\Gamma$ the set $K \times \mathbb{R}^q$ is admissible, we obtain a canonical map $J\colon \operatorname{colim}_{K \subset \mathrm{B}\Gamma} \mathrm{K}^{\mathrm{lf}}_*(K \times \mathbb{R}^q) \to \Gamma\mathrm{K}^{\mathrm{lf}}_*(\mathbb{R}^q)$. The q–fold iteration of the suspension isomorphism from Proposition 2.3 yields an isomorphism $s^q\colon \Gamma\mathrm{K}^{\mathrm{lf}}_*(\mathbb{R}^q) \cong \mathrm{K}_{*-q}(\mathrm{B}\Gamma)$. An analogous argument as in the proof of Proposition 2.3 produces an isomorphism $t^q\colon \operatorname{colim}_{K \subset \mathrm{B}\Gamma}(K \times \mathbb{R}^q) \cong \mathrm{K}_{*-q}(\mathrm{B}\Gamma)$ such that $t^q = s^q \circ J$. In particular, this shows that J must be an isomorphism.

For each $K \subseteq \mathrm{B}\Gamma$, the restriction $\iota^!\colon \mathrm{K}^{\mathrm{lf}}_*(K \times \mathbb{R}^q) \to \mathrm{K}^{\mathrm{lf}}_*(K \times B_\varepsilon(0))$ is induced by the map on $K \times \mathbb{R}^q$ that is the identity on $K \times B_\varepsilon(0)$ and takes $K \times (\mathbb{R}^q \setminus B_\varepsilon(0))$ to infinity in the one-point compactification of $K \times B_\varepsilon(0)$. Since this map induces a homotopy equivalence between the one-point compactifications, this implies that $\iota^!\colon \mathrm{K}^{\mathrm{lf}}_*(K \times \mathbb{R}^q) \to \mathrm{K}^{\mathrm{lf}}_*(K \times B_\varepsilon(0))$ is an isomorphism. □

Corollary 2.4 implies that classes in $\Gamma\mathrm{K}^{\mathrm{lf}}_*(\mathbb{R}^q)$ (and thus their images in $\Gamma\mathcal{C}_*(\mathbb{R}^q)$) depend only on the restrictions to arbitrarily small open subsets. A very similar localization property was exhibited by Schick and Zadeh [11] and is at the heart of their approach to the multipartitioned manifold index theorem. Analogously, our approach to the theorem in the next subsection crucially relies on the localization property from Corollary 2.4.

2.2 Multipartitioned manifolds

Let $f\colon X \to Y$ be a proper map, $u\colon X \to \mathrm{B}\Gamma$ classifying a covering $p\colon \widetilde{X} \to X$. Then the map $u \times f\colon X \to \mathrm{B}\Gamma \times Y$ induces a map $(u \times f)_*\colon \mathrm{K}^{\mathrm{lf}}_*(X) \to \Gamma\mathrm{K}^{\mathrm{lf}}_*(Y)$. If f is also coarse, then the Γ–equivariant map $\widetilde{u} \times (f \circ p)\colon \widetilde{X} \to \mathrm{E}\Gamma \times Y$ induces a map $(\widetilde{u} \times (f \circ p))_*\colon \mathrm{K}_*(\mathrm{C}^*(\widetilde{X})^\Gamma) \to \Gamma\mathcal{C}_*(Y)$.

Definition 2.5 A complete Riemannian manifold X is called q–multipartitioned by a closed submanifold $M \subseteq X$ via a continuous coarse map $f\colon X \to \mathbb{R}^q$ if f is smooth near $f^{-1}(0)$ and $0 \in \mathbb{R}^q$ is a regular value with $f^{-1}(0) = M$.

Definition 2.6 Let X be a complete spin m–manifold that is q–multipartitioned by $M \subseteq X$ via $f\colon X \to \mathbb{R}^q$. Fix a Γ–covering $p\colon \widetilde{X} \to X$ which is classified by a map $u\colon X \to \mathrm{B}\Gamma$. Consider the lifted map $\widetilde{u}\colon \widetilde{X} \to \mathrm{E}\Gamma$. Then we define the *higher partitioned manifold index* of X to be

$$\alpha^{f,u}_{\mathrm{PM}}(X) := (\widetilde{u} \times (f \circ p))_*(\mathrm{Ind}^\Gamma(\not{D}_{\widetilde{X}})) \in \Gamma\mathcal{C}_m(\mathbb{R}^q).$$

Furthermore, if M is a closed spin manifold and $v\colon M \to \mathrm{B}\Gamma$ a continuous map, then we set $\alpha^v(M) := \mu(v_*[\not{D}_M]) \in \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$, where $\mu\colon \mathrm{K}_*(\mathrm{B}\Gamma) \to \mathrm{K}_*(\mathrm{C}^*\Gamma)$ is the assembly map. If v classifies the universal covering of M this yields the Rosenberg index $\alpha^\Gamma(M)$.

Theorem 2.7 (multipartitioned manifold index theorem) *In the setup of Definition 2.6 we have*

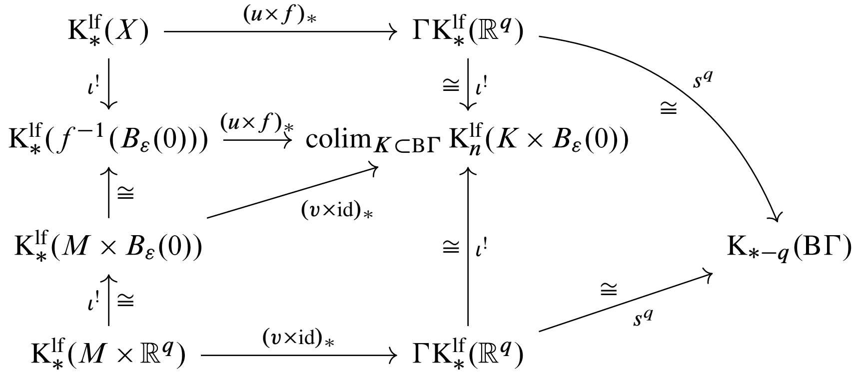
$$\sigma^q(\alpha^{f,u}_{\mathrm{PM}}(X)) = \alpha^{u|_M}(M) \in \mathrm{K}_{m-q}(\mathrm{C}^*_\epsilon\Gamma),$$

where $\sigma^q\colon \Gamma\mathcal{C}_*(\mathbb{R}^q) \to \mathrm{K}_{*-q}(\mathrm{C}^*_\epsilon\Gamma)$ is the q–fold iteration of the suspension isomorphism from Proposition 2.3.

Proof We have $\sigma^q(\alpha^{f,u}_{\mathrm{PM}}(X)) = \mathrm{Ind}^\Gamma(s^q(u \times f)_*([\not{D}_X]))$ by Proposition 2.3. We first deal with the product situation $X = M \times \mathbb{R}^q$ and $u = v \circ \mathrm{pr}_1$. In this special case we have $[\not{D}_X] = [\not{D}_M] \times [\not{D}_{\mathbb{R}^q}]$ and the statement follows from an iterated application of the product formula from Proposition 2.3:

$$\sigma^q(\alpha^{f,u}_{\mathrm{PM}}(X)) = \mathrm{Ind}^\Gamma\big(s^q\big(v_*([\not{D}_M]) \times [\not{D}_{\mathbb{R}^q}]\big)\big) = \mathrm{Ind}^\Gamma(v_*[\not{D}_M]) = \alpha^v(M).$$

In the general case we may assume without loss of generality that there exists $\varepsilon > 0$ such that $f^{-1}(B_\varepsilon(0)) \cong M \times B_\varepsilon(0)$ isometrically. Furthermore, we consider the following commutative diagram, where we set $v := u|_M$ and make extensive use of Proposition 2.3 and Corollary 2.4:



Since $f^{-1}(B_\varepsilon(0)) \cong M\times B_\varepsilon(0)$, the class $[\not{D}_X]\in \mathrm{K}^{\mathrm{lf}}_m(X)$ goes to $[\not{D}_M]\times[\not{D}_{\mathbb{R}^q}]\in \mathrm{K}^{\mathrm{lf}}_m(M\times\mathbb{R}^q)$ following the left vertical maps in the diagram from top to bottom. Thus the diagram implies $(u\times f)_*([\not{D}_X]) = v_*([\not{D}_M])\times[\not{D}_{\mathbb{R}^q}]\in \Gamma\mathrm{K}^{\mathrm{lf}}_m(\mathbb{R}^q)$. This reduces the general case to the product situation, which has already been established. □

Corollary 2.8 *If* $\alpha^{u|_M}(M)\neq 0$ *in the setup of Definition 2.6, then* $\mathrm{Ind}^\Gamma(\not{D}_{\widetilde{X}})\neq 0$. *In this case the Riemannian metric on* X *does not have uniform positive scalar curvature.*

2.3 Fiber bundles and codimension one

We are now almost ready to prove Theorems 1.5 and 1.7. Before doing that, we state the result of Dranishnikov which is needed for Theorem 1.5.

Theorem 2.9 [1, Theorem 3.5] *Let* $\widetilde{B}$ *be the universal covering of a closed aspherical* q*–manifold* B *with* $\mathrm{asdim}(\pi_1(B))<\infty$. *Then there exists* $k\in\mathbb{N}$ *and a proper Lipschitz map* $g\colon \widetilde{B}\times\mathbb{R}^k\to\mathbb{R}^{q+k}$ *of degree* 1.

Proof of Theorem 1.5 By Theorem 2.9, we may assume that there exists a proper Lipschitz map $g\colon \widetilde{B}\to\mathbb{R}^q$ of degree 1 (if necessary, replace the entire bundle by its product with the k–torus $S^1\times\cdots\times S^1$). Since Lipschitz functions can be approximated by smooth Lipschitz functions (see for example [5]), we may suppose without loss of generality that g is smooth. In addition, we may assume that $0\in\mathbb{R}^q$ is a regular value by Sard's theorem. Now consider the covering $\overline{M}\twoheadrightarrow M$ with $\pi_1(\overline{M})=\Lambda=\pi_1(N)$. The bundle projection $\pi\colon M\to B$ lifts to a Γ/Λ–equivariant smooth map $\overline{\pi}\colon \overline{M}\to\widetilde{B}$. Let $N' := (g\circ\overline{\pi})^{-1}(0)$. Then $\overline{M}$ is q–multipartitioned by N' via $f := g\circ\overline{\pi}$. Let $u\colon \overline{M}\to \mathrm{B}\Lambda$ be the map that classifies the Λ–covering $p\colon \widetilde{M}\to\overline{M}$, where

$\widetilde{M}$ is the universal covering of M. Since g has degree 1 and each fiber of $\bar{\pi}$ is a copy of N inside $\bar{M}$ over each of which p restricts to the universal covering, we have that $\alpha^{u|_{N'}}(N') = \alpha^{\Lambda}(N) \in \mathrm{K}_{m-q}(\mathrm{C}^*\Lambda)$. Now consider the homomorphism $\Psi\colon \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma) \to \mathrm{K}_{*-q}(\mathrm{C}^*_\epsilon\Lambda)$ given by the composition

$$\Psi\colon \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma) \cong \mathrm{K}_*(\mathrm{C}^*(\widetilde{M})^\Gamma) \to \mathrm{K}_*(\mathrm{C}^*(\widetilde{M})^\Lambda) \xrightarrow{\tilde{u}\times(f\circ p)} \Lambda\mathcal{C}_*(\mathbb{R}^q) \xrightarrow{\sigma^q} \mathrm{K}_{*-q}(\mathrm{C}^*_\epsilon\Lambda),$$

where the second map is induced by the inclusion $\mathrm{C}^*(\widetilde{M})^\Gamma \subseteq \mathrm{C}^*(\widetilde{M})^\Lambda$ that just forgets part of the equivariance. We have

$$\Psi(\alpha^\Gamma(M)) = \sigma^q(\alpha^{f,u}_{\mathrm{PM}}(\bar{M})) = \alpha^{u|_{N'}}(N') = \alpha^\Lambda(N),$$

where the first equality is by definition of $\alpha^{f,u}_{\mathrm{PM}}(\bar{M})$ and the second equality is due to Theorem 2.7 applied to $f = g \circ \bar{\pi}\colon \bar{M} \to \mathbb{R}^q$ and $u\colon \bar{M} \to \mathrm{B}\Lambda$. Since Ψ is a homomorphism this concludes the proof. □

Proof of Theorem 1.7 The following is very similar to the previous proof. We again consider the covering $\bar{M} \to M$ such that $\pi_1\bar{M} = \Lambda$. With the right choice of basepoints it is possible to lift the inclusion $N \hookrightarrow M$ to an embedding $N \hookrightarrow \bar{M}$. Since $N \hookrightarrow \bar{M}$ has codimension one with trivial normal bundle and is an isomorphism on π_1, it follows that $\bar{M} \setminus N$ has precisely two connected components. Hence $\bar{M}$ is partitioned (or 1–multipartitioned in our terminology above) by N via a map $f\colon \bar{M} \to \mathbb{R}$ which is essentially the distance function from N. Let $\widetilde{M}$ be the universal covering of M and $u\colon \bar{M} \to \mathrm{B}\Lambda$ the map that classifies the Λ–covering $p\colon \widetilde{M} \to \bar{M}$. Again we obtain a map

$$\Psi\colon \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma) \cong \mathrm{K}_*(\mathrm{C}^*(\widetilde{M})^\Gamma) \to \mathrm{K}_*(\mathrm{C}^*(\widetilde{M})^\Lambda) \xrightarrow{\tilde{u}\times(f\circ p)} \Lambda\mathcal{C}_*(\mathbb{R}) \xrightarrow{\sigma} \mathrm{K}_{*-1}(\mathrm{C}^*_\epsilon\Lambda)$$

such that $\Psi(\alpha^\Gamma(M)) = \alpha^\Lambda(N)$. □

3 Higher $\widehat{\mathrm{A}}$ obstructions via submanifolds

Geometric Setup 3.1 In addition to Geometric Setup 1.1, let $N = N_1 \cap \cdots \cap N_k$, where $N_1, \ldots, N_k \subseteq M$ are closed submanifolds with trivial normal bundle that intersect mutually transversely.[1] Let d be the maximum of the codimensions of the submanifolds N_i for $i \in \{1, \ldots, k\}$. Denote by $u\colon M \to \mathrm{B}\Gamma$ a classifying map of the universal covering and let $v := u \circ \iota\colon N \to \mathrm{B}\Gamma$. Moreover, let $w\colon N \to \mathrm{B}\Lambda$ be a classifying map of the universal covering of N.

[1]To be precise, this means that the inclusion $N_1 \times \cdots \times N_k \hookrightarrow M^k$ is transverse to the diagonal embedding $\triangle\colon M \hookrightarrow M^k$ in the usual sense.

We follow the notation of [7] and let $\Lambda^*(\mathrm{B}\Gamma)$ denote the subring of $\mathrm{H}^*(\mathrm{B}\Gamma;\mathbb{Q})$ generated by cohomology classes of degree at most 2.

Proposition 3.2 *Let $\epsilon \in \{\mathrm{red}, \max\}$. In Geometric Setup 3.1 suppose that the induced map in relative homology satisfies*

(2) $$(u, \mathrm{id}_N)_*\colon \mathrm{H}_k(M, N) \to \mathrm{H}_k(v) \text{ is injective for } 2 \leqslant k \leqslant d.$$

Assume furthermore that one of the following conditions holds:

- (a) *We have $\epsilon = \max$, $d \leqslant 2$ and there exists $x \in \Lambda^*(\mathrm{B}\Gamma)$ such that the higher $\widehat{\mathrm{A}}$-genus $\langle \widehat{\mathrm{A}}(\mathrm{T}N) \cup v^*(x), [N]\rangle$ does not vanish.*
- (b) *The group Γ satisfies the (maximal, if $\epsilon = \max$) strong Novikov conjecture and there exists $x \in \mathrm{H}^*(\mathrm{B}\Gamma;\mathbb{Q})$ such that the higher $\widehat{\mathrm{A}}$-genus $\langle \widehat{\mathrm{A}}(\mathrm{T}N) \cup v^*(x), [N]\rangle$ does not vanish.*

Then $\alpha^\Gamma(M) \in \mathrm{K}_*(\mathrm{C}^*_\epsilon\Gamma)$ does not vanish. In particular, M does not admit a metric of positive scalar curvature.

Proof Let $\eta_i \in \mathrm{H}^*(M;\mathbb{Q})$ denote the Poincaré dual of $N_i \subseteq M$. Since N_i has trivial normal bundle the restriction of η_i to N_i vanishes. In particular, $\iota^*\eta_i = 0 \in \mathrm{H}^*(N;\mathbb{Q})$, so there exists $\tilde{\eta}_i \in \mathrm{H}^*(M, N;\mathbb{Q})$ that restricts to $\eta_i \in \mathrm{H}^*(M;\mathbb{Q})$. By the upper bound on the codimensions, the degree of η_i is at most d for each $i \in \{1, \dots, k\}$. Note that $u\colon M \to \mathrm{B}\Gamma$ is 2–connected and thus $(u, \mathrm{id}_N)_*\colon \mathrm{H}_1(M, N) \to \mathrm{H}_1(v)$ is an isomorphism by the Hurewicz theorem and the long exact sequence associated to the triple $N \hookrightarrow M \xrightarrow{u} \mathrm{B}\Gamma$. Together with (2) this implies that there exists $\tilde{\xi}_i \in \mathrm{H}^*(v;\mathbb{Q})$ such that $(u, \mathrm{id}_N)^*\tilde{\xi}_i = \tilde{\eta}_i$ for all $i \in \{1, \dots, k\}$. Restricting these to $\mathrm{B}\Gamma$, we get $\xi_i \in \mathrm{H}^*(\mathrm{B}\Gamma;\mathbb{Q})$ such that $u^*\xi_i = \eta_i$. We have that $\eta = \eta_1 \cup \dots \cup \eta_k = u^*(\xi)$ is the Poincaré dual of $N = N_1 \cap \dots \cap N_k$, where $\xi := \xi_1 \cup \dots \cup \xi_k$. For each $x \in \mathrm{H}^*(\mathrm{B}\Gamma;\mathbb{Q})$, we then compute

$$\begin{aligned}
\langle \widehat{\mathrm{A}}(\mathrm{T}N) \cup v^*(x), [N]\rangle &= \langle \widehat{\mathrm{A}}(\mathrm{T}N) \cup \widehat{\mathrm{A}}(\nu(N \hookrightarrow M)) \cup v^*(x), [N]\rangle \\
&= \langle \iota^*\widehat{\mathrm{A}}(\mathrm{T}M) \cup v^*(x), [N]\rangle \\
&= \langle \widehat{\mathrm{A}}(\mathrm{T}M) \cup u^*(x) \cup \eta, [M]\rangle \\
&= \langle \widehat{\mathrm{A}}(\mathrm{T}M) \cup u^*(x \cup \xi), [M]\rangle \\
&= \langle u^*(x \cup \xi), \mathrm{ch}([\not{D}_M])\rangle,
\end{aligned}$$

where triviality of the normal bundle $\nu(N \hookrightarrow M)$ is used in the first equality. In other words, the particular higher $\widehat{\mathrm{A}}$-genus of N we started with can be rewritten as a higher $\widehat{\mathrm{A}}$-genus of M.

In case (a), this implies that $\langle z, \mathrm{ch}(u_*[\not{D}_M])\rangle \neq 0$, where $z := x \cup \xi \in \Lambda^*(\mathrm{B}\Gamma)$. Hence by [7, Theorem 1.2], this shows that $\alpha^\Gamma(M) = \mu(u_*([\not{D}_M])) \neq 0 \in \mathrm{K}_*(\mathrm{C}^*_{\max}\Gamma)$. In case (b), the computation simply shows that $0 \neq u_*([\not{D}_M]) \in \mathrm{K}_*(\mathrm{B}\Gamma) \otimes \mathbb{Q}$. Hence by the postulated rational injectivity of the (maximal, if $\epsilon = \max$) assembly map, the higher index does not vanish. □

It remains to put forward some further (sufficient) conditions for the homological condition (2). For instance, we find it conceptually appealing to consider the square

$$\begin{array}{ccc} N & \xrightarrow{\iota} & M \\ \downarrow{\scriptstyle w} & & \downarrow{\scriptstyle u} \\ \mathrm{B}\Lambda & \xrightarrow{j} & \mathrm{B}\Gamma \end{array}$$

and ask the induced map in relative homology $\mathrm{H}_*(M, N) \to \mathrm{H}_*(\mathrm{B}\Gamma, \mathrm{B}\Lambda)$ to be an equivalence up to a certain degree. Indeed, as it turns out in the lemma below, this is an easy sufficient condition for (2). Moreover, $\mathrm{H}_*(M, N) \to \mathrm{H}_*(\mathrm{B}\Gamma, \mathrm{B}\Lambda)$ being an isomorphism up to degree 2 and surjective in degree 3 is equivalent to surjectivity of $\pi_2(N) \to \pi_2(M)$. The latter is precisely the condition that we have already encountered in Remark 1.6.

Lemma 3.3 *Suppose that in Geometric Setup 3.1 one of the following conditions holds:*

- (a) *The map $\pi_2(N) \to \pi_2(M)$ is surjective and $d = 2$.*
- (a′) *The map $\mathrm{H}_k(M, N) \to \mathrm{H}_k(\mathrm{B}\Gamma, \mathrm{B}\Lambda)$ is an isomorphism for $2 \leqslant k \leqslant d$ and surjective for $k = d + 1$.*
- (b) *The homotopy groups $\pi_k(M)$ vanish for $2 \leqslant k \leqslant d$.*

Then the condition (2) *from the statement of Proposition 3.2 is satisfied.*

Moreover, for $d = 2$ the conditions (a) *and* (a′) *are equivalent.*

Proof We first show that for $d = 2$, (a) and (a′) are equivalent. Indeed, consider the following diagram of homotopy cofiber sequences:

$$\begin{array}{ccccc} N & \xrightarrow{\iota} & M & \longrightarrow & C_\iota \\ \downarrow{\scriptstyle w} & & \downarrow{\scriptstyle u} & & \downarrow \\ \mathrm{B}\Lambda & \xrightarrow{j} & \mathrm{B}\Gamma & \longrightarrow & C_j \\ \downarrow & & \downarrow & & \downarrow \\ C_w & \longrightarrow & C_u & \longrightarrow & C \end{array}$$

Since w and u are 2–connected by construction, it follows by the Hurewicz theorem that $\mathrm{H}_k(C_w) = \mathrm{H}_k(C_u) = 0$ for $k = 1, 2$ and that $\mathrm{H}_3(C_w) \cong \pi_3(w)$ and $\mathrm{H}_3(C_u) \cong \pi_3(u)$. In particular, looking at the lower horizontal sequence in the diagram, we see that we always have $\mathrm{H}_k(C) = 0$ for $k = 1, 2$. Moreover, since $\mathrm{B}\Gamma$ and $\mathrm{B}\Lambda$ are aspherical, we have $\pi_3(u) \cong \pi_2(M)$ and $\pi_3(w) \cong \pi_2(N)$. Thus surjectivity of $\pi_2(N) \to \pi_2(M)$ is equivalent to surjectivity of $\pi_3(w) \cong \mathrm{H}_3(C_w) \to \mathrm{H}_3(C_u) \cong \pi_3(u)$, which, in turn, is equivalent to $\mathrm{H}_3(C) = 0$ since we always have $H_2(C_w) = 0$. Finally, turning to the right vertical sequence of the diagram, the vanishing of $H_3(C)$ is equivalent to (a$'$) for $d = 2$ (since we have always $\mathrm{H}_k(C) = 0$ for $k = 1, 2$).

To see that (a$'$) implies (2), we just note that the map $\mathrm{H}_k(M, N) \to \mathrm{H}_k(\mathrm{B}\Gamma, \mathrm{B}\Lambda)$ factors as $\mathrm{H}_k(M, N) \to \mathrm{H}_k(v) \to \mathrm{H}_k(\mathrm{B}\Gamma, \mathrm{B}\Lambda)$.

To see that (b) implies (2), consider the long exact sequence of the triple $N \hookrightarrow M \xrightarrow{u} \mathrm{B}\Gamma$:

$$\cdots \to \mathrm{H}_{k+1}(u) \to \mathrm{H}_k(M, N) \to \mathrm{H}_k(v) \to \mathrm{H}_k(u) \to \cdots .$$

If $\pi_k(M) = 0$ for $2 \leqslant k \leqslant d$, then $u\colon M \to \mathrm{B}\Gamma$ is $(d+1)$–connected and hence $\mathrm{H}_k(u) = 0$ for $k \leqslant d + 1$. In particular, $\mathrm{H}_k(M, N) \to \mathrm{H}_k(v)$ is even an isomorphism for $k \leqslant d$. □

Finally, Theorems 1.9 and 1.10 follow immediately now by combining cases (a) and (b) from Proposition 3.2 (applied to $x = 1 \in \mathrm{H}^0(\mathrm{B}\Gamma)$) with cases (a) and (b) from Lemma 3.3, respectively.

Acknowledgements This article was written during the author's doctoral studies at the University of Göttingen, where he was supported by the German Research Foundation (DFG) through the Research Training Group 1493 "Mathematical structures in modern quantum physics".

References

[1] **A N Dranishnikov**, *On hypereuclidean manifolds*, Geom. Dedicata 117 (2006) 215–231 MR

[2] **A Engel**, *Rough index theory on spaces of polynomial growth and contractibility*, preprint (2015) arXiv

[3] **A Engel**, *Wrong way maps in uniformly finite homology and homology of groups*, preprint (2016) arXiv

[4] **G Gong**, **Q Wang**, **G Yu**, *Geometrization of the strong Novikov conjecture for residually finite groups*, J. Reine Angew. Math. 621 (2008) 159–189 MR

[5] **R E Greene**, **H Wu**, *C^∞ approximations of convex, subharmonic, and plurisubharmonic functions*, Ann. Sci. École Norm. Sup. 12 (1979) 47–84 MR

[6] **B Hanke**, **D Pape**, **T Schick**, *Codimension two index obstructions to positive scalar curvature*, Ann. Inst. Fourier (Grenoble) 65 (2015) 2681–2710 MR

[7] **B Hanke**, **T Schick**, *The strong Novikov conjecture for low degree cohomology*, Geom. Dedicata 135 (2008) 119–127 MR

[8] **N Higson**, **J Roe**, *Analytic K–homology*, Oxford Univ. Press (2000) MR

[9] **N Higson**, **J Roe**, **G Yu**, *A coarse Mayer–Vietoris principle*, Math. Proc. Cambridge Philos. Soc. 114 (1993) 85–97 MR

[10] **J Roe**, *Index theory, coarse geometry, and topology of manifolds*, CBMS Regional Conference Series in Math. 90, Amer. Math. Soc., Providence, RI (1996) MR

[11] **T Schick**, **M E Zadeh**, *Large scale index of multi-partitioned manifolds*, preprint (2013) arXiv

[12] **P Siegel**, *Homological calculations with the analytic structure group*, PhD thesis, Pennsylvania State University (2012) Available at `https://search.proquest.com/docview/1356843373?`

[13] **P Siegel**, *The Mayer–Vietoris sequence for the analytic structure group*, preprint (2012) arXiv

[14] **R Zeidler**, *Secondary large-scale index theory and positive scalar curvature*, PhD thesis, Georg-August-Universität, Göttingen (2016) Available at `http://hdl.handle.net/11858/00-1735-0000-0028-8826-7`

Mathematisches Institut, Westfälische Wilhelms-Universität Münster
Münster, Germany

`math@rzeidler.eu`

`http://wwwmath.uni-muenster.de/u/rudolf.zeidler/`

Received: 3 November 2016 Revised: 6 February 2017

msp

Algebraic & Geometric Topology 17 (2017) 3095–3136

An algebraic model for rational SO(3)–spectra

MAGDALENA KĘDZIOREK

Greenlees established an equivalence of categories between the homotopy category of rational SO(3)–spectra and the derived category $d\mathcal{A}(\mathrm{SO}(3))$ of a certain abelian category. In this paper we lift this equivalence of homotopy categories to the level of Quillen equivalences of model categories. Methods used in this paper provide the first step towards obtaining an algebraic model for the toral part of rational G–spectra, for any compact Lie group G.

55N91, 55P42, 55P60

1 Introduction

Modelling the category of rational G–spectra This paper is a contribution to the study of G–equivariant cohomology theories and gives a complete analysis for one class of theories, namely rational SO(3)–equivariant cohomology theories. To start with, let G be a compact Lie group. Recall that G–equivariant cohomology theories are represented by G–spectra, so the category of G–equivariant cohomology theories is equivalent to the homotopy category of G–spectra. The category of G–spectra is quite complicated, with rich structures coming from two sources: topology and the group actions, and one cannot expect a complete analysis of either cohomology theories or spectra integrally.

For a compact Lie group G, the category of rational G–spectra is the category of G–spectra, but with the model structure that is a left Bousfield localisation of the stable model structure at the rational sphere spectrum; see for example Barnes [1, Section 2.2]. Thus the weak equivalences are maps which become isomorphisms after applying the rational homotopy group functors, ie $\pi_*^H(-)\otimes\mathbb{Q}$ for all closed subgroups H in G.

Rationalising the category of G–spectra reduces topological complexity, simplifying it greatly. At the same time interesting equivariant behaviour remains. In order to understand this behaviour, we try to find a purely algebraic description of the category, that is an algebraic model category which is Quillen equivalent to the category of rational G–spectra. As a result, the homotopy category of the algebraic model is equivalent to the rational stable G–homotopy category via triangulated equivalences. Moreover all the homotopy information, such as homotopy limits, in both is the same.

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3095

The conjecture by Greenlees states that for any compact Lie group G there is a nice graded abelian category $\mathcal{A}(G)$ such that the category $d\mathcal{A}(G)$ of differential objects in $\mathcal{A}(G)$ with a certain model structure is Quillen equivalent to the category of rational G–spectra:

$$G\text{–}\mathrm{Sp}_{\mathbb{Q}} \simeq_Q d\mathcal{A}(G).$$

If we find such $d\mathcal{A}(G)$ we say that $d\mathcal{A}(G)$ is an *algebraic model* for rational G–spectra.

Existing work There are several examples of specific Lie groups G for which an algebraic model has been given. Firstly, when G is trivial, it was shown in Shipley [22, Theorem 1.1] that rational spectra are monoidally Quillen equivalent to chain complexes of $\mathbb{Q}$–modules. An algebraic model for rational G–spectra for finite G is described in Schwede and Shipley [21, Example 5.1.2] and simplified in Barnes [2] and Kędziorek [16]. An algebraic model for rational torus-equivariant spectra was presented in Greenlees and Shipley [11], whereas a slightly different approach in Barnes, Greenlees, Kędziorek and Shipley [6] gives a *symmetric monoidal* algebraic model for $SO(2)$. This was recently used by Barnes [4] to provide an algebraic model for rational $O(2)$–spectra.

However, there is no algebraic model known for the whole category of rational G–spectra for an arbitrary compact Lie group G. A first step in this direction, a model for rational G–spectra over an exceptional subgroup (see Definition 5.1) for any compact Lie group G, was provided in [16]. This result is used in Section 5.

The group SO(3) The group SO(3) is the group of rotations of $\mathbb{R}^3$. This is the natural next candidate to analyse on the way to understanding the behaviour of $d\mathcal{A}(G)$ for an arbitrary compact Lie group G. Notice that SO(3) is significantly more complicated than all groups considered so far, since it is the first group where the maximal torus is not normal in the whole group. Dealing with this complication shows a method to provide an algebraic model for a part of rational G–spectra called *toral* for any compact Lie group G. The toral part models those G–spectra whose geometric isotropy is a set of subgroups of the maximal torus and corresponds to cohomology theories with toral support. We discuss this further in Remark 3.29.

Main result Let G be $SO(3)$. In this paper we work with orthogonal G–spectra; see Mandell and May [18, Definition 2.6]. By Barnes [3, Theorem 4.4], the category of rational SO(3)–orthogonal spectra splits into three parts: toral, dihedral and exceptional. This uses idempotents of the rational Burnside ring $\mathsf{A}(SO(3))_{\mathbb{Q}}$ (see Section 2.3), and reflects a similar splitting at the level of homotopy categories.

The toral part models rational SO(3)–spectra with geometric isotropy in the family of subconjugates of the maximal torus $SO(2)$ in $SO(3)$. The dihedral part models rational SO(3)–spectra with geometric isotropy in the collection of subgroups $\mathcal{D}$, which consists of all dihedral subgroups of order greater than 4 and $O(2)$. The last part, which we call the exceptional part, models rational SO(3)–spectra with geometric isotropy in the collection of subgroups $\mathcal{E}$, which consists of all remaining subgroups (see Section 2.1). Thus we are able to work with each of these three parts separately to obtain an algebraic model for rational SO(3)–spectra.

The main result of this paper is as follows.

Main Theorem *There is a zig-zag of Quillen equivalences between rational* SO(3)*–orthogonal spectra and the algebraic category* $d\mathcal{A}(\mathrm{SO}(3))$.

The category $d\mathcal{A}(\mathrm{SO}(3))$, which we call the *algebraic model for rational* SO(3)*–spectra*, is a product of three parts, which reflects the splitting of the category of rational SO(3)–spectra

$$d\mathcal{A}(\mathrm{SO}(3)) \cong d\mathcal{A}(\mathrm{SO}(3),\mathcal{T}) \times \mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D})) \times \prod_{(H),H\in\mathcal{E}} \mathsf{Ch}(\mathbb{Q}[W_{\mathrm{SO}(3)}H]).$$

Here $d\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ is the *algebraic model for the toral part* described in Section 3.2, $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$ is the *algebraic model for the dihedral part* described in Section 4.1 and $\mathsf{Ch}(\mathbb{Q}[W_{\mathrm{SO}(3)}H])$ is the *algebraic model for the rational* SO(3)*–spectra over an exceptional subgroup* H discussed in Section 5.1. Since $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ is a graded abelian category we use the notation $d\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ for differential objects in there. We use the notation $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$ for differential graded objects (ie chain complexes) in $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$, since $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$ doesn't have a grading.

The Main Theorem follows from Proposition 2.6 and Theorems 3.36, 4.11 and 5.4.

Contribution of this paper The new idea in this paper concerns the toral part in Section 3. Since the maximal torus is not normal in SO(3) the algebraic model for the toral part gets more complicated than that for the torus (see Greenlees [9] and Barnes, Greenlees, Kędziorek and Shipley [6]) or $O(2)$ (see Barnes [4]). To control these complications we use the following method. We consider the restriction–coinduction adjunction between the toral part of rational SO(3)–spectra and the toral part of rational $O(2)$–spectra. Here $O(2)$ is the normaliser of the maximal torus in SO(3). This adjunction is a Quillen adjunction, but not a Quillen equivalence.

However, the cellularisation principle of Greenlees and Shipley [12] (see Section 2.2.2 for the definition of cellularisation) gives a Quillen equivalence between the toral part of rational SO(3)–spectra and a certain cellularisation of the toral part of rational $O(2)$–spectra; see Theorem 3.28. Now it is enough to cellularise the algebraic model for the toral part of rational $O(2)$–spectra and simplify this category (see Section 3.4) to obtain the model for the toral part of rational SO(3)–spectra.

The idea of using the restriction–coinduction adjunction between the toral part of rational G–spectra and the toral part of rational $N_G\mathbb{T}$–spectra (where $\mathbb{T}$ is the maximal torus in G) together with the cellularisation principle allows one to provide an algebraic model for the toral part of rational G–spectra, for any compact Lie group G; see Barnes, Greenlees and Kędziorek [5].

The method to obtain the algebraic model for the dihedral part of rational SO(3)–spectra is a slight alteration of the method for the dihedral part for rational $O(2)$–spectra from [4] and is presented in Section 4.2. Some changes in the proof from [4] are needed to take into account the fact that our dihedral part excludes subgroups conjugate to D_2 and D_4 (for reasons explained in Section 2.1), whereas the dihedral part of $O(2)$–spectra contains them. However, the idea of the proof remains the same.

Finally, an algebraic model of the exceptional part is an application of the methods from Kędziorek [16]. We point out that this is the only part of the paper that considers monoidal structures and gives a monoidal algebraic model.

Outline of the paper This paper is structured as follows. In Section 2 we present some general results about subgroups of SO(3), its rational Burnside ring $\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$ and the idempotents used to split the category of rational SO(3)–spectra into three parts: toral, dihedral and exceptional (Proposition 2.6). Section 3 is the heart of this paper. It contains the description of the algebraic model for the toral part of rational SO(3)–spectra. It also presents Quillen equivalences used in obtaining this algebraic model from the algebraic model for toral rational $O(2)$–spectra. Section 4 contains the algebraic model for the dihedral part. Finally, in Section 5 we recall the results from [16] to give an algebraic model for the exceptional part of rational SO(3)–spectra.

Notation We will stick to the convention of drawing the left adjoint above (or to the left of) the right one in any adjoint pair. We use the notation $G-\mathrm{Sp}$ for the category of G–equivariant orthogonal spectra.

Acknowledgements This work is based on a part of my PhD thesis (under the supervision of John Greenlees) and I would like to thank John Greenlees and David Barnes for many useful discussions and comments.

2 General results for SO(3)

We start this part by considering the closed subgroups of SO(3) in Section 2.1. We discuss the space $\mathcal{F}(G)/G$, which is the orbit space of all closed subgroups with finite index in their normaliser, where the topology is induced from the Hausdorff metric; see [17, Section V.2]. In Section 2.2 we recall two ways of changing a given stable model structure: left Bousfield localisation at an object and cellularisation. We will use these techniques repeatedly throughout the paper. In Section 2.3 we discuss the idempotents of the rational Burnside ring $\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$ and the induced splitting of rational SO(3)–orthogonal spectra. The main part of Section 2.3 consists of the analysis of two adjunctions: the induction–restriction and restriction–coinduction adjunctions in relation to localisations of categories of equivariant spectra at idempotents.

2.1 Closed subgroups of SO(3)

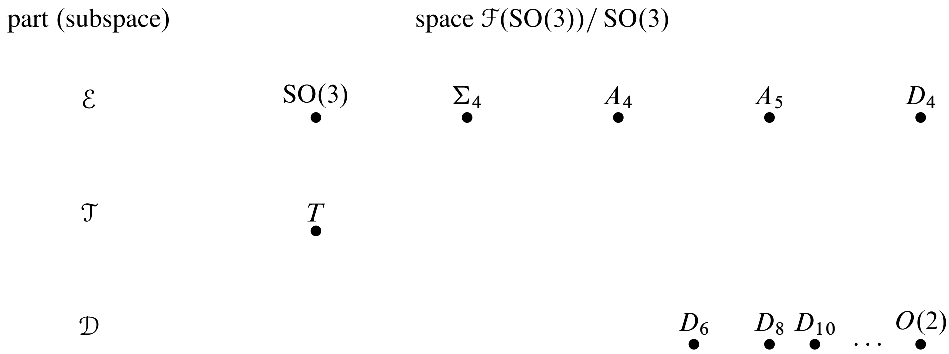
Recall that $\mathrm{SO}(3)$ is the group of rotations of $\mathbb{R}^3$. We choose a maximal torus T in $\mathrm{SO}(3)$ with rotation axis the z–axis. We divide the closed subgroups of G into three types: *toral* $\mathcal{T}$, *dihedral* $\mathcal{D}$ and *exceptional* $\mathcal{E}$. This division is motivated by the choice of idempotents in the rational Burnside ring for $\mathrm{SO}(3)$ that we will use to split the category of rational SO(3)–spectra.

The toral part consists of all tori in $\mathrm{SO}(3)$ and all cyclic subgroups of these tori. Note that for any natural number n there is one conjugacy class of subgroups from the toral part of order n in $\mathrm{SO}(3)$.

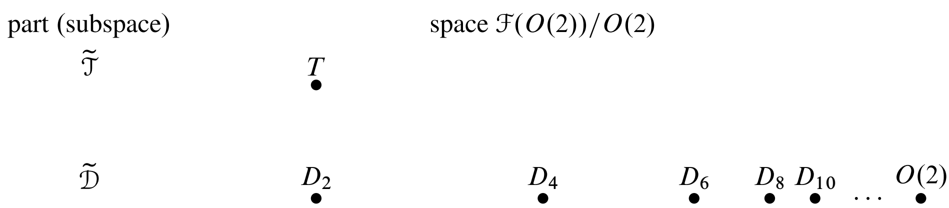
The dihedral part consists of all dihedral subgroups D_{2n} (dihedral subgroups of order $2n$) of $\mathrm{SO}(3)$ where n is greater than 2, together with all subgroups $O(2)$. Note that $O(2)$ is the normaliser for itself in $\mathrm{SO}(3)$. Moreover, there is only one conjugacy class of a dihedral subgroup D_{2n} for each n greater than 2, and the normaliser of D_{2n} in $\mathrm{SO}(3)$ is D_{4n} for $n > 2$. Notice that we excluded subgroups in the conjugacy classes of D_2 and D_4 from this part. Conjugates of D_2 are excluded from the dihedral part, since D_2 is conjugate to C_2 in $\mathrm{SO}(3)$ and that subgroup is already taken into account in the toral part. Conjugates of D_4 are excluded from the dihedral part since their normalisers in $\mathrm{SO}(3)$ are Σ_4 (symmetries of a cube), thus their Weyl groups Σ_4/D_4 are of order 6, whereas all other finite dihedral subgroups D_{2n}, $n > 2$ have Weyl groups of order 2. For simplicity we decided to treat D_4 separately and put it into the exceptional part.

There are five conjugacy classes of subgroups which we call exceptional, namely $\mathrm{SO}(3)$ itself, the rotation group Σ_4 of a cube, the rotation group A_4 of a tetrahedron, the rotation group A_5 of a dodecahedron and the dihedral group D_4 of order 4. Normalisers of these exceptional subgroups are as follows: Σ_4 is equal to its normaliser, A_5 is equal to its normaliser and the normaliser of A_4 is Σ_4, as is the normaliser of D_4.

Consider the space $\mathcal{F}(\mathrm{SO}(3))/\mathrm{SO}(3)$ of conjugacy classes of subgroups of $\mathrm{SO}(3)$ with finite index in their normalisers. The topology on this space is induced by the Hausdorff metric. We will use this space for choosing idempotents of the rational Burnside ring in Section 2.3. The topology on $\mathcal{E}$ is discrete, $\mathcal{T}$ consists of one point T and $\mathcal{D}$ forms a sequence of points converging to $O(2)$, as shown in the following diagram:



Before we go any further we recall the space $\mathcal{F}(O(2))/O(2)$. It consists of two parts: toral and dihedral. To distinguish between these parts and their analogues for SO(3) we choose the notation $\widetilde{\mathcal{T}}$ for the toral part of $O(2)$ and $\widetilde{\mathcal{D}}$ for the dihedral part of $O(2)$ (note that in [4] the notation without tilde was used for the toral and dihedral parts of $O(2)$). We will stick to this new notation convention throughout the paper. The toral part is just one point T corresponding to the maximal torus and all its subgroups. The dihedral part corresponds to all dihedral subgroups together with $O(2)$ and we present it below:



The only difference in the dihedral parts for $O(2)$ and SO(3) is captured by the fact that the dihedral part for $O(2)$ is a disjoint union of $\mathcal{D}$ and two points (corresponding to D_2 and D_4, respectively). At a first glance the toral part for SO(3) looks the same as the toral part for $O(2)$. However, for SO(3) it contains information about D_2 (since D_2 is conjugate to C_2 in SO(3)), whereas for $O(2)$ it does not. These differences will become significant in Section 2.3.

2.2 Left Bousfield localisation and cellularisation

In this section we briefly recall two ways of changing a given stable model structure: left Bousfield localisation at an object and cellularisation. We will repeatedly use them in the rest of the paper.

2.2.1 Left Bousfield localisation at an object For details on left Bousfield localisation at an object we refer the reader to [18, Section IV.6]. We recall the following result:

Theorem 2.1 [18, Chapter IV, Theorem 6.3] *Suppose E is a cofibrant object in $G{-}\mathrm{Sp}$ or a cofibrant based G–space. Then there exists a new model structure on the category $G{-}\mathrm{Sp}$, where a map $f\colon X \to Y$ is*

- *a weak equivalence if it is an E–equivalence, ie $\mathrm{Id}_E \wedge f\colon E \wedge X \to E \wedge Y$ is a weak equivalence;*
- *a cofibration if it is a cofibration with respect to the stable model structure;*
- *a fibration if it has the right lifting property with respect to all trivial cofibrations.*

The E–fibrant objects Z are the E–local objects, ie those such that $[f, Z]^G\colon [Y, Z]^G \to [X, Z]^G$ is an isomorphism for all E–equivalences f. E–fibrant approximation gives Bousfield localisation $\lambda\colon X \to L_E X$ of X at E.

We use the notation $L_E(G{-}\mathrm{Sp})$ for the model category described above and will refer to it as a *left Bousfield localisation of the category of G–spectra at E*. If E and F are cofibrant objects in $G{-}\mathrm{Sp}$ then the localisation first at E and then at F is the same model category as the localisation at $E \wedge F$ (and $F \wedge E$).

Recall that an E–equivalence between E–local objects is a weak equivalence (see [13, Theorems 3.2.13 and 3.2.14]).

In this paper we use the above definition with $X \in G{-}\mathrm{Sp}$ of the form $eS_{\mathbb{Q}}$ (for various e) where e is an idempotent of a rational Burnside ring $\mathrm{A}(G)_{\mathbb{Q}}$ and $S_{\mathbb{Q}}$ is a rational sphere spectrum (see [2, Section 5] for construction of the rational sphere spectrum $S_{\mathbb{Q}}$). Since we use idempotents of a rational Burnside ring, all our localisations are smashing (see [19] for definition of a smashing localisation). Thus they preserve homotopically compact generators (see Definition 2.5) since the fibrant replacement preserves infinite coproducts.

2.2.2 Cellularisation A cellularisation of a model category is a right Bousfield localisation at a set of objects. Such a localisation exists by [13, Theorem 5.1.1] whenever the model category is right proper and cellular. When we are in a stable context the results of [7] can be used.

In this section we recall the notion of cellularisation when $\mathcal{C}$ is a stable model category and some basic definitions and results.

Definition 2.2 Let $\mathcal{C}$ be a stable model category and K a stable set of objects of $\mathcal{C}$, ie a set such that the class of K–cellular objects of $\mathcal{C}$ is closed under desuspension (note that the class is always closed under suspension). We call K a set of *cells*. We say that a map $f\colon A \to B$ of $\mathcal{C}$ is a *K–cellular equivalence* if the induced map

$$[k, f]_*^{\mathcal{C}}\colon [k, A]_*^{\mathcal{C}} \to [k, B]_*^{\mathcal{C}}$$

is an isomorphism of graded abelian groups for each $k \in K$. An object $Z \in \mathcal{C}$ is said to be *K–cellular* if

$$[Z, f]_*^{\mathcal{C}}\colon [Z, A]_*^{\mathcal{C}} \to [Z, B]_*^{\mathcal{C}}$$

is an isomorphism of graded abelian groups for any K–cellular equivalence f.

Definition 2.3 A *right Bousfield localisation* or *cellularisation* of $\mathcal{C}$ with respect to a set of objects K is a model structure $K-\text{cell}-\mathcal{C}$ on $\mathcal{C}$ such that

- the weak equivalences are K–cellular equivalences,
- the fibrations of $K-\text{cell}-\mathcal{C}$ are the fibrations of $\mathcal{C}$,
- the cofibrations of $K-\text{cell}-\mathcal{C}$ are defined via left lifting property.

By [13, Theorem 5.1.1], if $\mathcal{C}$ is a right proper, cellular model category and K a set of objects in $\mathcal{C}$, then the cellularisation of $\mathcal{C}$ with respect to K, $K-\text{cell}-\mathcal{C}$, exists and is a right proper model category. The cofibrant objects of $K-\text{cell}-\mathcal{C}$ are called K–cofibrant and are precisely the K–cellular and cofibrant objects of $\mathcal{C}$.

The cellularisation of a proper, cellular, stable model category at a stable set of cofibrant objects K is very well behaved (see [7, Theorem 5.9]), in particular it is proper, cellular and stable. Left properness follows from [7, Proposition 5.8].

There is another important property we will often want the cells to satisfy, which makes right localisation behave in an even more tractable manner; see [7, Section 9]. This property is variously called smallness, compactness or finiteness. We choose to call it *homotopical compactness*, since there are several different meanings of compactness in the literature.

Definition 2.4 [21, Definition 2.1.2] An object X in a stable model category $\mathcal{C}$ is *homotopically compact* if for any family of objects $\{A_i\}_{i\in I}$ the canonical map

$$\bigoplus_{i\in I}[X, A_i]^{\mathcal{C}} \to \Big[X, \coprod_{i\in I} A_i\Big]^{\mathcal{C}}$$

is an isomorphism in the homotopy category of $\mathcal{C}$.

Recall that a homotopy category of a stable model category is triangulated; see Definition 7.1.1 of [14]. In this setting we can make the following definition after Definition 2.1.2 of [21].

Definition 2.5 Let $\mathcal{C}$ be a triangulated category with infinite coproducts. A full triangulated subcategory of $\mathcal{C}$ (with shift and triangles induced from $\mathcal{C}$) is called *localising* if it is closed under coproducts in $\mathcal{C}$. A set $\mathcal{P}$ of objects of $\mathcal{C}$ is called a *set of generators* if the only localising subcategory of $\mathcal{C}$ containing objects of $\mathcal{P}$ is the whole of $\mathcal{C}$. An object of a stable model category is called a generator if it is a generator when considered as an object of the homotopy category.

Using [21, Lemma 2.2.1] it is routine to check that if K consists of homotopically compact objects of $\mathcal{C}$ then K is a set of generators for $K-\text{cell}-\mathcal{C}$. Hence we know a set of generators for each of our cellularisations.

Notice that derived functors of both left and right Quillen equivalences preserve homotopically compact objects.

2.3 Idempotents, splitting and reductions

By the results of tom Dieck [8, Propositions 5.6.4 and 5.9.13] there is an isomorphism of rings

$$\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}} = C(\mathcal{F}(\mathrm{SO}(3))/\,\mathrm{SO}(3), \mathbb{Q}).$$

Here $\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$ is the rational Burnside ring for SO(3) and $C(\mathcal{F}(\mathrm{SO}(3))/\,\mathrm{SO}(3), \mathbb{Q})$ denotes the ring of continuous functions on the orbit space $\mathcal{F}(\mathrm{SO}(3))/\,\mathrm{SO}(3)$ with values in the discrete space $\mathbb{Q}$.

Thus it is clear that idempotents of the rational Burnside ring of SO(3) correspond to the characteristic functions on subspaces of the orbit space $\mathcal{F}(\mathrm{SO}(3))/\,\mathrm{SO}(3)$ discussed in Section 2.1 which are both open and closed.

In this paper we use the following idempotents in the rational Burnside ring of SO(3): $e_{\mathcal{T}}$ corresponding to the characteristic function of the toral part $\mathcal{T}$, ie the conjugacy class of the torus T; $e_{\mathcal{D}}$ corresponding to the characteristic function of the dihedral part $\mathcal{D}$; and $e_{\mathcal{E}}$ corresponding to the characteristic function of the exceptional part $\mathcal{E}$. Since $\mathcal{E}$ is a disjoint union of five points, it is in fact a sum of five idempotents, one for every (conjugacy class of a) subgroup in the exceptional part: $e_{\mathrm{SO}(3)}$, e_{Σ_4}, e_{A_4}, e_{A_5} and e_{D_4}. We use a simplified notation e_H to mean $e_{(H)_{\mathrm{SO}(3)}}$ here.

Analogously, we will use the notation $e_{\widetilde{\mathcal{T}}}$ for the idempotent in the rational Burnside ring of $O(2)$ corresponding to the toral part $\widetilde{\mathcal{T}}$ and $e_{\widetilde{\mathcal{D}}}$ for the idempotent corresponding to the dihedral part $\widetilde{\mathcal{D}}$ of $O(2)$.

For an idempotent $e \in \mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$ and a rational sphere spectrum $S_{\mathbb{Q}}$ (see [2, Section 5] for the construction) we define $eS_{\mathbb{Q}}$ to be the homotopy colimit (a mapping telescope) of the diagram

$$S_{\mathbb{Q}} \xrightarrow{e} S_{\mathbb{Q}} \xrightarrow{e} S_{\mathbb{Q}} \xrightarrow{e} \cdots .$$

We ask for this spectrum to be cofibrant either by choosing a good construction of homotopy colimit, or by cofibrantly replacing the result in the stable model structure for SO(3)–spectra. Now, by [18, Chapter IV, Theorem 6.3] (see also Theorem 2.1) the following left Bousfield localisations exist:

$$L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp}), \quad L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp}), \quad L_{e_{\mathcal{E}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp}).$$

Also, $L_{e_H S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp})$ exists for any exceptional subgroup $H \in \mathcal{E}$.

The first step on the way towards an algebraic model for rational SO(3)–spectra is to split this category using the above idempotents of the Burnside ring $\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$. By [3, Theorem 4.4] we get the following decomposition.

Proposition 2.6 *The adjunction*

$$\begin{array}{c}\mathrm{SO}(3)\text{–}\mathrm{Sp}_{\mathbb{Q}} \\ \Delta \big\downarrow \big\uparrow \Pi \\ L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp}) \times L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp}) \times L_{e_{\mathcal{E}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp})\end{array}$$

is a strong monoidal Quillen equivalence, where $\mathrm{SO}(3)\text{–}\mathrm{Sp}_{\mathbb{Q}}$ *denotes the category of rational* SO(3) *orthogonal spectra, the left adjoint is the diagonal functor and the right adjoint is the product.*

The main idea is to relate each of these localised categories to corresponding ones for simpler groups. Thus we recall that an inclusion $i\colon H \to G$ of a subgroup H into a group G induces two adjoint pairs at the level of orthogonal spectra, induction–restriction and restriction–coinduction (see [18, Section V.2]):

$$G\text{–}\mathrm{Sp} \;\underset{F_H(G_+,-)}{\overset{G_+\wedge_H -}{\leftrightarrows}}\; \xrightarrow{\quad i^* \quad} N\text{–}\mathrm{Sp}$$

These are both Quillen pairs with respect to the usual stable model structures on both sides. On the way to obtaining an algebraic model for rational SO(3)–spectra we will relate both the toral and dihedral parts of this category to the corresponding parts for rational $O(2)$–spectra. The natural choice of adjunction between SO(3)–spectra and

$O(2)$–spectra would be the induction and restriction functors. However, this turns out not to be a Quillen adjunction between the toral parts, as we discuss below.

Proposition 2.7 *Suppose $e_{\mathcal{T}}$ is the idempotent in $\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$ corresponding to the characteristic function of the toral part $\mathcal{T}$ (ie all subconjugates of the maximal torus of $\mathrm{SO}(3)$) and $e_{\widetilde{\mathcal{T}}}$ is the idempotent in $\mathsf{A}(O(2))_{\mathbb{Q}}$ corresponding to the characteristic function of the toral part $\widetilde{\mathcal{T}}$ (ie all subconjugates of the maximal torus of $O(2)$). Then*

$$i^*\colon L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \rightleftarrows L_{e_{\widetilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp}) \ :\mathrm{SO}(3)_+ \wedge_{O(2)} -$$

is not a Quillen adjunction.

Proof It is enough to show that $\mathrm{SO}(3)_+ \wedge_{O(2)} -$ does not preserve acyclic cofibrations. This argument is the same as the one in [16, Proposition 4.5], since D_2 is conjugate to C_2 in SO(3) and thus $i^*(e_{\mathcal{T}}) \neq e_{\widetilde{\mathcal{T}}}$. □

Although the adjunction above does not behave well with respect to these model structures, the one with restriction and coinduction does, as is shown in Proposition 2.12 below.

Proposition 2.8 *Suppose $e_{\mathcal{D}}$ is the idempotent of $\mathsf{A}(\mathrm{SO}(3))_{\mathbb{Q}}$ corresponding to all dihedral subgroups of order greater than 4 and all subgroups isomorphic to $O(2)$. Then*

$$i^*\colon L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \rightleftarrows L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp}) \ :\mathrm{SO}(3)_+ \wedge_{O(2)} -$$

is a Quillen adjunction.

Proof The proof follows the same pattern as the proof of [16, Proposition 4.4]. It was a Quillen adjunction before localisation by [18, Chapter V, Proposition 2.3] so the left adjoint preserves cofibrations. It preserves acyclic cofibrations as $\mathrm{SO}(3)_+ \wedge_{O(2)} -$ preserved acyclic cofibrations before localisation and we have a natural (in $O(2)$–spectra X) isomorphism

$$(\mathrm{SO}(3)_+ \wedge_{O(2)} X) \wedge e_{\mathcal{D}}S_{\mathbb{Q}} \cong \mathrm{SO}(3)_+ \wedge_{O(2)} (X \wedge i^*(e_{\mathcal{D}}S_{\mathbb{Q}})). \qquad \square$$

It turns out that the other adjunction — restriction and coinduction adjunction — gives a Quillen pair under general conditions on localisations.

Lemma 2.9 [16, Lemma 4.6] *Suppose G is any compact Lie group, $i\colon H \to G$ is an inclusion of a subgroup and V is an open and closed set in $\mathcal{F}(G)/G$. Then the adjunction*

$$i^*\colon L_{e_V S_{\mathbb{Q}}}(G\text{–Sp}) \rightleftarrows L_{i^*(e_V)S_{\mathbb{Q}}}(H\text{–Sp}) \ :F_H(G_+,-)$$

is a Quillen pair. We use the notation e_V here for the idempotent corresponding to the characteristic function on V.

In the next sections we will repeatedly use this lemma, mainly in situations where after a further localisation of the right-hand side we get a Quillen equivalence. To prepare for that, we distinguish the following two cases.

Corollary 2.10 *Let $\mathcal{D}$ denote the dihedral part of $\mathrm{SO}(3)$ and $e_{\mathcal{D}}$ the corresponding idempotent. Let $i\colon O(2) \to \mathrm{SO}(3)$ be an inclusion. Then*

$$i^*\colon L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \rightleftarrows L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp}) \ :F_{O(2)}(\mathrm{SO}(3)_+,-)$$

is a Quillen adjunction.

Remark 2.11 Note that the idempotent on the right-hand side $i^*(e_{\mathcal{D}})$ corresponds to the dihedral part of $O(2)$ excluding all subgroups D_2 and D_4. Thus $i^*(e_{\mathcal{D}}) = i^*(e_{\mathcal{D}})e_{\widetilde{\mathcal{D}}}$.

Proposition 2.12 *Let $i\colon O(2) \to \mathrm{SO}(3)$ be an inclusion. Then*

$$i^*\colon L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \rightleftarrows L_{e_{\widetilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp}) \ :F_{O(2)}(\mathrm{SO}(3)_+,-)$$

is a strong monoidal Quillen adjunction, where the idempotent on the right-hand side corresponds to the family of all subgroups of $O(2)$ subconjugate to a maximal torus $\mathrm{SO}(2)$ in $O(2)$.

Proof This follows from Lemma 2.9 and the composition of Quillen adjunctions

$$L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \underset{F_{O(2)}(\mathrm{SO}(3)_+,-)}{\overset{i^*}{\rightleftarrows}} L_{i^*(e_{\mathcal{T}})S_{\mathbb{Q}}}(O(2)\text{–Sp}) \underset{\mathrm{Id}}{\overset{\mathrm{Id}}{\rightleftarrows}} L_{e_{\widetilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp}) .$$

Note that $i^*(e_{\mathcal{T}}S_{\mathbb{Q}})$ has nontrivial geometric fixed points not only for all cyclic subgroups of $O(2)$ and $\mathrm{SO}(2)$, but also for D_2, as D_2 is conjugate to C_2 in $\mathrm{SO}(3)$. To ignore that and take into account only the toral part we use the fact that $e_{\widetilde{\mathcal{T}}}i^*(e_{\mathcal{T}}) = e_{\widetilde{\mathcal{T}}}$, which implies that the identity adjunction above is a Quillen pair. □

3 The toral part

In this section we use results from [6] and [4] to obtain an algebraic model for the toral part of rational $\mathrm{SO}(3)$–spectra. The first paper establishes a zig-zag of symmetric monoidal Quillen equivalences between rational $\mathrm{SO}(2)$–spectra, while the second one lifts this comparison to one compatible with the $W = O(2)/\mathrm{SO}(2)$–action to obtain an algebraic model for the toral part of rational $O(2)$–spectra.

We begin by describing the category $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ in Section 3.1 and $d\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ in Section 3.2. Then we proceed to establish the comparison between the toral part of rational SO(3)–orthogonal spectra and its algebraic model, $d\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$.

3.1 Categories $\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ and $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$

Before we are ready to describe the category $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ we have to introduce the category $\mathcal{A}(O(2),\widetilde{\mathcal{T}})$. We give a short description of $\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ as a category on the objects of $\mathcal{A}(\mathrm{SO}(2))$ with W–action. Recall that $W = O(2)/\mathrm{SO}(2)$ is the group of order 2.

Material in this section is based on [9] and [4, Section 3].

Definition 3.1 Let $\mathcal{F}$ denote the family of all finite cyclic subgroups in $O(2)$. Then we define a ring in the category of graded $\mathbb{Q}[W]$–modules

$$\mathcal{O}_{\mathcal{F}} := \prod_{H\in\mathcal{F}} \mathbb{Q}[c_H]$$

where each c_H has degree -2 and w (the nontrivial element of W) acts on each c_H by -1. For simplicity we set $c := c_1$.

We use the notation $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ for the colimit

$$\mathrm{colim}_k\, \mathcal{O}_{\mathcal{F}}[c^{-1}, c_{C_2}^{-1}, \dots, c_{C_k}^{-1}]$$

of localisations, where the maps in the colimit are the inclusions. $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ is an $\mathcal{O}_{\mathcal{F}}$–module using the inclusion

$$\mathcal{O}_{\mathcal{F}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}.$$

Notice that we can perform a similar construction on the ring $\widetilde{\mathcal{O}_{\mathcal{F}}} := (1-e_1)\mathcal{O}_{\mathcal{F}}$ and call it $\widetilde{\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}}$, where e_1 is the projection on the first factor in the ring $\mathcal{O}_{\mathcal{F}}$. Then another way to define $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ is as $\mathbb{Q}[c,c^{-1}]\times\widetilde{\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}}$. This last description of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$ will be useful when we compare this model to the one for the toral part of rational SO(3)–spectra.

Definition 3.2 An object of $\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ consists of a triple (M,V,β) where M is an $\mathcal{O}_{\mathcal{F}}$–module in $\mathbb{Q}[W]$–modules, V is a graded rational vector space with a W–action and β is a map of $\mathcal{O}_{\mathcal{F}}$–modules (in $\mathbb{Q}[W]$–modules)

$$\beta\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$$

such that

($\star$) $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} \beta$ is an isomorphism of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$–modules in $\mathbb{Q}[W]$–modules.

A morphism between two such objects $(\alpha, \phi)\colon (M, V, \beta) \to (M', V', \beta')$ consists of a map of $\mathcal{O}_{\mathcal{F}}$–modules $\alpha\colon M \to M'$ and a map of graded $\mathbb{Q}[W]$–modules such that the relevant square commutes.

Instead of modules over $\mathcal{O}_{\mathcal{F}}$ in $\mathbb{Q}[W]$–modules we can consider modules over $\mathcal{O}_{\mathcal{F}}[W]$ in $\mathbb{Q}$–modules, where $\mathcal{O}_{\mathcal{F}}[W]$ is a group ring with a twisted W–action (namely $wc_H = -c_H w$). We will use this description in the next section. Similarly, $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}[W]$ denotes a group ring with a twisted W–action.

Definition 3.3 An object of $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ is an object of $\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ equipped with a differential, or in other words it consists of a triple (M, V, β) where M is an $\mathcal{O}_{\mathcal{F}}$–module in $\mathsf{Ch}(\mathbb{Q}[W])$, V is an object of $\mathsf{Ch}(\mathbb{Q}[W])$ and β is a map of $\mathcal{O}_{\mathcal{F}}$–modules (in $\mathsf{Ch}(\mathbb{Q}[W])$)

$$\beta\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$$

such that

$(\star)$ $\quad$ $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} \beta$ is an isomorphism of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$–modules in $\mathsf{Ch}(\mathbb{Q}[W])$.

A morphism in this category is a morphism in $\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ which commutes with the differentials.

We proceed to discuss the properties of the category $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$. Firstly, all limits and colimits exist in $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$, by an argument analogous to [6, Definition 2.2.1].

The existence of a model structure on $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ follows from [9, Appendix B].

Theorem 3.4 *There is a stable, proper model structure on the category $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ where the weak equivalences are the homology isomorphisms. The cofibrations are the injections and the fibrations are defined via the right lifting property. We call this model structure the* ***injective model structure****.*

The existence of another, monoidal, model category structure on $d(\mathcal{A}(O(2), \widetilde{\mathcal{T}}))$ was established in [4]. However, since we are not considering monoidality of the algebraic model in this paper, the injective model structure on $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ is enough for our purposes.

3.2 Categories $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ and $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$

Looking at the toral parts of the spaces of subgroups of SO(3) and $O(2)$ we see that the stabiliser of the trivial subgroup is connected in SO(3), while it is not in $O(2)$.

This is a consequence of the fact that the maximal torus is not normal in SO(3) and it is the main ingredient capturing the difference between the algebraic models for the toral part of rational SO(3)–spectra and the toral part of rational $O(2)$–spectra.

Let us denote by $\mathcal{F}_{\mathrm{SO}(3)}$ the family of all finite cyclic subgroups in SO(3). Then we use the simplified notation $\mathcal{O}_{\overline{\mathcal{F}}} := \mathcal{O}_{\mathcal{F}_{\mathrm{SO}(3)}}$, by which we mean a graded ring

$$\mathbb{Q}[d] \times \prod_{(H)\in\mathcal{F}_{\mathrm{SO}(3)}, H\neq 1} \mathbb{Q}[c_{(H)}]$$

where d is in degree -4 and all $c_{(H)}$ are in degree -2. The nontrivial element $w \in W$ acts on it by fixing d and sending $c_{(H)}$ to $-c_{(H)}$ for all subgroups $H \in \overline{\mathcal{F}}_{\mathrm{SO}(3)}$, $H \neq 1$.

We define the ring $\mathcal{O}_{\overline{\mathcal{F}}}[W]$ as a product of $\mathbb{Q}[d]$ (with trivial W–action) and a group ring $(1-e_1)\mathcal{O}_{\overline{\mathcal{F}}}[W]$ with the twisted W–action, that is $wc_{(H)} = -c_{(H)}w$ for $H \in \mathcal{F}_{\mathrm{SO}(3)}$, $H \neq 1$.

Recall that c was the element of the first factor of the ring $\mathcal{O}_{\mathcal{F}}$ (see Definition 3.1). There is an adjunction

$$\mathbb{Q}\text{–mod} \underset{(-)^W}{\overset{\mathrm{Triv}}{\rightleftarrows}} \mathbb{Q}[W]\text{–mod}$$

where $(\mathbb{Q}[c])^W = \mathbb{Q}[d]$ (recall that $\mathbb{Q}[c]$ is the $\mathbb{Q}[W]$–module with W–action given by $wc = -c$). Thus using for example [20, Section 3.3] we have the adjunction

$$\mathbb{Q}[d]\text{–mod in } \mathbb{Q}\text{–mod} \underset{(-)^W}{\overset{\mathbb{Q}[c]\otimes_{\mathbb{Q}[d]}-}{\rightleftarrows}} \mathbb{Q}[c]\text{–mod in } \mathbb{Q}[W]\text{–mod}.$$

This extends to give the following result.

Proposition 3.5 *There is an adjunction*

$$\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} -\colon \mathcal{O}_{\overline{\mathcal{F}}}[W]\text{–mod} \rightleftarrows \mathcal{O}_{\mathcal{F}}[W]\text{–mod} \ :(-)^W \times \mathrm{Id}.$$

Proof The unit of this adjunction is the identity and the counit is the natural inclusion. □

We can compose this adjunction with the usual restriction–induction adjunction

$$\mathcal{O}_{\mathcal{F}}[W]\text{–mod} \underset{\mathrm{res}}{\overset{\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes_{\mathcal{O}_{\mathcal{F}}}-}{\rightleftarrows}} \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}[W]\text{–mod}$$

to get the adjunction

$$\mathcal{O}_{\overline{\mathcal{F}}}[W]\text{–mod} \underset{U}{\overset{\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes_{\mathcal{O}_{\overline{\mathcal{F}}}}-}{\rightleftarrows}} \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}[W]\text{–mod} \tag{3-1}$$

in $\mathbb{Q}$–modules.

We define the category $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ as follows.

Definition 3.6 An object in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ consists of a triple (M, V, β) where M is an $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module in $\mathbb{Q}$–modules, V is a graded rational vector space with a W–action and β is a map of $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules

$$\beta\colon M \to U(\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V)$$

such that the adjoint (using (3-1)) satisfies

($\star$) $\quad \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$ is an isomorphism of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}[W]$–modules.

A morphism between two such objects $(\alpha, \phi)\colon (M, V, \beta) \to (M', V', \beta')$ consists of a map of $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules $\alpha\colon M \to M'$ and a map of graded $\mathbb{Q}[W]$–modules such that the relevant square commutes.

Notice that the condition on the map β implies that the image of $e_1 M$ must lie in $(\mathbb{Q}[c, c^{-1}] \otimes V)^W$, ie in W–fixed points. From now on we will abuse the notation slightly and leave out the functor U (3-1) in the codomain of β in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

Remark 3.7 There are no idempotents in the category $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$; however, the category of $\mathcal{O}_{\overline{\mathcal{F}}}$–modules can be split, for example as $\mathbb{Q}[d]$–mod $\times\, (1-e_1)\mathcal{O}_{\overline{\mathcal{F}}}$–mod. We will use that property in the proof of Proposition 3.9.

Definition 3.8 An object of $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ consists of an $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module M equipped with a differential and a chain complex of $\mathbb{Q}[W]$–modules V together with a map of $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules $\gamma\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$ which commutes with differentials. A differential on a $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module M consists of maps $d_n\colon M_n \to M_{n-1}$ such that $d_{n-1} \circ d_n = 0$ and $\overline{c} d_n = d_{n-2}\overline{c}$, where $\overline{c}$ consists of elements $c_{(H)}$ on the H–factor, for all $(H) \in \overline{\mathcal{F}}$, $H \neq 1$, and 0 on the first factor, and where $\overline{d} d_n = d_{n-4}\overline{d}$; here $\overline{d}$ is d on the first factor and 0 everywhere else in the product.

A morphism in this category is a morphism in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ which commutes with the differentials.

We proceed to study the adjunction relating $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ and $\mathcal{A}(O(2), \widetilde{\mathcal{T}})$.

Proposition 3.9 *We have the following adjunction, where the adjoints are defined in the proof:*

$$F\colon \mathcal{A}(\mathrm{SO}(3), \mathcal{T}) \rightleftarrows \mathcal{A}(O(2), \widetilde{\mathcal{T}}) \;:R.$$

Proof Take $X = (\gamma\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V)$ in $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$. Then define

$$F(X) := (\overline{\gamma}\colon \mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V),$$

where $\overline{\gamma}$ is the adjoint of γ (since $\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} -$ is a left adjoint from $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules to $\mathcal{O}_{\mathcal{F}}[W]$–modules; see Proposition 3.5). It is easy to see that this construction gives an object in $\mathcal{A}(O(2), \widetilde{\mathcal{T}})$, ie that it satisfies the condition $(\star)$ from Definition 3.2. Since $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} \overline{\gamma}$ agrees with $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} \gamma$ in $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}[W]$–modules, it is an isomorphism by condition $(\star)$ from Definition 3.6.

Now take $Y = (\delta \colon N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U)$ in $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$. Then define

$$R(Y) := \big(\delta \circ i \colon (e_1 N)^W \times (1 - e_1)N \to N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U\big),$$

where i is the inclusion.

To see that $R(Y) \in \mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, we show the adjoint condition $(\star)$ from Definition 3.6 holds for $\delta \circ i$.

Thus we want to show that

$$\overline{\delta \circ i} \colon \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} ((e_1 N)^W \times (1 - e_1)N) \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U$$

is an isomorphism of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}[W]$ modules.

Notice that we have a natural map

$$\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} (\varepsilon_N) \colon \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} ((e_1 N)^W \times (1 - e_1)N) \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} (N)$$

where ε is the counit of the adjunction from Proposition 3.5.

After applying e_1, the map $e_1 \varepsilon_N$ is an isomorphism for finitely generated modules N. Since every module is a colimit of finitely generated ones and $\otimes$ commutes with colimits, $e_1 \varepsilon_N$ is an isomorphism for any N. Since ε_N is an isomorphism away from e_1 it is an isomorphism. To complete the argument notice that the diagram

$$\begin{array}{ccc} \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\overline{\mathcal{F}}}} ((e_1 N)^W \times (1 - e_1)N) & \xrightarrow{\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} (\varepsilon_N)} & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes_{\mathcal{O}_{\mathcal{F}}} (N) \\ & \searrow^{\overline{\delta \circ i}} & \downarrow \overline{\delta} \\ & & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U \end{array}$$

commutes, where $\overline{\delta}$ is the adjoint of δ (see Proposition 3.5).

It is easy to see that this is an adjoint pair, since the unit is the identity and the counit is the pair of maps $(\varepsilon, \mathrm{Id})$ and the identity on graded $\mathbb{Q}[W]$–modules. Here ε is the counit of the adjunction in Proposition 3.5. □

Proposition 3.10 *All small limits and colimits exist in* $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

Proof Suppose we have a diagram of objects $M_i \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V_i$ in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ indexed by a category I. The colimit of this diagram is

$$\mathrm{colim}_i\, M_i \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes (\mathrm{colim}_i\, V_i).$$

If the diagram is finite, than the limit is formed in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ in a similar way:

$$\lim_i M_i \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes (\lim_i V_i).$$

To construct infinite limits in a category $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ we use the same method as in [6, Definition 2.2.1]. However, since we don't use the construction of infinite limits anywhere in this paper, we skip the technicalities.

Verifying that these constructions define limits and colimits in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ is routine. □

Let $g\mathbb{Q}[W]$–mod denote the category of graded $\mathbb{Q}[W]$–modules. Recall that an $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module M is $\overline{\mathcal{F}}$–finite if it is a direct sum of its submodules $e_{(H)}M$:

$$M = \bigoplus_{(H)\in\overline{\mathcal{F}}} e_{(H)}M,$$

and let tors–$\mathcal{O}_{\overline{\mathcal{F}}}[W]^f$–mod denote the category of $\overline{\mathcal{F}}$–finite torsion $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules. We define two functors relating $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ to some simpler categories, which will allow us to create two classes of injective objects in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

Definition 3.11 Define the functor e: $g\mathbb{Q}[W]\text{–mod} \to \mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ by

$$e(V) := (P \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V),$$

where

$$e_1 P = \mathbb{Q}[d, d^{-1}] \otimes V^+ \oplus \Sigma^2 \mathbb{Q}[d, d^{-1}] \otimes V^- \quad \text{and} \quad (1-e_1)P = (1-e_1)\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V.$$

Here V^+ is the W–fixed part of V, V^- is the -1 eigenspace and Σ is the suspension. The structure map is essentially just an inclusion.

Define a functor f: tors–$\mathcal{O}_{\overline{\mathcal{F}}}[W]^f\text{–mod} \to \mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ by

$$f(N) := (N \to 0).$$

The domain for this functor was chosen so that $f(N) \in \mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, that is, it satisfies condition $(\star)$ from Definition 3.6.

Proposition 3.12 *For any object* $X = (\gamma\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes U)$ *in* $\mathcal{A} = \mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, *any* V *in* $\mathbb{Q}[W]$–mod *and any* N *in* tors–$\mathcal{O}_{\overline{\mathcal{F}}}[W]^f$–mod, *we have natural isomorphisms*

$$\mathrm{Hom}_{\mathcal{A}}(X, e(V)) = \mathrm{Hom}_{\mathbb{Q}[W]}(U, V),$$
$$\mathrm{Hom}_{\mathcal{A}}(X, f(N)) = \mathrm{Hom}_{\mathcal{O}_{\overline{\mathcal{F}}}[W]}(M, N).$$

Remark 3.13 This proposition implies that an object $e(V)$ is injective for any V and that if N is an injective $\overline{\mathcal{F}}$–finite torsion $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module then $f(N)$ is also injective.

Lemma 3.14 *The category* $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ *is a (graded) abelian category of injective dimension* 1. *Moreover it is split, ie every object* X *of* $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ *has a splitting* $X = X_+ \oplus X_-$ *such that* $\mathrm{Hom}(X_\delta, Y_\epsilon) = 0$ *and* $\mathrm{Ext}(X_\delta, Y_\epsilon) = 0$ *if* $\delta \neq \epsilon$ *and* $(\Sigma X)_+ = \Sigma(X_-)$ *and* $(\Sigma X)_- = \Sigma(X_+)$.

Proof The category $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ is enriched in abelian groups and by construction of all limits and colimits we can conclude that it is an abelian category.

For an object $X = (\gamma\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V)$ we construct the injective resolution of length 1 as follows. Let $TM := \ker\gamma$, which is torsion. Thus, since $\mathbb{Q}[d]$ and all $\mathbb{Q}[c_{(H)}][W]$ are of injective dimension 1, there is an injective resolution of $\overline{\mathcal{F}}$–finite torsion $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules

$$0 \to TM \to I' \to J' \to 0,$$

where I', J' are injective $\overline{\mathcal{F}}$–finite torsion $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules.

Let us use simplified notation below. Let P denote the $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module from the definition of $e(V)$ (see Definition 3.11).

If Q is the image of γ then $J'' = P/Q$ is divisible and an $\overline{\mathcal{F}}$–finite torsion $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–module and hence injective. We form a diagram of $\mathcal{O}_{\overline{\mathcal{F}}}[W]$–modules

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & TM & \longrightarrow & M & \longrightarrow & Q & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & I' & \longrightarrow & I' \oplus P & \longrightarrow & P & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & J' & \longrightarrow & J' \oplus J'' & \longrightarrow & J'' & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

where the middle vertical column is obtained using the horseshoe lemma (see for example [23, Lemma 2.2.8]), since left and right vertical columns are injective resolutions of TM and Q, respectively. Thus we get a diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M & \longrightarrow & I' \oplus P & \longrightarrow & J' \oplus J'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\gamma} & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V & \longrightarrow & \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V & \longrightarrow & 0 & \longrightarrow & 0 \end{array}$$

which is the required resolution of $\gamma\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$ in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

Finally, the splitting is given by taking the even- and odd-graded parts. This satisfies the required conditions since the resolution above of an object X_δ is entirely in parity δ. □

3.3 Model category $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$

In this section we will concentrate on the model category $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ and we will investigate its properties. First notice that all constructions from the previous section (limits and colimits, adjoints F and R) pass naturally to the category $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

By the results of the previous section and [9, Proposition 4.1.3] we can construct the derived category of $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ by taking objects with differential in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ and inverting the homology isomorphisms.

Theorem 3.15 *There is an injective model structure on the category* $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ *where weak equivalences are homology isomorphisms and cofibrations are monomorphisms.*

Proof Since the category $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ is abelian of injective dimension 1 we can use the construction from [9, Appendix A]. □

We call $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ with the injective model structure the *algebraic model for toral rational* SO(3)*–spectra*.

To show that the injective model structure is right proper in Proposition 3.19 we need to introduce a class of objects in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ called *wide spheres*. This class generalises the images of representation spheres from rational SO(3)–spectra in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, hence the name.

Definition 3.16 Define $\underline{c}^{2n}$ to be an element of the form $(c^{2n}, c^{2n}, c^{2n}, \dots)$ in $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$. Notice that we can view $\underline{c}^{2n}$ as an element of the form $(d^n, c^{2n}, c^{2n}, \dots)$ in $\mathcal{O}_{\overline{\mathcal{F}}}$ if $n > 0$.

For $n > 0$ define $\underline{c}^{2n+1}$ to be an element of the form $(c^{2n+1}, c^{2n+1}, c^{2n+1}, c^{2n+1}, \dots)$ in $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}$.

Definition 3.17 A *wide sphere* in $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ is an object $P=(S\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes T)$ where T is a graded $\mathbb{Q}[W]$–module which is finitely generated as a $\mathbb{Q}$–module on elements $t_1,\dots,t_d$, where every t_i is either W–fixed or W acts on t_i by -1 and $\deg(t_i)=k_i$. The module S is an $\mathcal{O}_{\overline{\mathcal{F}}}$–submodule of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes T$ generated by elements $\underline{c}^{a_i}\otimes t_1,\dots,\underline{c}^{a_d}\otimes t_d$ where a_i is either even if t_i is W–fixed or odd if W acts on t_i by -1, and an element $\sum_{i=1}^{d}\sigma_i\otimes t_i$ of $\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes T$. It is also required that the structure map be the inclusion. We denote by $\mathcal{P}$ the set of isomorphism classes of wide spheres.

We want to show that there are enough wide spheres in $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$, ie for any $X\in\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ there exists an epimorphism from some coproduct of wide spheres to X.

Proposition 3.18 *There are enough wide spheres in* $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$.

Proof We need to show that for any object $X=(\beta\colon N\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes U)$ in $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ and any $n\in N$ there exists a wide sphere P and a map $P\to X$ such that n is in the image and for any $u\in U$ there exists a wide sphere $\overline{P}$ and a map $\overline{P}\to X$ such that u is in the image. Since the adjoint of β is an isomorphism it is enough to show the above condition for any $n\in N$.

Take $X=(\beta\colon N\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes U)$ in $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ and $n\in N$. Then $\beta(n)=\sum_{i=1}^{d}\sigma_i\otimes t_i$. We may assume that for every i, either t_i is W–fixed or W acts on t_i by -1. Then notice that since $e_1\beta(n)$ is W–fixed, $e_1\sigma_i$ will be of the form c^{2k} if t_i was W–fixed or c^{2k+1} if W acts on t_i by -1 (k is some integer here).

For each i, there exist $p_i\in N$ such that $\beta(p_i)=\underline{c}^{2b_i}\otimes t_i$ if t_i was W–fixed or $\beta(p_i)=\underline{c}^{2b_i+1}\otimes t_i$ if W acts on t_i by -1. Set $f=(\underline{c})^{2b_1+\cdots+2b_d}$. We may assume that the b_i were large enough that $\sigma_i\underline{c}^{2b_1+\cdots+2b_d}/\underline{c}^{2b_i}$ is in $\mathcal{O}_{\overline{\mathcal{F}}}$ if t_i was W–fixed and $\sigma_i\underline{c}^{-1}\underline{c}^{2b_1+\cdots+2b_d}/\underline{c}^{2b_i}$ is in $\mathcal{O}_{\overline{\mathcal{F}}}$ if W acts on t_i by -1.

Now we have

$$\begin{aligned}\beta\Big(\sum\nolimits^{+}\sigma_i\underline{c}^{2b_1+\cdots+2b_d}/\underline{c}^{2b_i}\,p_i+\sum\nolimits^{-}\sigma_i\underline{c}^{-1}\underline{c}^{2b_1+\cdots+2b_d}/\underline{c}^{2b_i}\,p_i\Big)&=\sum_{i=1}^{d}\sigma_i f\otimes t_i\\&=\beta(fn),\end{aligned}$$

where $\sum^{+}$ denotes the sum over all t_i which are W–fixed and $\sum^{-}$ denotes the sum over all the others.

Since the adjoint of β is an isomorphism there exists an element $\underline{c}^{2b}$ such that

$$\underline{c}^{2b}\Big(\sum\nolimits^{+}\sigma_i\underline{c}^{2b_1+\cdots+2b_d}/\underline{c}^{2b_i}\,p_i+\sum\nolimits^{-}\sigma_i\underline{c}^{-1}\underline{c}^{2b_1+\cdots+2b_d}/\underline{c}^{2b_i}\,p_i\Big)=\underline{c}^{2b}fn.$$

We take $\underline{c}^{2b}$ to be the smallest such element.

We take a wide sphere $P = (S \to \mathcal{E}^{-1}_{\mathcal{O}_{\mathcal{F}}} \otimes T)$ where T is a $\mathbb{Q}$–vector space generated by t_i for $i = 1, \dots, d$, $\deg(t_i) = k_i$ and S is an $\mathcal{O}_{\mathcal{F}}$ submodule of $\mathcal{E}^{-1}_{\mathcal{O}_{\mathcal{F}}} \otimes T$ generated by $\sum_{i=1}^{d} \sigma_i \otimes t_i$ and $\underline{c}^{2b} f \otimes t_i$ if t_i is W–fixed and $\underline{c}^{2b-1} f \otimes t_i$ if W acts on t_i by -1. The structure map is the inclusion.

To finish the proof we get a map from P to X by sending $\sum_{i=1}^{d} \sigma_i \otimes t_i$ to n and $\underline{c}^{2b} f \otimes t_i$ to $\underline{c}^{2b} \underline{c}^{2b_1 + \cdots + 2b_d} / \underline{c}^{2b_i} p_i$ if t_i is W–fixed and $\underline{c}^{2b-1} f \otimes t_i$ to $\underline{c}^{2b-1} \underline{c}^{2b_1 + \cdots + 2b_d} / \underline{c}^{2b_i} p_i$ if W acts on t_i by -1.

The elements $\underline{c}^{2b}$ and f are needed to ensure that the relation between n and the p_i's after applying β is replicated in the wide sphere. □

Proposition 3.19 *The injective model structure on $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ is proper.*

Proof Since cofibrations are the monomorphism it is left proper. To show that it is right proper, notice that among trivial cofibrations there are maps $0 \to D^n \otimes P$, for any $P \in \mathcal{P}$, where $D^n \otimes P$ denotes an object built from P and ΣP with the differential being the identity map from the suspension of P to P. Recall that $\mathcal{P}$ denotes the set of isomorphism classes of wide spheres. Since there are enough wide spheres, the fibrations are in particular surjections. Right properness follows from the fact that in $\mathbb{Q}[W]$–mod and $\mathcal{O}_{\mathcal{F}}[W]$–mod pullbacks along surjections of homology isomorphisms are homology isomorphisms. □

Corollary 3.20 *The category $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ is a Grothendieck category.*

Proof Directed colimits are exact in $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, since they are in R–modules, for any ring R. Thus it remains to show that there is a (categorical) generator. We take $J := \bigoplus_{P \in \mathcal{P}} P$, where $\mathcal{P}$ is the set of all wide spheres. By Proposition 3.18, $\mathrm{Hom}(J, -)$ is faithful and thus J is a categorical generator. □

Next we define a set of objects which will be generators for the homotopy category of $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ with the injective model structure. Before we were considering categorical generators, but from now on the meaning of the word *generator* is as in Definition 2.5. Recall that if $\beta\colon M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$ is an object in $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, then M is in particular a module over $\mathcal{O}_{\mathcal{F}}[W]$ (which is an infinite product over conjugacy classes of cyclic subgroups in $\mathrm{SO}(3)$; see beginning of Section 3.2).

Definition 3.21 We define a set $\mathcal{K}$ in $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ to consist of all suspensions and desuspensions of the following objects:

- For the trivial subgroup, we take
$$\sigma_1 := (\mathbb{Q}_1 \to 0),$$
where $\mathbb{Q}$ is at the place indexed by the trivial subgroup and all other factors are 0.
- For every $H \in \overline{\mathcal{F}}$, $H \neq 1$, we take
$$\sigma_H := (\mathbb{Q}[W]_{(H)} \to 0),$$
where $\mathbb{Q}[W]$ is at the place indexed by the conjugacy class of a subgroup H and all other factors are 0.
- For the torus, we take
$$\sigma_T := (M \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W]),$$
where $e_1 M = \mathbb{Q}[d] \oplus \Sigma^2\mathbb{Q}[d]$, $(1-e_1)M = (1-e_1)\mathcal{O}_{\mathcal{F}}$. Here the map is the inclusion.

It remains to show that the set of cells $\mathcal{K}$ is a set of generators for the injective model structure on $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

Theorem 3.22 *The set $\mathcal{K}$ is a set of homotopically compact generators for the category $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ with the injective model structure.*

Proof First note that
$$\sigma_T = (\mathcal{O}_{\overline{\mathcal{F}}} \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}) \oplus (N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \widetilde{\mathbb{Q}}),$$
where $e_1 N = \Sigma^2\mathbb{Q}[d]$ and $(1-e_1)N = (1-e_1)\mathcal{O}_{\mathcal{F}} \otimes \widetilde{\mathbb{Q}}$ (here $\widetilde{\mathbb{Q}}$ denotes $\mathbb{Q}$ with the action of w by -1), and both structure maps are inclusions. We call the first summand S^0 and the second σ_T^-. Therefore it is enough to show that all suspensions and desuspensions of σ_1, σ_H, σ_T^-, S^0 for all $H \in \mathcal{F}$, $H \neq 1$ form a set of generators. We will call this set $\mathcal{L}$.

All cells are homotopically compact since they are compact and fibrant replacement commutes with direct sums.

We will show that if $[\sigma, X]_*^{\mathcal{A}} = 0$ for all $\sigma \in \mathcal{L}$ then $H_*(X) = 0$ and thus X is weakly equivalent to 0. By Lemma 3.14, [9, Lemma 4.2.4] and [4, Theorem 3.8] we can use the following Adams short exact sequence to calculate the maps in the derived category of $\mathcal{A} = \mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ from X to Y in $d\mathcal{A}$:
$$0 \to \mathrm{Ext}_{\mathcal{A}}(\Sigma H_*(X), H_*(Y)) \to [X, Y]_*^{\mathcal{A}} \to \mathrm{Hom}_{\mathcal{A}}(H_*(X), H_*(Y)) \to 0.$$

Observe that for every $X \in d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$, where

$$X = (\gamma\colon P \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V),$$

we have the fibre sequence

$$\widehat{X} \to X \to e(V),$$

where $e(V)$ is the functor described before Proposition 3.12 and $\widehat{X}$ is the fibre of the map $X \to e(V)$.

By definition, the structure map of $e(V)$ is an inclusion, and thus it is a torsion–free object. To simplify the notation, let

$$E\overline{\mathcal{F}}_+ = (\Sigma^{-2}\mathbb{Q}[d, d^{-1}]/\mathbb{Q}[d] \to 0) \oplus \bigoplus_{\substack{(H)\in\overline{\mathcal{F}} \\ H\neq 1}} ((\Sigma^{-2}\mathbb{Q}[c_{(H)}, c_{(H)}^{-1}]/\mathbb{Q}[c_{(H)}]) \to 0).$$

We call the H–summand in the above formula α_H. Then

$$\widehat{X} \simeq E\overline{\mathcal{F}}_+ \otimes X.$$

Now observe that every summand α_H in $E\overline{\mathcal{F}}_+$ is built as a sequential colimit from suspensions of $\alpha_H^n = (\mathbb{Q}[c_{(H)}]/{c_{(H)}}^n \to 0)$ and inclusions, or if it is the first summand α_1 it is built as a sequential colimit of $\alpha_1^n = (\mathbb{Q}[d]/d^n \to 0)$ and inclusions, and thus

$$[\sigma_K, \widehat{X}]_*^{\mathcal{A}} = [\sigma_K, E\overline{\mathcal{F}}_+ \otimes X]_*^{\mathcal{A}} \cong \Big[\sigma_K, \bigoplus_{(H)}(\alpha_H \otimes X)\Big]_*^{\mathcal{A}} \cong \bigoplus_i [\sigma_K, \alpha_H \otimes X]_*^{\mathcal{A}},$$

where the last isomorphism follows since σ_K is a homotopically compact object. For all H, α_H^n is a strongly dualisable object (by [9, Corollary 2.3.7 and Lemma 2.4.3]), and thus we can proceed:

$$\begin{aligned} [\sigma_K, \alpha_H \otimes X]_*^{\mathcal{A}} &\cong [\sigma_K, \mathrm{colim}_n\, \alpha_H^n \otimes X]_*^{\mathcal{A}} \\ &\cong \mathrm{colim}_i [\sigma_K, \mathrm{Hom}(D(\alpha_H^n), X)]_*^{\mathcal{A}} \\ &\cong \mathrm{colim}_i [D(\alpha_H^n) \otimes \sigma_K, X], \end{aligned} \tag{3-2}$$

since $D(\alpha_H^n) \otimes \sigma_K = 0$ if $K \neq H$ and every $D(\alpha_H^n) \otimes \sigma_H$ is finitely built from σ_H and by assumption $[\sigma, X] = 0$ for all $\sigma \in \mathcal{L}$, we have that $[D(\alpha_H^n) \otimes \sigma_H, X] = 0$ and thus $[\sigma_H, \widehat{X}]_*^{\mathcal{A}} = 0$ for all $H \in \overline{\mathcal{F}}$.

Now take X to be an object in $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ and assume that $[\sigma, X]_*^{\mathcal{A}} = 0$ for all $\sigma \in \mathcal{L}$. By the calculation above it follows that $[\sigma_H, \widehat{X}]_*^{\mathcal{A}} = 0$ for all $H \in \mathcal{F}$.

From the Adams short exact sequence we get that

$$\mathrm{Hom}_{\mathcal{A}}(H_*(\sigma_H), H_*(\widehat{X})) = \mathrm{Hom}_{\mathcal{A}}(\sigma_H, H_*(\widehat{X})) = e_{(H)} H_*(\widehat{X}) = 0.$$

Since $H_*(\widehat{X}) = \bigoplus_{(H)\in\overline{\mathcal{F}}} e_H H_*(\widehat{X})$ we conclude that $\widehat{X}$ is weakly equivalent to 0 and thus $[S^0, \widehat{X}]^{\mathcal{A}}_* = 0$ and $[\sigma_T^-, \widehat{X}]^{\mathcal{A}}_* = 0$.

Now, by the fibre sequence and the fact that every fibre sequence induces a long exact sequence on $[E, -]$ we deduce that $[\sigma, e(V)]^{\mathcal{A}}_* = 0$ for every $\sigma \in \mathcal{L}$. From the Adams short exact sequence it follows that

$$\begin{aligned}\mathrm{Hom}_{\mathcal{A}}(H_*(S^0), H_*(e(V))) &= \mathrm{Hom}_{\mathcal{A}}(S^0, H_*(e(V))) = H_*^+(e(V)) = 0,\\ \mathrm{Hom}_{\mathcal{A}}(H_*(\sigma_T^-), H_*(e(V))) &= \mathrm{Hom}_{\mathcal{A}}(\sigma_T^-, H_*(e(V)) = H_*^-(e(V)) = 0,\end{aligned}$$

where $H_*^+(e(V))$ is the W–fixed part of $H_*(e(V))$ and $H_*^-(e(V))$ denotes the -1 eigenspace. Since $H_*(e(V)) = H_*^+(e(V)) \oplus H_*^-(e(V))$ we get that $e(V)$ is weakly equivalent to 0. Since the fibre sequence induces a long exact sequence in homology we conclude that $H_*(X) = 0$ and thus X is weakly equivalent to 0, which finishes the proof. □

We finish this section by relating $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ and $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$.

Lemma 3.23 *The adjunction*

$$d\mathcal{A}(\mathrm{SO}(3), \mathcal{T}) \underset{R}{\overset{F}{\rightleftarrows}} d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$$

is a Quillen pair when we equip both categories with the injective model structures, where F and R are defined as in the proof of Proposition 3.9.

Proof The left adjoint is exact, so it preserves cofibrations (monomorphisms) and homology isomorphisms. □

Theorem 3.24 *The adjunction*

$$d\mathcal{A}(\mathrm{SO}(3), \mathcal{T}) \underset{R}{\overset{F}{\rightleftarrows}} F(\mathcal{K})\text{–cell–}d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$$

is a Quillen equivalence, where $\mathcal{K}$ is given in Definition 3.21.

Proof We cellularise the left-hand side of the adjunction in Lemma 3.23 at the set $\mathcal{K}$ and the right one at $F(\mathcal{K})$. The left-hand side is then just $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ by Theorem 3.22. Thus to use the cellularisation principle [12, Theorem 2.1] we need to prove that the derived unit is an isomorphism for every element of $\mathcal{K}$. Since the right adjoint preserves all weak equivalences it is enough to show that the categorical unit is a weak equivalence. However, we already know that the unit of this adjunction is the identity (it was shown in the proof of Proposition 3.9). It remains to show that the elements of the set $F(\mathcal{K})$ are homotopically compact in $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ with the injective model structure. This follows from the fact that R preserves coproducts (notice that one component of R is taking W–fixed points and over $\mathbb{Q}$ this is isomorphic to taking W–orbits; the other components of R are identities). This finishes the proof. $\square$

In the next section we will compare the cells coming from the topological generators (see Proposition 3.27) with the ones used for cellularising $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$. For these two sets of cells to agree we now change the set of cells used for cellularising $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$. We introduce the following Quillen self-equivalence (which is also an equivalence of categories) of $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ with the injective model structure. Use the notation $\widetilde{\mathbb{Q}}$ for the $\mathbb{Q}[W]$–module $\mathbb{Q}$ with nontrivial W–action. We denote by $-\otimes\widetilde{\mathbb{Q}}$ a self-adjoint functor on $d\mathcal{A}(O(2),\widetilde{\mathcal{T}})$ defined as

$$-\otimes\widetilde{\mathbb{Q}}(\beta\colon M\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes V):=(\beta\otimes\widetilde{\mathbb{Q}}\colon M\otimes\widetilde{\mathbb{Q}}\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes(V\otimes\widetilde{\mathbb{Q}}).$$

Thus, below, we use the notation $\widetilde{F}$ to denote $-\otimes\widetilde{\mathbb{Q}}\circ F$ and $\widetilde{R}$ to denote $R\circ-\otimes\widetilde{\mathbb{Q}}$.

The final result of this section follows from Theorem 3.24.

Corollary 3.25 *The following is a Quillen equivalence, where $\mathcal{K}$ is as in Definition 3.21 and $d\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ is considered with the injective model structure:*

$$d\mathcal{A}(\mathrm{SO}(3),\mathcal{T}) \underset{\widetilde{R}}{\overset{\widetilde{F}}{\rightleftarrows}} \widetilde{F}(\mathcal{K})\text{–cell–}d\mathcal{A}(O(2),\widetilde{\mathcal{T}}).$$

Remark 3.26 Let us calculate the cells from $\widetilde{F}(\mathcal{K})$ (ignoring suspensions as they work in the same way in both categories):

$$\widetilde{F}(\sigma_1)=\widetilde{F}(\mathbb{Q}_1\to 0)=\widetilde{\mathbb{Q}}\oplus\Sigma^{-2}\mathbb{Q}\to 0,$$

where c sends $\widetilde{\mathbb{Q}}$ to $\mathbb{Q}$ (both copies of $\mathbb{Q}$ are in the place corresponding to the trivial subgroup) and

$$\widetilde{F}(\sigma_{(H)})=\widetilde{F}(\mathbb{Q}[W]_{(H)}\to 0)=\mathbb{Q}[W]_{H},\to 0$$

where the left $\mathbb{Q}[W]$ is in the place corresponding to (H) and the resulting $\mathbb{Q}[W]$ is in the place corresponding to H. This holds for all $(H)\in\overline{\mathcal{F}}$ except for $H=1$. For the torus we have

$$\widetilde{F}(\sigma_{(T)})=\widetilde{F}(M\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes\mathbb{Q}[W])=\Sigma^2\widetilde{\mathbb{Q}}+\mathcal{O}_{\mathcal{F}}\otimes\mathbb{Q}[W]\to\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}}\otimes\mathbb{Q}[W],$$

where c acts on $\widetilde{\mathbb{Q}}$ in degree 2 ($\widetilde{\mathbb{Q}}$ is in the place corresponding to the trivial subgroup) by sending it to $\mathbb{Q}\subseteq\mathbb{Q}[W]$ in degree 0 and the map is the inclusion.

3.4 Restriction to the toral part of rational $O(2)$–spectra

The idea for the comparison is to restrict the toral part of rational SO(3)–spectra to the toral part of rational $O(2)$–spectra using the functor i^* as a left adjoint. Recall that the adjunction $(\mathrm{SO}(3)_+ \wedge_{O(2)} -, i^*)$ is not a Quillen pair for the model categories localised at the idempotents corresponding to the toral parts; see Proposition 2.7.

We use the proof from [4] giving an algebraic model for the toral part of rational $O(2)$–spectra, cellularising every step of the zig-zag of Quillen equivalences presented there. This way we obtain an algebraic model for the toral part of rational $O(2)$–spectra cellularised at the derived images of generators for $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\mathrm{-Sp})$. This gives an algebraic model; however, it is not very explicit. We finish this section by simplifying this category in Theorem 3.35 and showing that it is Quillen equivalent to $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ with the injective model structure.

We start by establishing generators for the toral part of rational SO(3)–spectra. We used the notation $\mathcal{K}$ in Definition 3.21 for the generators on the algebraic side. We will use the notation K for the generators on the topological side. We will end this section by showing that the derived images of the topological generators $\mathrm{im}(K)$ are precisely the algebraic generators $\mathcal{K}$ in $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

Proposition 3.27 *A set K consisting of all suspensions and desuspensions of one* SO(3)*–spectrum*

$$\sigma_n = \mathrm{SO}(3)_+ \wedge_{C_n} e_{C_n} S^0$$

for every natural $n > 0$ and an SO(3)*–spectrum* $\mathrm{SO}(3)/\mathrm{SO}(2)_+$ *is a set of cofibrant homotopically compact generators for the category* $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\mathrm{-Sp})$.

Proof First consider a set L consisting of all suspensions and desuspensions of one SO(3)–spectrum $\mathrm{SO}(3)/C_{n+}$ for every natural $n > 0$ and an SO(3)–spectrum $\mathrm{SO}(3)/\mathrm{SO}(2)_+$. All objects in L are homotopically compact in $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\mathrm{-Sp})$ since they are in $\mathrm{SO}(3)\mathrm{-Sp}$ and fibrant replacement in $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\mathrm{-Sp})$ commutes with coproducts. This is a set of generators for $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\mathrm{-Sp})$ by [18, Chapter IV, Proposition 6.7]. Since

$$\mathrm{SO}(3)/C_{n+} = \bigvee_{C_m \subseteq C_n} \sigma_m,$$

which is a consequence of [9, Lemma 2.1.5], the set K is a set of homotopically compact generators for $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\mathrm{-Sp})$. □

Next we restrict to the toral part of rational $O(2)$–spectra.

Theorem 3.28 *The adjunction*

$$i^*\colon L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \rightleftarrows i^*(K)\text{–cell–}L_{e_{\tilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp}) \ :F_{O(2)}(\mathrm{SO}(3)_+,-)$$

is a Quillen equivalence, where the idempotent on the right-hand side corresponds to the family of all cyclic subgroups of $O(2)$*.*

Proof The fact that this is a Quillen adjunction follows from Proposition 2.12 and the cellularisation principle [12, Theorem 2.1] for K and $i^*(K)$. Since K was a set of generators for the category $L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$, the cellularisation with respect to K will not change this model structure.

All cells from K are homotopically compact and cofibrant by Proposition 3.27. We need to check that their images with respect to i^* are homotopically compact in $L_{e_{\tilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp})$, ie suspension spectra of $\mathrm{SO}(3)/C_{n+}$ for all n and $\mathrm{SO}(3)/\mathrm{SO}(2)_+$ as toral $O(2)$–spectra. It is enough to show that they are homotopically compact as $O(2)$–spectra, which follows from the fact that a smooth, compact G–manifold admits a structure of a finite G–CW complex [15, Theorem I] and a suspension spectrum of a finite G–CW complex is homotopically compact. It thus follows that the images of the summands σ_n are also homotopically compact and cofibrant in $L_{e_{\tilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp})$.

It remains to show that the components of the derived unit maps at generators are weak equivalences. For this, it is enough to check the induced map on the level of homotopy categories. This is equivalent to showing that the derived functor Li^* is an isomorphism on hom-sets. This holds by [10, Theorem 6.1], which states that if $X \cong e_{\mathcal{T}}X$ then Li^* is an isomorphism

$$[X,Y]^{\mathrm{SO}(3)} \to e_{\tilde{\mathcal{T}}}[i^*X, i^*Y]^{O(2)},$$

which implies that

$$Li^*\colon [X,Y]^{L_{e_{\mathcal{T}}}\mathrm{SO}(3)} \cong [e_{\mathcal{T}}X, e_{\mathcal{T}}Y]^{\mathrm{SO}(3)} \to e_{\tilde{\mathcal{T}}}[i^*(e_{\mathcal{T}}X), i^*(e_{\mathcal{T}}Y)]^{O(2)} \cong [i^*X, i^*Y]^{L_{e_{\tilde{\mathcal{T}}}}O(2)}$$

is an isomorphism, where the superscript $L_{e_{\mathcal{T}}}\mathrm{SO}(3)$ was used to mean the homotopy category of $L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$. Similarly, the superscript $L_{e_{\tilde{\mathcal{T}}}}O(2)$ was used to mean the homotopy category of $L_{e_{\tilde{\mathcal{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp})$. Thus the adjunction is a Quillen equivalence. □

Remark 3.29 The result above generalises to any compact Lie group G. The restriction–coinduction adjunction is a Quillen equivalence between the toral part of rational G–spectra and a certain cellularisation of the toral part of rational N–spectra, where N is the normaliser of the maximal torus in G. This is used in [5] to provide an algebraic model for the toral part of rational G–spectra for any compact Lie group G.

Since the Quillen equivalence above provides a link between the toral part of rational SO(3)–spectra and the toral part of rational $O(2)$–spectra we use the result of [4].

Theorem 3.30 [4, Corollary 4.22] *There is a zig-zag of Quillen equivalences between* $L_{e_{\widetilde{\mathcal{T}}} S_{\mathbb{Q}}}(O(2)\text{–Sp})$ *and* $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$*, where* $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ *is considered with the dualisable model structure.*

To provide an algebraic model for rational SO(3)–spectra we need to cellularise every step of the zig-zag from [4, Section 4] with respect to derived images of $i^*(K)$ from Theorem 3.28. Cellularisation preserves Quillen equivalences and gives the following result.

Theorem 3.31 *There is a zig-zag of Quillen equivalences between* $L_{e_{\mathcal{T}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$ *and* $\mathrm{im}(K)\text{–cell–}d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$*, where* $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ *is considered with the dualisable model structure. Here* $\mathrm{im}(K)$ *denotes the derived image under the zig-zag of Quillen equivalences described in* [4, Section 4] *of the set of cells* K *described in Proposition 3.27.*

Theorem 3.31 already gives an algebraic model for the toral part of rational SO(3)–spectra. However, it is not easy to work with a cellularisation of a model category. Thus we show that the model above is Quillen equivalent to the simpler, algebraic category $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ described in Section 3.2. To do this, we first switch to the cellularisation of the injective model structure.

Lemma 3.32 *The identity adjunction between*

$$\mathrm{im}(K)\text{–cell–}d\mathcal{A}(O(2), \widetilde{\mathcal{T}}) \quad \textit{and} \quad \mathrm{im}(K)\text{–cell–}d\mathcal{A}(O(2), \widetilde{\mathcal{T}}),$$

where one $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ *is equipped with the dualisable model structure and the other is equipped with the injective model structure, is a Quillen equivalence.*

Proof The result follows from the fact that the identity adjunction was a Quillen equivalence between $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ with the dualisable model structure and $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ with the injective model structure. □

Lemma 3.33 *The set of homology of elements of* $\mathrm{im}(K)$ *consists of the same objects as* $\widetilde{F}(\mathcal{K})$*, where* $\mathcal{K}$ *is the set described in Definition 3.21 and* $\mathrm{im}(K)$ *denotes the derived image under the zig-zag of Quillen equivalences described in* [4, Section 4] *of the set of cells* K *described in Proposition 3.27.*

Proof First we show that, for every $n>1$, σ_n is weakly equivalent in $L_{e_{\widetilde{\mathfrak{T}}}S_{\mathbb{Q}}}(O(2)\text{–Sp})$ to $O(2)\wedge_{C_n} e_{C_n}S^0$. The map is induced by the inclusion of $O(2)$ into $\mathrm{SO}(3)$ and we will show that it induces an isomorphism on all π_*^H for $H\in\widetilde{\mathfrak{T}}$. We will use the notation $N=O(2)$ and $G=\mathrm{SO}(3)$ below. We have

$$\begin{aligned}\pi_*^H(N\wedge_{C_n} e_{C_n}S^0) &= [N/H_+, F_{C_n}(N_+, S^{L_N(C_n)}\wedge e_{C_n}S^0)]^N\\ &= [N/H_+, S^{L_N(C_n)}\wedge e_{C_n}S^0]^{C_n}.\end{aligned}$$

Here $L_N(C_n)$ is the tangent C_n–representation at the identity coset of N/C_n and thus is the 1–dimensional trivial representation. Since the codomain has only geometric fixed points for $H=C_n$ we get a nonzero result only for $H=C_n$:

$$[\Phi^{C_n}(N/C_{n+}), \Phi^{C_n}(S^{L_N(C_n)})]=[S^1\vee S^1, S^1]=\Sigma(\mathbb{Q}[W]).$$

Similarly we have

$$\begin{aligned}\pi_*^H(G\wedge_{C_n} e_{C_n}S^0) &= [G/H_+, F_{C_n}(G_+, S^{L_G(C_n)}\wedge e_{C_n}S^0)]^G\\ &= [G/H_+, S^{L_G(C_n)}\wedge e_{C_n}S^0]^{C_n},\end{aligned}$$

and since the codomain has only geometric fixed points for $H=C_n$ we get a nonzero result only for $H=C_n$:

$$[\Phi^{C_n}(G/C_{n+}), \Phi^{C_n}(S^{L_G(C_n)})]=[S^1\vee S^1, S^1]=\Sigma(\mathbb{Q}[W]).$$

Notice that $L_G(C_n)$ is 3–dimensional, but it has a 1–dimensional C_n–fixed subspace.

The images of the cells in $\mathcal{A}(O(2),\widetilde{\mathfrak{T}})$ are therefore

$$\mathrm{im}(G\wedge_{C_n} e_{C_n}S^0)=\mathrm{im}(N\wedge_{C_n} e_{C_n}S^0)=(\Sigma\mathbb{Q}[W]_{C_n}\to 0)$$

by [9, Example 5.8.1], where $\Sigma\mathbb{Q}[W]$ is in the place C_n.

Now we will use the functors $\pi_*^{\mathcal{A}}$ described in [9, Theorem 5.6.1 and Lemma 5.6.2]. Since $\mathrm{SO}(3)_+$ is free we get

$$\begin{aligned}\pi_*^{\mathcal{A}}(\mathrm{SO}(3)_+) &= (\pi_*^T(\mathrm{SO}(3)_+)\to 0)\\ &= (\pi_*(\Sigma\,\mathrm{SO}(3)/T_+)\to 0)\\ &= (\pi_*(\Sigma S(\mathbb{R}^3)_+)\to 0)\\ &= (\Sigma^3\widetilde{\mathbb{Q}}\oplus\Sigma\mathbb{Q}\to 0),\end{aligned}$$

where $\Sigma^3\widetilde{\mathbb{Q}}\oplus\Sigma\mathbb{Q}$ is in the place corresponding to the trivial subgroup 1 and c sends $\widetilde{\mathbb{Q}}$ in degree 3 to $\mathbb{Q}$ in degree 1.

Finally, $\mathrm{SO}(3)/T_+ = S(\mathbb{R}^3)_+$ is built as an $O(2)$–space from the cells

$$N/T_+ \vee N/D_{2+} \cup N_+ \wedge e^1.$$

Thus the cofibre sequence

$$N_+ \to N/T_+ \vee N/D_{2+} \to G/T_+$$

gives the long exact sequence

$$\cdots \to (\Sigma\mathbb{Q}[W] \to 0) \to (\mathcal{O}_{\mathcal{F}}[W] \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W]) \oplus (\Sigma\mathbb{Q} \to 0) \to \mathrm{im}(G/T_+) \to \cdots$$

and hence

$$\mathrm{im}(G/T_+) = \Sigma^2\widetilde{\mathbb{Q}} + \mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W] \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W],$$

where c acts on $\widetilde{\mathbb{Q}}$ in degree 2 ($\widetilde{\mathbb{Q}}$ is in the place corresponding to the trivial subgroup) by sending it to $\mathbb{Q} \subseteq \mathbb{Q}[W]$ in degree 0 and the map is the inclusion.

These images are exactly the cells (up to suspension) in $\widetilde{F}(\mathcal{K})$ (see Remark 3.26), which finishes the proof. □

Remark 3.34 It remains to show that the derived images in $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ of generators described in Definition 3.21 are formal, that is, they are weakly equivalent to their homology in $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$. We claim it's clear for $(\Sigma\mathbb{Q}[W]_{C_n} \to 0)$, where $\Sigma\mathbb{Q}[W]$ is in the place C_n. It is also clear for $(\Sigma^3\widetilde{\mathbb{Q}} \oplus \Sigma\mathbb{Q} \to 0)$, where $\Sigma^3\widetilde{\mathbb{Q}} \oplus \Sigma\mathbb{Q}$ is in the place corresponding to the trivial subgroup 1 and c sends $\widetilde{\mathbb{Q}}$ in degree 3 to $\mathbb{Q}$ in degree 1.

To show that $A = (\Sigma^2\widetilde{\mathbb{Q}} + \mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W] \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W])$ is formal (where c acts on $\widetilde{\mathbb{Q}}$ in degree 2 ($\widetilde{\mathbb{Q}}$ is in the place corresponding to the trivial subgroup) by sending it to $\mathbb{Q} \subseteq \mathbb{Q}[W]$ in degree 0 and the structure map is the inclusion) we proceed as follows. Suppose $X = (N \to \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V) \in d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ such that $H_*(X) \cong A$. We want to construct a map $A \to X$ in $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ which is a weak equivalence. We proceed in two parts, using the fact that $\mathbb{Q}[W] \cong \mathbb{Q} \oplus \widetilde{\mathbb{Q}}$ and a $\mathbb{Q}[W]$–map from $\mathbb{Q}[W]$ is determined by the image of $1 \in \mathbb{Q}$ and the image of $1 \in \widetilde{\mathbb{Q}}$.

First, we choose an anti-fixed cycle x in $e_1 N$ representing 1 in $\Sigma^2\widetilde{\mathbb{Q}}$. This determines $c(x) \in \mathbb{Q}$ which represents 1 in homology of $e_1 N$ (it also determines all higher powers of c applied to x). Now we choose a fixed cycle $\overline{x} \in (1-e_1)N$ in degree 0 representing $\overline{1}$ in homology (where $\overline{1}$ is 1 on all places of the infinite product except the first one, where it's 0); $\overline{x}$ is fixed by W. It follows that $(c(x), \overline{x})$ is a cycle in N representing constant (and fixed by W) 1 in $H_0(N)$. The element $(c(x), \overline{x})$ maps to an element $1 \otimes b \in \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$, which represents $1 \otimes 1$ in degree 0 of $H_*(\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V) \cong \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes H_*(V) \cong \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W]$.

Second, we choose an anti-fixed cycle y in N in degree 0 representing a constant element 1 in $H_0(N)$ which is W–anti-fixed. Element y maps into an element $1 \otimes k \in \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V$ representing the anti-fixed $1 \otimes 1$ in degree 0 of $H_*(\mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes V) \cong \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes H_*(V) \cong \mathcal{E}^{-1}\mathcal{O}_{\mathcal{F}} \otimes \mathbb{Q}[W]$. The choices of $x, \bar{x}$ and y determine a map in $d\mathcal{A}(O(2), \widetilde{\mathcal{T}})$ which is clearly a homology isomorphism.

Theorem 3.35 *The adjunction*

$$\widetilde{F}\colon \mathrm{d}\mathcal{A}(\mathrm{SO}(3), \mathcal{T}) \rightleftarrows \mathrm{im}(K)\text{–cell–}d\mathcal{A}(O(2), \widetilde{\mathcal{T}}) \ :\widetilde{R}$$

defined after Theorem 3.24 is a Quillen equivalence, where both categories (before cellularisation on the right) are equipped with the injective model structure. Here $\mathrm{im}(K)$ *denotes the derived image under the zig-zag of Quillen equivalences described in* [4, Section 4] *of the set of cells* K *described in Proposition 3.27.*

Proof It is enough to show that $\mathrm{im}(K)$ consists of the same objects (up to a weak equivalence) as $\widetilde{F}(\mathcal{K})$, where $\mathcal{K}$ is the set described in Definition 3.21, which we established in Lemma 3.33 and Remark 3.34. The result follows then from Corollary 3.25. □

We summarise the results of this section.

Theorem 3.36 *There is a zig-zag of Quillen equivalences between* $L_{e_{\mathcal{T}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$ *and* $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$.

4 The dihedral part

The algebraic model for the dihedral part of rational SO(3)–spectra is almost identical to the algebraic model of the dihedral part of rational $O(2)$–spectra presented in [4, Section 5]. The difference comes from two things. First, in SO(3) every dihedral subgroup of order 2, namely D_2, is conjugate to cyclic subgroups C_2 and thus is already taken into account in the toral part. Second, the normaliser of D_4 in SO(3) is a subgroup Σ_4. For those reasons we exclude subgroups conjugate to D_2 and subgroups conjugate to D_4 from the dihedral part $\mathcal{D}$. Excluding D_2 and D_4 from the dihedral part $\mathcal{D}$ allows us to deduce that the information captured by subgroups of SO(3) that are in $\mathcal{D}$ is the same as that captured by subgroups of $O(2)$ that are in $\widetilde{\mathcal{D}} \setminus \{D_2, D_4\}$; see Proposition 4.8. This leads to the reduction of the dihedral part of rational SO(3)–spectra to the (part of the) dihedral part of rational $O(2)$–spectra in Theorem 4.9.

We know from [10] that the model for the homotopy category of the dihedral part of rational SO(3)–spectra is of the form of certain sheaves over an orbit space for $\mathcal{D}$, denoted further by $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$. Section 4.1 discusses this category as well as the category of chain complexes in $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$; $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$. In Section 4.2 we present the comparison between the dihedral part of rational SO(3)–spectra and its algebraic model $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$.

4.1 Categories $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ and $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$

First we recall the construction of $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ (see also [10]), then we present the model structure on $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$ and recall a set of homotopically compact generators for this model category.

Material in this section is based on [4, Section 5.1]. There is a slight difference between the definition of $\mathcal{A}(O(2), \widetilde{\mathcal{D}})$ presented there ($\mathcal{A}(O(2), \mathcal{D})$ is the notation used in [4] for this category) and $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ below, namely we start indexing modules from $k = 3$, which corresponds to $D_6 = D_{2k}$. Indexing in [4] starts from 1.

Let W be the group of order two.

Definition 4.1 Define a category $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ as follows.

An object M consists of a $\mathbb{Q}$–module M_∞, a collection $M_k \in \mathbb{Q}[W]$–mod for $k > 2$ and a map (called the germ map) of $\mathbb{Q}[W]$–modules $\sigma_M\colon M_\infty \to \mathrm{colim}_{n>2} \prod_{k\geqslant n} M_k$, where the W–action on M_∞ is trivial.

A map $f\colon M \to N$ in $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ consists of a map $f_\infty\colon M_\infty \to N_\infty$ of $\mathbb{Q}$–modules and a collection of maps of $\mathbb{Q}[W]$–modules $f_k\colon M_k \to N_k$ which commute with germ maps σ_M and σ_N:

$$\begin{array}{ccc} M_\infty & \xrightarrow{\sigma_M} & \mathrm{colim}_{n>2} \prod_{k\geqslant n} M_k \\ {\scriptstyle f_\infty}\downarrow & & \downarrow{\scriptstyle \mathrm{colim}_{n>2} \prod_{k\geqslant n} f_k} \\ N_\infty & \xrightarrow{\sigma_N} & \mathrm{colim}_{n>2} \prod_{k\geqslant n} N_k \end{array}$$

Definition 4.2 Define $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$ to be the category of chain complexes in $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ and $g\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$ to be the category of graded objects in $\mathcal{A}(\mathrm{SO}(3), \mathcal{D})$.

An object M of $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$ consists of a rational chain complex M_∞, a collection of chain complexes of $\mathbb{Q}[W]$–modules M_k for $k > 2$ and a germ map of chain complexes of $\mathbb{Q}[W]$–modules $\sigma_M\colon M_\infty \to \mathrm{colim}_{n>2} \prod_{k\geqslant n} M_k$, where the W–action on M_∞ is trivial.

Note that we used a chain complex notation here, unlike for the toral part, where we used $d\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$ to mean differential objects in $\mathcal{A}(\mathrm{SO}(3), \mathcal{T})$. The difference between these two is that $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$ is not a graded category, and we introduce a grading taking chain complexes in $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$. On the other hand, $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$ is already graded, and we are interested in differential objects in $\mathcal{A}(\mathrm{SO}(3),\mathcal{T})$.

Remark 4.3 Since the only difference between our definition of $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$ and the one for $\mathcal{A}(O(2),\widetilde{\mathcal{D}})$ lies in index k, all constructions for $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$ are analogous to the ones for $\mathcal{A}(O(2),\widetilde{\mathcal{D}})$ presented in [4].

It is helpful to consider several adjoint pairs involving the category $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$. They are used to get a model structure on $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$.

Definition 4.4 [4, Definition 5.9] Let $A\in\mathsf{Ch}(\mathbb{Q})$, $X\in\mathsf{Ch}(\mathbb{Q}[W])$ and $M\in\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$. For a natural number $k>2$ we define the following functors:

- $i_k\colon \mathsf{Ch}(\mathbb{Q}[W])\to\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$, given by $(i_k(X))_\infty=0$ and $(i_k(X))_n=0$ for $n\neq k$ and $(i_k(X))_k=X$.
- $p_k\colon \mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))\to\mathsf{Ch}(\mathbb{Q}[W])$, given by $p_k(M)=M_k$.
- $c\colon \mathsf{Ch}(\mathbb{Q})\to\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$, given by $(cA)_k=A$, $(cA)_\infty=A$, and where σ_{cA} is the diagonal map into the product.

Then (i_k,p_k), (p_k,i_k) and $(c,\boxplus^W)$ are adjoint pairs, where the functor $\boxplus^W$ is given in [4, Definition 5.6].

The category $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$ is bicomplete by [4, Lemma 5.7] so we can proceed to define a model structure on it.

Proposition 4.5 [4, Proposition 5.10] *There exists a model structure on the category $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$ where f is a weak equivalence or fibration if f_∞ and each of the f_k are weak equivalences or fibrations, respectively. This model structure is cofibrantly generated and proper.*

We call this model structure the *projective model structure* on $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$. By [4, Proposition 5.10] the generating cofibrations are of the form $cI_{\mathbb{Q}}$ and $i_kI_{\mathbb{Q}[W]}$ for $k\geqslant 3$ and generating acyclic cofibrations are of the form $cJ_{\mathbb{Q}}$ and $i_kJ_{\mathbb{Q}[W]}$ for $k\geqslant 3$. Here $I_{\mathbb{Q}}$ and $J_{\mathbb{Q}}$ denote generating cofibrations and generating trivial cofibrations, respectively, for the projective model structure on $\mathsf{Ch}(\mathbb{Q})$, and $I_{\mathbb{Q}[W]}$, $J_{\mathbb{Q}[W]}$ denote generating cofibrations and generating trivial cofibrations, respectively, for the projective model structure on $\mathsf{Ch}(\mathbb{Q}[W])$ (for details see [14, Definition 2.3.3]).

We finish this section by giving a set of homotopically compact generators (recall Definitions 2.5 and 2.4) for $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$.

Lemma 4.6 [4, Lemma 5.11] *The set of objects $\mathcal{G}_a$ consisting of $i_k\mathbb{Q}[W]$ for $k \geqslant 3$ and $c\mathbb{Q}$ is a set of homotopically compact, cofibrant and fibrant generators for the category $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3), \mathcal{D}))$ with the projective model structure.*

4.2 Comparison

First we give homotopically compact, cofibrant generators for $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$. We stick to the convention of writing e_H for $e_{(H)_{\mathrm{SO}(3)}}$.

Lemma 4.7 *The set*

$$\widehat{\mathcal{G}} := \{\mathrm{SO}(3)/O(2)_+\} \cup \{e_{D_{2n}}\mathrm{SO}(3)/D_{2n+} \mid n > 2\}$$

is a set of homotopically compact, cofibrant generators for $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$.

Proof The proof is the same as the proof of [4, Lemma 5.14]. □

To finish the discussion about generators, we show that the restriction functor

$$i^*\colon L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \to L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$$

preserves generators up to weak equivalence.

Proposition 4.8 *Recall that $i^*(e_{\mathcal{D}})$ is the idempotent in $\mathsf{A}(O(2))_{\mathbb{Q}}$ corresponding to the characteristic function on subgroups D_{2n} for $n > 2$ and $O(2)$.*

1. *The map $f\colon O(2)/O(2)_+ \to i^*(\mathrm{SO}(3)/O(2)_+)$ induced by inclusion $O(2) \to \mathrm{SO}(3)$ is a weak equivalence in $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$.*
2. *The map $f_{2n}\colon e_{D_{2n}}O(2)/D_{2n+} \to i^*(e_{D_{2n}}\mathrm{SO}(3)/D_{2n+})$ for $n > 2$ induced by the inclusion $O(2) \to \mathrm{SO}(3)$ is a weak equivalence in $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$.*

Proof To show that the map $f\colon O(2)/O(2)_+ \to i^*(\mathrm{SO}(3)/O(2)_+)$ is a weak equivalence in the given model structure, we need to show that $i^*(e_{\mathcal{D}})f$ is an equivariant rational π_*–isomorphism. Thus we need to check that for all subgroups $H \leq O(2)$ the H–geometric fixed points map

$$\Phi^H(i^*(e_{\mathcal{D}})f)\colon \Phi^H(i^*(e_{\mathcal{D}})O(2)/O(2)_+) \to \Phi^H(i^*(e_{\mathcal{D}})i^*(\mathrm{SO}(3)/O(2)_+))$$

is a nonequivariant rational π_*–isomorphism.

Since taking geometric fixed points commutes with smash product and suspensions, for every subgroup $H \notin (\widetilde{\mathcal{D}} \setminus \{D_2, D_4\})$, $\Phi^H(i^*(e_{\mathcal{D}})f)$ is a trivial map between trivial objects. For $H = O(2)$ the map is an identity on S^0 since $O(2)$ is its own normaliser in SO(3). For other $H \in (\widetilde{\mathcal{D}} \setminus \{D_2, D_4\})$ it is an identity on S^0 since, for each n, there is just one conjugacy class of D_{2n} subgroups in $O(2)$ (and if $g \in \mathrm{SO}(3)$ and $g \notin O(2)$ then $g^{-1} D_{2n} g \not\subset O(2)$).

Part (2) follows the same pattern, however the domain and codomain of the map f_{2n} are already local in $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$, so $f \cong i^*(e_{\mathcal{D}})f$. Since the idempotent used is $e_{D_{2n}}$ the only nontrivial geometric fixed points will be for the subgroup $H = D_{2n}$. The result follows from the fact that $N_{O(2)} D_{2n} = N_{\mathrm{SO}(3)} D_{2n}$, which implies that the map on geometric fixed points for D_{2n} is the identity on D_{4n}/D_{2n+}, and that finishes the proof. □

To give an algebraic model for the dihedral part of rational SO(3)–spectra we firstly use the restriction–coinduction adjunction in the next theorem to move to a certain part of rational $O(2)$–spectra. Then we show that this part of rational $O(2)$–spectra is a localisation of the dihedral part of rational $O(2)$–spectra from [4]. As a result, the method presented in [4] of obtaining an algebraic model for this part applies in our case almost verbatim.

Theorem 4.9 *Let* $i\colon O(2) \to \mathrm{SO}(3)$ *be an inclusion. Then the adjunction*

$$i^*\colon L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp}) \rightleftarrows L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp}) : F_{O(2)}(\mathrm{SO}(3)_+, -)$$

is a Quillen equivalence. (Note that the idempotent on the right-hand side corresponds to the set of all dihedral subgroups of order greater than 4 together with $O(2)$.)

Proof This is a Quillen adjunction by Corollary 2.10 and moreover i^* is a right Quillen functor by Proposition 2.8.

We will use [14, Corollary 1.3.16(c)]. To show that this adjunction is a Quillen equivalence first notice that $F_{O(2)}(\mathrm{SO}(3)_+, -)$ preserves and reflects weak equivalences between fibrant objects. For any fibrant $X \in L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$ and $H \in \widetilde{\mathcal{D}} \setminus \{D_2, D_4\}$ we have natural isomorphisms

$$[\mathrm{SO}(3)/H_+, F_{O(2)}(\mathrm{SO}(3)_+, X)] \cong [i^* \mathrm{SO}(3)/H_+, X] \cong [O(2)/H_+, X],$$

where the second one follows from Proposition 4.8. Since weak equivalences between fibrant objects are detected by H–homotopy groups, $F_{O(2)}(\mathrm{SO}(3)_+, -)$ preserves and reflects weak equivalences between fibrant objects.

We need to show that the derived unit

$$Y \to F_{O(2)}(\mathrm{SO}(3)_+, \widehat{f} i^*(Y))$$

is a weak equivalence on cofibrant objects in $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3){-}\mathrm{Sp})$. It is enough to check that the induced map

$$[X,Y]^{L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)-\mathrm{Sp})} \cong [X, e_{\mathcal{D}}Y]^{\mathrm{SO}(3)} \to [X, F_{O(2)}(\mathrm{SO}(3)_+, \widehat{f}i^*(e_{\mathcal{D}}Y))]^{\mathrm{SO}(3)}$$

is an isomorphism for every generator X of $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3){-}\mathrm{Sp})$ (see Lemma 4.7 for the set of generators). This map fits into the commuting diagram below:

$$\begin{array}{ccc} [X, e_{\mathcal{D}}Y]^{\mathrm{SO}(3)} & & \\ \downarrow & \overset{i^*}{\searrow} & \\ [X, F_{O(2)}(\mathrm{SO}(3)_+, \widehat{f}i^*(e_{\mathcal{D}}Y))]^{\mathrm{SO}(3)} & \xrightarrow{\cong} & [i^*X, \widehat{f}i^*(e_{\mathcal{D}}Y)]^{O(2)} \end{array}$$

Since the horizontal map is an isomorphism it is enough to show that i^* is an isomorphism on hom sets, where the domain is a generator for $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3){-}\mathrm{Sp})$. We do this by using the second Quillen adjunction between these two categories, namely $(\mathrm{SO}(3)_+ \wedge_{O(2)} -, i^*)$.

Let η denote the categorical unit of the adjunction $(\mathrm{SO}(3)_+ \wedge_{O(2)} -, i^*)$. The map η on cofibrant generators is of the form

$$\eta_{e_H O(2)/H_+}\colon e_H O(2)/H_+ \to e_H i^*(\mathrm{SO}(3)/H_+),$$

induced by an inclusion $O(2) \to \mathrm{SO}(3)$. By Proposition 4.8 this is a weak equivalence in $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2){-}\mathrm{Sp})$ for all H in $\mathcal{D}$ and thus $-\circ\eta$ induces an isomorphism in the homotopy category. We have the commuting diagram

$$\begin{array}{ccc} [e_H\, \mathrm{SO}(3)/H_+, e_{\mathcal{D}}Y]^{\mathrm{SO}(3)} & & \\ \cong\downarrow & \overset{i^*}{\searrow} & \\ [e_H O(2)/H_+, i^*(e_{\mathcal{D}}Y)]^{O(2)} & \xleftarrow[-\circ\eta]{} & [i^*(e_H\, \mathrm{SO}(3)/H_+), i^*(e_{\mathcal{D}}Y)]^{O(2)} \end{array}$$

where H above denotes a finite dihedral subgroup or $O(2)$ (when H is $O(2)$ we understand e_H as $e_{\mathcal{D}}$).

It follows that i^* is an isomorphism on hom sets and thus the derived unit of the adjunction where i^* is the left adjoint is a weak equivalence in $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3){-}\mathrm{Sp})$, which finishes the proof. □

To obtain the algebraic model for rational SO(3)–spectra it is enough to get one for $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2){-}\mathrm{Sp})$. We use the comparison method presented in [4] for the dihedral part of rational $O(2)$–spectra in this case.

Theorem 4.10 *There is a zig-zag of Quillen equivalences from* $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$ *to* $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$.

Proof Notice that $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$ is a localisation of the dihedral part of rational $O(2)$–spectra $L_{e_{\widetilde{\mathcal{D}}}S_{\mathbb{Q}}}(O(2)\text{–Sp})$ at an idempotent $i^*(e_{\mathcal{D}})$, since $i^*(e_{\mathcal{D}})e_{\widetilde{\mathcal{D}}} = i^*(e_{\mathcal{D}})$. The set

$$\widetilde{\mathcal{G}} := \{O(2)/O(2)_+\} \cup \{e_{D_{2n}}O(2)/D_{2n+} \mid n > 2\}$$

is a set of homotopically compact, cofibrant generators for $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$ by the same argument as in [4, Lemma 5.14].

Thus it is enough to use the proof of [4, Theorem 5.18] based on the tilting theorem of Schwede and Shipley [21, Theorem 5.1.1] restricted to the set of generators $\widetilde{\mathcal{G}}$ for $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$ on one hand and the set of generators $\mathcal{G}_a$ (see Lemma 4.6) on the algebraic side. This shows that $L_{i^*(e_{\mathcal{D}})S_{\mathbb{Q}}}(O(2)\text{–Sp})$ is Quillen equivalent to the category $\mathsf{Ch}(\mathcal{A}(\mathrm{SO}(3),\mathcal{D}))$. □

Theorem 4.9 and Theorem 4.10 give the algebraic model for the dihedral part of rational SO(3)–spectra.

Theorem 4.11 *There is a zig-zag of Quillen equivalences between* $L_{e_{\mathcal{D}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$ *and* $\mathcal{A}(\mathrm{SO}(3),\mathcal{D})$.

5 The exceptional part

The last part of rational SO(3)–spectra, $L_{e_{\mathcal{E}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$, captures the behaviour of conjugacy classes of five subgroups: $\mathrm{SO}(3)$, Σ_4, A_4, A_5 and D_4; see Section 2.1.

Definition 5.1 [16, Definition 2.1] Recall that a subgroup H of G is *exceptional* if three conditions are satisfied:

- there is an idempotent $e_{(H)} \in \mathsf{A}(G)_{\mathbb{Q}}$ corresponding to the conjugacy class of H,
- the Weyl group $N_G H/H$ of H is finite, and
- H does not contain any subgroup K such that H/K is a (nontrivial) torus.

All subgroups in this part satisfy the definition above, hence the name *exceptional part*.

Recall that the stable model structure on G–spectra is a monoidal model structure satisfying the monoid axiom. Thus any left Bousfield localisation at a cofibrant object E

of a category of G–spectra is again a monoidal model category (by a straightforward check of the pushout-product axiom and the definition of E–weak equivalence). It also satisfies the monoid axiom, since $E \wedge -$ commutes with transfinite compositions and pushouts. By [3, Theorem 4.4] we have the following result.

Proposition 5.2 *There is a strong symmetric monoidal Quillen equivalence*

$$\triangle\colon L_{e_{\mathcal{E}}S_{\mathbb{Q}}}\,\mathrm{SO}(3)\text{–}\mathrm{Sp}_{\mathbb{Q}} \rightleftarrows \textstyle\prod_{(H),H\in\mathcal{E}} L_{e_{(H)_{\mathrm{SO}(3)}}S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–}\mathrm{Sp})\ :\Pi.$$

First we recall some details on what will be the building block of the algebraic model for the exceptional part, ie the category $\mathsf{Ch}(\mathbb{Q}[W_G H])$ of chain complexes of $\mathbb{Q}[W_G H]$–modules, and then we summarise the monoidal comparison from [16].

5.1 The category $\mathsf{Ch}(\mathbb{Q}[W])$

Suppose W is a finite group. The category of chain complexes of left $\mathbb{Q}[W]$–modules can be equipped with the projective model structure, where weak equivalences are homology isomorphisms and fibrations are levelwise surjections. This model structure is cofibrantly generated by [14, Section 2.3].

Note that $\mathbb{Q}[W]$ is not generally a commutative ring, however it is a Hopf algebra with cocommutative coproduct given by $\Delta\colon \mathbb{Q}[W] \to \mathbb{Q}[W]\otimes\mathbb{Q}[W]$, $g \mapsto g\otimes g$. This allows us to define an associative and commutative tensor product on $\mathsf{Ch}(\mathbb{Q}[W])$, namely tensor over $\mathbb{Q}$, where the W–action on the $X\otimes_{\mathbb{Q}} Y$ is diagonal. The unit is a chain complex with $\mathbb{Q}$ at the level 0 with trivial W–action and zeros everywhere else and it is cofibrant in the projective model structure. The monoidal product defined this way is closed, where the internal hom is given by a formula for an internal hom in $\mathbb{Q}$–modules with W–action given by conjugation.

By [2, Proposition 4.3] the category $\mathsf{Ch}(\mathbb{Q}[W])$ is a monoidal model category satisfying the monoid axiom.

5.2 Monoidal comparison

The following result is the main theorem of [16].

Theorem 5.3 *Suppose G is any compact Lie group. Then there is a zig-zag of symmetric monoidal Quillen equivalences from $L_{e_{(H)_G}S_{\mathbb{Q}}}(G\text{–}\mathrm{Sp})$ of rational G–spectra over an exceptional subgroup H to $\mathsf{Ch}(\mathbb{Q}[W_G H])$ equipped with the projective model structure.*

We apply the result above for $G = \mathrm{SO}(3)$ to get the algebraic model for the exceptional part of rational $\mathrm{SO}(3)$–spectra.

Theorem 5.4 *There is a zig-zag of symmetric monoidal Quillen equivalences from* $L_{e_{\mathcal{E}} S_{\mathbb{Q}}}(\mathrm{SO}(3)\text{–Sp})$ *to* $\prod_{(H), H \in \mathcal{E}} \mathsf{Ch}(\mathbb{Q}[W_{\mathrm{SO}(3)} H])$

Proof This follows from Proposition 5.2 and Theorem 5.3. □

Below we present a short sketch of steps in the monoidal comparison for rational G–spectra over an exceptional subgroup to outline general ideas. We refer the reader to [16] for all the details.

Fix an exceptional subgroup H in G. First, using the restriction–coinduction adjunction, we move from the category $L_{e_{(H)_G} S_{\mathbb{Q}}}(G\text{–Sp})$ to the category $L_{e_{(H)_N} S_{\mathbb{Q}}}(N\text{–Sp})$, where N denotes the normaliser $N_G H$. The second step is to use the fixed point–inflation adjunction between $L_{e_{(H)_N} S_{\mathbb{Q}}}(N\text{–Sp})$ and $L_{e_1 S_{\mathbb{Q}}}(W\text{–Sp})$, where W denotes the Weyl group N/H. Recall that W is finite, as H is an exceptional subgroup of G. Next we use the restriction of universe to pass from $L_{e_1 S_{\mathbb{Q}}}(W\text{–Sp})$ to the category $\mathrm{Sp}[W]$ of rational orthogonal spectra with W–action. We then pass to symmetric spectra with W–action using the forgetful functor from orthogonal spectra and then to $H\mathbb{Q}$–modules with W–action in symmetric spectra. From here we use [22, Theorem 1.1] to get to $\mathsf{Ch}(\mathbb{Q})[W]$, the category of rational chain complexes with W–action, which is equivalent as a monoidal model category to $\mathsf{Ch}(\mathbb{Q}[W])$, the category of chain complexes of $\mathbb{Q}[W]$–modules. That gives an algebraic model which is compatible with the monoidal product, ie this zig-zag of Quillen equivalences induces a strong monoidal equivalence on the level of homotopy categories.

References

[1] **D Barnes**, *Rational equivariant spectra*, PhD thesis, University of Sheffield (2008) Available at `http://arxiv.org/abs/0802.0954`

[2] **D Barnes**, *Classifying rational G–spectra for finite G*, Homology Homotopy Appl. 11 (2009) 141–170 MR

[3] **D Barnes**, *Splitting monoidal stable model categories*, J. Pure Appl. Algebra 213 (2009) 846–856 MR

[4] **D Barnes**, *Rational $O(2)$–equivariant spectra*, Homology Homotopy Appl. 19 (2017) 225–252 MR

[5] **D Barnes**, **J P C Greenlees**, **M Kędziorek**, *An algebraic model for rational toral G–spectra for any compact Lie group G* In preparation

[6] **D Barnes**, **J P C Greenlees**, **M Kędziorek**, **B Shipley**, *Rational* SO(2)*–equivariant spectra*, Algebr. Geom. Topol. 17 (2017) 983–1020 MR

[7] **D Barnes**, **C Roitzheim**, *Stable left and right Bousfield localisations*, Glasg. Math. J. 56 (2014) 13–42 MR

[8] **T tom Dieck**, *Transformation groups and representation theory*, Lecture Notes in Math. 766, Springer (1979) MR

[9] **J P C Greenlees**, *Rational S^1–equivariant stable homotopy theory*, Mem. Amer. Math. Soc. 661, Amer. Math. Soc., Providence, RI (1999) MR

[10] **J P C Greenlees**, *Rational* SO(3)*–equivariant cohomology theories*, from "Homotopy methods in algebraic topology" (J P C Greenlees, R R Bruner, N Kuhn, editors), Contemp. Math. 271, Amer. Math. Soc., Providence, RI (2001) 99–125 MR

[11] **J P C Greenlees**, **B Shipley**, *An algebraic model for rational torus-equivariant spectra*, preprint (2011) arXiv

[12] **J P C Greenlees**, **B Shipley**, *The cellularization principle for Quillen adjunctions*, Homology Homotopy Appl. 15 (2013) 173–184 MR

[13] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR

[14] **M Hovey**, *Model categories*, Mathematical Surveys and Monographs 63, Amer. Math. Soc., Providence, RI (1999) MR

[15] **S Illman**, *Existence and uniqueness of equivariant triangulations of smooth proper G–manifolds with some applications to equivariant Whitehead torsion*, J. Reine Angew. Math. 524 (2000) 129–183 MR

[16] **M Kędziorek**, *An algebraic model for rational G–spectra over an exceptional subgroup*, preprint (2015) arXiv

[17] **L G Lewis, Jr**, **J P May**, **M Steinberger**, **J E McClure**, *Equivariant stable homotopy theory*, Lecture Notes in Math. 1213, Springer (1986) MR

[18] **M A Mandell**, **J P May**, *Equivariant orthogonal spectra and S–modules*, Mem. Amer. Math. Soc. 755, Amer. Math. Soc., Providence, RI (2002) MR

[19] **D C Ravenel**, *Localization with respect to certain periodic homology theories*, Amer. J. Math. 106 (1984) 351–414 MR

[20] **S Schwede**, **B Shipley**, *Equivalences of monoidal model categories*, Algebr. Geom. Topol. 3 (2003) 287–334 MR

[21] **S Schwede**, **B Shipley**, *Stable model categories are categories of modules*, Topology 42 (2003) 103–153 MR

[22] **B Shipley**, *$H\mathbb{Z}$–algebra spectra are differential graded algebras*, Amer. J. Math. 129 (2007) 351–379 MR

[23] **C A Weibel**, *An introduction to homological algebra*, Cambridge Studies in Advanced Math. 38, Cambridge Univ. Press (1994) MR

Max Planck Institute for Mathematics
Bonn, Germany

kedziorek@mpim-bonn.mpg.de

Received: 28 November 2016 Revised: 23 March 2017

msp

Algebraic & Geometric Topology 17 (2017) 3137–3187

Betti numbers and stability for configuration spaces via factorization homology

BEN KNUDSEN

Using factorization homology, we realize the rational homology of the unordered configuration spaces of an arbitrary manifold M, possibly with boundary, as the homology of a Lie algebra constructed from the compactly supported cohomology of M. By locating the homology of each configuration space within the Chevalley–Eilenberg complex of this Lie algebra, we extend theorems of Bödigheimer, Cohen and Taylor and of Félix and Thomas, and give a new, combinatorial proof of the homological stability results of Church and Randal-Williams. Our method lends itself to explicit calculations, examples of which we include.

57R19; 17B56, 55R80

1 Introduction

We study the configuration space $B_k(M)$ of k unordered points in a manifold M, defined as

$$B_k(M) = \mathrm{Conf}_k(M)_{\Sigma_k} := \{(x_1, \dots, x_k) \in M^k \mid x_i \neq x_j \text{ for } i \neq j\}/\Sigma_k,$$

where the permutation group Σ_k acts by permuting the x_i. Our main theorem concerns the homology of these spaces.

Theorem 1.1 *Let M be an n–manifold. There is an isomorphism of bigraded vector spaces*

$$\bigoplus_{k \geq 0} H_*(B_k(M); \mathbb{Q}) \cong H^{\mathcal{L}}\big(H_c^{-*}(M; \mathcal{L}(\mathbb{Q}^w[n-1]))\big).$$

Here H_c^{-*} denotes compactly supported cohomology, $\mathbb{Q}^w$ is the orientation sheaf of M, $H^{\mathcal{L}}$ denotes Lie algebra homology and $\mathcal{L}$ is the free graded Lie algebra functor. The auxiliary grading on the left is by cardinality of the configuration and on the right by powers of the Lie generator.

Our methods apply equally to the calculation of the twisted homology of configuration spaces and of the homology of certain relative configuration spaces defined for manifolds

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3137

with boundary; precise statements may be found in Theorem 4.5 and Theorem 4.9, respectively. All results and arguments herein are valid over an arbitrary field of characteristic zero.

The study of configuration spaces is classical. To name some highlights, the space $B_k(\mathbb{R}^2)$ is a classifying space for the braid group on k strands (see Artin [2]); the space $\mathrm{Conf}_k(\mathbb{R}^n)$ has the homotopy type of the space of k–ary operations of the little n–cubes operad and so plays a central role in the theory of n–fold loop spaces (see eg Cohen, Lada and May [17], May [43] and Segal [51]); certain spaces of labeled configurations provide models for more general types of mapping spaces (see Bödigheimer [8], McDuff [44], Salvatore [48], Segal [51]); and, according to a striking theorem of Longoni and Salvatore [39], the homotopy type of $B_k(M)$ is not an invariant of the homotopy type of M.

As this last fact indicates, configuration spaces depend in subtle ways on the structure of the background manifold. On the other hand, the homology of these spaces has often been shown to be surprisingly simple, provided one is willing to work over a field of characteristic zero. Indeed, Bödigheimer, Cohen and Taylor [10] show that the Betti numbers of $B_k(M)$ are determined by those of M when M is of odd dimension, and Félix and Thomas [24] show that, in the even-dimensional case, the Betti numbers of $B_k(M)$ are determined by the rational cohomology ring of M, as long as M is closed, orientable and nilpotent. We recover extensions of these results as immediate consequences of Theorem 1.1.

Corollary 1.2 *The groups $H_*(B_k(M);\mathbb{Q})$ depend only on n and*

- *the graded abelian group $H_*(M;\mathbb{Q})$ if n is odd, or*
- *the cup product $H_c^{-*}(M;\mathbb{Q}^w)^{\otimes 2} \to H_c^{-*}(M;\mathbb{Q})$ if n is even.*

The computational power of Theorem 1.1 lies in the bigrading, which permits one to isolate the homology of a single configuration space within the Chevalley–Eilenberg complex computing the appropriate Lie homology. Employing this strategy, we show that the chain complexes computing $H_*(B_k(M);\mathbb{Q})$ exhibited in [10] and [24] are isomorphic to subcomplexes of the Chevalley–Eilenberg complex; precise statements appear in Section 4.3. Better yet, in dealing with the entire Chevalley–Eilenberg complex at once, one is able to perform computations for all k simultaneously; see Section 6.

Another important aspect of the study of configuration spaces is the phenomenon of *homological stability*. As k tends to infinity, the Betti numbers of $B_k(M)$ are eventually constant, despite the absence of a map of spaces $B_k(M) \to B_{k+1}(M)$ in general; see Church [14], Church, Eilenburg and Farb [15], Randall-Williams [46] and Cantero and Palmer [13]. Here too, characteristic zero is special.

Regarding stability, we prove the following.

Theorem 1.3 *Let M be a connected n–manifold with $n > 1$. The cap product with the unit in $H^0(M;\mathbb{Q})$ induces a map*

$$H_*(B_{k+1}(M);\mathbb{Q}) \to H_*(B_k(M);\mathbb{Q})$$

that is

- *an isomorphism for $* < k$ and a surjection for $* = k$ when M is an orientable surface, and*
- *an isomorphism for $* \leq k$ and a surjection for $* = k+1$ in all other cases.*

The sense in which the homology of configuration spaces forms a coalgebra, so that the cap product is defined, will be explained in Section 5. We lack a conceptual explanation for the exceptional behavior in dimension 2, as it emerges from our argument solely as a numerical/combinatorial coincidence.

This result improves on the stable range of Church [14] and very slightly on that of Randal-Williams [46]. As in the former work, our stable range can be further improved if the low-degree Betti numbers of M vanish. As the example of the Klein bottle shows, the bound $* \leq k$ is sharp in the sense that no better stable range holds for all manifolds that are not orientable surfaces. When M is open, the surjectivity statement is proven in [46]; to the author's knowledge, the result is new for compact manifolds.

Conceptually, we think of Theorem 1.1 as providing an explanation and organizing principle for the behavior of configuration spaces in characteristic zero. The germ of our approach, and the source of the connection to Lie algebras, is the calculation, due to Arnol'd and Cohen, of the homology of the ordered configuration spaces of $\mathbb{R}^n$, which is the fundamental result of the subject; see Arnol'd [1] and Cohen, Lada and May [17]. Specifically, for $n \geq 2$, the homology groups of the spaces $\mathrm{Conf}_k(\mathbb{R}^n)$ form a shifted version of the operad governing Poisson algebras, with the shifted Lie bracket given by the fundamental class of $\mathrm{Conf}_2(\mathbb{R}^n) \simeq S^{n-1}$; see Sinha [52] for a beautiful geometric discussion of this identification. Locally, then, configuration spaces enjoy a rich algebraic structure; factorization homology, our primary tool in this work, provides a means of assembling this structure across coordinate patches of a general manifold, globalizing the calculation of Arnol'd and Cohen. Theorem 1.1 is the natural output of this procedure.

At a more formal level, we rely on the fact that the factorization homology of M, with coefficients taken in a certain free algebra, can be computed in two different ways.

On the one hand, according to Proposition 3.1, it has an expression in terms of the configuration spaces of M. On the other hand, the free algebra may be thought of as a kind of enveloping algebra, and a calculation of the author's in [35] identifies the same invariant as Lie algebra homology. On the face of it, these calculations only coincide for framed manifolds; we show that they agree in general in characteristic zero.

In keeping with our metamathematical goal of making the case for factorization homology as a computational tool, we do not focus on the technical underpinnings of the theory. The interested reader may find these in Ayala and Francis [5; 4; 3], Ayala, Francis and Tanaka [7; 6], Francis [25] and Lurie [42].

The paper is split into seven sections. In Sections 2–3, we review the basics of factorization homology and discuss calculations thereof in several cases of interest. Theorem 1.1 and its variants are proved in Section 4 assuming several deferred results, and the classical results alluded to above follow. In Section 5, we discuss coalgebraic phenomena arising from configuration spaces, which lead us to the proof of Theorem 1.3 and one of the missing ingredients in the main theorem. Finally, Section 6 is concerned with explicit computations, and Section 7 supplies the remaining missing ingredients.

Conventions (1) In accordance with the bulk of the literature on factorization homology, we work in an ∞–categorical context, where for us an ∞–category will always mean a quasicategory. The standard references here are Lurie [40; 42], but we will need to ask only very little of the vast theory developed therein, and the reader may obtain a sense of the arguments and results by substituting "homotopy colimit" for "colimit" everywhere, for example.

(2) Every manifold is smooth and may be embedded as the interior of a compact manifold with boundary (such an embedding is not part of the data). We view manifolds as objects of the ∞–category $\mathcal{M}\mathrm{fld}_n$, the topological nerve of the topological category of n–manifolds and smooth embeddings, which is symmetric monoidal under disjoint union.

(3) Our homology theories are valued in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$, the underlying ∞–category of the category of $\mathbb{Q}$–chain complexes equipped with the standard model structure. With the single exception of Theorem 2.1, $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ is understood to be symmetric monoidal under tensor product.

(4) The homology of a chain complex V is written $H(V)$, while the homology of a space X is written $H_*(X)$. Hence $H_*(X) = H(C_*(X))$. If $\mathfrak{g}$ is a differential graded Lie algebra, then $H(\mathfrak{g})$ is a graded Lie algebra.

(5) Chain complexes are homologically graded. If V is a chain complex, $V[k]$ is the chain complex with $(V[k])_n = V_{n-k}$, and, for $x \in V$, the corresponding element

in $V[k]$ is denoted $\sigma^k x$. Cohomology is concentrated in negative degrees; to reinforce this point, we write $H^{-*}(X)$ for the graded vector space whose degree-$-k$ part is the k^{th} cohomology group of X; for example,

$$H^{-*}(S^n;\mathbb{Q}) \cong \begin{cases} \mathbb{Q} & \text{if } * \in \{-n, 0\}, \\ 0 & \text{otherwise.} \end{cases}$$

(6) If X is a space and V is a chain complex, the *tensor of* X *with* V is the chain complex

$$X \otimes V := C_*(X) \otimes V.$$

(7) If (X, A) is a pair of spaces, the quotient of X by A is the pointed space X/A defined as the pushout in the following diagram:

$$\begin{array}{ccc} A & \longrightarrow & X \\ \downarrow & & \downarrow \\ \mathrm{pt} & \longrightarrow & X/A \end{array}$$

In particular, we have $X/\varnothing = X_+$.

(8) If X is an object of the ∞–category $\mathcal{C}$ with an action of the group G, then X_G and X^G denote the G–coinvariants and G–invariants of X, respectively, which are objects of $\mathcal{C}$. When $\mathcal{C}$ is topological spaces or chain complexes, this object coincides in the homotopy category with *homotopy* coinvariants.

Acknowledgments I am grateful to John Francis for suggesting this project, and for his help and patience.

I would like to thank Lee Cohn, Elden Elmanto, Boris Hanin, Sander Kupers, Jeremy Miller, Martin Palmer and Dylan Wilson for their comments on earlier versions of this paper, Joel Specter and Eric Wofsey for suggested examples, Bruno Vallette for an enlightening lecture, and Jordan Ellenberg and Aaron Mazel-Gee for crucial conversations leading to significant improvements of the results.

After writing the paper, I learned that some of the same results are accessible through the work of Ezra Getzler by first passing to symmetric group invariants in the complex constructed in Getzler [28; 29], and then invoking Proposition 7.6 below. I am grateful to Dan Petersen for bringing these papers to my attention.

This paper is derived from my PhD thesis [36]. Revision was undertaken during visits to the Mathematisches Forschungsinstitut Oberwolfach and the Hausdorff Research Institute for Mathematics in Bonn.

Finally, I would like to thank the referees for their comments.

2 Factorization homology

2.1 Homology theories

In this section, we review the basic notions of factorization homology, also known as topological chiral homology. The primary reference is [5]. As there, our point of view is that factorization homology is a natural theory of homology for manifolds. To illustrate in what sense this is so, we first recall the classical characterization of ordinary homology, phrased in a way that invites generalization.

Theorem 2.1 (Eilenberg–Steenrod axioms) *Let V be a chain complex. There is a symmetric monoidal functor $C_*(-;V)$ from spaces with disjoint union to chain complexes with direct sum, called* ***singular homology with coefficients in V***, *which is characterized up to natural equivalence by the following properties:*

(1) $C_*(\mathrm{pt};V) \simeq V$;

(2) *the natural map*

$$C_*(X_1;V) \bigoplus_{C_*(X_0;V)} C_*(X_2;V) \to C_*(X;V)$$

is an equivalence, where X is the pushout of the diagram of cofibrations

$$X_1 \hookleftarrow X_0 \hookrightarrow X_2.$$

Property (2), a local-to-global principle equivalent to the usual excision axiom, is the reason that homology is computable and hence useful.

Of course, ordinary homology is a homotopy invariant. In the study of manifolds, the equivalence relation of interest is often finer than homotopy equivalence, and one could hope for a theory better suited to such geometric investigations. To discover what form this theory might take, let us contemplate a generic symmetric monoidal functor $(\mathcal{M}\mathrm{fld}_n, \sqcup) \to (\mathcal{C}\mathrm{h}_{\mathbb{Q}}, \otimes)$. By analogy with Theorem 2.1, we ask that this functor be determined by its value on $\mathbb{R}^n$, the basic building block in the construction of n–manifolds. Unlike a point, however, Euclidean space has interesting internal structure.

Definition 2.2 An *n–disk algebra* in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ is a symmetric monoidal functor

$$A\colon (\mathcal{D}\mathrm{isk}_n, \sqcup) \to (\mathcal{C}\mathrm{h}_{\mathbb{Q}}, \otimes),$$

where $\mathcal{D}\mathrm{isk}_n \subseteq \mathcal{M}\mathrm{fld}_n$ is the full subcategory spanned by manifolds diffeomorphic to $\coprod_k \mathbb{R}^n$ for some $k \in \mathbb{Z}_{\geq 0}$.

In other words, $\mathcal{D}\mathrm{isk}_n$ is the (nerve of the) category of operations associated to the endomorphism operad of the manifold $\mathbb{R}^n$, and an n–disk algebra is an algebra over this operad. In contrast, the endomorphism operad of a point in topological spaces is the commutative operad, and every chain complex is canonically and essentially uniquely a commutative algebra in $(\mathcal{C}\mathrm{h}_{\mathbb{Q}}, \oplus)$.

Taking the extra structure of $\mathbb{R}^n$ into account, [5, Theorem 3.24] provides an analogous classification theorem.

Theorem 2.3 (Ayala and Francis) *Let A be an n–disk algebra. There is a symmetric monoidal functor $\int_{(-)} A$ from n–manifolds with disjoint union to chain complexes with tensor product, called* ***factorization homology with coefficients in A****, which is characterized up to natural equivalence by the following properties:*

(1) $\int_{\mathbb{R}^n} A \simeq A$ *as n–disk algebras;*

(2) *the natural map*

$$\int_{M_1} A \bigotimes_{\int_{M_0\times\mathbb{R}} A} \int_{M_2} A \to \int_M A$$

is an equivalence, where M is obtained as the collar-gluing of the diagram of embeddings $M_1 \hookleftarrow M_0 \times \mathbb{R} \hookrightarrow M_2$.

Just as the functor of singular chains is but one model for ordinary homology, factorization homology may be constructed in several equivalent ways. The construction that we will favor is as follows.

Let $A\colon \mathcal{D}\mathrm{isk}_n \to \mathcal{C}\mathrm{h}_{\mathbb{Q}}$ be an n–disk algebra. Then factorization homology with coefficients in A is the left Kan extension in the following diagram of ∞–categories:

$$\begin{array}{ccc} \mathcal{D}\mathrm{isk}_n & \xrightarrow{A} & \mathcal{C}\mathrm{h}_{\mathbb{Q}} \\ \downarrow & \nearrow_{\int_{(-)} A} & \\ \mathcal{M}\mathrm{fld}_n & & \end{array}$$

Explicitly, it may be calculated as the colimit

$$\int_M A \simeq \operatorname{colim}(\mathcal{D}\mathrm{isk}_{n/M} \to \mathcal{D}\mathrm{isk}_n \xrightarrow{A} \mathcal{C}\mathrm{h}_{\mathbb{Q}}).$$

Remark 2.4 Since $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ admits sifted colimits and $\otimes$ distributes over them, Theorem 3.2.3 of [4] guarantees that the left Kan extension and the symmetric monoidal left Kan extension exist and coincide.

2.2 Variant: framed manifolds

The category $\mathcal{D}\mathrm{isk}_n$ is closely related to the classical operad E_n of little n–cubes. To make this connection, we recall that a *framing* of an n–manifold M is a nullhomotopy of its tangent classifier

$$M \xrightarrow{TM \simeq *} BO(n).$$

With the corresponding notion of *framed embedding* between framed manifolds in hand, one obtains an ∞–category $\mathcal{M}\mathrm{fld}_n^{\mathrm{fr}}$ of framed n–manifolds; see [5, Definition 2.7].

Definition 2.5 A *framed n–disk algebra* in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ is a symmetric monoidal functor $A\colon (\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}, \sqcup) \to (\mathcal{C}\mathrm{h}_{\mathbb{Q}}, \otimes)$, where $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}} \subseteq \mathcal{M}\mathrm{fld}_n^{\mathrm{fr}}$ is the full subcategory spanned by framed manifolds diffeomorphic to $\bigsqcup_k \mathbb{R}^n$ for some $k \in \mathbb{Z}_{\geq 0}$.

As before, the factorization homology of a framed n–manifold with coefficients in a framed n–disk algebra is defined as the left Kan extension from $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$. Indeed, the whole theory carries over into the context of topological manifolds equipped with a microtangential B–structure arising from a map $B \to B\mathrm{Top}(n)$. In this paper, we will only make use of the cases $B = BO(n)$, corresponding to smooth manifolds (see [5, Example 2.11 and Remark 3.29]), and $B = *$, corresponding to framed manifolds.

Now, the topological operad E_n has an associated ∞–operad (see [42, Section 2.1]), and [6, Example 2.11] asserts an equivalence

$$\mathrm{Alg}_{\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}}(\mathcal{C}) \xrightarrow{\sim} \mathrm{Alg}_{E_n}(\mathcal{C})$$

for any symmetric monoidal ∞–category $\mathcal{C}$. Moreover, this equivalence induces a further equivalence

$$\mathrm{Alg}_{\mathcal{D}\mathrm{isk}_n}(\mathcal{C}) \xrightarrow{\sim} \mathrm{Alg}_{E_n}(\mathcal{C})^{O(n)}.$$

Informally, an n–disk algebra is an E_n–algebra with an action of $O(n)$ compatible with the action on E_n given by rotating disks. In the language of [49], n–disk algebras are algebras for the *semidirect product* $E_n \rtimes O(n)$.

Remark 2.6 The reader is cautioned not to confuse the framed n–disk algebras employed here with the "framed E_n–algebras" that occur elsewhere in the literature. These algebras carry an action of $\mathrm{SO}(n)$ and yield homology theories for *oriented* manifolds.

2.3 Free algebras

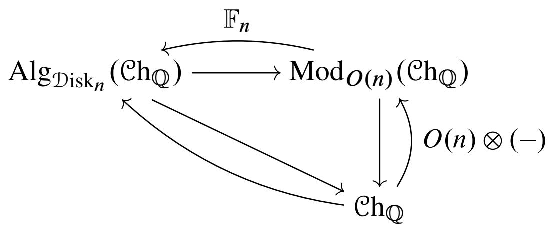
We introduce several functors that will be important for us in what follows. The reference here is [3].

Within the ∞–category $\mathcal{D}\mathrm{isk}_n$ there is a Kan complex with a single vertex, the object $\mathbb{R}^n$, whose endomorphisms are $\mathrm{Emb}(\mathbb{R}^n, \mathbb{R}^n) \simeq O(n)$, so that we may identify this Kan complex with $BO(n)$. Restricting to this subcategory defines a forgetful functor

$$\mathrm{Alg}_{\mathcal{D}\mathrm{isk}_n}(\mathcal{C}\mathrm{h}_{\mathbb{Q}}) \to \mathrm{Fun}(BO(n), \mathcal{C}\mathrm{h}_{\mathbb{Q}}) \xrightarrow{\sim} \mathrm{Mod}_{O(n)}(\mathcal{C}\mathrm{h}_{\mathbb{Q}}).$$

The latter symbol denotes the ∞–category of chain complexes equipped with an action of $C_*(O(n); \mathbb{Q})$, which we refer to simply as *$O(n)$–modules*. This functor admits a left adjoint $\mathbb{F}_n$, the free n–disk algebra generated by an $O(n)$–module.

Evaluation on $\mathbb{R}^n$ defines a still more forgetful functor, which we think of as associating to an algebra its underlying chain complex. The situation is summarized in the following commuting diagram of adjunctions, in which the straight arrows are right and the bent arrows left adjoints:



In particular, for a chain complex V, the free n–disk algebra on V is naturally equivalent to $\mathbb{F}_n(O(n) \otimes V)$. More generally, there is the following description.

Proposition 2.7 *There is a natural equivalence*

$$\mathbb{F}_n(K) \xrightarrow{\sim} \bigoplus_{k \geq 0} \Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big) \otimes_{\Sigma_k \ltimes O(n)^k} K^{\otimes k} \Big),$$

where K is an $O(n)$–module.

Proof The map is supplied by the universal property of the free algebra. In the case $K = O(n) \otimes V$, it is an equivalence, since $\mathbb{F}_n(K)$ is now the free n–disk algebra on the chain complex V, so that

$$\begin{aligned}
\mathbb{F}_n(K) &\simeq \bigoplus_{k \geq 0} \Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big) \otimes_{\Sigma_k} V^{\otimes k} \Big) \\
&\cong \bigoplus_{k \geq 0} \Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big) \otimes_{\Sigma_k \ltimes O(n)^k} \big(O(n)^k \otimes V^{\otimes k} \big) \Big) \\
&\cong \bigoplus_{k \geq 0} \Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big) \otimes_{\Sigma_k \ltimes O(n)^k} K^{\otimes k} \Big).
\end{aligned}$$

Since a general $O(n)$–module may be expressed as a split geometric realization of free $O(n)$–modules, and since $\mathbb{F}_n$, as a left adjoint, preserves geometric realizations, it suffices to show that the right-hand side shares this property. But both $\mathrm{Mod}_{O(n)}(\mathcal{C}\mathrm{h}_{\mathbb{Q}})$ and $\mathrm{Alg}_{\mathcal{D}\mathrm{isk}_n}(\mathcal{C}\mathrm{h}_{\mathbb{Q}})$ are monadic over $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$, so on both sides the geometric realization is computed in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$, and the right-hand side clearly preserves colimits in chain complexes. □

In the framed case, $\mathrm{Emb}^{\mathrm{fr}}(\mathbb{R}^n, \mathbb{R}^n)$ is contractible, so there is only the one forgetful functor

$$\mathrm{Alg}_{\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}}(\mathcal{C}\mathrm{h}_{\mathbb{Q}}) \to \mathcal{C}\mathrm{h}_{\mathbb{Q}},$$

whose left adjoint, the free framed n–disk algebra functor, is denoted $\mathbb{F}_n^{\mathrm{fr}}$.

By restriction along the natural inclusion $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}} \to \mathcal{D}\mathrm{isk}_n$, any n–disk algebra is in particular a framed n–disk algebra, and there is an equivalence of $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebras

$$\mathbb{F}_n(V) \simeq \mathbb{F}_n^{\mathrm{fr}}(V),$$

where V is a chain complex considered as a trivial $O(n)$–module.

3 Calculations

3.1 Frame bundles

The object of this section is twofold. First, we compute the factorization homology of the free n–disk algebra generated by an $O(n)$–module K. Second, for suitable K, we interpret this calculation in terms of the homology of configuration spaces.

For a manifold M, let $\mathrm{Fr}_M \to M$ denote the corresponding principal $O(n)$–bundle. Since $\mathrm{Conf}_k(M)$ is an open submanifold of M^k, its structure group is canonically reducible to $O(n)^k$, and we denote the corresponding principal $O(n)^k$–bundle by $\mathrm{Conf}_k^{\mathrm{fr}}(M)$.

Proposition 3.1 *There is a natural equivalence*

$$\int_M \mathbb{F}_n(K) \xrightarrow{\sim} \bigoplus_{k\geq 0} \big(\mathrm{Conf}_k^{\mathrm{fr}}(M) \otimes_{\Sigma_k \ltimes O(n)^k} K^{\otimes k}\big),$$

where K is an $O(n)$–module.

Proof The natural map

$$\operatorname*{colim}_{\mathcal{D}\mathrm{isk}_{n/M}} \big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big)\big) \xrightarrow{\sim} \mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, M\big)$$

is an equivalence by [42, page 726], so we have

$$\begin{aligned}\int_M \mathbb{F}_n(K) &\simeq \operatorname*{colim}_{\mathcal{D}\mathrm{isk}_{n/M}} \Big(\bigoplus_{k\geq 0}\Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big) \otimes_{\Sigma_k \ltimes O(n)^k} K^{\otimes k}\Big)\Big)\\ &\simeq \bigoplus_{k\geq 0}\Big(\operatorname*{colim}_{\mathcal{D}\mathrm{isk}_{n/M}} \Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, -\big)\Big) \otimes_{\Sigma_k \ltimes O(n)^k} K^{\otimes k}\Big)\\ &\simeq \bigoplus_{k\geq 0}\Big(\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, M\big) \otimes_{\Sigma_k \ltimes O(n)^k} K^{\otimes k}\Big).\end{aligned}$$

To conclude, we note that evaluation at the origin defines a projection

$$\mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, M\big) \to \mathrm{Conf}_k(M),$$

and the natural derivative map $\mathrm{Emb}\big(\bigsqcup_k \mathbb{R}^n, M\big) \to \mathrm{Conf}_k^{\mathrm{fr}}(M)$ covering the identity is an equivalence of $O(n)^k$–spaces over $\mathrm{Conf}_k(M)$. □

Remark 3.2 This proposition is a special case of a calculation carried out in the more general context of *zero-pointed manifolds* in [3, Theorem 2.4.1]. We have included this simplified argument for the reader's convenience.

It will be important in what follows to be able to identify the summand of this object corresponding to a particular choice of k.

Definition 3.3 The *cardinality grading* of the functor $\int_M \mathbb{F}_n(K)$ is the grading corresponding to the direct sum decomposition of Proposition 3.1.

Note that this grading corresponds to the grading induced on the colimit by the cardinality grading of the functor $\mathbb{F}_n(K)$.

We will be most interested in this calculation for particularly simple choices of $O(n)$–module K.

Corollary 3.4 *There is a natural equivalence*

$$\int_M \mathbb{F}_n(\mathbb{Q}) \xrightarrow{\sim} \bigoplus_{k\geq 0} C_*(B_k(M);\mathbb{Q}).$$

Proposition 3.1 can also be used to study the twisted homology of $B_k(M)$. To pursue this direction, we must first identify the orientation cover $\widetilde{B_k(M)}$ of $B_k(M)$. For this we note that the orientation cover

$$\widetilde{\mathrm{Conf}_k(M)} \to \mathrm{Conf}_k(M) \to B_k(M)$$

has structure group $\Sigma_k \times C_2$ when considered as a bundle over $B_k(M)$; that the automorphism corresponding to $-1 \in C_2$ reverses orientation; and that the automorphism corresponding to $\tau \in \Sigma_k$ reverses orientation if $\mathrm{sgn}(\tau) = -1$ and n is odd and preserves orientation otherwise. Therefore, the action of the subgroup

$$H := \{(\tau, \mathrm{sgn}(\tau)^n) \mid \tau \in \Sigma_k\} < \Sigma_k \times C_2$$

is orientation-preserving, and we deduce the following proposition.

Proposition 3.5 *$\widetilde{B_k(M)} \cong \widetilde{\mathrm{Conf}_k(M)}_H$ as covers of $B_k(M)$.*

For a chain complex V, let V^{sgn} denote the sign representation of C_2 on V, and $V^{\det}$ the $O(n)$–module obtained from the latter by restriction along the determinant $O(n) \to C_2$. Recall that, for an n–manifold N, the homology of N twisted by the orientation character may be computed as the homology of the complex

$$C_*(N; \mathbb{Q}^w) := \widetilde{N} \otimes_{C_2} \mathbb{Q}^{\mathrm{sgn}} \cong \mathrm{Fr}_N \otimes_{O(n)} \mathbb{Q}^{\det}.$$

Proposition 3.6 *Let M be an n–manifold.*

(1) *If n is even, there is a natural equivalence*

$$\int_M \mathbb{F}_n(\mathbb{Q}^{\det}) \xrightarrow{\sim} \bigoplus_{k \geq 0} C_*(B_k(M); \mathbb{Q}^w).$$

(2) *If n is odd, there is a natural equivalence*

$$\int_M \mathbb{F}_n(\mathbb{Q}^{\det}[1]) \xrightarrow{\sim} \bigoplus_{k \geq 0} C_*(B_k(M); \mathbb{Q}^w)[k].$$

Proof (1) We have that

$$\begin{aligned}
\mathrm{Conf}_k^{\mathrm{fr}}(M) \otimes_{\Sigma_k \ltimes O(n)^k} (\mathbb{Q}^{\det})^{\otimes k} &\cong \mathrm{Conf}_k^{\mathrm{fr}}(M) \otimes_{\Sigma_k \ltimes O(nk)} \mathbb{Q}^{\det} \\
&\cong \widetilde{\mathrm{Conf}_k(M)} \otimes_{\Sigma_k \times C_2} \mathbb{Q}^{\mathrm{sgn}} \\
&\cong \widetilde{\mathrm{Conf}_k(M)}_{\Sigma_k} \otimes_{C_2} \mathbb{Q}^{\mathrm{sgn}} \\
&\cong \widetilde{B_k(M)} \otimes_{C_2} \mathbb{Q}^{\mathrm{sgn}},
\end{aligned}$$

where we used the commutativity of the diagram

$$\begin{array}{ccc}
O(n)^k & \longrightarrow & O(nk) \\
{\scriptstyle \det^k}\downarrow & & \downarrow{\scriptstyle \det} \\
C_2^k & \xrightarrow{\text{multiply}} & C_2
\end{array}$$

and the fact that $H = \Sigma_k \times \{1\}$ when n is even. The claim follows after summing over k and applying Proposition 3.1.

(2) Similarly, we have that

$$\begin{aligned}\mathrm{Conf}^{\mathrm{fr}}_k(M) \otimes_{\Sigma_k \ltimes O(n)^k} (\mathbb{Q}^{\det}[1])^{\otimes k} &\cong \mathrm{Conf}^{\mathrm{fr}}_k(M) \otimes_{\Sigma_k \ltimes O(nk)} (\mathbb{Q}^{\det} \otimes \mathbb{Q}[1]^{\otimes k}) \\ &\cong \widetilde{\mathrm{Conf}_k(M)} \otimes_{\Sigma_k \times C_2} (\mathbb{Q}^{\mathrm{sgn}} \otimes \mathbb{Q}[1]^{\otimes k}) \\ &\cong \widetilde{\mathrm{Conf}_k(M)}_H \otimes_{C_2} \mathbb{Q}^{\mathrm{sgn}}[k] \\ &\cong \widetilde{B_k(M)} \otimes_{C_2} \mathbb{Q}^{\mathrm{sgn}}[k],\end{aligned}$$

where we used that $\mathbb{Q}^{\mathrm{sgn}} \otimes \mathbb{Q}[1]^{\otimes k}$ is a trivial H–module and $[\Sigma_k \times C_2 : H] = 2$. □

3.2 Commutative algebras

We now consider a calculation of factorization homology in a certain degenerate case, which is a slight generalization of that considered in [5, Proposition 5.1]. We will make use of this calculation in the next section.

Restriction of embeddings defines a map $\mathrm{Emb}\big(\coprod_k \mathbb{R}^n, \mathbb{R}^n\big) \to \prod_k \mathrm{Emb}(\mathbb{R}^n, \mathbb{R}^n) \simeq \prod_k O(n)$, which assemble to form a symmetric monoidal functor

$$\pi\colon \mathcal{D}\mathrm{isk}_n \to BO(n)^{\sqcup},$$

where $BO(n)^{\sqcup}$ is the ∞–category obtained as the nerve of the topological category with objects the natural numbers and morphism spaces given by

$$\mathrm{Map}_{BO(n)^{\sqcup}}(r,s) = \coprod_{f\colon \langle r\rangle \to \langle s\rangle} \prod_{i=1}^{s} O(n)^{f^{-1}(i)},$$

which is symmetric monoidal under addition. For more on this and related ∞–categories, the reader may consult [42, Section 2.4.3]. For us, the relevance of this object is the following consequence of [42, Theorem 2.4.3.18].

Theorem 3.7 (Lurie) *There is an equivalence*

$$\mathrm{Fun}^{\otimes}(BO(n)^{\sqcup}, \mathcal{C}\mathrm{h}_{\mathbb{Q}}) \xrightarrow{\sim} \mathrm{Mod}_{O(n)}(\mathrm{Alg}_{\mathrm{Com}}(\mathcal{C}\mathrm{h}_{\mathbb{Q}})).$$

This result motivates our next definition.

Definition 3.8 A *commutative refinement* of an n–disk algebra A is a factorization

$$\begin{array}{ccc} \mathcal{D}\mathrm{isk}_n & \xrightarrow{A} & \mathcal{C}\mathrm{h}_{\mathbb{Q}} \\ {\scriptstyle \pi}\downarrow & \nearrow\!{\scriptstyle \text{(dashed)}} & \\ BO(n)^{\sqcup} & & \end{array}$$

through a symmetric monoidal functor $BO(n)^{\sqcup} \to \mathcal{C}\mathrm{h}_{\mathbb{Q}}$.

By the previous theorem, a commutative refinement endows the underlying object of A with the structure of a commutative algebra for which the n–disk algebra structure maps are homomorphisms. More formally, we obtain a factorization

$$\begin{array}{ccc} \mathcal{D}\mathrm{isk}_n & \xrightarrow{A} & \mathcal{C}\mathrm{h}_{\mathbb{Q}} \\ {\scriptstyle A_{\mathrm{Com}}}\downarrow & \nearrow & \\ \mathrm{Alg}_{\mathrm{Com}}(\mathcal{C}\mathrm{h}_{\mathbb{Q}}) & & \end{array}$$

of A through the forgetful functor.

Example 3.9 By the Künneth theorem, the functor $H\colon \mathcal{C}\mathrm{h}_{\mathbb{Q}} \to \mathcal{C}\mathrm{h}_{\mathbb{Q}}$ is symmetric monoidal, whence the homology of an n–disk algebra is canonically an n–disk algebra. Since H factors through the discrete ∞–category of graded vector spaces, we have a symmetric monoidal factorization

$$\begin{array}{ccc} \mathcal{D}\mathrm{isk}_n & \xrightarrow{H(A)} & \mathcal{C}\mathrm{h}_{\mathbb{Q}} \\ {\scriptstyle \pi}\downarrow & & \\ BO(n)^{\sqcup} & & \\ \downarrow & \nearrow\!{\scriptstyle \text{(dashed)}} & \\ BC_2^{\sqcup} & & \end{array} \qquad \text{with } \pi_0\colon \mathcal{D}\mathrm{isk}_n \to BC_2^{\sqcup}$$

through the homotopy category of $\mathcal{D}\mathrm{isk}_n$, so that $H(A)$ is canonically commutative.

Definition 3.10 Let X be a topological space and B a commutative algebra. The *tensor* of X and B is the colimit

$$X \otimes B = \mathrm{colim}\big(X \to \mathrm{pt} \xrightarrow{B} \mathrm{Alg}_{\mathrm{Com}}(\mathcal{C}\mathrm{h}_{\mathbb{Q}})\big)$$

of the constant functor from X, viewed as an ∞–groupoid, with value B.

Remark 3.11 When $X = S^1$, this construction has the homotopy type of the Hochschild chains of A. In general, one recovers Pirashvili's higher Hochschild homology.

Let $\mathcal{D}\mathrm{isk}^1_{n/M}$ denote the full subcategory of $\mathcal{D}\mathrm{isk}_{n/M}$ spanned by those arrows $\coprod_k \mathbb{R}^n \to M$ with $k = 1$.

Proposition 3.12 *Suppose that A admits a commutative refinement. There is a natural equivalence*

$$\mathrm{Fr}_M \otimes_{O(n)} A \simeq \mathrm{colim}\big(\mathcal{D}\mathrm{isk}^1_{n/M} \to \mathcal{D}\mathrm{isk}_n \xrightarrow{A_{\mathrm{Com}}} \mathrm{Alg}_{\mathrm{Com}}(\mathcal{C}\mathrm{h}_{\mathbb{Q}})\big).$$

Proof Since the colimit is the left Kan extension to a point, and since Kan extensions compose, we may write

$$\underset{\mathcal{D}\mathrm{isk}^1_{n/M}}{\mathrm{colim}}\ A_{\mathrm{Com}} \simeq \underset{BO(n)}{\mathrm{colim}}\, \pi_! A_{\mathrm{Com}} \simeq (\pi_! A_{\mathrm{Com}})_{O(n)},$$

so that it suffices to identify $\pi_! A_{\mathrm{Com}}$.

Since the projection $\mathcal{D}\mathrm{isk}_{n/M} \to \mathcal{D}\mathrm{isk}_n$ is a left fibration, so is $\pi\colon \mathcal{D}\mathrm{isk}^1_{n/M} \to BO(n)$; in particular, this functor is a co-Cartesian fibration, which implies that the inclusion $\pi^{-1}(\mathrm{pt}) \to \pi_{/\mathrm{pt}}$ of the fiber over the basepoint into the overcategory is a right adjoint and hence final. Therefore, we have

$$\pi_! A_{\mathrm{Com}} = \underset{\pi_{/\mathrm{pt}}}{\mathrm{colim}}\, A_{\mathrm{Com}} \simeq \underset{\pi^{-1}(\mathrm{pt})}{\mathrm{colim}}\ A_{\mathrm{Com}} = \pi^{-1}(\mathrm{pt}) \otimes A$$

According to [5, Corollary 2.13], the ∞–category $\mathcal{D}\mathrm{isk}^1_{n/M}$ is equivalent to the Kan complex M, and the map $\pi\colon \mathcal{D}\mathrm{isk}^1_{n/M} \to BO(n)$ coincides under this identification with the classifying map for the tangent bundle of M. In particular, the fiber of this map is $O(n)$–equivalent to Fr_M, which completes the proof. □

Proposition 3.13 *Suppose that A admits a commutative refinement. There is a natural equivalence*

$$\int_M A \simeq \mathrm{Fr}_M \otimes_{O(n)} A.$$

Proof By the previous proposition, it suffices to show that the inclusion $\mathcal{D}\mathrm{isk}^1_{n/M} \to \mathcal{D}\mathrm{isk}_{n/M}$ and the forgetful functor $\mathrm{Alg}_{\mathrm{Com}}(\mathcal{C}\mathrm{h}_{\mathbb{Q}}) \to \mathcal{C}\mathrm{h}_{\mathbb{Q}}$ induce equivalences

$$\underset{\mathcal{D}\mathrm{isk}^1_{n/M}}{\mathrm{colim}}\ A_{\mathrm{Com}} \xrightarrow{\sim} \underset{\mathcal{D}\mathrm{isk}_{n/M}}{\mathrm{colim}}\ A_{\mathrm{Com}} \xrightarrow{\sim} \underset{\mathcal{D}\mathrm{isk}_{n/M}}{\mathrm{colim}}\ A$$

when A is commutative.

Since $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ is $\otimes$–presentable (see [5, Definition 3.4]), the second equivalence follows from [5, Corollary 3.22], which asserts that $\mathcal{D}\mathrm{isk}_{n/M}$ is sifted, and [42, Corollary 3.2.3.2], which implies that the forgetful functor from commutative algebras preserves sifted colimits.

The first equivalence holds whenever M is framed by [5, Proposition 5.1], since in this case the diagram

$$\begin{array}{ccc} \mathcal{D}\mathrm{isk}^{\mathrm{fr}}_{n/M} & \xrightarrow{\sim} & \mathcal{D}\mathrm{isk}_{n/M} \\ \downarrow & & \downarrow \\ \mathcal{D}\mathrm{isk}^{\mathrm{fr}}_{n} & \longrightarrow & \mathcal{D}\mathrm{isk}_{n} \\ \downarrow & & \downarrow \\ \mathrm{Fin} = \{e\}^{\sqcup} & \longrightarrow & BO(n)^{\sqcup} \end{array}$$

commutes. In particular, the equivalence holds for $M = \coprod_k \mathbb{R}^n$, and we conclude that

$$A \simeq \mathrm{Fr}_{(-)} \otimes_{O(n)} A$$

as n–disk algebras. Therefore, the claim will be established once we are assured that the expression on the right satisfies condition (2) of Theorem 2.3. For this, we note that the functor $\mathrm{Fr}_{(-)}$ takes a collar-gluing of manifolds to a pushout of $O(n)$–spaces, and that the functor $- \otimes_{O(n)} A$ preserves colimits of $O(n)$–spaces. □

3.3 A spectral sequence

We employ a certain "commutative-to-noncommutative" spectral sequence in the proof of Theorem 1.1. For technical reasons, it will be convenient to restrict our attention to n–disk algebras valued in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}^{\geq 0}$, the full subcategory of chain complexes concentrated in nonnegative homological degree. This restriction is not essential.

Proposition 3.14 *Let M be an n–manifold and A an n–disk algebra in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}^{\geq 0}$. There is a natural first-quadrant spectral sequence*

$$E^2_{p,q} \cong H_{p,q}\big(\mathrm{Fr}_M \otimes_{O(n)} H(A)\big) \implies H_{p+q}\bigg(\int_M A\bigg),$$

with differential d^r of bidegree $(-r, r-1)$.

The nature of the bigrading will become clear in the proof.

To construct this spectral sequence, we employ a rigidified version of the overcategory $\mathcal{D}\mathrm{isk}_{n/M}$, denoted $\mathrm{Disj}(M)$ following [42, Chapter 5], which is the poset of those open subsets of M diffeomorphic to $\coprod_k \mathbb{R}^n$ for some k. We refer the reader to [42, Proposition 5.5.2.13] for the proof of the following result.

Proposition 3.15 *There is a final functor* $N(\mathrm{Disj}(M)) \to \mathcal{D}\mathrm{isk}_{n/M}$.

Thus, by [40, Proposition 4.1.1.8], the factorization homology of M may be computed as a colimit over the nerve of the ordinary category $\mathrm{Disj}(M)$. Having achieved this simplification, we proceed as follows. Using the fact that $\mathcal{C}\mathrm{h}_{\mathbb{Q}}^{\geq 0}$ arises from a combinatorial simplicial model category, [40, Proposition 4.2.4.4] implies that any functor $N(\mathrm{Disj}(M)) \to \mathcal{C}\mathrm{h}_{\mathbb{Q}}^{\geq 0}$ of ∞–categories is equivalent in the ∞–category of functors to one coming from a functor of ordinary categories. Having chosen such a "straightening" of A, which we abusively denote by A, [40, Theorem 4.2.4.1] now guarantees that the homotopy colimit of A coincides with the ∞–categorical colimit.

Proof of Proposition 3.14 From the discussion of the previous paragraph and [47, Corollary 5.1.3], we have equivalences

$$\int_M A \simeq \operatorname*{hocolim}_{\mathrm{Disj}(M)} A \simeq B(\mathrm{pt}, \mathrm{Disj}(M), A),$$

where $B(\mathrm{pt}, \mathrm{Disj}(M), A)$ denotes the realization of the simplicial chain complex given in simplicial degree p by

$$B_p(\mathrm{pt}, \mathrm{Disj}(M), A) = \bigoplus_{U_p \to \cdots \to U_0 \to M} A(U_p)$$

(here we use for a second time the fact that the model structure on nonnegatively graded chain complexes is simplicial). Filtering by skeleta in the usual way, we obtain a spectral sequence

$$E^1_{p,q} = \bigoplus_{U_p \to \cdots \to U_0 \to M} H_q(A(U_p)) \implies H_{p+q}\left(\int_M A\right),$$

with the differential d^1 given by the alternating sum of the face maps (see [50, Proposition 5.1], for example, which treats the case of a simplicial space). In other words, the E^1 page is the (graded) chain complex associated to the (graded) simplicial chain complex $B_\bullet(\mathrm{pt}, \mathrm{Disj}(M), H(A))$ via the Dold–Kan correspondence, so that, invoking Proposition 3.13, we have natural isomorphisms

$$E^2_{p,q} \cong H_{p,q}(B(\mathrm{pt}, \mathrm{Disj}(M), H(A))) \cong H_{p,q}\left(\int_M H(A)\right) \cong H_{p,q}(\mathrm{Fr}_M \otimes_{O(n)} H(A)). \quad \square$$

Remark 3.16 Horel discusses a version of this spectral sequence in [33, Section 5].

3.4 Enveloping algebras

In this section, we outline the place of Lie algebras in the theory of factorization homology, the general reference for which is [35].

It has long been known that configuration spaces are intimately related to Lie algebras; see [17; 18; 16], for example. To see the connection, suppose that A is a $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebra in chain complexes, with $n \geq 2$. Part of the structure of such an object is a multiplication map

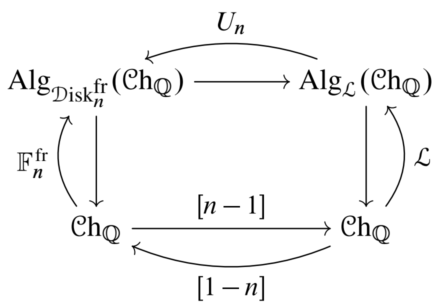
$$m\colon \mathrm{Emb}^{\mathrm{fr}}\big(\textstyle\coprod_2 \mathbb{R}^n, \mathbb{R}^n\big) \otimes A^{\otimes 2} \to A,$$

and since the homology of $\mathrm{Emb}^{\mathrm{fr}}\big(\coprod_2 \mathbb{R}^n, \mathbb{R}^n\big) \simeq \mathrm{Conf}_2(\mathbb{R}^n) \simeq S^{n-1}$ is concentrated in degrees 0 and $n-1$, this multiplication encodes two maps

$$m_0\colon A^{\otimes 2} \to A \quad \text{and} \quad m_{n-1}\colon A^{\otimes 2} \to A[1-n]$$

defining a commutative multiplication on A and a Lie bracket on $A[n-1]$, again up to homotopy. The Jacobi identity for m_{n-1} follows from the three-term or Yang–Baxter relations in $H_*(\mathrm{Conf}_3(\mathbb{R}^n))$ (see [20]), and $O(n)$, acting on S^{n-1} by degree ± 1 maps, interchanges it with the opposite bracket.

The fact that this discussion illustrates is the existence of a forgetful functor from $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebras to Lie algebras at the level of ∞–categories. Indeed, according to [35], there is the following commuting diagram of adjunctions:



Here $\mathcal{L}$ denotes the free Lie algebra functor.

The $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebra $U_n(\mathfrak{g})$ is known as the *n–enveloping algebra* of $\mathfrak{g}$; see [31, Section 4.6] for a discussion of the identification between U_1 and the usual universal enveloping algebra. The factorization homology of these algebras is computed in [35].

Theorem 3.17 (Knudsen) *There is a natural equivalence*

$$\int_M U_n(\mathfrak{g}) \xrightarrow{\sim} C^{\mathcal{L}}(\mathfrak{g}^{M^+}).$$

We pause briefly to explain the terms of the theorem.

(1) The ∞–category of differential graded Lie algebras has limits and is therefore cotensored over pointed spaces; we denote by $\mathfrak{g}^X$ the cotensor of the pointed space X with the Lie algebra $\mathfrak{g}$. A model for this object is provided by [32, Lemma 4.8.3].

Proposition 3.18 *Let X be a pointed finite CW complex. There is a natural equivalence*

$$\mathfrak{g}^X \simeq A_{\mathrm{PL}}(X) \otimes \mathfrak{g}.$$

Here A_{PL} denotes the functor of reduced piecewise-linear de Rham forms (see [22, Section 10(c)], for example), and the right-hand side carries the canonical Lie bracket on the tensor product of a nonunital commutative algebra and a Lie algebra, which is defined by the formula

$$[a \otimes v, b \otimes w] = (-1)^{|v||b|} ab \otimes [v, w].$$

(2) The symbol $C^{\mathcal{L}}$ denotes the functor of Lie algebra chains. This coaugmented cocommutative coalgebra is defined abstractly via the monadic bar construction against the free Lie algebra monad, but it has a concrete incarnation as the *Chevalley–Eilenberg complex*

$$\mathrm{CE}(\mathfrak{g}) = (\mathrm{Sym}(\mathfrak{g}[1]), d_{\mathfrak{g}} + D),$$

where D is defined as a coderivation by specifying that

$$D(\sigma x \wedge \sigma y) = (-1)^{|x|} \sigma[x, y].$$

See [27, Section 6] for a discussion of the comparison between the monadic bar construction and the Chevalley–Eilenberg complex. We remark that $\mathrm{CE}(\mathfrak{g})$ is a coaugmented cocommutative differential graded coalgebra, and the resulting coproduct on $H^{\mathcal{L}}(\mathfrak{g})$ coincides with the one inherited from the monadic bar construction; indeed, both are induced by the diagonal $\mathfrak{g} \to \mathfrak{g} \oplus \mathfrak{g}$, which is a map of Lie algebras.

The equivalence of Theorem 3.17 specializes to a natural equivalence

$$U_n(\mathfrak{g}) \simeq C^{\mathcal{L}}(\mathfrak{g}^{(\mathbb{R}^n)^+})$$

of $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebras. In this way, Theorem 3.17 can be thought of as identifying an n–disk algebra refinement of the $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebra $U_n(\mathfrak{g})$, so that the expression $\int_M U_n(\mathfrak{g})$ is sensible for manifolds M that are not necessarily framed.

Returning to the discussion that began this section, if A is now an n–disk algebra rather than merely a $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebra, then A determines a shifted Lie algebra in $O(n)$–modules, but now with $O(n)$ acting on the suspension coordinates. A full discussion of this phenomenon and the corresponding enveloping algebra is beyond the scope of this paper. Since the analogue of Theorem 3.17 is true in that context, we will content ourselves with making it our definition.

As a matter of notation, if X is a pointed $O(n)$–space and $\mathfrak{g}$ a Lie algebra in $O(n)$–modules, we denote the $O(n)$–invariants of $\mathfrak{g}^X$ by $\mathrm{Map}^{O(n)}(X, \mathfrak{g})$.

Definition 3.19 Let $\mathfrak{g}$ be a Lie algebra in $O(n)$–modules. The *n–enveloping algebra* of $\mathfrak{g}$ is the n–disk algebra

$$U_n(\mathfrak{g}) = C^{\mathcal{L}}(\mathrm{Map}^{O(n)}(\mathrm{Fr}_{(\mathbb{R}^n)^+}, \mathfrak{g})).$$

Here we take the frame bundle of the one-point compactification to be the cofiber

$$\mathrm{Fr}_{M^+} = \mathrm{cofib}\big(\mathrm{Fr}_{\overline{M}}|_{\partial \overline{M}} \to \mathrm{Fr}_{\overline{M}}\big)$$

of $O(n)$–spaces, where $\overline{M}$ is a compact n–manifold with boundary whose interior is M; see [4, Definition 4.5.1] for a more invariant interpretation of this object.

A choice of framing of $\mathbb{R}^n$ trivializes $\mathrm{Fr}_{(\mathbb{R}^n)^+}$, inducing an equivalence

$$\mathrm{Map}^{O(n)}(\mathrm{Fr}_{(\mathbb{R}^n)^+}, \mathfrak{g}) \simeq \mathfrak{g}^{(\mathbb{R}^n)^+},$$

which is even equivariant for the diagonal action of $O(n)$ on the target, so this definition specializes via Theorem 3.17 to our earlier one when $\mathfrak{g}$ is an ordinary Lie algebra.

The corresponding factorization homology calculation is the following.

Proposition 3.20 *There is a natural equivalence*

$$\int_M U_n(\mathfrak{g}) \xrightarrow{\sim} C^{\mathcal{L}}(\mathrm{Map}^{O(n)}(\mathrm{Fr}_{M^+}, \mathfrak{g}))$$

for M an n–manifold and $\mathfrak{g}$ a Lie algebra in $O(n)$–modules.

Proof Since $C^{\mathcal{L}}$, as a left adjoint, preserves colimits, it suffices to exhibit an equivalence of Lie algebras

$$\int_M \mathfrak{g}^{(\mathbb{R}^n)^+} \xrightarrow{\sim} \mathrm{Map}^{O(n)}(\mathrm{Fr}_{M^+}, \mathfrak{g}),$$

which is supplied by the argument of [5, Proposition 5.13], since sifted colimits of Lie algebras are computed in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ by [41, Proposition 2.1.16]. □

We close this section with a definition of a grading that will play an important role in what follows. Let $\mathfrak{g}$ be a differential graded Lie algebra with a *weight decomposition* as a direct sum of complexes $\mathfrak{g} = \bigoplus_k \mathfrak{g}(k)$ with the property that $[v, w] \in \mathfrak{g}(r+s)$ when $v \in \mathfrak{g}(r)$ and $w \in \mathfrak{g}(s)$.

Example 3.21 A free Lie algebra has a canonical weight decomposition

$$\mathcal{L}(V) = \bigoplus_{k \geq 0} \mathcal{L}(k) \otimes_{\Sigma_k} V^{\otimes k}.$$

Example 3.22 If $\mathfrak{g}$ has a weight decomposition, then $A_{\mathrm{PL}}(X) \otimes \mathfrak{g}$ carries a canonical weight decomposition for any space X.

Such a decomposition induces a *weight grading* on the underlying graded vector space of $\mathrm{Sym}(\mathfrak{g}[1])$ of the Chevalley–Eilenberg complex. In fact, since we have assumed that the bracket and differential of $\mathfrak{g}$ each respect the weight decomposition, the Chevalley–Eilenberg differential applied to a monomial of pure weight k again has pure weight k, so that $\mathrm{CE}(\mathfrak{g})$ is a bicomplex. In this way, a weight decomposition of $\mathfrak{g}$ induces a weight grading on $H(U_n(\mathfrak{g}))$.

4 Configuration spaces

4.1 The main result

In this section, we prove Theorem 1.1 assuming the validity of several results, discussion of which is postponed for the sake of continuity, as the proofs involve different techniques from those used thus far.

As a preliminary step, we have the following basic pair of observations.

Proposition 4.1 (1) *Let K be an $O(n)$–module and $\underline{K}$ its underlying chain complex. There is a natural equivalence of framed n–disk algebras*

$$\mathbb{F}_n^{\mathrm{fr}}(\underline{K}) \simeq \mathbb{F}_n(K).$$

(2) *Let $\mathfrak{g}$ be a Lie algebra in $O(n)$–modules and $\underline{\mathfrak{g}}$ its underlying Lie algebra. There is a natural equivalence of framed n–disk algebras*

$$U_n(\underline{\mathfrak{g}}) \simeq U_n(\mathfrak{g}).$$

Proof A choice of framing for $\mathbb{R}^n$ induces an $O(n)$–equivariant homotopy equivalence

$$\mathrm{Emb}^{\mathrm{fr}}\big(\textstyle\bigsqcup_k \mathbb{R}^n, \mathbb{R}^n\big) \times O(n)^k \xrightarrow{\sim} \mathrm{Emb}\big(\textstyle\bigsqcup_k \mathbb{R}^n, \mathbb{R}^n\big),$$

whence from Proposition 2.7 we have

$$\begin{aligned}
\mathbb{F}_n(K) &\cong \bigoplus_{k\geq 0}\Big(\mathrm{Emb}^{\mathrm{fr}}\big(\textstyle\bigsqcup_k \mathbb{R}^n, \mathbb{R}^n\big) \times O(n)^k\Big) \otimes_{\Sigma^k \ltimes O(n)^k} K^{\otimes k} \\
&\cong \bigoplus_{k\geq 0}\Big(\Big(\mathrm{Emb}^{\mathrm{fr}}\big(\textstyle\bigsqcup_k \mathbb{R}^n, \mathbb{R}^n\big) \times O(n)^k\Big) \otimes_{O(n)^k} K^{\otimes k}\Big)_{\Sigma_k} \\
&\cong \bigoplus_{k\geq 0} \mathrm{Emb}^{\mathrm{fr}}\big(\textstyle\bigsqcup_k \mathbb{R}^n, \mathbb{R}^n\big) \otimes_{\Sigma_k} K^{\otimes k} \\
&\cong \mathbb{F}_n^{\mathrm{fr}}(\underline{K}).
\end{aligned}$$

This proves (1), and (2) is immediate from Definition 3.19. □

Remark 4.2 Thinking topologically, the generic example of an n–disk algebra in spaces is an n–fold loop space on an $O(n)$–space X; see [49] or [54]. In this context, the statement is that, *as an n–fold loop space*, the homotopy type of $\Omega^n X$ does not depend on the action of $O(n)$ on X.

Connecting (1) and (2) is the following formal observation, which amounts to the statement that left adjoints compose.

Proposition 4.3 *Let V be a chain complex. There is a natural equivalence*

$$\mathbb{F}_n^{\mathrm{fr}}(V) \xrightarrow{\sim} U_n(\mathcal{L}(V[n-1]))$$

of framed n–disk algebras, where $\mathcal{L}$ is the free Lie algebra functor.

This observation is a generalization of the familiar fact that the universal enveloping algebra of the free Lie algebra on a set of generators S is free on S as an associative algebra; however, equipped with the involution given by its Hopf algebra antipode, the universal enveloping algebra of the free Lie algebra on S is *not* the free algebra-with-involution on S. This classical fact illustrates the $n=1$ case of the general phenomenon that the free n–disk algebra on the trivial $O(n)$–module V is *not* the n–enveloping algebra of the free Lie algebra on V. As the following proposition shows, the $O(n)$–action must be twisted to restore the equivalence.

Proposition 4.4 *Let K be an $O(n)$–module. There is a natural equivalence*

$$\mathbb{F}_n(K) \xrightarrow{\sim} U_n\big(\mathcal{L}((\mathbb{R}^n)^+ \otimes K[-1])\big)$$

of n–disk algebras, where $(\mathbb{R}^n)^+ \otimes K$ carries the diagonal $O(n)$–action.

Proof First, we note that the unit

$$K \to ((\mathbb{R}^n)^+ \otimes K)^{(\mathbb{R}^n)^+}$$

of the tensor/cotensor adjunction is an equivalence of $O(n)$–modules. Indeed, it suffices to verify this in the case $K = \mathbb{Q}$, in which case the map induces the isomorphism $\mathbb{Q} \cong (\mathbb{Q}^{\det})^{\otimes 2}$ in homology.

Now, composing this unit map with the natural inclusions

$$((\mathbb{R}^n)^+ \otimes K)^{(\mathbb{R}^n)^+} \to \mathcal{L}((\mathbb{R}^n)^+ \otimes K[-1])^{(\mathbb{R}^n)^+}[1] \to C^{\mathcal{L}}\big(\mathcal{L}((\mathbb{R}^n)^+ \otimes K[-1])^{(\mathbb{R}^n)^+}\big)$$

of $O(n)$–modules, we obtain a map of n–disk algebras

$$\mathbb{F}_n(K) \to U_n\big(\mathcal{L}((\mathbb{R}^n)^+ \otimes K[-1])\big)$$

from the universal property of the free algebra. It will suffice to show that this map is an equivalence upon passing to underlying $\mathcal{D}\mathrm{isk}_n^{\mathrm{fr}}$–algebras, which follows from the previous two propositions and the (nonequivariant) equivalence $\mathbb{Q}[n] \simeq \widetilde{C}_*((\mathbb{R}^n)^+;\mathbb{Q})$. □

Proof of Theorem 1.1 An equivalence of n–disk algebras induces an equivalence on passing to factorization homology. Using the indicated results, we obtain equivalences

$$\begin{aligned}
\bigoplus_{k\geq 0} C_*(B_k(M);\mathbb{Q}) &\simeq \int_M \mathbb{F}_n(\mathbb{Q}) && (3.4)\\
&\simeq \int_M U_n\big(\mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)[-1]\big)\big) && (4.4)\\
&\simeq C^{\mathcal{L}}\Big(\mathrm{Map}^{O(n)}\big(\mathrm{Fr}_{M^+}, \mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)[-1]\big)\big)\Big) && (3.20)\\
&\simeq C^{\mathcal{L}}\big(\mathrm{Map}^{O(n)}(\mathrm{Fr}_{M^+}, \mathcal{L}(\mathbb{Q}^{\det}[n-1]))\big) && (7.1)\\
&\simeq C^{\mathcal{L}}\big(\mathrm{Map}^{C_2}(\widetilde{M}^+, \mathcal{L}(\mathbb{Q}^{\mathrm{sgn}}[n-1]))\big)\\
&\simeq C^{\mathcal{L}}\big(H_c^{-*}(M, \mathcal{L}(\mathbb{Q}^w[n-1]))\big). && (7.5)
\end{aligned}$$

Applying Proposition 3.14 to this equivalence of algebras, we obtain an isomorphism of spectral sequences. The weight and cardinality gradings of the two algebras pass to factorization homology, so that these spectral sequences are each trigraded. According to Proposition 5.4, the isomorphism preserves the extra grading on E^2 and hence on E^∞. □

4.2 Variations

In this section, we discuss the corresponding results for twisted homology and manifolds with boundary.

Theorem 4.5 *Let M be an n–manifold.*

(1) *If n is even, there is an isomorphism of bigraded vector spaces*

$$\bigoplus_{k\geq 0} H_*(B_k(M);\mathbb{Q}^w) \cong H^{\mathcal{L}}\big(H_c^{-*}(M;\mathcal{L}(\mathbb{Q}[n-1]))\big).$$

(2) *If n is odd, there is an isomorphism of bigraded vector spaces*

$$\bigoplus_{k\geq 0} H_*(B_k(M);\mathbb{Q}^w)[k] \cong H^{\mathcal{L}}\big(H_c^{-*}(M;\mathcal{L}(\mathbb{Q}[n]))\big).$$

Proof We imitate the proof of Theorem 1.1. In the even case, we have

$$\begin{aligned}
\bigoplus_{k\geq 0} C_*(B_k(M);\mathbb{Q}^w) &\simeq \int_M \mathbb{F}_n(\mathbb{Q}^{\det}) && (3.6)\\
&\simeq \int_M U_n\big(\mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)\otimes\mathbb{Q}^{\det}[-1]\big)\big) && (4.4)\\
&\simeq C^{\mathcal{L}}\Big(\mathrm{Map}^{O(n)}\big(\mathrm{Fr}_{M^+},\mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)\otimes\mathbb{Q}^{\det}[-1]\big)\big)\Big) && (3.20)\\
&\simeq C^{\mathcal{L}}\big(\mathrm{Map}^{O(n)}\big(\mathrm{Fr}_{M^+},\mathcal{L}((\mathbb{Q}^{\det})^{\otimes 2}[n-1])\big)\big) && (7.1)\\
&\simeq C^{\mathcal{L}}\big(\mathrm{Map}^{O(n)}(\mathrm{Fr}_{M^+},\mathcal{L}(\mathbb{Q}[n-1]))\big)\\
&\simeq C^{\mathcal{L}}(\mathcal{L}(\mathbb{Q}[n-1])^{M^+})\\
&\simeq C^{\mathcal{L}}\big(H_c^{-*}(M,\mathcal{L}(\mathbb{Q}[n-1]))\big), && (7.5)
\end{aligned}$$

and the odd case is essentially identical. The same argument as in the proof of Theorem 1.1 shows that the resulting isomorphism is bigraded. □

Now, if M is a manifold with boundary, then $B_k(M)\simeq B_k(\mathring{M})$, since configuration spaces are isotopy functors. A more interesting configuration space in this context is the *relative configuration space*

$$B_k(M,\partial M) := \frac{B_k(M)}{\{(x_1,\ldots,x_k)\mid x_i\in\partial M \text{ for some } i\}}.$$

From the point of view of factorization homology, the natural setting in which to study these spaces is that of the *zero-pointed manifolds* of [4], a class of pointed spaces that are manifolds away from the basepoint. Indeed, if M is a manifold with boundary, then $M/\partial M$ is naturally a zero-pointed manifold.

The algebraic counterpart of a basepoint is an *augmentation*.

Definition 4.6 An *augmented n–disk algebra* is an n–disk algebra A together with a map of n–disk algebras $\epsilon\colon A\to\mathbb{Q}$.

Example 4.7 The free n–disk algebra $\mathbb{F}_n(K)$ is naturally augmented via the unique map of $O(n)$–modules $K\to 0$.

Example 4.8 The n–enveloping algebra $U_n(\mathfrak{g})$ is naturally augmented via the unique map of Lie algebras $\mathfrak{g}\to 0$.

The theory of factorization homology for zero-pointed n–manifolds with coefficients in augmented n–disk algebras is expounded at length in [4] and [3]. For us, what is important is that, if M is a manifold with boundary, then the factorization homology of $M/\partial M$ is defined for any choice of augmented n–disk algebra; moreover, if $\partial M = \varnothing$, then $M/\partial M = M_+$, and the factorization homology of the zero-pointed manifold M_+ with coefficients in $\epsilon\colon A \to \mathbb{Q}$ is equivalent to the factorization homology of M with coefficients in A defined previously.

Our arguments go through in this more general context.

Theorem 4.9 *Let M be an n–manifold with boundary. There is an isomorphism of bigraded vector spaces*

$$\bigoplus_{k\geq 0} \widetilde{H}_*(B_k(M,\partial M);\mathbb{Q}) \cong H^{\mathcal{L}}\big(H_c^{-*}(M;\mathcal{L}(\mathbb{Q}^w[n-1]))\big).$$

Proof We explain the adjustments necessary in the proof of Theorem 1.1. First, [3, Theorem 2.4.1] guarantees the equivalence

$$\bigoplus_{k\geq 0} \widetilde{C}_*(B_k(M,\partial M);\mathbb{Q}) \simeq \int_{M/\partial M} \mathbb{F}_n(\mathbb{Q}).$$

Second, it is immediate from its definition that the map of Proposition 4.4 is a map of augmented n–disk algebras, so that

$$\int_{M/\partial M} \mathbb{F}_n(\mathbb{Q}) \simeq \int_{M/\partial M} U_n\big(\mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)[-1]\big)\big).$$

The proof of Proposition 3.20 translates verbatim into the zero-pointed context, so that we have the further equivalence

$$\int_{M/\partial M} U_n\big(\mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)[-1]\big)\big) \simeq C^{\mathcal{L}}\Big(\mathrm{Map}^{O(n)}\big(\mathrm{Fr}_{M^+}, \mathcal{L}\big(\widetilde{C}_*((\mathbb{R}^n)^+)[-1]\big)\big)\Big).$$

The remainder of the proof goes through unchanged. □

Remark 4.10 When M has boundary, there are two obvious candidates for the orientation sheaf of M, namely the ordinary and the exceptional pushforwards of the orientation sheaf of the interior of M. We intend the former here.

4.3 Formulas

In this section, we use Theorem 1.1 and the Chevalley–Eilenberg complex to reproduce and extend the classical results on the rational homology of configuration spaces alluded to in the introduction.

We remind the reader that the free Lie algebra on $\mathbb{Q}^w[r]$ is given as a graded vector space by

$$\mathcal{L}(\mathbb{Q}^w[r]) \cong \begin{cases} \mathbb{Q}^w[r] \oplus \mathbb{Q}[2r] & \text{for } r \text{ odd,} \\ \mathbb{Q}^w[r] & \text{for } r \text{ even.} \end{cases}$$

When r is odd, the only nonvanishing bracket is the isomorphism $(\mathbb{Q}^w[r])^{\otimes 2} \cong \mathbb{Q}[2r]$.

Corollary 4.11 *If n is odd, there is an isomorphism*

$$H_*(B_k(M);\mathbb{Q}) \cong \mathrm{Sym}^k(H_*(M;\mathbb{Q})).$$

Proof Since n is odd, the Lie algebra in question is abelian, so that the Chevalley–Eilenberg complex has no differential, and the weight grading coincides with the usual grading of the symmetric algebra. The claim follows after replacing shifted, twisted, compactly supported cohomology with homology using Poincaré duality. □

This result is [10, Theorem C] as formulated in dual form in [23, Theorem 4], in which the isomorphism on cohomology is shown to be an isomorphism of algebras.

Corollary 4.12 *If n is even, $H_*(B_k(M);\mathbb{Q})$ is isomorphic to the homology of the complex*

$$\left(\bigoplus_{i=0}^{\lfloor k/2 \rfloor} \mathrm{Sym}^{k-2i}(H_c^{-*}(M;\mathbb{Q}^w)[n]) \otimes \mathrm{Sym}^i(H_c^{-*}(M;\mathbb{Q})[2n-1]),\, D \right),$$

where the differential D is defined as a coderivation by the equation

$$D(\sigma^n\alpha \wedge \sigma^n\beta) = (-1)^{(n-1)|\beta|}\sigma^{2n-1}(\alpha \smile \beta).$$

Proof It suffices by Theorem 1.1 to identify the complex in question with the weight-k part of the Chevalley–Eilenberg complex for $\mathfrak{g} = H_c^{-*}(M;\mathcal{L}(\mathbb{Q}^w[n-1]))$, which as a graded vector space is given by

$$\mathrm{Sym}(\mathfrak{g}[1]) \cong \mathrm{Sym}(H_c^{-*}(M;\mathbb{Q}^w)[n]) \otimes \mathrm{Sym}(H_c^{-*}(M;\mathbb{Q})[2n-1]),$$

with differential determined as a coderivation by the bracket of $\mathfrak{g}$, which is none other than the shifted cup product shown above, with the sign determined by the usual Koszul rule of signs. Since the cogenerators of the first tensor factor have weight 1 and those of the second tensor factor weight 2, the subcomplex of total weight k is exactly the sum shown above. □

When M is closed, orientable and nilpotent, we recover the linear and Poincaré dual of [24, Theorem A], as formulated in [23, Theorem 1]. When M is a once-punctured surface, we recover [9, Theorem C].

Remark 4.13 The proofs of Theorem 1 and the even-dimensional half of Theorem 3 of [23] rely crucially on the results of [24] and thereby on the hypotheses of compactness, orientability and nilpotence. At the time of writing, these hypotheses do not appear in the statements of the theorems.

It follows from our results, however, that these theorems are true at the stated level of generality. Indeed, by [23, Theorem 6], the Σ_k–invariants of the E_1 page of the Cohen–Taylor–Totaro spectral sequence (see [18] and [53]) coincide with the linear dual of the complex exhibited in Corollary 4.12.

An analogous spectral sequence in the nonorientable case, possibly with twisted coefficients, is available due to [28] and [29]; see also [45].

We leave it to the reader to formulate the analogous results on twisted homology and those concerning the homology of the relative configuration spaces $B_k(M, \partial M)$, which follow in the same way from Theorems 4.5 and 4.9, respectively. To the author's knowledge, the computation in the twisted case is new in all cases except when M is orientable and n is even, so that $B_k(M)$ is orientable, and the computation in the relative case is new in all cases except when $\partial M = \varnothing$.

5 Coalgebraic structure

5.1 Primitives and weight

Our present goal is to supply the first of the missing ingredients in the proof of the main theorem, namely the identification of the cardinality and weight gradings at the level of homology (see Definition 3.3 and the end of Section 3.4 for definitions of these gradings). We make this identification locally on M in this section and globalize in the following section using a spectral sequence argument.

Let K be an $O(n)$–module. We define the following maps:

(1) ι: $K \to \mathbb{F}_n(K)$ is the map of $O(n)$–modules given by the unit of the free/forgetful adjunction;

(2) η: $\mathbb{Q} \to \mathbb{F}_n(K)$ is the unit of $\mathbb{F}_n(K)$ as an n–disk algebra;

(3) $\delta\colon \mathbb{F}_n(K) \to \mathbb{F}_n(K) \otimes \mathbb{F}_n(K)$ is the map of n–disk algebras induced by the composite

$$K \xrightarrow{\Delta} K \oplus K \xrightarrow{\eta\otimes\iota+\iota\otimes\eta} \mathbb{F}_n(\mathbb{Q}) \otimes \mathbb{F}_n(\mathbb{Q}),$$

where Δ is the diagonal and we have tacitly employed the canonical identifications $K \otimes \mathbb{Q} \cong K \cong \mathbb{Q} \otimes K$;

(4) δ_M and η_M are the maps on factorization homology induced by δ and η, respectively.

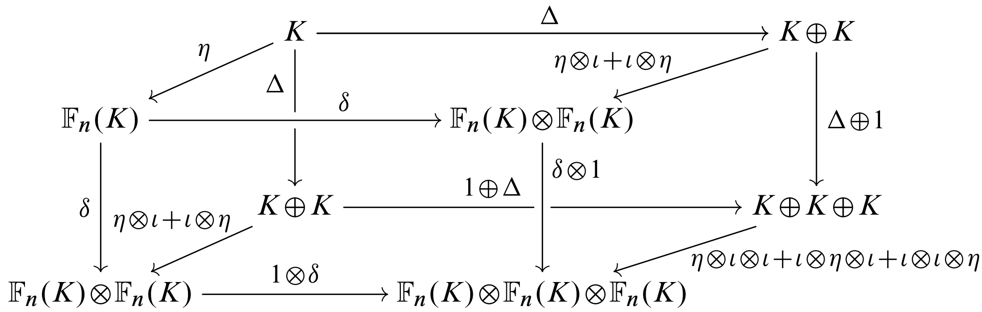
Note that we have suppressed the choice of K from the notation.

Although we will only use the case $M = \mathbb{R}^n$ here, we record the following result for its inherent interest.

Proposition 5.1 *The maps $H(\delta_M)$ and $H(\eta_M)$ endow $H\big(\int_M \mathbb{F}_n(K)\big)$ with the structure of a coaugmented cocommutative coalgebra.*

Proof The functor $\int_M$ is symmetric monoidal in the algebra variable by [42, Theorem 5.5.3.2], so it suffices to verify the claim in the case $M = \mathbb{R}^n$. The required axioms all follow from the universal property of the free algebra; we spell out the argument for coassociativity, leaving the remainder to the reader.

Consider the following cubical diagram:



It will suffice to show that the square diagram given by the front face of the cube commutes in the ∞–category of n–disk algebras, since this square witnesses coassociativity after applying factorization homology and passing to the homotopy category of chain complexes. Applying the universal property of the free algebra, the required commutativity is equivalent to commutativity as a diagram of $O(n)$–modules after precomposing with η. By a standard diagram chase, it suffices to verify that the remaining five faces each commute:

- The left and top face commute by the definition of δ.
- The back face commutes by the universal property of the direct sum, considered as the ∞–categorical product.
- The right and bottom face commute by the definition of δ and the universal property of the direct sum, considered as the ∞–categorical coproduct. $\square$

Although we have defined this coalgebra structure in abstract terms, it has an appealing geometric interpretation, which we discuss in Section 5.2 below.

When $M = \mathbb{R}^n$, the same homology is also an algebra, and even commutative for $n \geq 2$. Since δ is a map of n–disk algebras, $H(\mathbb{F}_n(\mathbb{Q}))$ inherits the structure of a *bialgebra*, and in fact a Hopf algebra, although we will not make use of the antipode.

For the duration of this section, we make the abbreviation

$$\mathfrak{g}(K) := \mathcal{L}((\mathbb{R}^n)^+ \otimes K[-1])^{(\mathbb{R}^n)^+}.$$

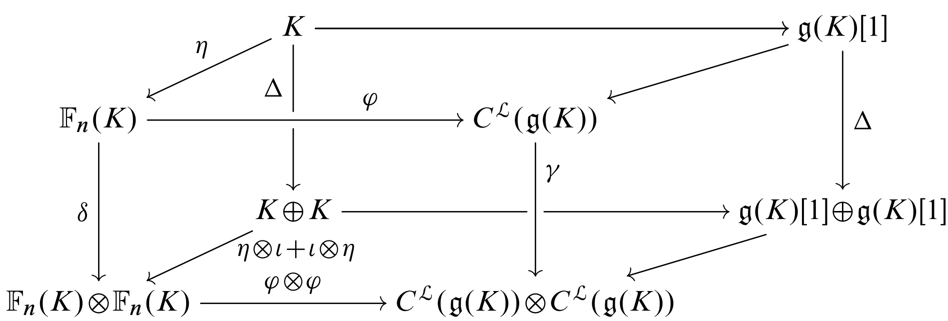
Proposition 5.2 *The isomorphism on homology induced by the equivalence of Proposition 4.4 is an isomorphism of bialgebras.*

Proof Denote by φ the equivalence

$$\mathbb{F}_n(K) \xrightarrow{\sim} C^{\mathcal{L}}(\mathfrak{g}(K))$$

of Proposition 4.4. Since φ is a map of n–disk algebras, the induced map on homology is a map of algebras; therefore, it will suffice to show that this map is also a map of coalgebras.

Consider the cubical diagram



where γ denotes the comultiplication on Lie algebra chains. As before, we wish to show that the front face commutes in the ∞–category of n–disk algebras, and, as before, this reduces to checking the commutativity of the remaining five faces in the ∞–category of $O(n)$–modules:

- The left face commutes by the definition of δ.
- The back face commutes by functoriality of the diagonal.
- The top face commutes by the definition of φ.
- The bottom face commutes by the definition of φ and the universal property of the direct sum, considered as the categorical coproduct.
- The right face commutes because the functor $C^{\mathcal{L}}$ is Cartesian monoidal. □

This bialgebra is a familiar one, and the various components of its structure interact predictably with the bigradings.

Proposition 5.3 (1) *There are isomorphisms*

$$H^{\mathcal{L}}(\mathfrak{g}(K)) \cong \operatorname{Sym}\big(H(\mathfrak{g}(K))[1]\big) \cong H(\mathbb{F}_n(K))$$

of graded bialgebras, where Sym *is equipped with the standard product and coproduct.*

(2) *The product in $H(\mathbb{F}_n(K))$ preserves the cardinality grading.*

(3) *The coproduct in $H(\mathbb{F}_n(K))$ preserves the cardinality grading.*

(4) *The product in $H^{\mathcal{L}}(\mathfrak{g}(K))$ preserves the weight grading.*

(5) *The coproduct in $H^{\mathcal{L}}(\mathfrak{g}(K))$ preserves the weight grading.*

Proof (1) We note that $\mathfrak{g}(K)$ is a formal Lie algebra, since the pointed space $(\mathbb{R}^n)^+$ is formal; moreover, since $H(\mathfrak{g}(K))$ is abelian, there is no differential in the Chevalley–Eilenberg complex, so we have isomorphisms of coaugmented coalgebras

$$H^{\mathcal{L}}(\mathfrak{g}(K)) \cong H^{\mathcal{L}}\big(H(\mathfrak{g}(K))\big) \cong \operatorname{Sym}\big(H(\mathfrak{g}(K))[1]\big).$$

From the discussion of Section 3.4, the product of $H^{\mathcal{L}}(\mathfrak{g}(K))$ is the map induced on Lie algebra homology by the n–disk algebra structure map of $\mathfrak{g}(K)$ corresponding to any embedding $\mathbb{R}^n \sqcup \mathbb{R}^n \to \mathbb{R}^n$, and any such structure map induces the fold map

$$H(\mathfrak{g}(K)) \oplus H(\mathfrak{g}(K)) \xrightarrow{+} H(\mathfrak{g}(K))$$

at the level of homology. Likewise, the coproduct is induced by the diagonal, and we recognize the standard bialgebra structure on Sym. The second isomorphism now follows by Proposition 5.2.

(2) The cardinality grading is natural, and the product is the map induced on homology by the n–disk algebra structure map corresponding to any embedding $\mathbb{R}^n \sqcup \mathbb{R}^n \to \mathbb{R}^n$.

(3) Since the coproduct preserves the cardinality grading on generators by definition, the claim follows from (2) and the fact that $H(\mathbb{F}_n(K))$ is a bialgebra.

(4) The claim is immediate from the definition of the weight grading.

(5) Since the coproduct preserves the cardinality grading on generators by definition, the claim follows from (4) and the fact that $H^{\mathcal{L}}(\mathfrak{g}(K))$ is a bialgebra. □

The desired identification of bigradings now follows easily.

Proposition 5.4 *In the case $M = \mathbb{R}^n$, the isomorphisms of Theorems 1.1, 4.5 and 4.9 are isomorphisms of bigraded vector spaces.*

Proof We present the argument for Theorem 1.1, the others being essentially identical.

We follow the convention that a subscript indicates homological degree, a generator decorated with a tilde has weight 2 and an unadorned generator has weight 1. There is an isomorphism of bialgebras $H(\mathbb{F}_n(\mathbb{Q})) \cong \mathrm{Sym}(V_n)$, where

$$V_n = \begin{cases} \mathbb{Q}\langle x_0\rangle & \text{for } n \text{ odd,} \\ \mathbb{Q}\langle x_0, \tilde{y}_{n-1}\rangle & \text{for } n \text{ even.} \end{cases}$$

Identifying both sides of the isomorphism of Theorem 1.1 with $\mathrm{Sym}(V_n)$, Proposition 5.2 permits us to view this isomorphism as an automorphism f of this graded bialgebra. Now, as a morphism of graded coalgebras, f takes primitives to primitives, so that there is an induced map $f|_{V_n}\colon V_n \to V_n$ of graded vector spaces, which we claim is a bigraded isomorphism. In the case of odd n, the claim is implied by the injectivity of $f|_{V_n}$, while in the even case we note that, for degree reasons, $f(x)$ is a scalar multiple of x and $f(\tilde{y})$ is a scalar multiple of $\tilde{y}$. By injectivity, this scalar is nonzero, and we conclude that $f|_{V_n}$ is a bigraded isomorphism.

Now, since f is also a map of algebras, we have $f(x_1 \cdots x_r) = f(x_1) \cdots f(x_r)$, which, together with the previous paragraph, shows that f preserves weight on monomials. Since monomials form a bihomogeneous basis and f is linear, the proof is complete. □

5.2 Interlude: splitting configurations

Configuration spaces of different cardinalities are interrelated by splitting and forgetting maps inherited from the Cartesian product via the embedding $\mathrm{Conf}_k(M) \to M^k$. This rich structure invites an inductive way of thinking that appears in one form or another in essentially every classical approach to these spaces; see [1] and [21] for the origins of this approach and [14] for a modern implementation.

In the setting of factorization homology, the importance of these splitting maps is that they assemble to form a coproduct, a shadow of which we have seen in the previous section, endowing $\mathbb{F}_n(K)$ with the structure of an n–disk algebra *in cocommutative coalgebras*. We will not need the full force of this statement, nor will we need the geometric interpretation of this coalgebra structure; nevertheless, we devote the remainder of this section to elucidating this interpretation, both for its general interest and for the motivation it provides for our proof of homological stability.

Remark 5.5 The constructions of this section are valid in more general stable settings than chain complexes, including the symmetric monoidal ∞–category of spectra with smash product. We intend to return to this setting in future work.

The basic ingredient is the collection of natural transformations

$$s_{i,j}\colon \mathrm{Conf}_k^{\mathrm{fr}} \to \mathrm{Conf}_i^{\mathrm{fr}} \times \mathrm{Conf}_j^{\mathrm{fr}},$$

defined whenever $i+j=k$, which make the diagram

$$\begin{array}{ccc} \mathrm{Conf}_k^{\mathrm{fr}}(M) & \xrightarrow{(s_{i,j})_M} & \mathrm{Conf}_i^{\mathrm{fr}}(M)\times \mathrm{Conf}_j^{\mathrm{fr}}(M) \\ \downarrow & & \downarrow \\ \prod_k \mathrm{Fr}_M & \xrightarrow{\cong} & \prod_i \mathrm{Fr}_M \times \prod_j \mathrm{Fr}_M \end{array}$$

commute; in other words,

$$(s_{i,j})_M(x_1,\dots,x_k) = ((x_1,\dots,x_i),(x_{i+1},\dots,x_k)).$$

Given an $O(n)$–module K, we have maps

$$s_{i,j}^K\colon \mathrm{Conf}_k^{\mathrm{fr}}\otimes K^{\otimes k} \xrightarrow{\delta_{i,j}\otimes 1} (\mathrm{Conf}_i^{\mathrm{fr}}\times \mathrm{Conf}_j^{\mathrm{fr}})\otimes K^{\otimes k} \xrightarrow{\simeq} \mathrm{Conf}_i^{\mathrm{fr}}\otimes K^{\otimes i}\otimes \mathrm{Conf}_j^{\mathrm{fr}}\otimes K^{\otimes j},$$

which are $(\Sigma_i\times\Sigma_j)\ltimes O(n)^k$–equivariant. Taking $O(n)^k$–coinvariants and using that induction is right adjoint to restriction for the inclusion $\Sigma_i\times\Sigma_j\to\Sigma_k$, we obtain by adjunction a Σ_k–equivariant map

$$\tilde{s}_{i,j}^K\colon \mathrm{Conf}_k^{\mathrm{fr}}\otimes_{O(n)^k} K^{\otimes k} \to \mathrm{Ind}_{\Sigma_i\times\Sigma_j}^{\Sigma_k}(\mathrm{Conf}_i^{\mathrm{fr}}\otimes_{O(n)^i} K^{\otimes i}\otimes \mathrm{Conf}_j^{\mathrm{fr}}\otimes_{O(n)^j} K^{\otimes j}).$$

Finally, taking Σ_k–coinvariants and summing over k, i and j, we obtain a map

$$s^K\colon \bigoplus_{k\geq 0}(\mathrm{Conf}_k^{\mathrm{fr}}\otimes_{\Sigma_k\ltimes O(n)^k} K^{\otimes k}) \to \bigoplus_k \bigoplus_{i+j=k}\big(\mathrm{Conf}_i^{\mathrm{fr}}\otimes_{\Sigma_i\ltimes O(n)^i} K^{\otimes i}\otimes \mathrm{Conf}_j^{\mathrm{fr}}\otimes_{\Sigma_j\ltimes O(n)^j} K^{\otimes j}\big).$$

Collecting terms and restricting to $\mathcal{D}\mathrm{isk}_n$, we recognize this as a monoidal natural transformation

$$s^K\colon \mathbb{F}_n(K) \to \mathbb{F}_n(K) \otimes \mathbb{F}_n(K).$$

The proof of homological stability given in the next section is completely internal to the Chevalley–Eilenberg complex, but the motivation behind it comes from thinking of the symmetric coproduct, given by splitting monomials in all possible ways, as corresponding to this geometric coproduct, given by splitting configurations in all possible ways. To see the connection, we recall that, in the approach of [14], stability is induced by the transfer maps

$$\begin{array}{ccc} H_*(\mathrm{Conf}_{k+1}(M);\mathbb{Q}) & \xrightarrow{\sum_i (p_i)_*} & H_*(\mathrm{Conf}_k(M);\mathbb{Q}) \\ \downarrow & & \downarrow \\ H_*(B_{k+1}(M);\mathbb{Q}) & \xrightarrow{\mathrm{tr}} & H_*(B_k(M);\mathbb{Q}) \end{array}$$

where p_i denotes the projection that forgets x_i. In terms of our splitting maps, we have a factorization

$$\begin{array}{ccc} \mathrm{Conf}_{k+1}(M) & \xrightarrow{p_i} & \mathrm{Conf}_k(M) \\ {\scriptstyle\sigma_i}\downarrow & & \uparrow \\ \mathrm{Conf}_{k+1}(M) & \xrightarrow{s_{1,k}} & M \times \mathrm{Conf}_k(M) \end{array}$$

where σ_i denotes the permutation that moves x_i to the first position while maintaining the relative order of the remaining points, and the unmarked arrow is the projection. The composite $s_{1,k}\sigma_i$ is a component of the coproduct defined above, and the projection away from the M factor corresponds at the level of homology to evaluating against the unit in $H^0(M;\mathbb{Q})$. Together, these observations suggest that homological stability should be induced taking a *cap product*. We realize this idea in the next section.

5.3 Stability

This section assembles the proof of Theorem 1.3. Throughout, unless otherwise noted, M will be connected, without boundary and of dimension $n > 1$. For the sake of brevity, we make the abbreviation

$$\mathfrak{g}_M = H_c^{-*}(M;\mathcal{L}(\mathbb{Q}^w[n-1])).$$

Let $\lambda \in H^0(M)$ denote the multiplicative unit. We view this cohomology class as a functional on $H_0(M) \cong H_c^0(M;\mathbb{Q}^w)[n]$ and hence, extending by zero, on $\mathrm{CE}(\mathfrak{g}_M)$, since the former is canonically a summand of the underlying bigraded vector space of the latter. Thus we may contemplate the *cap product* with this element, denoted $\lambda \frown (-)$, which is defined as the composite

$$\mathrm{CE}(\mathfrak{g}_M) \cong \mathbb{Q} \otimes \mathrm{CE}(\mathfrak{g}_M) \xrightarrow{\lambda\otimes\gamma} \mathrm{CE}(\mathfrak{g}_M)^\vee \otimes \mathrm{CE}(\mathfrak{g}_M) \otimes \mathrm{CE}(\mathfrak{g}_M) \xrightarrow{\langle -,-\rangle\otimes \mathrm{id}} \mathrm{CE}(\mathfrak{g}_M).$$

Denote by $p \in H^n_c(M;\mathbb{Q}^w) \subset \mathrm{CE}(\mathfrak{g}_M)$ the Poincaré dual of a point in M, which is well-defined since M is connected. Extend the set $\{1, p\}$ once and for all to a bihomogeneous basis $\mathcal{B}$ for $\mathfrak{g}_M[1]$. Then the set of nonzero monomials in elements of $\mathcal{B}$ form a bihomogeneous basis for $\mathrm{CE}(\mathfrak{g}_M)$, and, under the resulting identification of this vector space with its dual, λ is identified with the dual functional to p. Since p is closed of degree 0 and weight 1, we conclude the following:

Proposition 5.6 $\lambda \frown (-)$ *is a chain map of degree* 0 *and weight* -1.

There is a simple formula describing this map. Here and throughout, when we speak of divisibility, multiplication and differentiation in the Chevalley–Eilenberg complex, we refer only to the formal manipulation of bigraded polynomials; in particular, $\mathrm{CE}(\mathfrak{g}_M)$ is *not* in general a differential graded algebra.

Proposition 5.7 *The formula*

$$\lambda \frown x = \frac{dx}{dp}$$

holds for all $x \in \mathrm{CE}(\mathfrak{g}_M)$.

Proof Both sides are linear, so the claim is equivalent to the equality

$$\lambda \frown p^r y = rp^{r-1} y$$

whenever $r \geq 0$ and y is a monomial in elements of $\mathcal{B} \setminus \{p\}$. There are now two cases.

The first case is when y is a scalar, in which case we may assume by linearity that $y = 1$, so that $x = p^r$, and

$$\begin{aligned}\gamma(x) &= \gamma(p)^r \\ &= (p \otimes 1 + 1 \otimes p)^r \\ &= \sum_{i=0}^{r} \binom{r}{i} p^i \otimes p^{r-i},\end{aligned}$$

so that

$$\lambda \frown x = \sum_{i=0}^{r} \binom{r}{i} \langle \lambda, p^i \rangle p^{r-i} = rp^{r-1}.$$

The second is when y is a monomial in $\mathcal{B}\setminus\{1,p\}$, in which case we may write

$$\gamma(y) = y\otimes 1 + 1\otimes y + \sum_j y_j\otimes y_j'$$

with y_j and y_j' monomials in $\mathcal{B}\setminus\{1,p\}$. Then we have

$$\begin{aligned}\gamma(p^r y) &= \gamma(p)^r\gamma(y)\\ &= (p\otimes 1 + 1\otimes p)^r\Big(y\otimes 1 + 1\otimes y + \sum_j y_j\otimes y_j'\Big)\\ &= \sum_{i=0}^r \binom{r}{i}\Big(p^i y\otimes p^{r-i} + p^i\otimes p^{r-i}y + \sum_j p^i y_j\otimes p^{r-i}y_j'\Big),\end{aligned}$$

whence

$$\lambda\frown p^r y = \sum_{i=0}^r\binom{r}{i}\Big(\langle\lambda, p^i y\rangle p^{r-i} + \langle\lambda, p^i\rangle p^{r-i}y + \sum_j\langle\lambda, p^i y_j\rangle p^{r-i}y_j'\Big) = rp^{r-1}y,$$

since $p^i y$ is not a scalar multiple of p for any i, nor is $p^i y_j$ a scalar multiple of p for any (i,j). □

Corollary 5.8 *The chain map $\lambda\frown(-)$ is surjective.*

Proof It suffices to show that a general monomial in elements of $\mathcal{B}$ lies in the image. Such a monomial may be written as $p^r y$ with $r\geq 0$ and y a monomial in elements of $\mathcal{B}\setminus\{p\}$. We than have

$$\frac{d}{dp}\Big(\frac{1}{r+1}p^{r+1}y\Big) = p^r y. \qquad \square$$

The central observation behind our approach to stability is the following.

Proposition 5.9 *Let x be a nonzero monomial in $\mathrm{CE}(\mathfrak{g}_M)$. Then x is divisible by p provided either*

- $\mathrm{wt}(x) > |x| + 1$ *and M is an orientable surface, or*
- $\mathrm{wt}(x) > |x|$ *and M is not an orientable surface.*

Proof Suppose $\mathrm{wt}(x) > |x|$, and write $x = x_1\cdots x_r$ with $x_i\in\mathcal{B}$. Then $\mathrm{wt}(x_j) > |x_j|$ for some j. Since $x_j\in\mathfrak{g}_M[1]$, the weight of this element is either 1 or 2.

In the first case, $x_j\in H_c^{-*}(M;\mathbb{Q}^w)[n]$, and we have

$$|x_j| < \mathrm{wt}(x_j) = 1 \quad\text{implies}\quad |x_j| = 0,$$

since $H_c^{-*}(M;\mathbb{Q}^w)[n]$ is concentrated in degrees $0 \leq * \leq n$. But $H_c^n(M;\mathbb{Q}^w)$ is one-dimensional on the class p, so x_j is a scalar multiple of p.

In the second case, $x_j \in H_c^{-*}(M;\mathbb{Q})[2n-1]$, and we have $|x_j| < 2$. Because $H_c^{-*}(M;\mathbb{Q})[2n-1]$ is concentrated in degrees $n-1 \leq * \leq 2n-1$, we conclude that $x_j = 0$ provided $n \neq 2$ (recall that we have already assumed $n > 1$). Thus $x = 0$, which is a contradiction. This proves the claim when M is not a surface.

If M is a nonorientable surface, then $H_c^2(M;\mathbb{Q}) \cong H_0(M;\mathbb{Q}^w) = 0$, and therefore $H_c^{-*}(M;\mathbb{Q})[2n-1]$ is concentrated in degrees 2 and 3. Thus, in this case as well, we have a contradiction.

Assume now that M is an orientable surface and $\mathrm{wt}(x) > |x|+1$. As before, write $x = x_1 \cdots x_r$ and choose x_j with $\mathrm{wt}(x_j) > |x_j|$, and assume that x_j is not a scalar multiple of p. Then by the argument above, $\mathrm{wt}(x_j) = 2$, so $|x_j| = 1$, since $H_c^{-*}(M;\mathbb{Q})[3]$ is concentrated in degrees $1 \leq * \leq 3$.

Now, the monomial $x' = x_1 \cdots \hat{x}_j \cdots x_r$ has the property that

$$\mathrm{wt}(x') = \mathrm{wt}(x) - 2 > |x| - 1 = |x'|,$$

so there is some x_i with $i \neq j$ and $\mathrm{wt}(x_i) > |x_i|$. If x_i is a scalar multiple of p, we are finished; otherwise, repeating the same argument shows that x_i has degree 1. But $H_c^2(M;\mathbb{Q}) \cong H_0(M;\mathbb{Q})$ is one-dimensional, so that x_i is a scalar multiple of x_j, and x is divisible by x_j^2. Since x_j is of odd degree, this implies that $x = 0$, which is a contradiction. □

We are now equipped to prove Theorem 1.3. Denote by $C(k)$ the subcomplex of the Chevalley–Eilenberg complex spanned by the weight-k monomials. Taking the cap product with 1 restricts to a map $\Phi_k\colon C(k+1) \to C(k)$, and we aim to show that this map induces an isomorphism in homology in the specified range.

Recall that the r^{th} *brutal truncation* of a chain complex V is the chain complex $\tau_{\leq r} V$ whose underlying graded vector space is

$$(\tau_{\leq r} V)_i = \begin{cases} V_i & \text{if } i \leq r, \\ 0 & \text{otherwise.} \end{cases}$$

and whose differential is the restriction of the differential of V. Truncation is a functor on chain complexes in the obvious way.

We make use of the following elementary fact.

Proposition 5.10 *Let $f\colon V \to W$ be a surjective chain map such that $\tau_{\leq r} f$ is a chain isomorphism. Then f is a homology isomorphism through degree r and a homology surjection in degree $r+1$.*

Proof From the definition of the brutal truncation, it is immediate that f is a homology isomorphism through degree $r-1$. Moreover, f induces a bijection on r–cycles and an injection on r–boundaries.

To show that f is a homology isomorphism in degree r, it suffices to show that $f^{-1}(w)$ is a boundary if $w \in W_r$ is a boundary. Write $w = du$; then, by surjectivity, there is some $\tilde{u} \in V_r$ such that $f(\tilde{u}) = u$, and

$$f(d\tilde{u}) = df(\tilde{u}) = du = w \quad \text{implies} \quad f^{-1}(w) = d\tilde{u},$$

as desired.

To show that f is a homology surjection in degree $r+1$, let $v \in W_{r+1}$ be a cycle. By surjectivity, $v = f(\tilde{v})$, and it will suffice to show that $\tilde{v}$ is a cycle, for which we have

$$f(d\tilde{v}) = df(\tilde{v}) = dv = 0 \quad \text{implies} \quad d\tilde{v} = f^{-1}(0) = 0. \qquad \square$$

Proof of Theorem 1.3 Assume first that M is not an orientable surface. By the previous proposition and Corollary 5.8, we are reduced to showing that $\tau_{\leq k}\Phi_k$ is a chain isomorphism. To see this, let $x \in \tau_{\leq k}C(k+1)$ be a monomial. Then $\mathrm{wt}(x) > |x|$, so that $x = p^r y$ with $r > 0$ and y a monomial in $\mathcal{B} \setminus \{p\}$ by Proposition 5.9. By Proposition 5.7, $\Phi_k(x) = rp^{r-1}y$, so $\tau_{\leq k}\Phi_k$ maps distinct elements of our preferred basis for $C(k+1)$ to nonzero scalar multiples of distinct elements of our preferred basis for $C(k)$, which implies that $\tau_{\leq k}\Phi_k$ is injective. But Φ and hence Φ_k are surjective by Corollary 5.8, so $\tau_{\leq k}\Phi_k$ is as well.

Assume now that M is an orientable surface. For the same reason, we are reduced to showing that $\tau_{k-1}\Phi_k$ is a chain isomorphism, which is accomplished by the same argument, using the other half of Proposition 5.9. $\square$

Remark 5.11 Let $\mathbb{K}$ denote the Klein bottle. As shown in Section 6,

$$\dim H_*(B_k(\mathbb{K});\mathbb{Q}) = \begin{cases} 1 & i \in \{0,1,2,k+1\}, \\ 2 & 3 \leq i \leq k, \\ 0 & \text{else}. \end{cases}$$

In particular, $H_{k+1}(B_{k+1}(\mathbb{K});\mathbb{Q}) \not\cong H_{k+1}(B_k(\mathbb{K});\mathbb{Q})$, so our bound is sharp in the sense that no better stable range holds for all manifolds that are not orientable surfaces.

Remark 5.12 If M is orientable and $H_*(M;\mathbb{Q}) = 0$ for $1 \leq * \leq r-1$, then $H_c^{-*}(M;\mathbb{Q}) = 0$ for $n-r+1 \leq -* \leq n-1$, and the argument of Proposition 5.9 shows that a monomial x is divisible by p provided its weight is greater than $\frac{|x|}{r}+1$. This improved estimate leads to an improved stable range, as in [14, Proposition 4.1].

Remark 5.13 In [37], factorization homology is used to obtain homological stability results for various constructions on open manifolds. The approach there is through certain "partial algebras" and appears unrelated to ours.

6 Examples

We now present a selection of computations illustrating the following general procedure for determining the rational homology of the configuration space of k points in an n–manifold M:

(1) compute the compactly supported cohomology of M, twisted if necessary;
(2) compute the Lie algebra homology of $H_c^{-*}(M;\mathcal{L}(\mathbb{Q}^w[n-1]))$;
(3) count basis elements of weight k.

It is worth noting that the Chevalley–Eilenberg complex allows one to obtain answers simultaneously for all k, reducing an infinite sequence of computations to one.

The computations of this section are all relatively elementary, and one can do better with more effort. In [19], this approach is used to determine the Betti numbers of $B_k(\Sigma)$ for every surface Σ.

Convention In the following examples, a variable decorated with a tilde has weight 2, while an unadorned variable has weight 1.

6.1 Punctured euclidean space

As a warm-up and base case, we recover the classical computation of $H_*(B_k(\mathbb{R}^n);\mathbb{Q})$. Since there are no cup products in the compactly supported cohomology of $\mathbb{R}^n$, there are no differentials in the corresponding Chevalley–Eilenberg complex. Thus $H_*(B_k(\mathbb{R}^n);\mathbb{Q})$ is identified with the subspace of $\mathbb{Q}[x]$ spanned by x^k when n is odd, while for n even, the identification is with the subspace of

$$\mathbb{Q}[x]\otimes\Lambda[\tilde{x}],\quad |x|=0,\ |\tilde{x}|=n-1,$$

spanned by elements of weight k, a basis for which is given by $\{x^k, x^{k-2}\tilde{x}\}$. We conclude, for all $k>1$, that

$$H_*(B_k(\mathbb{R}^n);\mathbb{Q})\cong\begin{cases}\mathbb{Q} & \text{for } n \text{ odd},\\ \mathbb{Q}\oplus\mathbb{Q}[n-1] & \text{for } n \text{ even}.\end{cases}$$

Now, choose $\bar{p}=\{p_1,\dots,p_m\}\in\mathbb{R}^n$. There is a homotopy equivalence $(\mathbb{R}^n\setminus\bar{p})^+\simeq S^n\vee(S^1)^{\vee m}$, so that $H_c^{-*}(\mathbb{R}^n\setminus\bar{p};\mathbb{Q})\cong\mathbb{Q}^m[-1]\oplus\mathbb{Q}[-n]$. There are no cup products, so there can be no differentials.

If n is odd, Theorem 1.1 identifies $H_*(\mathbb{R}^n \setminus \bar{p}; \mathbb{Q})$ with the weight-k part of

$$\mathbb{Q}[x, y_1, \dots, y_m], \quad |x| = 0,\ |y_i| = n-1,$$

and an easy induction now shows that

$$\dim H_*(B_k(\mathbb{R}^n \setminus \bar{p}); \mathbb{Q}) = \begin{cases} \binom{m+i-1}{i} & \text{for } * = i(n-1), 0 \le i \le k, \\ 0 & \text{otherwise.} \end{cases}$$

(It is helpful to recall that $\binom{m+i-1}{i}$ is the number of ways to choose i not-necessarily-distinct elements from a set of m elements.)

If n is even, then the corresponding vector space is the weight-k part of

$$\mathbb{Q}[x, \tilde{y}_1, \dots, \tilde{y}_m] \otimes \Lambda[\tilde{x}, y_1, \dots, y_m], \quad |x| = 0,\ |y_i| = |\tilde{x}| = n-1,\ |\tilde{y}_i| = 2n-2.$$

Counting inductively in terms of less punctured Euclidean spaces, one finds that

$$H_*(B_k(\mathbb{R}^n \setminus \bar{p}); \mathbb{Q}) \cong \bigoplus_{l=0}^{k} \bigoplus_{j_1+\cdots+j_m=l} H_{*-l(n-1)}(B_{k-l}(\mathbb{R}^n); \mathbb{Q}),$$

from which it follows easily that

$$\dim H_*(B_k(\mathbb{R}^n \setminus \bar{p}); \mathbb{Q}) = \begin{cases} \binom{m+i-1}{m-1} + \binom{m+i-2}{m-1} & \text{for } * = i(n-1),\ 0 \le i < k, \\ \binom{m+k-1}{m-1} & \text{for } * = k(n-1), \\ 0 & \text{otherwise.} \end{cases}$$

(It is helpful to recall that $\binom{m+i-1}{m-1}$ is the number of ways to write i as the sum of m nonnegative integers.)

It should be clear from this example that Theorem 1.1 reduces calculations to counting problems whenever n is odd or the relevant compactly supported cohomology has no cup products.

6.2 Punctured torus

Since $H_c^{-*}(T^2 \setminus \mathrm{pt}; \mathbb{Q}) \cong \widetilde{H}^{-*}(T^2; \mathbb{Q})$, the relevant Lie algebra is isomorphic to

$$\mathfrak{h} \oplus \mathbb{Q}\langle \tilde{a}, \tilde{b}, c \rangle,$$

where $\mathfrak{h} = \mathbb{Q}\langle a, b, \tilde{c} \rangle$ as a vector space,

$$|a| = |b| = |\tilde{c}| = 0, \quad |\tilde{a}| = |\tilde{b}| = 1, \quad |c| = -1,$$

and the bracket is defined by the equation

$$[a, b] = \tilde{c}.$$

The Lie homology of $\mathfrak{h}$ is calculated by the complex

$$(\Lambda[x, y, \tilde{z}], d(xy) = \tilde{z}),$$

(where for ease of notation we have set $x = \sigma a$ and so on), a basis for the homology of which is easily seen to be given by the image in homology of the set $\{1, x, y, x\tilde{z}, y\tilde{z}, xy\tilde{z}\}$. Thus we have an identification of $H_*(B_k(T^2 \setminus \mathrm{pt}); \mathbb{Q})$ with the weight-k part of

$$\mathbb{Q}\langle 1, x, y, x\tilde{z}, y\tilde{z}, xy\tilde{z}\rangle \otimes \mathbb{Q}[\tilde{x}, \tilde{y}, z], \quad |z| = 0,\ |x| = |y| = |\tilde{z}| = 1,\ |\tilde{x}| = |\tilde{y}| = 2.$$

Counting, we find that

$$\dim H_*(B_k(T^2 \setminus \mathrm{pt}); \mathbb{Q}) = \begin{cases} \frac{3i-1}{2} + 1 & \text{for } * = 2i + 1 < k, \\ \frac{3i}{2} + 1 & \text{for } * = 2i < k, \\ k + 1 & \text{for } * = k \text{ odd}, \\ \frac{k}{2} + 1 & \text{for } * = k \text{ even}, \\ 0 & \text{otherwise}. \end{cases}$$

An amusing comparison can be seen by taking $k = 2$ in the above formula, which yields

$$H_*(B_2(T^2 \setminus \mathrm{pt}); \mathbb{Q}) \cong \mathbb{Q} \oplus \mathbb{Q}^2[1] \oplus \mathbb{Q}^2[2].$$

On the other hand, from the preceding example, one calculates that

$$H_*(B_2(\mathbb{R}^2 \setminus \{p_1, p_2\}); \mathbb{Q}) \cong \mathbb{Q} \oplus \mathbb{Q}^3[1] \oplus \mathbb{Q}^3[2].$$

Thus, despite the fact that the punctured torus and the twice-punctured plane are homotopy equivalent, having $S^1 \vee S^1$ as a common deformation retract, their configuration spaces are not homotopy equivalent.

6.3 Real projective space

Let n be even, so that $\mathbb{RP}^n$ is nonorientable. Then, as a ring, $H_c^{-*}(\mathbb{RP}^n; \mathbb{Q}) \cong \mathbb{Q}$, and the Lie homology of interest is $H^{\mathcal{L}}(\mathcal{L}(\mathbb{Q}[n-1])) \cong \mathbb{Q} \oplus \mathbb{Q}[n]$, whence, for $k > 1$,

$$H_*(B_k(\mathbb{RP}^n); \mathbb{Q}^w) = 0.$$

As for the untwisted homology, we note that $H_c^{-*}(\mathbb{RP}^n; \mathbb{Q}^w) \cong \mathbb{Q}[-n]$ by Poincaré duality, so that the cup product map $H_c^{-*}(\mathbb{RP}^n; \mathbb{Q}^w)^{\otimes 2} \to H_c^{-*}(\mathbb{RP}^n; \mathbb{Q})$ is trivial for degree reasons. Thus

$$H_c^{-*}(\mathbb{RP}^n; \mathcal{L}(\mathbb{Q}^w[n-1])) \cong \mathbb{Q}[-1] \oplus \mathbb{Q}[2n-2]$$

is abelian, so that $H_*(B_k(\mathbb{RP}^n);\mathbb{Q})$ is isomorphic to the weight-k part of

$$\mathbb{Q}[x]\otimes\Lambda[\tilde{y}],\quad |x|=0,\ |\tilde{y}|=2n-1.$$

Hence for all $k>1$,

$$H_*(B_k(\mathbb{RP}^n);\mathbb{Q})\cong\mathbb{Q}\oplus\mathbb{Q}[2n-1].$$

See [55] for an alternate method of computation in the case $n=2$.

6.4 Klein bottle, twisted

Let $\mathbb{K}$ denote the Klein bottle. Then $H_c^{-*}(\mathbb{K};\mathbb{Q})\cong\mathbb{Q}\oplus\mathbb{Q}[-1]$, with the generator in degree zero acting as a unit for the multiplication. As a vector space, the Lie algebra in question is $\mathfrak{g}:=\mathbb{Q}\langle a,\tilde{a},b,\tilde{b}\rangle$, where $|b|=0$, $|a|=|\tilde{b}|=1$ and $|\tilde{a}|=2$, and the bracket is defined by the equations

$$[a,a]=\tilde{a},\quad [a,b]=-\tilde{b}.$$

The subspace spanned by $\{b,\tilde{b}\}$ is an ideal realizing $\mathfrak{g}$ as an extension

$$0\to\mathbb{Q}\langle b,\tilde{b}\rangle\to\mathfrak{g}\to\mathcal{L}(\mathbb{Q}\langle a\rangle)\to 0,$$

so that we may avail ourselves of the Lyndon–Hochschild–Serre spectral sequence

$$E^2_{p,q}\cong H_p^{\mathcal{L}}\big(\mathcal{L}(\mathbb{Q}\langle a\rangle);H_q^{\mathcal{L}}(\mathbb{Q}\langle b,\tilde{b}\rangle)\big)\implies H^{\mathcal{L}}_{p+q}(\mathfrak{g}).$$

There are no differentials for degree reasons, and the E^2 page is computed as the homology of the complex

$$0\to\mathbb{Q}\langle a\rangle[1]\otimes\mathrm{Sym}(\mathbb{Q}\langle b,\tilde{b}\rangle[1])\to\mathrm{Sym}(\mathbb{Q}\langle b,\tilde{b}\rangle[1])\to 0,$$

where the differential is the action of a. It follows that a basis for $H^{\mathcal{L}}(\mathfrak{g})$ is given by $\{\sigma a\otimes(\sigma\tilde{b})^i,\sigma b\otimes(\sigma\tilde{b})^j\mid i,j\ge 0\}$. Counting monomials of weight k, we find that

$$H_*(B_k(\mathbb{K});\mathbb{Q}^w)\cong\begin{cases}\mathbb{Q}[k]\oplus\mathbb{Q}[k+1] & \text{for } k \text{ odd},\\ 0 & \text{for } k \text{ even}.\end{cases}$$

6.5 Nonorientable surfaces

Let $N_h=(\mathbb{RP}^2)^{\#h}$. Using the method of the previous example, one could proceed to obtain a general formula for the twisted homology of $B_k(N_h)$. Here we will determine the corresponding untwisted homology. We have

$$H_c^{-*}(N_h;\mathbb{Q})\cong\mathbb{Q}\oplus\mathbb{Q}[-1]^{h-1},\quad H_c^{-*}(N_h;\mathbb{Q}^w)\cong\mathbb{Q}[-1]^{h-1}\oplus\mathbb{Q}[-2],$$

so that there can be no cup products. Thus $H_*(B_k(N_h);\mathbb{Q})$ is the weight-k part of

$$\mathbb{Q}[x,\tilde{y}_1,\dots,\tilde{y}_{h-1}]\otimes\Lambda[\tilde{z},w_1,\dots,w_{h-1}],\quad |x|=0,\ |w_i|=1,\ |\tilde{y}_i|=2,\ |\tilde{z}|=3.$$

Counting inductively as in the example of punctured Euclidean space, we find that

$$H_*(B_k(N_h);\mathbb{Q})\cong\bigoplus_{l=0}^{k}\bigoplus_{j_1+\cdots+j_{h-1}=l}H_{*-l}(B_{k-l}(\mathbb{RP}^2);\mathbb{Q}),$$

from which it follows that

$$\dim H_*(B_k(N_h);\mathbb{Q})=\begin{cases}\binom{h+*-2}{h-2}+\binom{h+*-5}{h-2} & \text{for } *\le k,\\ \binom{h+*-5}{h-2} & \text{for } *=k+1,\\ 0 & \text{otherwise.}\end{cases}$$

6.6 Open and closed Möbius band

Let $\mathbb{M}$ denote the closed Möbius band. Then since $\mathbb{M}$ has the same compactly supported cohomology ring as the Klein bottle, our earlier calculation shows that

$$\widetilde{H}_*(B_k(\mathbb{M},\partial\mathbb{M});\mathbb{Q}^w)\cong\begin{cases}\mathbb{Q}[k]\oplus\mathbb{Q}[k+1] & \text{for } k \text{ odd},\\ 0 & \text{for } k \text{ even.}\end{cases}$$

On the other hand, by Poincaré duality, we have $H_c^{-*}(\mathbb{M};\mathbb{Q}^w)=0$, and hence $H_c^{-*}(\mathbb{M};\mathcal{L}(\mathbb{Q}^w[1]))\cong H^{-*}(\mathbb{M};\mathbb{Q})[2]$ is abelian, and $\widetilde{H}_*(B_k(\mathbb{M},\partial\mathbb{M});\mathbb{Q})$ is the weight-k part of

$$\mathbb{Q}[\tilde{x}]\otimes\Lambda[\tilde{y}],\quad |\tilde{x}|=2,\ |\tilde{y}|=3,$$

so

$$\widetilde{H}_*(B_k(\mathbb{M},\partial\mathbb{M});\mathbb{Q})\cong\begin{cases}0 & \text{for } k \text{ odd},\\ \mathbb{Q}[k]\oplus\mathbb{Q}[k+1] & \text{for } k \text{ even.}\end{cases}$$

The situation with the corresponding open manifold is quite different. We have $H_c^{-*}(\mathring{\mathbb{M}};\mathbb{Q})=0$ since $(\mathring{\mathbb{M}})^+\cong\mathbb{RP}^2$, so

$$H_*(B_k(\mathring{\mathbb{M}});\mathbb{Q}^w)=0$$

for all $k>1$. On the other hand, $H_c^{-*}(\mathring{\mathbb{M}};\mathbb{Q}^w)\cong\mathbb{Q}[-1]\oplus\mathbb{Q}[-2]$ by Poincaré duality, so that $H_*(B_k(\mathring{\mathbb{M}});\mathbb{Q})$ is the weight-k part of

$$\mathbb{Q}[x]\otimes\Lambda[y],\quad |x|=0,\ |y|=1,$$

whence

$$H_*(B_k(\mathring{\mathbb{M}});\mathbb{Q})\cong\mathbb{Q}[0]\oplus\mathbb{Q}[1]$$

for all $k\ge 1$.

7 Two formality results

In this final section, we supply the remaining two ingredients in the proof of Theorem 1.1. Although unrelated to each other, these formality statements may be of independent interest.

7.1 The $O(n)$–equivariant sphere

Since the reduced homology of S^n is one-dimensional, any choice of representative of a homology generator defines a quasi-isomorphism

$$C_*(S^n) \simeq \mathbb{Z} \oplus \mathbb{Z}[n].$$

The goal of this section is to prove that, rationally, this equivalence can be made $O(n)$–equivariant.

Theorem 7.1 *There is an equivalence of $O(n)$–modules*

$$C_*(S^n; \mathbb{Q}) \simeq \mathbb{Q} \oplus \mathbb{Q}^{\det}[n].$$

The proof has three main ingredients, the first of which is rational homotopy theory. We consider the Borel construction

$$\hat{\xi}\colon E\,\mathrm{SO}(n) \times_{\mathrm{SO}(n)} S^n \to B\,\mathrm{SO}(n)$$

where $\mathrm{SO}(n)$ acts on $S^n \cong (\mathbb{R}^n)^+$ by extension of its canonical action on $\mathbb{R}^n$. In other words, $\hat{\xi}$ is the fiberwise one-point compactification of the universal oriented n–plane bundle ξ. We denote by $E(\hat{\xi})$ the total space of this sphere bundle.

Sphere bundles over simply connected spaces admit particularly simple rational descriptions. According to [22, Sections 15(a)–(b)], we have the following commutative diagram, whose terms we will explain presently:

$$\begin{array}{ccccc}
(\mathrm{Sym}(W_n), d_1) & \xrightarrow{\sim} & A_{\mathrm{PL}}(S^n) & \xrightarrow{\sim} & C^{-*}(S^n; \mathbb{Q}) \\
\uparrow & & \uparrow & & \uparrow \\
(S \otimes \mathrm{Sym}(W_n), d_1 + d_2) & \xrightarrow{\sim} & A_{\mathrm{PL}}(E(\hat{\xi})) & \xrightarrow{\sim} & C^{-*}(E(\hat{\xi}); \mathbb{Q}) \\
\uparrow & & \uparrow & & \uparrow \\
S & \xrightarrow{\sim} & A_{\mathrm{PL}}(B\,\mathrm{SO}(n)) & \xrightarrow{\sim} & C^{-*}(B\,\mathrm{SO}(n); \mathbb{Q})
\end{array}$$

In this diagram,

(1) $S := H^{-*}(B\,\mathrm{SO}(n); \mathbb{Q})$ is a polynomial algebra,

(2) A_{PL} denotes the functor of PL de Rham forms,

(3) the horizontal arrows in the right-hand column are components of the natural quasi-isomorphism $\oint\colon A_{\mathrm{PL}} \to C^{-*}$ given by integrating forms over simplices,

(4) each term appearing in the leftmost column is a Sullivan model for the corresponding space, and

(5) W_n denotes the graded vector space

$$W_n = \begin{cases} \mathbb{Q}\langle x_{-n}\rangle & \text{for } n \text{ odd,} \\ \mathbb{Q}\langle x_{-n}, y_{-2n+1}\rangle & \text{for } n \text{ even,} \end{cases}$$

the differential d_1 is defined by the equation $d_1(y) = x^2$ and the differential d_2 is specified by its value on y, which is an element of P determined by the bundle $\hat{\xi}$.

We direct the reader to [11] for more on (1), and to [22, Sections 10(c), 10(e), 12, 15(b)], respectively, for more on (2)–(5). The reader is advised that, although we have maintained our convention of homological grading, the prevailing convention in rational homotopy theory is cohomological.

The second ingredient is the theory of A_∞–algebras and their modules, for which we refer the reader to [34]. The relevance here is that, according to [12, Section 3.1], the integration map $\oint$ extends to a map of A_∞–algebras (referred to in [12] as "strongly homotopic differential algebras"), so that $C^{-*}(E(\hat{\xi}); \mathbb{Q})$ becomes an A_∞-S–module via the bottom composite in the above diagram.

Proposition 7.2 *There is a quasi-isomorphism of A_∞-S–modules*

$$S \oplus S[-n] \xrightarrow{\sim} C^{-*}(E(\hat{\xi}); \mathbb{Q}).$$

Proof The fiberwise basepoint furnishes $\hat{\xi}$ with a section, and the Gysin sequence now implies that the top map in the commuting diagram

$$\begin{array}{ccc} S \oplus S[-n] & \longrightarrow & (S \otimes \mathrm{Sym}(W_n), d_1 + d_2) \\ \uparrow & & \uparrow \\ S & = & S \end{array}$$

is a quasi-isomorphism. Combining this diagram with the previous yields the result. □

The third ingredient is the Koszul duality between modules for the symmetric algebra S and modules for the exterior algebra Λ on the same generators with degrees shifted by 1. According to [11], there is a Hopf algebra isomorphism $\Lambda \cong H_*(\mathrm{SO}(n); \mathbb{Q})$, where the latter carries the Pontryagin product induced by the group structure of $\mathrm{SO}(n)$.

Koszul duality is the algebraic avatar of the correspondence between $\mathrm{SO}(n)$–spaces and spaces fibered over $B\mathrm{SO}(n)$ witnessed by the Borel construction. There are many variations on this theme; the relevant facts for our purposes are the following, which are extracted from [26, Theorem 1.2 and Proposition 3.1]; see also [30]. Our notation differs slightly from that in [26], and we maintain the terminology of A_∞–modules rather than "weak modules".

Theorem 7.3 (Franz; Goresky, Kottwitz and MacPherson) *There is a functor* $\boldsymbol{h}$ *from* A_∞-S*–modules to* A_∞-Λ*–modules with the following properties:*

1. *Let* $\pi\colon X \to B\mathrm{SO}(n)$ *be a space over* $B\mathrm{SO}(n)$*. Then the* A_∞-Λ*–modules* $\boldsymbol{h}(C^{-*}(X))$ *and* $C^{-*}(\mathrm{hofiber}(\pi))$ *are connected by a zig-zag of natural quasi-isomorphisms.*
2. *Let* V *be a graded vector space. Then* $\boldsymbol{h}(S\otimes V)\cong V$*, where* V *is regarded as a trivial* A_∞-Λ*–module.*

Proposition 7.4 *There is an equivalence of* $\mathrm{SO}(n)$*–modules*

$$C_*(S^n;\mathbb{Q})\simeq \mathbb{Q}\oplus\mathbb{Q}[n],$$

where the latter is regarded as a trivial $\mathrm{SO}(n)$*–module.*

Proof Both of the $\mathrm{SO}(n)$–modules in question are dualizable objects of $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$, so it suffices to exhibit an $\mathrm{SO}(n)$–equivalence $C^{-*}(\mathrm{SO}(n);\mathbb{Q})\simeq\mathbb{Q}\oplus\mathbb{Q}[-n]$ between the duals. By [42, Theorem 4.3.3.17], the homotopy category of the ∞–category of $\mathrm{SO}(n)$–modules coincides with the homotopy category obtained from the model category of $C_*(\mathrm{SO}(n);\mathbb{Q})$–modules equipped with the usual model structure on modules over a differential graded algebra. By [34, Section 4.3], this homotopy category in turn coincides with the full subcategory of the homotopy category of A_∞-$C_*(\mathrm{SO}(n);\mathbb{Q})$–modules spanned by the "homologically unital modules", so that, since the modules in question are homologically unital, it will suffice to produce to an isomorphism in the homotopy category of A_∞–modules. By [34, Section 6.2], it suffices to produce an isomorphism in the homotopy category of A_∞-Λ–modules after restricting along the A_∞–quasi-isomorphism $\Lambda\to C_*(\mathrm{SO}(n);\mathbb{Q})$ of [26]. For this, we apply the Koszul duality of Theorem 7.3 to the A_∞–quasi-isomorphism of Proposition 7.2, yielding the zig-zag of A_∞–quasi-isomorphisms

$$\mathbb{Q}\oplus\mathbb{Q}[-n]\simeq\boldsymbol{h}(S\oplus S[-n])\to\boldsymbol{h}\big(C^{-*}(E(\hat{\xi});\mathbb{Q})\big)\simeq C^{-*}(S^n;\mathbb{Q}).\qquad\square$$

Proof of Theorem 7.1 We explain the following diagram of $O(n)$–modules:

$$\mathbb{Q}\oplus\mathbb{Q}^{\det}[n]\to\mathbb{Q}[C_2]\oplus\mathbb{Q}[C_2][n]\simeq\mathbb{Q}[C_2]\otimes C_*(S^n;\mathbb{Q})\to C_*(S^n;\mathbb{Q}).$$

(1) Let e and σ denote the basis elements of $\mathbb{Q}[C_2]$ corresponding to the identity and generator, respectively. The left-hand map sends $1 \in \mathbb{Q}$ to $\frac{e+\sigma}{2}$ and $1 \in \mathbb{Q}^{\det}$ to $\frac{e-\sigma}{2}$. This is a map of C_2–modules and therefore of $O(n)$–modules, since $O(n)$ acts on both domain and codomain by restriction along the determinant.

(2) Fixing a choice of isomorphism $O(n) \cong C_2 \ltimes \mathrm{SO}(n)$, we obtain an isomorphism $C_*(O(n);\mathbb{Q}) \cong \mathbb{Q}[C_2] \otimes C_*(\mathrm{SO}(n);\mathbb{Q})$ of $O(n)$–modules. The middle equivalence is now obtained by applying the functor of induction from $\mathrm{SO}(n)$ to $O(n)$ to the equivalence of Proposition 7.4.

(3) The right-hand arrow is the counit of the induction-restriction adjunction.

Applying homology yields an isomorphism, completing the proof. □

7.2 Two-step nilpotent Lie algebras

In this section, we prove that the Lie algebras of interest to us are formal.

Proposition 7.5 *Let K be either $\mathbb{Q}$ or $\mathbb{Q}^{\mathrm{sgn}}$. For any $r \in \mathbb{Z}$ and any manifold M, the Lie algebra $\mathrm{Map}^{C_2}(\widetilde{M}^+, \mathcal{L}(K[r]))$ is formal.*

The proof will rely on the following technical result.

Proposition 7.6 *Let*

$$0 \to \mathfrak{h} \to \mathfrak{e} \to \mathfrak{g} \to 0$$

be an exact sequence of Lie algebras in $\mathcal{C}\mathrm{h}_{\mathbb{Q}}$ with $\mathfrak{g}$ and $\mathfrak{h}$ abelian. Assume that $\mathfrak{g}$ acts trivially on $\mathfrak{h}$ and that the underlying sequence of chain complexes splits. Then $\mathfrak{e}$ is formal.

Proof The hypotheses imply that the bracket on $\mathfrak{e} \cong \mathfrak{g} \oplus \mathfrak{h}$ is given by

$$[(g_1, h_1), (g_2, h_2)] = f(g_1, g_2)$$

for some (not uniquely defined) map $f\colon \mathrm{Sym}^2(\mathfrak{g}[1])[-2] \to \mathfrak{h}$, and the bracket on $H(\mathfrak{e}) \cong H(\mathfrak{g}) \oplus H(\mathfrak{h})$ is determined in the same way by f_*.

Choose quasi-isomorphisms $\varphi\colon \mathfrak{g} \to H(\mathfrak{g})$ and $\psi\colon \mathfrak{h} \to H(\mathfrak{h})$. Without loss of generality, we may assume that both maps induce the identity on homology. Let $\bar{\psi}$ be a quasi-inverse to ψ. Then $(\bar{\psi} \circ f_* \circ \varphi^{\wedge 2})_* = f_*$, so

$$\bar{\psi} \circ f_* \circ \varphi^{\wedge 2} - f = d_{\mathfrak{h}} G + G d_{\mathrm{Sym}}$$

for some homotopy operator $G\colon \mathrm{Sym}^2(\mathfrak{g}[1])[-2] \to \mathfrak{h}[-1]$.

Now, since $\mathfrak{g}$ is abelian and acts trivially, this equation may be written as

$$D(G) = \bar{\psi} \circ f_* \circ \varphi^{\wedge 2} - f,$$

where D denotes the differential in the Chevalley–Eilenberg cochain complex computing $H^*_{\mathcal{L}}(\mathfrak{g}, \mathfrak{h})$. Since extensions of $\mathfrak{g}$ by the module $\mathfrak{h}$ are classified by $H^2_{\mathcal{L}}(\mathfrak{g}, \mathfrak{h})$, it follows that f and $\bar{\psi} \circ f_* \circ \varphi^{\wedge 2}$ determine isomorphic extensions, so that we may take $f = \bar{\psi} \circ f_* \circ \varphi^{\wedge 2}$ after choosing a different splitting. But then $\psi \circ f = f_* \circ \varphi^{\wedge 2}$, so that the composite

$$\mathfrak{e} \xrightarrow{\cong} \mathfrak{g} \oplus \mathfrak{h} \xrightarrow{(\varphi,\psi)} H(\mathfrak{g}) \oplus (\mathfrak{h}) \xrightarrow{\cong} H(\mathfrak{e})$$

is a map of Lie algebras. Since it is also a quasi-isomorphism of chain complexes, the proof is complete. □

Proof of Proposition 7.5 The exact sequence

$$0 \to \mathfrak{h} \to \mathcal{L}(K[r]) \to K[r] \to 0$$

satisfies the hypotheses of Proposition 7.6, where

$$\mathfrak{h} = \begin{cases} K^{\otimes 2}[2r] & \text{for } r \text{ odd,} \\ 0 & \text{for } r \text{ even.} \end{cases}$$

By Proposition 3.18, we have

$$\operatorname{Map}^{C_2}(\widetilde{M}^+, \mathcal{L}(K[r]) \simeq (A_{\mathrm{PL}}(\widetilde{M}^+) \otimes \mathcal{L}(K[r]))^{C_2}.$$

Since the operations of tensoring with the commutative algebra $A_{\mathrm{PL}}(\widetilde{M}^+)$ and taking C_2 fixed points preserve the hypotheses of Proposition 7.6, the claim follows. □

Remark 7.7 Proposition 7.5 asserts that $A_{\mathrm{PL}}(M) \otimes \mathcal{L}(\mathbb{Q}[r])$ is formal whenever M is compact and orientable. When r is odd, this fact may be surprising at first glance, since M is not assumed to be formal.

A conceptual understanding of this phenomenon is afforded by the homotopy transfer theorem; see [38, Section 10.3], for example. Indeed, let A be any nonunital differential graded commutative algebra and $\mathfrak{g}$ any two-step nilpotent graded Lie algebra. Fixing an additive homotopy equivalence between A and $H(A)$, we obtain a transferred L_∞–algebra structure on $H(A) \otimes \mathfrak{g}$. The higher brackets of the transferred structure combine information about the Massey products of A and the Lie bracket of $\mathfrak{g}$.

In our case, using the fact that $\mathfrak{g}$ has no nontrivial iterated brackets, the explicit formulas of the homotopy transfer theorem show that these higher brackets all vanish, which implies that $A \otimes \mathfrak{g}$ is formal. In other words, although the Massey products in $H(A)$ may be nontrivial, they are damped out by the nilpotence of $\mathfrak{g}$.

References

[1] **V I Arnol'd**, *The cohomology ring of the colored braid group*, Math. Notes 5 (1969) 138–140 MR

[2] **E Artin**, *Theory of braids*, Ann. of Math. 48 (1947) 101–126 MR

[3] **D Ayala**, **J Francis**, *Poincaré/Koszul duality*, preprint (2014) arXiv

[4] **D Ayala**, **J Francis**, *Zero-pointed manifolds*, preprint (2014) arXiv

[5] **D Ayala**, **J Francis**, *Factorization homology of topological manifolds*, J. Topol. 8 (2015) 1045–1084 MR

[6] **D Ayala**, **J Francis**, **H L Tanaka**, *Factorization homology of stratified spaces*, Selecta Math. 23 (2017) 293–362 MR

[7] **D Ayala**, **J Francis**, **H L Tanaka**, *Local structures on stratified spaces*, Adv. Math. 307 (2017) 903–1028 MR

[8] **C-F Bödigheimer**, *Stable splittings of mapping spaces*, from "Algebraic topology" (H R Miller, D C Ravenel, editors), Lecture Notes in Math. 1286, Springer (1987) 174–187 MR

[9] **C-F Bödigheimer**, **F R Cohen**, *Rational cohomology of configuration spaces of surfaces*, from "Algebraic topology and transformation groups" (T tom Dieck, editor), Lecture Notes in Math. 1361, Springer (1988) 7–13 MR

[10] **C-F Bödigheimer**, **F Cohen**, **L Taylor**, *On the homology of configuration spaces*, Topology 28 (1989) 111–123 MR

[11] **A Borel**, *Sur l'homologie et la cohomologie des groupes de Lie compacts connexes*, Amer. J. Math. 76 (1954) 273–342 MR

[12] **A K Bousfield**, **V K A M Gugenheim**, *On PL de Rham theory and rational homotopy type*, Mem. Amer. Math. Soc. 179, Amer. Math. Soc. (1976) MR

[13] **F Cantero**, **M Palmer**, *On homological stability for configuration spaces on closed background manifolds*, preprint (2014) arXiv

[14] **T Church**, *Homological stability for configuration spaces of manifolds*, Invent. Math. 188 (2012) 465–504 MR

[15] **T Church**, **J S Ellenberg**, **B Farb**, *FI-modules and stability for representations of symmetric groups* (2012) arXiv

[16] **F R Cohen**, *On configuration spaces, their homology, and Lie algebras*, J. Pure Appl. Algebra 100 (1995) 19–42 MR

[17] **F R Cohen**, **T J Lada**, **J P May**, *The homology of iterated loop spaces*, Lecture Notes in Math. 533, Springer (1976) MR

[18] **F R Cohen**, **L R Taylor**, *Computations of Gel'fand–Fuks cohomology, the cohomology of function spaces, and the cohomology of configuration spaces*, from "Geometric applications of homotopy theory, I" (H R Miller, D C Ravenel, editors), Lecture Notes in Math. 657, Springer (1978) 106–143 MR

[19] **G Drummond-Cole**, **B Knudsen**, *Betti numbers of configuration spaces of surfaces*, J. London Math. Soc. (online publication July 2017)

[20] **E R Fadell**, **S Y Husseini**, *Geometry and topology of configuration spaces*, Springer (2001) MR

[21] **E Fadell**, **L Neuwirth**, *Configuration spaces*, Math. Scand. 10 (1962) 111–118 MR

[22] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Math. 205, Springer (2001) MR

[23] **Y Félix**, **D Tanré**, *The cohomology algebra of unordered configuration spaces*, J. London Math. Soc. 72 (2005) 525–544 MR

[24] **Y Félix**, **J-C Thomas**, *Rational Betti numbers of configuration spaces*, Topology Appl. 102 (2000) 139–149 MR

[25] **J Francis**, *The tangent complex and Hochschild cohomology of $\mathscr{E}_n$–rings*, Compos. Math. 149 (2013) 430–480 MR

[26] **M Franz**, *Koszul duality and equivariant cohomology*, Doc. Math. 11 (2006) 243–259 MR

[27] **B Fresse**, *Koszul duality of operads and homology of partition posets*, from "Homotopy theory: relations with algebraic geometry, group cohomology, and algebraic K–theory" (P Goerss, S Priddy, editors), Contemp. Math. 346, Amer. Math. Soc. (2004) 115–215 MR

[28] **E Getzler**, *The homology groups of some two-step nilpotent Lie algebras associated to symplectic vector spaces*, preprint (1999) arXiv

[29] **E Getzler**, *Resolving mixed Hodge modules on configuration spaces*, Duke Math. J. 96 (1999) 175–203 MR

[30] **M Goresky**, **R Kottwitz**, **R MacPherson**, *Equivariant cohomology, Koszul duality, and the localization theorem*, Invent. Math. 131 (1998) 25–83 MR

[31] **O Gwilliam**, *Factorization algebras and free field theories*, PhD thesis, Northwestern University (2012) MR Available at `https://search.proquest.com/docview/1018399857`

[32] **V Hinich**, *Homological algebra of homotopy algebras*, Comm. Algebra 25 (1997) 3291–3323 MR

[33] **G Horel**, *Higher Hochschild cohomology of the Lubin–Tate ring spectrum*, Algebr. Geom. Topol. 15 (2015) 3215–3252 MR

[34] **B Keller**, *Introduction to A–infinity algebras and modules*, Homology Homotopy Appl. 3 (2001) 1–35 MR

[35] **B Knudsen**, *Higher enveloping algebras*, preprint (2016) arXiv

[36] **B Knudsen**, *Higher enveloping algebras and configuration spaces of manifolds*, PhD thesis, Northwestern University (2016) MR Available at `https://search.proquest.com/docview/1799681505?pq-origsite=gscholar`

[37] **A Kupers**, **J Miller**, *Homological stability for topological chiral homology of completions*, Adv. Math. 292 (2016) 755–827 MR

[38] **J-L Loday**, **B Vallette**, *Algebraic operads*, Grundl. Math. Wissen. 346, Springer (2012) MR

[39] **R Longoni**, **P Salvatore**, *Configuration spaces are not homotopy invariant*, Topology 44 (2005) 375–380 MR

[40] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR

[41] **J Lurie**, *Derived algebraic geometry, X: Formal moduli problems*, preprint (2011) Available at `http://www.math.harvard.edu/~lurie/papers/DAG-X.pdf`

[42] **J Lurie**, *Higher algebra* (2014) Available at `http://www.math.harvard.edu/~lurie/papers/HA.pdf`

[43] **J P May**, *The geometry of iterated loop spaces*, Lectures Notes in Math. 271, Springer (1972) MR

[44] **D McDuff**, *Configuration spaces of positive and negative particles*, Topology 14 (1975) 91–107 MR

[45] **D Petersen**, *A spectral sequence for stratified spaces and configuration spaces of points*, Geom. Topol. 21 (2017) 2527–2555 MR

[46] **O Randal-Williams**, *Homological stability for unordered configuration spaces*, Q. J. Math. 64 (2013) 303–326 MR

[47] **E Riehl**, *Categorical homotopy theory*, New Mathematical Monographs 24, Cambridge Univ. Press (2014) MR

[48] **P Salvatore**, *Configuration spaces with summable labels*, from "Cohomological methods in homotopy theory" (J Aguadé, C Broto, C Casacuberta, editors), Progr. Math. 196, Birkhäuser, Basel (2001) 375–395 MR

[49] **P Salvatore**, **N Wahl**, *Framed discs operads and Batalin–Vilkovisky algebras*, Q. J. Math. 54 (2003) 213–231 MR

[50] **G Segal**, *Classifying spaces and spectral sequences*, Inst. Hautes Études Sci. Publ. Math. 34 (1968) 105–112 MR

[51] **G Segal**, *Configuration-spaces and iterated loop-spaces*, Invent. Math. 21 (1973) 213–221 MR

[52] **D P Sinha**, *The (non-equivariant) homology of the little disks operad*, from "Operads 2009" (J-L Loday, B Vallette, editors), Sémin. Congr. 26, Soc. Math. France, Paris (2013) 253–279 MR

[53] **B Totaro**, *Configuration spaces of algebraic varieties*, Topology 35 (1996) 1057–1067 MR

[54] **N Wahl**, *Ribbon braids and related operads*, PhD thesis, Oxford (2001) Available at `http://eprints.maths.ox.ac.uk/43/1/wahl.pdf`

[55] **J H Wang**, *On the braid groups for* $\mathbf{R}P^2$, J. Pure Appl. Algebra 166 (2002) 203–227 MR

Department of Mathematics, Harvard University
Cambridge, MA, United States

`knudsen@math.harvard.edu`

`http://scholar.harvard.edu/knudsen/`

Received: 8 December 2016 Revised: 9 December 2016

msp

Algebraic & Geometric Topology 17 (2017) 3189–3212

Presentably symmetric monoidal ∞–categories are represented by symmetric monoidal model categories

THOMAS NIKOLAUS
STEFFEN SAGAVE

We prove the theorem stated in the title. More precisely, we show the stronger statement that every symmetric monoidal left adjoint functor between presentably symmetric monoidal ∞-categories is represented by a strong symmetric monoidal left Quillen functor between simplicial, combinatorial and left proper symmetric monoidal model categories.

55U35; 18D10, 18G55

1 Introduction

The theory of ∞–categories has in recent years become a powerful tool for studying questions in homotopy theory and other branches of mathematics. It complements the older theory of Quillen model categories, and in many applications the interplay between the two concepts turns out to be crucial. In an important class of examples, the relation between ∞–categories and model categories is by now completely understood, thanks to work of Lurie [8, Appendix A.3] and Joyal [6], based on earlier results by Dugger [2]: On the one hand, every combinatorial simplicial model category $\mathcal{M}$ has an underlying ∞–category $\mathcal{M}_\infty$. This ∞–category $\mathcal{M}_\infty$ is *presentable*, ie it satisfies the set-theoretic smallness condition of being accessible and has all ∞–categorical colimits and limits. On the other hand, every presentable ∞–category is equivalent to the ∞–category associated with a combinatorial simplicial model category [8, Proposition A.3.7.6]. The presentability assumption is essential here since a sub-∞–category of a presentable ∞–category is in general not presentable, and does not come from a model category.

In many applications one studies combinatorial model categories $\mathcal{M}$ equipped with a symmetric monoidal product that is compatible with the model structure. The underlying ∞–category $\mathcal{M}_\infty$ of such a *symmetric monoidal model category* inherits the extra structure of a *symmetric monoidal ∞–category*; see Lurie [9, Example 4.1.3.6 and Proposition 4.1.3.10]. Since the monoidal product of $\mathcal{M}$ is a Quillen bifunctor,

Published: 19 September 2017 DOI: 10.2140/agt.2017.17.3189

$\mathcal{M}_\infty$ is an example of a *presentably symmetric monoidal ∞–category*, ie a symmetric monoidal ∞–category $\mathcal{C}$ which is presentable and whose associated tensor bifunctor $\otimes\colon \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ preserves colimits separately in each variable. In view of the above discussion, it is an obvious question whether every presentably symmetric monoidal ∞–category arises from a combinatorial symmetric monoidal model category. This was asked for example by Lurie [9, Remark 4.5.4.9]. The main result of the present paper is an affirmative answer to this question:

Theorem 1.1 *For every presentably symmetric monoidal ∞–category $\mathcal{C}$, there is a simplicial, combinatorial and left proper symmetric monoidal model category $\mathcal{M}$ whose underlying symmetric monoidal ∞–category is equivalent to $\mathcal{C}$.*

One can view this as a rectification result: the a priori weaker and more flexible notion of a symmetric monoidal ∞–category, which can encompass coherence data on all layers, can be rectified to a symmetric monoidal category where only coherence data up to degree 2 is allowed. An analogous result in the monoidal (but not symmetric monoidal) case is outlined in [9, Remark 4.1.4.9]. The symmetric result is significantly more complicated, as it is generally harder to rectify to a commutative structure than to an associative one. As we will see in Section 2.6 below, the theorem can actually be strengthened to a functorial version stating that symmetric monoidal left adjoint functors are represented by strong symmetric monoidal left Quillen functors.

The strategy of proof for Theorem 1.1 is as follows. Using localization techniques, we reduce the statement to the case of presheaf categories. By a result appearing in work of Pavlov and Scholbach [10], we can represent a symmetric monoidal ∞–category by an E_∞–algebra M in simplicial sets with the Joyal model structure. The main result of Kodjabachev and Sagave [7] implies that this E_∞–algebra can be rigidified to a strictly commutative monoid in the category of diagrams of simplicial sets indexed by finite sets and injections. We construct a chain of Quillen equivalences relating the contravariant model structure on sSet/M with a suitable contravariant model structure on objects over the commutative rigidification of M. The last step provides a symmetric monoidal model category, and employing a result by Gepner, Groth and Nikolaus [4] we show that it models the symmetric monoidal ∞–category of presheaves on M.

It is also worth noting that our proof of Theorem 1.1 does in fact provide a symmetric monoidal model category $\mathcal{M}$ with favorable properties: operad algebras in $\mathcal{M}$ inherit a model structure from $\mathcal{M}$, and weak equivalences of operads induces Quillen equivalences between the categories of operad algebras; see Theorem 2.5 below. In particular, there is a model structure on the category of commutative monoid objects in $\mathcal{M}$ which is Quillen equivalent to the lifted model structure on E_∞–objects in $\mathcal{M}$

and moreover models the ∞–category of commutative algebras in the ∞–category represented by $\mathcal{M}$. Hence, formally $\mathcal{M}$ behaves very much like symmetric spectra with the positive model structure.

1.2 Applications

Our main result allows one to abstractly deduce the existence of symmetric monoidal model categories that represent homotopy theories with only homotopy coherent symmetric monoidal structures. For example, it was unknown for a long time if there is a good point set level model for the smash product on the stable homotopy category. Since a presentably symmetric monoidal ∞–category that models the stable homotopy category can be established without referring to such a point set level model for the smash product, the existence of a model category of spectra with good smash product follows from our result. (Explicit constructions of such model categories of course predate the notion of presentably symmetric monoidal ∞–categories.)

But there are also examples where the question about the existence of symmetric monoidal models is open. One such example is the category of topological operads. It admits a tensor product, called the *Boardman–Vogt tensor product*, which controls the interchange of algebraic structures. The known symmetric monoidal point set level models for this tensor product cannot be derived, ie they do not give rise to a symmetric monoidal model category. However, for the underlying ∞–category of ∞–operads a presentably symmetric monoidal product is constructed by Lurie [9, Chapter 2.2.5]. In this case, our result allows to abstractly deduce the existence of a symmetric monoidal model category modeling operads with the Boardman–Vogt tensor product.

1.3 Organization

In Section 2 we show that Theorem 1.1 and its functorial enhancement can be reduced to the case of presheaf categories. In Section 3 we develop variants of the contravariant model structure that are compatible with the rigidification for E_∞–quasicategories recently developed by Kodjabachev and Sagave [7]. In the final Section 4 we prove that an instance of the contravariant model structure provides the desired result about presheaf categories.

Acknowledgments We would like to thank Gijs Heuts, Dmitri Pavlov and Markus Spitzweck for helpful discussions. Moreover, we would like to thank the referee for useful comments.

2 Reduction to presheaf categories

In this section we explain how Theorem 1.1 follows from a statement about presheaf categories that will be established in Section 3.

As defined by Lurie [9, Definition 2.0.0.7], a symmetric monoidal ∞–category is a cocartesian fibration of simplicial sets $\mathcal{C}^{\otimes} \to N(\mathcal{F}in_*)$ satisfying a certain condition. We explain in Proposition 4.1 below that a symmetric monoidal ∞–category can be represented by an E_∞–algebra in simplicial sets with the Joyal model structure. We also note that by [9, Example 4.1.3.6], every symmetric monoidal model category gives rise to a symmetric monoidal ∞–category, and every symmetric monoidal left Quillen functor induces a left adjoint symmetric monoidal functor between the respective ∞–categories.

Recall that an ∞–category $\mathcal{C}$ is called *presentable* if it is κ–accessible for some regular cardinal κ and admits all small colimits. In that case we can write $\mathcal{C}$ as an accessible localization of the category of presheaves $\mathcal{P}(\mathcal{C}^\kappa)$ on the full subcategory $\mathcal{C}^\kappa \subset \mathcal{C}$ of κ–compact objects. Here we denote the category of presheaves on an ∞–category $\mathcal{D}$ as $\mathcal{P}(\mathcal{D}) = \mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathcal{S})$, where $\mathcal{S} = N(\mathrm{Kan}^\Delta)$ is the ∞–category of spaces obtained as the homotopy coherent nerve of the simplicially enriched category of Kan complexes. Moreover $\mathcal{C}^\kappa$ is essentially small. Replacing $\mathcal{C}^\kappa$ by a small ∞–category $\mathcal{D}$ we see that every presentable ∞–category is equivalent to an accessible localization of the category of presheaves $\mathcal{P}(\mathcal{D})$ on some small ∞–category $\mathcal{D}$. For a detailed discussion of presentable ∞–categories and accessible localizations we refer the reader to [8, Chapter 5.5].

To study a symmetric monoidal analogue of this statement, we recall the following terminology from the introduction.

Definition 2.1 A symmetric monoidal ∞–category $\mathcal{C}$ is *presentably symmetric monoidal* if $\mathcal{C}$ is presentable and the associated tensor bifunctor $\otimes\colon \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ preserves colimits separately in each variable.

For every symmetric monoidal structure on an ∞–category $\mathcal{D}$, the ∞–category $\mathcal{P}(\mathcal{D})$ inherits a symmetric monoidal structure which by [9, Corollary 4.8.1.12] is uniquely determined by the following two properties:

- The tensor product makes $\mathcal{P}(\mathcal{D})$ into a presentably symmetric monoidal ∞–category.
- The Yoneda embedding $j\colon \mathcal{D} \to \mathcal{P}(\mathcal{D})$ can be extended to a symmetric monoidal functor.

We call this structure the *Day convolution symmetric monoidal structure*. It follows from [9, 4.8.1.10(4)] that it has the following universal property: for every presentably symmetric monoidal ∞–category $\mathcal{E}$, the Yoneda embedding $j\colon \mathcal{D} \to \mathcal{P}(\mathcal{D})$ induces an equivalence

$$\mathrm{Fun}^{L,\otimes}(\mathcal{P}(\mathcal{D}), \mathcal{E}) \to \mathrm{Fun}^{\otimes}(\mathcal{D}, \mathcal{E}).$$

Here $\mathrm{Fun}^{\otimes}$ denotes the ∞–category of symmetric monoidal functors and $\mathrm{Fun}^{L,\otimes}$ denotes the ∞–category of functors which are symmetric monoidal and in addition preserve all small colimits (or, equivalently, which are left adjoint).

In order to state our first structure result for presentably symmetric monoidal ∞–categories, let us recall the notion of a symmetric monoidal localization of a symmetric monoidal ∞–category $\mathcal{C}$. An accessible localization $L\colon \mathcal{C} \to \mathcal{C}$ is called *symmetric monoidal* if the full subcategory of local objects $\mathcal{C}^0 \subseteq \mathcal{C}$ admits a presentably symmetric monoidal structure such that the induced localization functor $L\colon \mathcal{C} \to \mathcal{C}^0$ admits a symmetric monoidal structure. In that case these symmetric monoidal structures are essentially unique. By [9, Proposition 2.2.1.9], the localization L is symmetric monoidal precisely if for every local equivalence $X \to Y$ in $\mathcal{C}$ and every object $Z \in \mathcal{C}$ the induced morphism $X \otimes Z \to Y \otimes Z$ is also a local equivalence. Note that this condition can be completely checked on the level of homotopy categories. See also [4, Section 3] for a discussion of symmetric monoidal localizations.

Proposition 2.2 *Every presentably symmetric monoidal ∞–category is an accessible, symmetric monoidal localization of the category of presheaves $\mathcal{P}(\mathcal{D})$ on some small, symmetric monoidal ∞–category $\mathcal{D}$.*

Proof Let $\mathcal{C}$ be a presentably symmetric monoidal ∞–category. Choose a regular cardinal κ such that $\mathcal{C}$ is κ accessible. By enlarging κ we can assume that the κ–compact objects $\mathcal{C}^\kappa \subset \mathcal{C}$ form a full symmetric monoidal subcategory. We can replace $\mathcal{C}^\kappa$ up to equivalence by a small, symmetric monoidal ∞–category $\mathcal{D}$ since it is essentially small. Then we find that $\mathcal{C}$ is an accessible localization of $\mathcal{P}(\mathcal{D})$. The inclusion $\mathcal{D} \simeq \mathcal{C}^\kappa \to \mathcal{P}(\mathcal{D})$ is by construction symmetric monoidal. We conclude that the localization functor $\mathcal{P}(\mathcal{D}) \to \mathcal{C}$ can be endowed with a symmetric monoidal structure with respect to the Day convolution symmetric monoidal structure, using the universal property of the Day convolution. By the description of symmetric monoidal localizations given above this finishes the proof. □

Following [1, Definition 1.21] (or rather [1, Corollary 2.7]), we say that a combinatorial model category is *tractable* if it admits a set of generating cofibrations with cofibrant domains.

Now assume that $\mathcal{M}$ is a simplicial, combinatorial, tractable and left proper symmetric monoidal model category. Denote the underlying symmetric monoidal ∞–category by $\mathcal{M}_\infty$. Let $L\colon \mathcal{M}_\infty \to \mathcal{M}_\infty$ be an accessible and symmetric monoidal localization. We say that a morphism $f\colon A \to B$ in $\mathcal{M}$ is

- a *local cofibration* if it is a cofibration in the original model structure on $\mathcal{M}$,
- a *local weak equivalence* if $L(\iota f)$ is an equivalence in $\mathcal{M}_\infty$, where ιf denotes the corresponding morphism in $\mathcal{M}_\infty$, and
- a *local fibration* if it has the right lifting property with respect to all morphisms in $\mathcal{M}$ which are simultaneously a cofibration and a weak equivalence.

Proposition 2.3 *The above choices of local cofibrations, local fibrations and local weak equivalences define a simplicial, combinatorial, tractable and left proper symmetric monoidal model structure. The underlying ∞–category of this model category $\mathcal{M}^{\mathrm{loc}}$ and the ∞–category of local objects $L\mathcal{M}_\infty \subseteq \mathcal{M}_\infty$ are equivalent as symmetric monoidal ∞–categories.*

Proof We use [8, Proposition A.3.7.3] to conclude that $\mathcal{M}^{\mathrm{loc}}$ exists and that it is a simplicial, combinatorial and left proper model category. By construction, it is a left Bousfield localization of $\mathcal{M}$. It remains to verify that the local model structure is symmetric monoidal. Since $\mathcal{M}$ is tractable, so is $\mathcal{M}^{\mathrm{loc}}$, and it follows from [1, Corollary 2.8] that we may assume that both the generating cofibrations of $\mathcal{M}^{\mathrm{loc}}$ and the generating acyclic cofibrations of $\mathcal{M}^{\mathrm{loc}}$ have cofibrant domains. To verify the pushout-product axiom, it therefore suffices to show that on the level of homotopy categories for an object $Z \in \mathrm{Ho}(\mathcal{M})$ and a local equivalence $X \to Y$ in $\mathrm{Ho}(\mathcal{M})$ the morphism of $X \otimes Z \to Y \otimes Z$ is a local equivalence as well (here the tensor is the tensor on the homotopy category, ie the derived tensor product). But this is true since the corresponding fact is true in the ∞–category $\mathcal{M}_\infty$ as discussed above.

By construction the ∞–category $L\mathcal{M}_\infty$ of local objects is modeled by the localized model structure $\mathcal{M}^{\mathrm{loc}}$. It remains to show that the two are equivalent as symmetric monoidal ∞–categories. To this end we just observe that the identity is a symmetric monoidal left Quillen functor $\mathcal{M} \to \mathcal{M}^{\mathrm{loc}}$. Thus the localized model structure endows $L\mathcal{M}_\infty$ with a symmetric monoidal structure such that the localization $\mathcal{M} \to \mathcal{M}^{\mathrm{loc}}$ is symmetric monoidal. But this was our defining property of the symmetric monoidal structure on $L\mathcal{M}_\infty$. □

The next proposition is the technical backbone of this paper and will be proven at the end of Section 3.

Proposition 2.4 *Let $\mathcal{D}$ be a small symmetric monoidal ∞–category. Then there exists a simplicial, combinatorial, tractable and left proper symmetric monoidal model category $\mathcal{M}$ whose underlying presentably symmetric monoidal ∞–category is symmetric monoidally equivalent to $\mathcal{P}(\mathcal{D})$ equipped with the Day convolution structure.*

We can now prove the main theorem from the introduction:

Proof of Theorem 1.1 Propositions 2.2 and 2.3 reduce the claim to the statement of Proposition 2.4. □

The following theorem establishes more properties of the symmetric monoidal model categories that are provided by our proof of Theorem 1.1.

Theorem 2.5 *Let $\mathcal{C}$ be a presentably symmetric monoidal ∞–category. Then the symmetric monoidal model category $\mathcal{M}$ of Theorem 1.1 can be chosen such that the following holds:*

(i) *For any operad $\mathcal{O}$ in* sSet, *the forgetful functor $\mathcal{M}[\mathcal{O}] \to \mathcal{M}$ from the category of $\mathcal{O}$–algebras in $\mathcal{M}$ creates a model structure on $\mathcal{M}[\mathcal{O}]$.*

(ii) *If $\mathcal{P} \to \mathcal{O}$ is a weak equivalence of operads, then the induced adjunction $\mathcal{M}[\mathcal{P}] \rightleftarrows \mathcal{M}[\mathcal{O}]$ is a Quillen equivalence. In particular, the categories of E_∞–objects and strictly commutative monoid objects in $\mathcal{M}$ are Quillen equivalent.*

(iii) *The ∞–category associated with the lifted model structure on commutative monoid objects in $\mathcal{M}$ is equivalent to the ∞–category of commutative algebra objects in the ∞–category $\mathcal{C}$.*

Proof Parts (i) and (ii) follow from our construction and Proposition 3.20 below. Part (iii) follows from [10, Theorem 7.10]. The *symmetric flatness* hypothesis needed for the latter theorem is verified in the proof of Proposition 3.20 below. □

2.6 Functoriality

We now provide a strengthening of our main result for functors. The methods and ideas are precisely the same as before, we only have to carefully keep track of the functoriality.

We first prove a slight generalization of Proposition 2.2. For the formulation, we say that a symmetric monoidal left adjoint functor $F\colon \mathcal{C} \to \mathcal{C}'$ between presentably symmetric monoidal ∞–categories is a *localization of a symmetric monoidal left adjoint functor*

$G\colon \mathcal{E} \to \mathcal{E}'$ if there is a commutative diagram of presentably symmetric monoidal ∞–categories

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{G} & \mathcal{E} \\ {\scriptstyle L}\downarrow & & \downarrow{\scriptstyle L'} \\ \mathcal{C} & \xrightarrow{F} & \mathcal{C}' \end{array}$$

in which the vertical functors L and L' are symmetric monoidal localizations. It is easy to see that once G and the localizations L and L' are given, G descends to a functor F if and only if it sends local equivalences to local equivalences. Moreover, F is determined up to equivalence by G in that case.

Lemma 2.7 *Let $F\colon \mathcal{C} \to \mathcal{C}'$ be a symmetric monoidal left adjoint functor between presentably symmetric monoidal ∞–categories. Then there exists a symmetric monoidal functor $f\colon \mathcal{D} \to \mathcal{D}'$ between small symmetric monoidal ∞–categories such that F is a localization of the left Kan extension $f_!\colon \mathcal{P}(\mathcal{D}) \to \mathcal{P}(\mathcal{D}')$.*

Proof First note that by [8, Proposition 5.4.7.7], every left adjoint functor $\mathcal{C} \to \mathcal{C}'$ preserves κ–compact objects for some κ, ie it restricts to a functor $F|_{\mathcal{C}^\kappa}\colon \mathcal{C}^\kappa \to (\mathcal{C}')^\kappa$. Since F is left adjoint, it is the left Kan extension of $F|_{\mathcal{C}^\kappa}$. This in turn implies that it is a localization of

$$(F|_{\mathcal{C}^\kappa})_!\colon \mathcal{P}(\mathcal{C}^\kappa) \to \mathcal{P}(\mathcal{C}'^\kappa).$$

Replacing the essentially small ∞–categories $\mathcal{C}^\kappa$ and $(\mathcal{C}')^\kappa$ by small categories proves the claim. □

In the proof of the next theorem we will use Proposition 4.3, which we state and prove in Section 4.

Theorem 2.8 *Let $F\colon \mathcal{C} \to \mathcal{C}'$ be a symmetric monoidal left adjoint functor between presentably symmetric monoidal ∞–categories. Then there exists a simplicial symmetric monoidal left adjoint functor $S\colon \mathcal{M} \to \mathcal{M}'$ between simplicial, combinatorial and left proper symmetric monoidal model categories $\mathcal{M}$ and $\mathcal{M}'$ such that the underlying functor $S_\infty\colon \mathcal{M}_\infty \to \mathcal{M}'_\infty$ is equivalent to F.*

Proof We first use Lemma 2.7 to conclude that there is a symmetric monoidal functor $f\colon \mathcal{D} \to \mathcal{D}'$ between small symmetric monoidal ∞–categories such that F is a localization of $f_!$. Using Proposition 4.3 below, we can realize $f_!$ as a left Quillen functor $S\colon \mathcal{M} \to \mathcal{M}'$ between symmetric monoidal model categories which model $\mathcal{P}(\mathcal{D})$ and $\mathcal{P}(\mathcal{D}')$. We now equip the categories $\mathcal{M}$ and $\mathcal{M}'$ with the local model structures which, by Proposition 2.3, correspond to the localization that give $\mathcal{C}$ and $\mathcal{C}'$. Since the functor $f_!$ descents to a local functor, it preserves local equivalences. Thus the functor S is also left Quillen with respect to the local model structures and the underlying functor of ∞–categories represents the functor F. □

3 The contravariant $\mathcal{I}$–model structure

In this section we set up the model structures that will be used in the proof of Proposition 2.4 and its functorial refinement Proposition 4.3.

3.1 The contravariant model structure

Let S be a simplicial set and let sSet/S be the category of objects over S. We recall from [8, Chapter 2.1.4] or [6, Section 8] that sSet/S admits a *contravariant* model structure where the cofibrations are the monomorphisms and the fibrant objects $X \to S$ are the *right fibrations*, ie the maps with the right lifting property with respect to the set of horn inclusions $\Lambda^n_i \subseteq \Delta^n$ for $0 < i \leq n$. As we will explain in Section 4, the contravariant model structure is relevant for our work because of its connection to presheaf categories coming from the straightening and unstraightening constructions [8, Chapter 2.2.1].

We will frequently use the following feature of the contravariant model structure:

Lemma 3.2 [8, Remark 2.1.4.12] *A morphism of simplicial sets $S \to T$ induces a Quillen adjunction $\mathrm{sSet}/S \rightleftarrows \mathrm{sSet}/T$ with respect to the contravariant model structures. If $S \to T$ is a Joyal equivalence of simplicial sets, then this adjunction is a Quillen equivalence.* □

For simplicial sets K and T, we consider the functor

$$K \times -\colon \mathrm{sSet}/T \to \mathrm{sSet}/K \times T \tag{3-1}$$

sending objects and morphisms in sSet/T to their product with id_K.

Lemma 3.3 *If $f\colon X \to Y$ is an acyclic cofibration in the contravariant model structure on sSet/T, then $K \times f$ is an acyclic cofibration in the contravariant model structure on $\mathrm{sSet}/K \times T$.*

We note that since we do not view $K \times -$ as an endofunctor of sSet/T by projecting away from K, this lemma is not implied by the fact that the contravariant model structure is simplicial.

Proof of Lemma 3.3 By [6, Lemma 8.16], the acyclic cofibrations in the contravariant model structure are characterized by the left lifting property with respect to the right fibrations between objects that are right fibrations relative to the base. Hence we have to prove that for every acyclic cofibration $U \to V$ in the contravariant model structure on sSet/T and for every commutative diagram

$$\begin{array}{ccc} K \times U & \longrightarrow & X \\ \downarrow & & \downarrow \\ K \times V & \longrightarrow & Y \\ \downarrow & & \downarrow \\ K \times T & \xrightarrow{=} & K \times T \end{array}$$

in sSet where the right-hand vertical maps are right fibrations, the upper square admits a lift $K \times V \to X$. Using the tensor/cotensor adjunction $(K \times -, (-)^K)$ on sSet, this is equivalent to finding a lift in the upper left-hand square in

$$\begin{array}{ccccc} U & \longrightarrow & T \times_{(K\times T)^K} X^K & \longrightarrow & X^K \\ \downarrow & & \downarrow & & \downarrow \\ V & \longrightarrow & T \times_{(K\times T)^K} Y^K & \longrightarrow & Y^K \\ \downarrow & & \downarrow & & \downarrow \\ T & \xrightarrow{=} & T & \longrightarrow & (K \times T)^K \end{array}$$

Since base change preserves right fibrations and the cotensor preserves right fibrations (by the dual of [8, Corollary 2.1.2.9]), the upper vertical map in the middle is a right fibration between right fibrations relative to T. □

Since $K \times -$ preserves contravariant cofibrations and all objects in sSet/T are cofibrant, Ken Brown's lemma and the preceding statement imply:

Corollary 3.4 *The functor* $K \times -\colon \mathrm{sSet}/T \to \mathrm{sSet}/K \times T$ *preserves contravariant weak equivalences.* □

3.5 The Joyal $\mathcal{I}$–model structure

Let $\mathcal{I}$ be the category with the finite sets $\boldsymbol{m} = \{1, \dots, m\}$ for $m \geq 0$ as objects and the injective maps as morphisms. An object $\boldsymbol{m}$ of $\mathcal{I}$ is *positive* if $|\boldsymbol{m}| \geq 1$, and $\mathcal{I}_+$ denotes the full subcategory of $\mathcal{I}$ spanned by the positive objects.

In the following, we briefly summarize the main results about the *Joyal $\mathcal{I}$–model structures* on the functor category $\mathrm{sSet}^{\mathcal{I}} = \mathrm{Fun}(\mathcal{I}, \mathrm{sSet})$ of $\mathcal{I}$–diagrams of simplicial sets from [7]. These results are motivated by (and largely derived from) the construction of the corresponding Kan model structures on $\mathrm{sSet}^{\mathcal{I}}$ in [12].

We say that a morphism f in $\mathrm{sSet}^{\mathcal{I}}$ is a *Joyal $\mathcal{I}$–equivalence* if $\mathrm{hocolim}_{\mathcal{I}} f$ is a Joyal equivalence in sSet. It is shown in [7, Proposition 2.3] that $\mathrm{sSet}^{\mathcal{I}}$ admits an *absolute* and a *positive* Joyal $\mathcal{I}$–model structure. In both cases, the weak equivalences are the Joyal $\mathcal{I}$–equivalences. An object X is fibrant in the absolute (resp. positive) model structure if each $\alpha\colon \boldsymbol{m} \to \boldsymbol{n}$ in $\mathcal{I}$ (resp. in $\mathcal{I}_+$) induces a weak equivalence of fibrant objects $\alpha_*\colon X(\boldsymbol{m}) \to X(\boldsymbol{n})$ in $\mathrm{sSet}_{\mathrm{Joyal}}$. In both cases, the $\mathcal{I}$–model structures arise as left Bousfield localizations of absolute or positive Joyal level model structures. Particularly, we will use that a Joyal $\mathcal{I}$–equivalence between positive $\mathcal{I}$–fibrant objects $X \to Y$ is a *positive Joyal level equivalence*, ie $X(\boldsymbol{m}) \to Y(\boldsymbol{m})$ is a Joyal equivalence for all $\boldsymbol{m}$ in $\mathcal{I}_+$. Finally, we note that by [7, Corollary 2.4], there are Quillen equivalences

$$\mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}} \underset{\mathrm{id}}{\overset{\mathrm{id}}{\rightleftarrows}} \mathrm{sSet}^{\mathcal{I}}_{\mathrm{abs}} \underset{\mathrm{const}_{\mathcal{I}}}{\overset{\mathrm{colim}_{\mathcal{I}}}{\rightleftarrows}} \mathrm{sSet}_{\mathrm{Joyal}}. \tag{3-2}$$

Concatenation of finite ordered sets induces a permutative monoidal structure on $\mathcal{I}$ with monoidal unit $\boldsymbol{0}$ and symmetry isomorphism the obvious block permutation. The functor category $\mathrm{sSet}^{\mathcal{I}}$ inherits a symmetric monoidal Day type convolution product $\boxtimes$ with monoidal unit $\mathcal{I}(\boldsymbol{0}, -)$ from the cartesian product in sSet and the concatenation in $\mathcal{I}$. Since $\mathrm{sSet}^{\mathcal{I}}$ is tensored over sSet, any operad $\mathcal{D}$ in sSet gives rise to a category $\mathrm{sSet}^{\mathcal{I}}[\mathcal{D}]$ of $\mathcal{D}$–algebras in $\mathrm{sSet}^{\mathcal{I}}$. The central feature of the positive model structure on $\mathrm{sSet}^{\mathcal{I}}$ is that without additional assumptions on $\mathcal{D}$, the forgetful functor $\mathrm{sSet}^{\mathcal{I}}[\mathcal{D}] \to \mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}}$ creates a *positive* model structure on $\mathrm{sSet}^{\mathcal{I}}[\mathcal{D}]$, where a map is weak equivalence or fibration if the underlying map in $\mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}}$ is [7, Theorem 3.1].

We say that an operad $\mathcal{E}$ in sSet is an E_∞–operad in $\mathrm{sSet}_{\mathrm{Joyal}}$ if Σ_n acts freely on the n^{th} space $\mathcal{E}(n)$ and $\mathcal{E}(n) \to *$ is a Joyal equivalence. If $\mathcal{E}$ is an E_∞–operad in $\mathrm{sSet}_{\mathrm{Joyal}}$, then the Joyal model structure on sSet lifts to a Joyal model structure on $\mathrm{sSet}[\mathcal{E}]$ by an argument analogous to the absolute case of [7, Theorem 3.1].

Theorem 3.6 [7, Theorem 1.2] *Let $\mathcal{E}$ be an E_∞–operad in* $\mathrm{sSet}_{\mathrm{Joyal}}$. *Then the canonical morphism $\Phi\colon \mathcal{E} \to \mathcal{C}$ to the commutativity operad and the composite adjunction in* (3-2) *induce a chain of Quillen equivalences*

$$\mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}}[\mathcal{C}] \underset{\Phi^*}{\overset{\Phi_*}{\leftrightarrows}} \mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}}[\mathcal{E}] \underset{\mathrm{const}_{\mathcal{I}}}{\overset{\mathrm{colim}_{\mathcal{I}}}{\rightleftarrows}} \mathrm{sSet}_{\mathrm{Joyal}}[\mathcal{E}].$$

The theorem leads to the following rigidification of E_∞–objects in $\mathrm{sSet}_{\mathrm{Joyal}}$ to $\mathcal{C}$–algebras in $\mathrm{sSet}^{\mathcal{I}}$, that is, to commutative monoids in $(\mathrm{sSet}^{\mathcal{I}}, \boxtimes)$.

Corollary 3.7 *Let M be an $\mathcal{E}$–algebra in $\mathrm{sSet}_{\mathrm{Joyal}}$. There exists a rigidification functor $(-)^{\mathrm{rig}}\colon \mathrm{sSet}^{\mathcal{I}}[\mathcal{E}] \to \mathrm{sSet}^{\mathcal{I}}[\mathcal{C}]$ and a natural chain of positive Joyal level equivalences between positive fibrant objects $\Phi^*(M^{\mathrm{rig}}) \leftarrow M^c \to \mathrm{const}_{\mathcal{I}} M$ in $\mathrm{sSet}^{\mathcal{I}}[\mathcal{E}]$.*

Proof This is analogous to the result about E_∞–spaces in [12, Corollary 3.7]: We let $M^c \xrightarrow{\sim} \mathrm{const}_{\mathcal{I}} M$ be a cofibrant replacement in $\mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}}[\mathcal{E}]$. Moreover, we let $\Phi_*(M^c) \to \Phi_*(M^c)^{\mathrm{fib}}$ be a fibrant replacement in $\mathrm{sSet}^{\mathcal{I}}_{\mathrm{pos}}[\mathcal{C}]$. Then the adjunction unit induces an $\mathcal{I}$–equivalence $M^c \to \Phi^*(\Phi_*(M^c)^{\mathrm{fib}})$. Since both objects are positive $\mathcal{I}$–fibrant, it is even a positive Joyal level equivalence. Hence $M^{\mathrm{rig}} = \Phi_*(M^c)^{\mathrm{fib}}$ has the desired property. □

3.8 The contravariant level and $\mathcal{I}$–model structures

Let $Z\colon \mathcal{I} \to \mathrm{sSet}$ be an $\mathcal{I}$–diagram of simplicial sets. We are interested in various model structures on the comma category $\mathrm{sSet}^{\mathcal{I}}/Z$ of objects over Z that are induced from the contravariant model structure. For this purpose, it is important to note that the category $\mathrm{sSet}^{\mathcal{I}}/Z$ can be obtained by assembling the comma categories $\mathrm{sSet}/Z(\boldsymbol{m})$ for varying $\boldsymbol{m}$. Indeed, every morphism $\alpha\colon \boldsymbol{m} \to \boldsymbol{n}$ in $\mathcal{I}$ induces an adjunction

$$\alpha_!\colon \mathrm{sSet}/Z(\boldsymbol{m}) \rightleftarrows \mathrm{sSet}/Z(\boldsymbol{n}) :\alpha^* \tag{3-3}$$

via composition with and base change along $\alpha_*\colon Z(\boldsymbol{m}) \to Z(\boldsymbol{n})$, and the adjunctions are compatible with the composition in $\mathcal{I}$. We also note that for every object $\boldsymbol{m}$ of $\mathcal{I}$, there is an adjunction

$$F_{\boldsymbol{m}}\colon \mathrm{sSet}/Z(\boldsymbol{m}) \rightleftarrows \mathrm{sSet}^{\mathcal{I}}/Z :\mathrm{Ev}_{\boldsymbol{m}} \tag{3-4}$$

with right adjoint $\mathrm{Ev}_{\boldsymbol{m}}(X \to Z) = X(\boldsymbol{m}) \to Z(\boldsymbol{m})$ and left adjoint

$$F_{\boldsymbol{m}}(K \to Z(\boldsymbol{m})) = \Bigl(\boldsymbol{n} \mapsto \coprod_{(\alpha\colon \boldsymbol{m} \to \boldsymbol{n}) \in \mathcal{I}} \alpha_!(K \to Z(\boldsymbol{m}))\Bigr).$$

A morphism $X \to Y$ in $\mathrm{sSet}^{\mathcal{I}}/Z$ is defined to be

- an absolute (resp. positive) contravariant level equivalence if for each object (resp. each positive object) $\boldsymbol{m}$ of $\mathcal{I}$, the morphism $X(\boldsymbol{m}) \to Y(\boldsymbol{m})$ is a contravariant weak equivalence in $\mathrm{sSet}/Z(\boldsymbol{m})$,
- an absolute (resp. positive) contravariant level fibration if for each object (resp. each positive object) $\boldsymbol{m}$ of $\mathcal{I}$, the morphism $X(\boldsymbol{m}) \to Y(\boldsymbol{m})$ is a fibration in the contravariant model structure on $\mathrm{sSet}/Z(\boldsymbol{m})$, and

- an absolute (resp. positive) contravariant cofibration if it has the left lifting property with respect to all morphisms that are absolute (resp. positive) contravariant level fibrations and equivalences.

Lemma 3.9 *These classes of maps define an* absolute *(resp. a* positive*)* contravariant level model structure *on* $\mathrm{sSet}^{\mathcal{I}}/Z$ *which is simplicial, combinatorial, tractable and left proper.*

Proof The key observation is that by Lemma 3.2, the adjunction (3-3) is a Quillen adjunction with respect to the contravariant model structures. With this observation, the existence of the absolute contravariant level model structure follows by a standard lifting argument using the adjunction

$$\prod_{\boldsymbol{m}\in\mathcal{I}} \mathrm{sSet}/Z(\boldsymbol{m}) \rightleftarrows \mathrm{sSet}^{\mathcal{I}}/Z$$

induced by the adjunctions $(F_{\boldsymbol{m}}, \mathrm{Ev}_{\boldsymbol{m}})$ from (3-4) and the product model structure on the codomain; compare [1, Theorem 2.28]. If $I_{Z(\boldsymbol{m})}$ is a set of generating cofibrations for $\mathrm{sSet}/Z(\boldsymbol{m})$, then $\{F_{\boldsymbol{m}}(i) \mid \boldsymbol{m} \in \mathcal{I}, i \in I_{Z(\boldsymbol{m})}\}$ is a set of generating cofibrations for the absolute contravariant level model structure, and similarly for the generating acyclic cofibrations. The model structure is obviously tractable, and it is simplicial and left proper since $\mathrm{sSet}/Z(\boldsymbol{m})$ is.

In the positive case, we index the above product by the objects of $\mathcal{I}_+$ instead. □

The contravariant model structure on $\mathrm{sSet}/Z(\boldsymbol{m})$ is cofibrantly generated and left proper. Since its cofibrations are the monomorphisms, we may use

$$I_{Z(\boldsymbol{m})} = \big\{(K \to Z(\boldsymbol{m})) \to (L \to Z(\boldsymbol{m})) \mid (K \to L) = (\partial\Delta^n \hookrightarrow \Delta^n)\big\}$$

as a set of generating cofibrations of $\mathrm{sSet}/Z(\boldsymbol{m})$. Let $W_{Z(\boldsymbol{m})}$ be the set of objects in $\mathrm{sSet}/Z(\boldsymbol{m})$ given by the domains and codomains of $I_{Z(\boldsymbol{m})}$. By [3, Proposition A.5], a map $U \to V$ of fibrant objects in the contravariant model structure on $\mathrm{sSet}/Z(\boldsymbol{m})$ is a contravariant weak equivalence if and only if the induced morphism of simplicial mapping spaces $\mathrm{Map}_{Z(\boldsymbol{m})}(K, U) \to \mathrm{Map}_{Z(\boldsymbol{m})}(K, V)$ is a weak homotopy equivalence of simplicial sets for every object $K \to Z(\boldsymbol{m})$ in $W_Z(\boldsymbol{m})$. For an object $K \to Z(\boldsymbol{m})$ in $W_Z(\boldsymbol{m})$ and a morphism $\alpha\colon \boldsymbol{m} \to \boldsymbol{n}$ in $\mathcal{I}$, we let

$$F_{\boldsymbol{n}}(\alpha_!(K)) \to F_{\boldsymbol{m}}(K)$$

be the morphism in $\mathrm{sSet}^{\mathcal{I}}/Z$ that is adjoint to the inclusion

$$\alpha_!(K) \hookrightarrow \coprod_{(\beta\colon \boldsymbol{m}\to\boldsymbol{n})\in\mathcal{I}} \beta_!(K) = \mathrm{Ev}_{\boldsymbol{n}}(F_{\boldsymbol{m}}(K))$$

of the summand indexed by α. We write

(3-5) $S^Z = \{F_{\boldsymbol{n}}(\alpha_!(A)) \to F_{\boldsymbol{m}}(A) \mid (\alpha\colon \boldsymbol{m} \to \boldsymbol{n}) \in \mathcal{I} \text{ and } (A \to Z(\boldsymbol{m})) \in W_Z(\boldsymbol{m})\}$

for the set of all such maps and let S^Z_+ be the subset of S^Z consisting those maps that come from $\alpha \in \mathcal{I}_+$.

Proposition 3.10 *The left Bousfield localization of the absolute (resp. positive) contravariant level model structure on* $\mathrm{sSet}^{\mathcal{I}}/Z$ *with respect to* S^Z *(resp.* S^Z_+*) exists. It is a simplicial, combinatorial, tractable and left proper model structure.* □

We refer to this model structure as the *absolute* (resp. *positive*) *contravariant $\mathcal{I}$–model structure*. The weak equivalences in these model structures are called *absolute* (resp. *positive*) $\mathcal{I}$–equivalences. The cofibrations are the same as in the respective level model structures. An object $X \to Z$ is absolute (resp. positive) contravariant $\mathcal{I}$–fibrant if is absolute (resp. positive) contravariant level fibrant an each $\alpha\colon \boldsymbol{m} \to \boldsymbol{n}$ in $\mathcal{I}$ (resp. in $\mathcal{I}_+$) induces a contravariant weak equivalence $X(\boldsymbol{m}) \to \alpha^*(X(\boldsymbol{n}))$ in $\mathrm{sSet}/Z(\boldsymbol{m})$.

The contravariant $\mathcal{I}$–model structures are homotopy invariant in level equivalences of the base:

Lemma 3.11 *Let $Z \to Z'$ be a morphism in* $\mathrm{sSet}^{\mathcal{I}}$*. Then the induced adjunction* $\mathrm{sSet}^{\mathcal{I}}/Z \rightleftarrows \mathrm{sSet}^{\mathcal{I}}/Z'$ *is a Quillen adjunction with respect to the absolute and positive contravariant $\mathcal{I}$–model structures. If $Z \to Z'$ is an absolute (resp. a positive) Joyal level equivalence, then it is a Quillen equivalence with respect to the absolute (resp. positive) contravariant $\mathcal{I}$–model structures.*

Proof We treat the absolute case; the positive case is similar. It is clear that the adjunction in question is a Quillen adjunction with respect to the absolute level model structure. Since $(Z \to Z')_!(S_Z)$ is a subset of $S_{Z'}$, there is an induced Quillen adjunction on the localizations. Using Lemma 3.2, it is also clear that an absolute Joyal level equivalence induces a Quillen equivalence with respect to the absolute contravariant level model structures. To see that it is a Quillen equivalence, we note that by adjunction, the $(Z \to Z')_!(S_Z)$–local objects coincide with the $S_{Z'}$–local objects. □

We write $(-)_{\mathcal{I}} = \mathrm{colim}_{\mathcal{I}}$ for the colimit over $\mathcal{I}$ and note that the adjunction

$$(-)_{\mathcal{I}}\colon \mathrm{sSet}^{\mathcal{I}} \rightleftarrows \mathrm{sSet} : \mathrm{const}_{\mathcal{I}}$$

induces adjunctions of overcategories

(3-6) $$\mathrm{sSet}^{\mathcal{I}}/Z \rightleftarrows \mathrm{sSet}^{\mathcal{I}}/\,\mathrm{const}_{\mathcal{I}}(Z_{\mathcal{I}}) \rightleftarrows \mathrm{sSet}/Z_{\mathcal{I}}.$$

Lemma 3.12 *Let Z be cofibrant and fibrant in the absolute Joyal $\mathcal{I}$–model structure on $\mathrm{sSet}^{\mathcal{I}}$. Then the composite adjunction $\mathrm{sSet}^{\mathcal{I}}/Z \rightleftarrows \mathrm{sSet}/Z_{\mathcal{I}}$ is a Quillen equivalence with respect to the absolute contravariant $\mathcal{I}$–model structure on $\mathrm{sSet}^{\mathcal{I}}/Z$ and the contravariant model structure on $\mathrm{sSet}/Z_{\mathcal{I}}$.*

Proof Since Z is cofibrant and fibrant, the Quillen equivalence (3-2) shows that the adjunction unit $Z \to \mathrm{const}_{\mathcal{I}}(Z_{\mathcal{I}})$ is an absolute Joyal level equivalence. Hence the first adjunction in (3-6) is a Quillen equivalence by Lemma 3.11. It follows from the definitions that the second adjunction is a Quillen adjunction whose right adjoint detects weak equivalences between fibrant objects. Hence it is sufficient to show that the derived adjunction unit is an absolute contravariant $\mathcal{I}$–equivalence. Let $X \to \mathrm{const}_{\mathcal{I}}(Z_{\mathcal{I}})$ be a cofibrant object in the absolute contravariant $\mathcal{I}$–model structure. A fibrant replacement $X \to X'$ and the adjunction counit of $(F_{\mathbf{0}}, \mathrm{Ev}_{\mathbf{0}})$ provide a chain of absolute contravariant $\mathcal{I}$–equivalences between cofibrant objects

$$X \overset{\sim}{\rightarrowtail} X' \overset{\sim}{\longleftarrow} F_{\mathbf{0}}\mathrm{Ev}_{\mathbf{0}}(X').$$

Since $\mathbf{0}$ is initial in $\mathcal{I}$, there is an isomorphism $F_{\mathbf{0}}\mathrm{Ev}_{\mathbf{0}}(X') \cong \mathrm{const}_{\mathcal{I}} X'(\mathbf{0})$. The claim follows because the evaluation of the adjunction unit of $((-)_{\mathcal{I}}, \mathrm{const}_{\mathcal{I}})$ on $\mathrm{const}_{\mathcal{I}} X'(\mathbf{0})$ is even an isomorphism and $\mathrm{const}_{\mathcal{I}}$ preserves weak equivalence between all objects. □

Proposition 3.13 *For every absolute Joyal $\mathcal{I}$–fibrant Z in $\mathrm{sSet}^{\mathcal{I}}$, the identity functors form a Quillen equivalence $(\mathrm{sSet}^{\mathcal{I}}/Z)_{\mathrm{pos}} \rightleftarrows (\mathrm{sSet}^{\mathcal{I}}/Z)_{\mathrm{abs}}$ with respect to the positive and absolute contravariant $\mathcal{I}$–model structures.*

Proof Let $Z^c \to Z$ be a cofibrant replacement in the absolute Joyal $\mathcal{I}$–model structure and let $Z^c \to \mathrm{const}_{\mathcal{I}}(Z^c_{\mathcal{I}})$ be the adjunction unit. Since these two maps are absolute Joyal level equivalences, Lemma 3.11 and the two out of three property for Quillen equivalences reduce the claim to the case where $Z = \mathrm{const}_{\mathcal{I}} T$ for a simplicial set T.

The category $\mathrm{sSet}^{\mathcal{I}}/\mathrm{const}_{\mathcal{I}} T$ is equivalent to the category $(\mathrm{sSet}/T)^{\mathcal{I}}$ of $\mathcal{I}$–diagrams in sSet/T. Under this equivalence, the absolute contravariant $\mathcal{I}$–model structure corresponds to the homotopy colimit model structure on $(\mathrm{sSet}/T)^{\mathcal{I}}$ provided by [3, Theorem 5.1]. The cited theorem implies that the weak equivalences in the absolute contravariant $\mathcal{I}$–model structure are the maps that induce contravariant weak equivalences under $\mathrm{hocolim}_{\mathcal{I}}\colon (\mathrm{sSet}/T)^{\mathcal{I}} \to \mathrm{sSet}/T$.

The argument for comparing the model structures now works as in [7, Proposition 2.3]: The inclusion $\mathcal{I}_+ \to \mathcal{I}$ is homotopy cofinal [12, Proof of Corollary 5.9], and hence every positive contravariant level equivalence is an $\mathrm{hocolim}_{\mathcal{I}}$–equivalence. Together with $S_+^{\mathrm{const}_{\mathcal{I}} T} \subset S^{\mathrm{const}_{\mathcal{I}} T}$, this shows that every positive contravariant $\mathcal{I}$–equivalence is an absolute contravariant $\mathcal{I}$–equivalence. For the converse, it suffices to show that a $\mathrm{hocolim}_{\mathcal{I}}$–equivalence of positive contravariant $\mathcal{I}$–fibrant objects is a positive contravariant $\mathcal{I}$–equivalence. Using again that $\mathcal{I}_+ \to \mathcal{I}$ is homotopy cofinal, this follows by restricting along $\mathcal{I}_+ \to \mathcal{I}$ and applying [3, Theorem 5.1(a)] in $(\mathrm{sSet}/T)^{\mathcal{I}_+}$. □

Corollary 3.14 *If Z is absolute Joyal cofibrant and positive Joyal $\mathcal{I}$–fibrant, then $\mathrm{sSet}^{\mathcal{I}}/Z \rightleftarrows \mathrm{sSet}/Z_{\mathcal{I}}$ is a Quillen equivalence with respect to the positive contravariant $\mathcal{I}$–model structure on $\mathrm{sSet}^{\mathcal{I}}/Z$ and the contravariant model structure on $\mathrm{sSet}/Z_{\mathcal{I}}$.*

Proof Since the derived adjunction unit $Z \to \mathrm{const}_{\mathcal{I}}((Z_{\mathcal{I}})^{\text{Joyal-fib}}) = Z'$ is a positive level equivalence, the adjunction $\mathrm{sSet}^{\mathcal{I}}/Z \rightleftarrows \mathrm{sSet}^{\mathcal{I}}/Z'$ is a Quillen equivalence with respect to the positive contravariant $\mathcal{I}$–model structure by Lemma 3.11. Because Z' is cofibrant and fibrant in the absolute Joyal $\mathcal{I}$–model structure, Proposition 3.13 and Lemma 3.12 show the claim. □

Let N be a commutative monoid object in $(\mathrm{sSet}^{\mathcal{I}}, \boxtimes)$. Then the overcategory $\mathrm{sSet}^{\mathcal{I}}/N$ inherits a symmetric monoidal product

$$(X \to N) \boxtimes (Y \to N) = (X \boxtimes Y \to N \boxtimes N \to N)$$

from the symmetric monoidal structure of N and the multiplication of N.

The following result is a key step in the proof of our main result:

Theorem 3.15 *Let $\mathcal{E}$ be an E_∞–operad in $\mathrm{sSet}_{\mathrm{Joyal}}$ and let M be an $\mathcal{E}$–algebra. Then there is a chain of Quillen equivalences of simplicial, combinatorial and left proper model categories*

$$\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}} \leftrightarrows \mathrm{sSet}^{\mathcal{I}}/M^{c} \rightleftarrows \mathrm{sSet}^{\mathcal{I}}/\mathrm{const}_{\mathcal{I}} M \rightleftarrows \mathrm{sSet}/M$$

relating sSet/M with the contravariant model structure and the symmetric monoidal model category $\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}}$ with the positive contravariant $\mathcal{I}$–model structure. The chain is natural with respect to M.

Proof Using the chain of positive level equivalences $M^{\mathrm{rig}} \leftarrow M^{c} \to \mathrm{const}_{\mathcal{I}} M$ from Corollary 3.7 and the fact that $\mathrm{const}_{\mathcal{I}} M \cong F_{\mathbf{0}} M$ is absolute Joyal $\mathcal{I}$–cofibrant, the chain of Quillen equivalences is a consequence of Lemma 3.11 and Corollary 3.14. It is shown in Corollary 3.19 that $\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}}$ satisfies the pushout product axiom. □

We need one more observation about the tensor product on $\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}}$. We call an object in $\mathrm{Ho}(\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}})$ representable if it corresponds to an object of the form $\Delta^0 \to M$ under the equivalence $\mathrm{Ho}(\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}}) \simeq \mathrm{Ho}(\mathrm{sSet}/M)$ induced by the chain of Quillen equivalences from Theorem 3.15. Note that these are precisely the objects which correspond to representable presheaves under the equivalence to presheaves on the ∞–category M.

Lemma 3.16 *The tensor product of two representables in* $\mathrm{Ho}(\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}})$ *is again representable.*

Proof It follows from the construction of M^{rig} and the chain of Quillen equivalences that the representables in $\mathrm{Ho}(\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}})$ are represented by the cofibrant objects of the form $F^{\mathcal{I}}_{\boldsymbol{k}}(\Delta^0) \to M$ with $\boldsymbol{k}$ an positive object of $\mathcal{I}$. Since $F^{\mathcal{I}}_{\boldsymbol{k}}(K) \boxtimes F^{\mathcal{I}}_{\boldsymbol{l}}(L) \cong F^{\mathcal{I}}_{\boldsymbol{k}\sqcup\boldsymbol{l}}(K \times L)$, this set of objects is closed under the monoidal product. □

3.17 Monoidal properties of the contravariant $\mathcal{I}$-model structure

The following proposition is the key tool for the homotopical analysis of the $\boxtimes$–product on $\mathrm{sSet}^{\mathcal{I}}/N$ for a commutative N. Both its statement and proof are analogous to [12, Proposition 8.2; 7, Proposition 2.6]:

Proposition 3.18 *Let N be a commutative monoid object in* $\mathrm{sSet}^{\mathcal{I}}$*. If $X \to N$ is absolute contravariant cofibrant, then* $X \boxtimes - \colon \mathrm{sSet}^{\mathcal{I}}/N \to \mathrm{sSet}^{\mathcal{I}}/N$ *preserves positive contravariant $\mathcal{I}$–equivalences between arbitrary objects.*

Proof We begin by showing that if $Y_1 \to Y_2$ is an absolute contravariant level equivalence in $\mathrm{sSet}^{\mathcal{I}}/N$, then so is $X \boxtimes Y_1 \to X \boxtimes Y_2$. For this, we use a cell induction argument and first consider the case $X = F_{\boldsymbol{m}}(K)$.

By [12, Lemma 5.6], the map $(F_{\boldsymbol{m}}(K) \boxtimes (Y_1 \to Y_2))(\boldsymbol{n})$ is isomorphic to

$$K \times \big(\mathrm{colim}_{\boldsymbol{m}\sqcup\boldsymbol{k}\to\boldsymbol{n}}(Y_1(\boldsymbol{k}) \to Y_2(\boldsymbol{k}))\big) \tag{3-7}$$

where the colimit is taken over the comma category $(\boldsymbol{m} \sqcup - \downarrow \boldsymbol{n})$. Since each connected component of this comma category has a terminal object, we can choose a set A of morphisms $\alpha\colon \boldsymbol{m} \sqcup \boldsymbol{k} \to \boldsymbol{n}$ such that (3-7) is isomorphic to

$$\coprod_{(\alpha\colon \boldsymbol{m}\sqcup\boldsymbol{k}\to\boldsymbol{n})\in A} K \times (Y_1(\boldsymbol{k}) \to Y_2(\boldsymbol{k})).$$

Using Corollary 3.4, it follows that each summand is a contravariant weak equivalence in $\mathrm{sSet}/(K \times N(\boldsymbol{k}))$. Composing with the map

$$K \times N(\boldsymbol{k}) \to N(\boldsymbol{m}) \times N(\boldsymbol{k}) \to N(\boldsymbol{n})$$

induced by the morphism $\alpha\colon \boldsymbol{k} \sqcup \boldsymbol{m} \to \boldsymbol{n}$ indexing the summand, it follows that each summand is a contravariant weak equivalence in $\mathrm{sSet}/N(\boldsymbol{n})$. Hence (3-7) is a contravariant weak equivalence in $\mathrm{sSet}/N(\boldsymbol{n})$.

Next we assume that $F_{\boldsymbol{m}}(K) \to F_{\boldsymbol{m}}(L)$ is a generating cofibration in $\mathrm{sSet}^{\mathcal{I}}/N$, that $X_{\alpha+1}$ is the pushout of $F_{\boldsymbol{m}}(L) \leftarrow F_{\boldsymbol{m}}(K) \to X_\alpha$ in $\mathrm{sSet}^{\mathcal{I}}/N$ and that $X_\alpha \boxtimes -$ preserves absolute contravariant level equivalences. By the above decomposition, $F_{\boldsymbol{m}}(K \to L) \boxtimes Y_i$ is a cofibration when evaluated at $\boldsymbol{n}$, and the gluing lemma in the left proper model category $\mathrm{sSet}/N(\boldsymbol{n})$ shows that $X_{\alpha+1} \boxtimes (Y_1 \to Y_2)$ is an absolute contravariant level equivalence in sSet/N. Since a general absolute contravariant cofibrant object X is a retract of a colimit of a sequence of maps of this form, it follows that $X \boxtimes -$ preserves absolute contravariant level equivalences.

We now turn to the statement of the proposition and assume that $Y_1 \to Y_2$ is a positive contravariant $\mathcal{I}$–equivalence. By applying the previous argument to cofibrant replacements of the Y_i, we may assume that the Y_i are absolute contravariant cofibrant. Let $Y_2 \rightarrowtail N^c \xrightarrow{\sim} \twoheadrightarrow N$ be a factorization in the absolute Joyal model structure. By Lemma 3.11, $Y_1 \to Y_2$ is a positive contravariant $\mathcal{I}$–equivalence in $\mathrm{sSet}^{\mathcal{I}}/N^c$. Since the induced map of colimits is a contravariant equivalence in $\mathrm{sSet}/(X_{\mathcal{I}} \times N^c_{\mathcal{I}})$ by Corollaries 3.4 and 3.14, another application of Corollary 3.14 shows that the induced map $X \boxtimes Y_1 \to X \boxtimes Y_2$ is a positive contravariant $\mathcal{I}$–equivalence in $\mathrm{sSet}^{\mathcal{I}}/(X \boxtimes N^c)$. Composing with $X \boxtimes N^c \to N \boxtimes N \to N$ shows that $X \boxtimes Y_1 \to X \boxtimes Y_2$ is a positive contravariant $\mathcal{I}$–equivalence in $\mathrm{sSet}^{\mathcal{I}}/N$. □

Corollary 3.19 *Let N be a commutative monoid object in $\mathrm{sSet}^{\mathcal{I}}$. The positive contravariant $\mathcal{I}$–model structure on $\mathrm{sSet}^{\mathcal{I}}/N$ satisfies the pushout product axiom and the monoid axiom as defined in* [13].

Proof The cofibration part of the pushout product axiom follows from Proposition 8.4 of [12]. As explained there, Proposition 3.18 implies the statement about the generating acyclic cofibrations.

For the monoid axiom, we have to show that transfinite composition of cobase changes of maps of the form $X \boxtimes (Y_1 \to Y_2)$ with $Y_1 \to Y_2$ an acyclic cofibration are contravariant $\mathcal{I}$–equivalences. Since $\mathrm{sSet}^{\mathcal{I}}/N$ is tractable, we may assume that also the generating acyclic cofibrations of the positive contravariant $\mathcal{I}$–model structure have cofibrant domains and codomains [1, Corollary 2.8]. Using Proposition 3.18 and a cofibrant replacement of X, it follows that $X \boxtimes (Y_1 \to Y_2)$ is a contravariant $\mathcal{I}$–equivalence. It is also an injective level cofibration, ie a cofibration when evaluated at any object $\boldsymbol{n}$ of $\mathcal{I}$. Using a cofibrant replacement in the absolute contravariant level model structure, it follows that cobase changes and transfinite compositions preserve morphisms that are both contravariant $\mathcal{I}$–equivalences and injective level cofibrations. □

The next proposition states that (any monoidal left Bousfield localization of) the positive contravariant $\mathcal{I}$–model structure on $\mathrm{sSet}^{\mathcal{I}}/N$ lifts to operad algebras in the best possible way.

Proposition 3.20 *Let N be a commutative monoid object in* $\mathrm{sSet}^{\mathcal{I}}$*, let $\mathcal{M}$ be a left Bousfield localization of the positive contravariant $\mathcal{I}$–model structure on* $\mathrm{sSet}^{\mathcal{I}}/N$*, and assume that $\mathcal{M}$ satisfies the pushout product axiom with respect to $\boxtimes$.*

(i) *For any operad $\mathcal{O}$ in* sSet*, the forgetful functor $\mathcal{M}[\mathcal{O}] \to \mathcal{M}$ from the category of $\mathcal{O}$–algebras in $\mathcal{M}$ creates a model structure on $\mathcal{M}[\mathcal{O}]$.*

(ii) *If $\mathcal{P} \to \mathcal{O}$ is a weak equivalence of operads, then the induced adjunction $\mathcal{M}[\mathcal{P}] \rightleftarrows \mathcal{M}[\mathcal{O}]$ is a Quillen equivalence.*

Proof The criteria given in [10, Theorems 5.10 and 7.5] reduce this to showing that $\mathcal{M}$ is *symmetric h–monoidal* and *symmetric flat* in the sense of [11, Definitions 4.2.4 and 4.2.7].

As a first step, we show that the levelwise cofibrations in $\mathcal{M}$ are h–cofibrations in the sense of [11, Definition 2.0.4], ie that cobase change along levelwise cofibrations preserves weak equivalences. For this it is sufficient that pushouts along levelwise cofibrations are homotopy pushouts in $\mathcal{M}$. Let $V \leftarrow U \to X$ be a diagram in $\mathcal{M}$ with $U \to V$ a levelwise cofibration. Let $U \to V' \to V$ be a factorization of $U \to V$ into a positive $\mathcal{I}$–cofibration $U \to V'$ and a positive level equivalence $V' \to V$. Then the induced map of pushouts $V' \amalg_U X \to V \amalg_U X$ is a positive level equivalence by a levelwise application of the left properness of the contravariant model structure. Hence $V \amalg_U X$ is a homotopy pushout.

By [11, Theorem 4.3.9(iii)], it is sufficient to verify symmetric h–monoidality on the generating (acyclic) cofibrations. For this we let

$$v_i = F^{\mathcal{I}}_{\boldsymbol{k_i}}(\partial\Delta^{m_i} \to \Delta^{m_i}) \quad \text{for } 1 \le i \le e \tag{3-8}$$

be a family of generating cofibrations of $\mathcal{M}$. (We drop the augmentation to N from the notation.) Let $(n_i)_{1\le i\le e}$ be a family of natural numbers. Then the iterated pushout product map

$$v = v_1^{\square n_1} \square \cdots \square v_e^{\square n_e} \tag{3-9}$$

is a $\Sigma_{(n_i)} = \Sigma_{n_1} \times \cdots \times \Sigma_{n_e}$–equivariant map. For every $\Sigma_{(n_i)}$–object Y in $\mathcal{M}$, there is an isomorphism

$$Y \boxtimes v \cong (Y \boxtimes F^{\mathcal{I}}_{\boldsymbol{k}}(*)) \times \iota,$$

where $\boldsymbol{k} = \boldsymbol{k_1}^{\sqcup n_1} \sqcup \cdots \sqcup \boldsymbol{k_e}^{\sqcup n_e}$ and

$$\iota = (\partial\Delta^{m_1} \to \Delta^{m_1})^{\square n_1} \square \cdots \square (\partial\Delta^{m_e} \to \Delta^{m_e})^{\square n_e}$$

is the iterated pushout product map in spaces. Hence $Y \square v$ is a levelwise cofibration of simplicial sets, and so is its quotient by the $\Sigma_{(n_i)}$–action. This verifies the cofibration part of the symmetric h–monoidality.

Next let $(v_i\colon V_i \to W_i)_{1\leq i\leq e}$ be a family of generating acyclic cofibrations for $\mathcal{M}$. We may assume that the V_i and W_i are positive cofibrant since $\mathrm{sSet}^{\mathcal{I}}/N$ and hence $\mathcal{M}$ is tractable. Let $v\colon V \to W$ be defined as in (3-9) and let Y be a $\Sigma_{(n_i)}$–object in $\mathcal{M}$. For the acyclic cofibration part of the symmetric h–monoidality, we have to show that $(Y \boxtimes v)_{\Sigma_{(n_i)}}$ is a weak equivalence in $\mathcal{M}$. Let $f\colon X \to Y$ be a cofibrant replacement in $\mathcal{M}$ and consider the diagram

$$\begin{array}{ccccc} X \boxtimes V & \xrightarrow{f\boxtimes V} & Y \boxtimes V & \underset{\sim}{\xleftarrow{p_V}} & (Y \boxtimes V)^{\mathrm{cof}} \\ {\scriptstyle X\boxtimes v}\downarrow & & \downarrow{\scriptstyle Y\boxtimes v} & & \downarrow{\scriptstyle g} \\ X \boxtimes W & \xrightarrow{f\boxtimes W} & Y \boxtimes W & \underset{\sim}{\xleftarrow{p_W}} & (Y \boxtimes W)^{\mathrm{cof}} \end{array}$$

where g is a replacement of $Y \boxtimes v$ by a map of cofibrant objects in the projective model structure on $\mathcal{M}^{\Sigma_{(n_i)}}$. The map $X \boxtimes v$ is a weak equivalence in $\mathcal{M}$ by the pushout product axiom in $\mathcal{M}$, and the maps $f \boxtimes V$ and $f \boxtimes W$ are positive $\mathcal{I}$–equivalences by Proposition 3.18. Hence $Y \boxtimes v$ and g are weak equivalences in $\mathcal{M}$. To see that $Y \boxtimes v$ becomes a weak equivalence after taking $\Sigma_{(n_i)}$–orbits, we first note that g induces a weak equivalence of $\Sigma_{(n_i)}$–orbits because it is a map of cofibrant objects. Hence it is sufficient to show that p_V and p_W induce a weak equivalence of $\Sigma_{(n_i)}$–orbits. Since these are actually positive contravariant level equivalences, it is sufficient to show that the $\Sigma_{(n_i)}$–action on $Y \boxtimes W$ is free in positive levels. The group Σ_{n_i}–acts freely on $W_i^{\boxtimes n_i}(\boldsymbol{m})$ because W_i is positive cofibrant [7, Lemma 2.9]. The fact that there is a morphism of $\Sigma_{(n_i)}$–spaces

$$(Y \boxtimes W)(\boldsymbol{m}) \to W(\boldsymbol{m}) \to (W_1^{\boxtimes n_1} \boxtimes \cdots \boxtimes W_e^{\boxtimes n_e})(\boldsymbol{m}) \to W_1^{\boxtimes n_1}(\boldsymbol{m}) \times \cdots \times W_e^{\boxtimes n_e}(\boldsymbol{m})$$

thus implies that $\Sigma_{(n_i)}$ act freely on $Y \boxtimes W(\boldsymbol{m})$. This completes the acyclic cofibration part of the symmetric h–monoidality.

For symmetric flatness, it is by [11, Theorem 4.3.9(ii)] sufficient to show that for a weak equivalence $y\colon Y \to Z$ in the projective model structure on $\mathcal{M}^{\Sigma_{(n_i)}}$ and for v as in (3-8) and (3-9), the map $(y \square v)_{\Sigma_{(n_i)}}$ is a weak equivalence in $\mathcal{M}$. Here $y \square v$ is the pushout product map in the square

$$\begin{array}{ccc} Y \boxtimes V & \xrightarrow{y\boxtimes V} & Z \boxtimes V \\ {\scriptstyle Y\boxtimes v}\downarrow & & \downarrow{\scriptstyle Z\boxtimes v} \\ Y \boxtimes W & \xrightarrow{y\boxtimes W} & Z \boxtimes W \end{array}$$

Replacing y by a weak equivalence of cofibrant objects in $\mathcal{M}^{\Sigma_{(n_i)}}$ and using Proposition 3.18 and the pushout product axiom in $\mathcal{M}$ shows that the vertical maps are weak equivalences in $\mathcal{M}$. Since $X \boxtimes v$ is a levelwise cofibration by [12, Proposition 7.1(vi)], it is an h–cofibration by the argument at the beginning of the proof. Hence $y \square v$ is a weak equivalence in $\mathcal{M}$ by two out of three. Arguing as in the previous step of the proof, the fact that $\Sigma_{(n_i)}$ acts freely on the positive levels of $Y \boxtimes W$ implies that $(y \square v)_{\Sigma_{(n_i)}}$ is a weak equivalence in $\mathcal{M}$. □

Remark 3.21 The argument given in the previous proof actually shows the stronger statement that the two assertions in the proposition hold for colored operads and for operads internal to $\mathcal{C}$.

4 E_∞ objects and symmetric monoidal ∞–categories

The goal of this section is to prove Proposition 2.4 and its functorial refinement Proposition 4.3.

The ∞–category $\mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty$ of small symmetric monoidal ∞–categories is equivalent to the ∞–category $\mathrm{CAlg}(\mathrm{Cat}_\infty)$ of commutative algebra objects in ∞–categories [9, Remark 2.4.2.6]. Now let $\mathcal{E}$ be an E_∞–operad in $\mathrm{sSet}_{\mathrm{Joyal}}$ in the above sense (for example, the Barratt–Eccles operad). We will use the following result about the rectification of commutative algebras in the ∞–categorical sense to operad algebras in the model category.

Proposition 4.1 *There is an equivalence of ∞–categories*

$$(\mathrm{sSet}_{\mathrm{Joyal}}[\mathcal{E}])_\infty \simeq \mathrm{CAlg}(\mathrm{Cat}_\infty) \tag{4-1}$$

relating the ∞–category associated with the model category of $\mathcal{E}$–algebras in $\mathrm{sSet}_{\mathrm{Joyal}}$ and $\mathrm{CAlg}(\mathrm{Cat}_\infty)$. For an object M in $\mathrm{sSet}_{\mathrm{Joyal}}[\mathcal{E}]$, the ∞–category represented by M is naturally equivalent to the underlying ∞–category of the associated commutative algebra in Cat_∞.

Proof This is essentially a consequence of [10, Theorem 7.10] (which is in turn based on [9, Theorem 4.5.3.7]). However, [10, Theorem 7.10] is not directly applicable since it is formulated in terms of simplicial model categories and simplicial operads, while $\mathcal{E}$ is an operad in $\mathrm{sSet}_{\mathrm{Joyal}}$. As explained in [10, Remark 7.12], this context requires a different argument for identifying the free $\mathcal{E}$–algebra $\mathcal{E}(X)$ on a cofibrant object X with its derived counterpart in $\mathrm{CAlg}(\mathrm{Cat}_\infty)$. To circumvent this problem, we note that under the chain of Quillen equivalences in Theorem 3.6, $\mathcal{E}(X)$ corresponds to the free commutative algebra on a positive cofibrant replacement of $\mathrm{const}_{\mathcal{I}}(X)$. Using [7, Lemma 2.9] in place of [9, Lemma 4.5.4.11(3)], the claim about $\mathcal{E}(X)$ follows as in part (e) of the proof of [9, Theorem 4.5.3.7]. □

We are now ready to give the proof of the key proposition from Section 2:

Proof of Proposition 2.4 Using the above discussion, we choose an $\mathcal{E}$–algebra M in sSet representing the given small symmetric monoidal ∞–category $\mathcal{D}$ and consider the model category $\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}}$ arising from Theorem 3.15. By Proposition 3.10 and Corollary 3.19, this is a simplicial, combinatorial, tractable and left proper symmetric monoidal model category. Let $\mathcal{C} = (\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}})_\infty$ be the presentably symmetric monoidal ∞–category associated with $\mathrm{sSet}^{\mathcal{I}}/M^{\mathrm{rig}}$. We will show that $\mathcal{C}$ and $\mathcal{P}(\mathcal{D})$ are equivalent as symmetric monoidal ∞–categories.

It is immediate from Theorem 3.15 that after forgetting the monoidal structure, $\mathcal{C}$ is equivalent to the underlying ∞–category of the contravariant model structure on sSet/M. The underlying ∞–category of sSet/M is equivalent to the ∞–category $\mathcal{P}(\mathcal{D})$ by means of the ∞–categorical Grothendieck construction [8, Theorem 2.2.1.2] and the fact that the underlying ∞–category of M is equivalent to the underlying ∞–category of $\mathcal{D}$. Note that all the involved equivalences, ie the equivalences coming from Theorem 3.15 as well as the Grothendieck construction, are pseudonatural in M, that is, natural in a 2–categorical sense. Thus, invoking [5, Appendix A], we conclude that the induced equivalence of ∞–categories

$$\Phi\colon \mathcal{P}(\mathcal{D}) \to \mathcal{C}$$

is natural in $\mathcal{D}$ in the ∞–categorical sense. Note however that this equivalence does not necessarily need to respect the symmetric monoidal structures.

We need to show that Φ is compatible with the symmetric monoidal structures on $\mathcal{P}(\mathcal{D})$ and $\mathcal{C}$. By the universal property of the Day convolution symmetric monoidal structure on $\mathcal{D}$ reviewed in Section 2, it suffices to equip the functor

$$\Psi = \Phi \circ j\colon \mathcal{D} \to \mathcal{C}$$

given by composition with the Yoneda embedding $j\colon \mathcal{D} \to \mathcal{P}(\mathcal{D})$ with a symmetric monoidal structure. The functor Ψ is also natural in $\mathcal{D}$ in the ∞–categorical sense. We denote the essential image of Ψ by $\Psi(\mathcal{D})$. By construction $\Psi(\mathcal{D})$ is a full subcategory of $\mathcal{C}$. It follows from Lemma 3.16 that $\Psi(\mathcal{D})$ is closed under tensor products in $\mathcal{C}$. Thus it inherits a symmetric monoidal structure from $\mathcal{C}$ such that the inclusion functor $\Psi(\mathcal{D}) \to \mathcal{C}$ is a symmetric monoidal functor.

To complete the proof, it is sufficient to show that the corestriction $\mathcal{D} \to \Psi(\mathcal{D})$ of Ψ is a symmetric monoidal functor. For this we use the equivalence (4-1) and the functoriality of the involved constructions to view the construction $\mathcal{D} \mapsto \Psi(\mathcal{D})$ as a functor

$$G\colon \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty \to \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty$$

This functor G comes with a natural equivalence $UG \simeq U$ given by Ψ, where $U\colon \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty \to \mathcal{C}\mathrm{at}_\infty$ is the canonical forgetful functor. The next lemma implies that G is canonically equivalent to the identity functor on $\mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty$ and that the equivalence refines Ψ. We conclude that for each $\mathcal{D}$, the functor Ψ refines to an equivalence $\mathcal{D} \simeq \Psi(\mathcal{D})$ of symmetric monoidal ∞–categories. □

Lemma 4.2 *Let $G\colon \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty \to \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty$ be a functor together with an equivalence $UG \simeq U$. Then the equivalence admits a canonical refinement to an equivalence $G \simeq \mathrm{id}$.*

Proof We first observe that G preserves limits and filtered colimits, since these are generated by the functor U. Together with the fact that $\mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty$ is presentable and the adjoint functor theorem, this shows that G is right adjoint. Denote the left adjoint of G by F. The equivalence $UG \simeq U$ implies that the diagram

$$\begin{array}{ccc} & \mathcal{C}\mathrm{at}_\infty & \\ {}^{\mathrm{Fr}}\swarrow & & \searrow^{\mathrm{Fr}} \\ \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty & \xrightarrow{\;F\;} & \mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty \end{array}$$

commutes, where Fr is the free symmetric monoidal category functor. Now we use that the functor Fr exhibits $\mathcal{S}\mathrm{ym}\mathcal{M}\mathrm{on}\mathcal{C}\mathrm{at}_\infty$ as the free presentable, preadditive category on Cat_∞ [4, Theorem 4.6]. Since F is left adjoint this implies that it has to be canonically equivalent to the identity. Thus also the right adjoint G is canonically equivalent to the identity. □

The proof of Proposition 2.4 in fact provides the following stronger statement:

Proposition 4.3 *For every symmetric monoidal functor $f\colon \mathcal{D} \to \mathcal{D}'$ between small ∞–categories there exists a symmetric monoidal, left Quillen functor between model categories $F\colon \mathcal{M} \to \mathcal{M}'$ such that $f_!\colon \mathcal{P}(\mathcal{D}) \to \mathcal{P}(\mathcal{D}')$ is symmetric monoidally equivalent to the underlying functor of F.* □

Proof We use Proposition 4.1 to represent f by a map of $\mathcal{E}$–algebras. Then we get the induced functor between model categories and our proof of Proposition 2.4 shows that this models the ∞–functor $f_!$. □

References

[1] **C Barwick**, *On left and right model categories and left and right Bousfield localizations*, Homology Homotopy Appl. 12 (2010) 245–320 MR

[2] **D Dugger**, *Combinatorial model categories have presentations*, Adv. Math. 164 (2001) 177–201 MR

[3] **D Dugger**, *Replacing model categories with simplicial ones*, Trans. Amer. Math. Soc. 353 (2001) 5003–5027 MR

[4] **D Gepner**, **M Groth**, **T Nikolaus**, *Universality of multiplicative infinite loop space machines*, Algebr. Geom. Topol. 15 (2015) 3107–3153 MR

[5] **R Haugseng**, **D Gepner**, **T Nikolaus**, *Lax colimits and free fibrations in ∞–categories*, Doc. Math. 22 (2017) 1225–1266

[6] **A Joyal**, *The theory of quasi-categories and its applications*, lecture notes, CRM (2008) Available at `http://tinyurl.com/JoyalQuasiCat`

[7] **D Kodjabachev**, **S Sagave**, *Strictly commutative models for E_∞ quasi-categories*, Homology Homotopy Appl. 17 (2015) 121–128 MR

[8] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR

[9] **J Lurie**, *Higher algebra*, book in progress, Harvard (2012) Available at `http://math.harvard.edu/~lurie`

[10] **D Pavlov**, **J Scholbach**, *Admissibility and rectification of colored symmetric operads*, preprint (2015) arXiv

[11] **D Pavlov**, **J Scholbach**, *Homotopy theory of symmetric powers*, preprint (2015) arXiv

[12] **S Sagave**, **C Schlichtkrull**, *Diagram spaces and symmetric spectra*, Adv. Math. 231 (2012) 2116–2193 MR

[13] **S Schwede**, **B E Shipley**, *Algebras and modules in monoidal model categories*, Proc. London Math. Soc. 80 (2000) 491–511 MR

Max Planck Institut für Mathematik
Bonn, Germany

IMAPP, Radboud University Nijmegen
Nijmegen, The Netherlands

`thoni@mpim-bonn.mpg.de`, `s.sagave@math.ru.nl`

`http://people.mpim-bonn.mpg.de/thoni/`, `http://www.math.ru.nl/~sagave/`

Received: 31 January 2017

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in AGT is available on the AGT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 17 Issue 5 (pages 2565–3212) 2017

1472-2747(2017)17:5;1-R