Algebraic & Geometric Topology

Volume 23 (2023)

Issue 5 (pages 1935–2414)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

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

The subscription price for 2023 is US $650/year for the electronic version, and $940/year (+$70, 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) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, 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 **23**:5 (2023) 1935–1957
DOI: 10.2140/agt.2023.23.1935
Published: 25 July 2023

Splitting Madsen–Tillmann spectra II: The Steinberg idempotents and Whitehead conjecture

TAKUJI KASHIWABARA
HADI ZARE

We show that there is a splitting of the spectrum $\Sigma^{-n} D(n)$ off the Madsen–Tillmann spectrum $\mathrm{MTO}(n) = \mathrm{BO}(n)^{-\gamma_n}$ compatible with the classic splitting of $M(n)$ off $\mathrm{BO}(n)_+$, localized at the prime $p = 2$. For $n = 2$, together with our previous splitting result on Madsen–Tillmann spectra, this shows that $\mathrm{MTO}(2)$ is homotopy equivalent to $\mathrm{BSO}(3)_+ \vee \Sigma^{-2} D(2)$. We also discuss its implication for characteristic classes.

55P42, 55P47, 55R40, 57R20; 55R35, 55S12, 55S15, 57N70

Dedicated to the memory of Stephen A Mitchell

1 Introduction

The Madsen–Tillmann spectrum $\mathrm{MTO}(n)$ is defined to be the Thom spectrum of the virtual bundle $-\gamma_n$ over $\mathrm{BO}(n)$, where γ_n is the universal n–plane bundle; see Galatius, Tillmann, Madsen and Weiss [4] — see also Galatius and Randal-Williams [3, Section 1.1.2] for the general construction of Madsen–Tillman spectra. It is known that these spectra filter the spectrum MO; i.e. there is a sequence

$$S^0 = \mathrm{MTO}(0) \to \Sigma\mathrm{MTO}(1) \to \cdots \to \Sigma^{n-1}\mathrm{MTO}(n-1) \xrightarrow{\iota_n} \Sigma^n\mathrm{MTO}(n) \to \cdots, \tag{1}$$

where ι_n is induced by the inclusion $O(n-1) \subset O(n)$, with the property that

$$\operatorname{hocolim} \Sigma^n \mathrm{MTO}(n) \cong \mathrm{MO}$$

[4, remark after (3.4)].[1] Furthermore, the cofiber of the successive stages is homotopy equivalent to $\mathrm{BO}(n)_+$; i.e. we have a cofibration sequence

$$(2) \qquad \cdots \to \Sigma^{-1}\mathrm{MTO}(n-1) \to \mathrm{MTO}(n) \xrightarrow{\omega_{O(n)}} \mathrm{BO}(n)_+ \xrightarrow{\tau} \mathrm{MTO}(n-1) \to \cdots$$

[4, (3.3)], where $\omega_{O(n)}$ is the map induced by the "embedding" of $-\gamma_n$ into the 0–dimensional trivial bundle, X_+ is the union of X with a disjoint base point, τ is the Becker–Schultz–Mann–Miller–Miller transfer [1, Section 2; 10, 3.7] — see also [6, Section 2.3] — associated to the inclusion $O(n-1) \subset O(n)$. In other words, the spectrum MO can be built up from pieces $\mathrm{BO}(n)_+$.

We have shown in our previous work that localized away from 2, $\mathrm{MTO}(2n) \simeq \mathrm{BO}(2n)_+$ and $\mathrm{MTO}(2n+1) \simeq *$ for all $n \geq 0$ [6, Theorem 1.1.B], reducing essentially the study of $\mathrm{MTO}(n)$'s to 2–local problems. Thus we will work at the prime $p = 2$. So throughout the paper homology and cohomology are taken with $\mathbb{Z}/2$ coefficients unless otherwise stated. We work most of the time in the 2–local stable homotopy category whose objects are 2–local spectra and morphisms are homotopy classes of maps of spectra; consequently, by commutative we mean homotopy commutative. We identify a spectrum with its 2–localization. We note that when both sides of a morphism in this category are of finite type then inducing an isomorphism in $\mathbb{Z}/2$–cohomology implies an isomorphism of 2–local spectra; we shall use this reasoning freely throughout the paper. We identify a (pointed) space X with its suspension spectrum $\Sigma^\infty X$ unless otherwise stated. In the literature, sometimes a space X is identified with $\Sigma^\infty X_+$, which explains notational discrepancies the reader may find between the current paper and results we quote. We use the same letter to denote a map $f\colon X \to Y$ and its suspensions $f\colon \Sigma^k X \to \Sigma^k Y$ with $k \in \mathbb{Z}$. For a spectrum E, we shall write $\Omega^\infty E = \operatorname{colim} \Omega^i E_i$ for the infinite loop space associated to E and $\Omega^\infty_0 E$ denotes it base point component corresponding to $0 \in \pi_0 E$, noting that if E is 0–connected then $\Omega^\infty E = \Omega^\infty_0 E$. For a pointed space X the standard notations $QX = \Omega^\infty(\Sigma^\infty X)$ and $Q_0 X = \Omega^\infty_0(\Sigma^\infty X)$ will be used.

[1] In [4] this is said to be the colimit. The use of the word colimit can be justified by the fact that this is actually the colimit on the level of underlying point set at each n if one considers spectrum X as a collection of spaces X_n and structure maps $\Sigma X_n \to X_{n+1}$. However, this is clearly not the colimit in the category of spectra, so we avoid the use of this term.

At the prime 2, Randal-Williams computed $H_*(\Omega_0^\infty \mathrm{MTO}(i))$ for $i = 1$ and 2 [16, Theorems A and B]. Combining the two theorems, we get an exact sequence of Hopf algebras

$$H_*(Q_0\mathrm{BO}(2)_+) \to H_*(Q_0\mathrm{BO}(1)_+) \to H_*(Q_0\mathrm{BO}(0)_+) \to \mathbb{Z}/2, \tag{3}$$

where the (Hopf) kernel of the first two maps are isomorphic to $H_*(\Omega_0^\infty \mathrm{MTO}(i))$ for $i = 2$ and 1, respectively. Thus a natural question to ask was whether this exact sequence could be extended further to the left with $H_*(\Omega_0^\infty \mathrm{MTO}(i))$ isomorphic to the kernel of each stage. We showed that this was impossible in [6, Proposition 1.11]. So a new question to ask, then, is to what extent we can generalize [16, Theorems A and B]. This question leads to a search for another sequence of spectra with the beginning as in (1). It turns out that there indeed is such a sequence, well known to stable homotopy theorists. For a *space* X, denote by $\mathrm{Sp}^k(X)$ the k^{th} symmetric product of X, that is the quotient of X^k by the obvious action of the symmetric group Σ_k. It is easy to show that this induces a functor in the stable category which we still denote by Sp^k. Define the spectrum $D(n)$ as the cofiber of the diagonal map $\mathrm{Sp}^{2^{n-1}} S^0 \to \mathrm{Sp}^{2^n} S^0$; see Mitchell and Priddy [14, Section 4.2]. We have $D(0) = S^0$ and $D(1) \cong \Sigma\mathrm{MTO}(1)$ [14, Proposition 4.4]. Furthermore, Mitchell and Priddy defined a map $\iota_n\colon D(n-1) \to D(n)$ [14, Proposition 4.3]; thus we get a sequence

$$S^0 = D(0) \to D(1) \to \cdots \to D(n-1) \xrightarrow{\iota_n} D(n) \to \cdots. \tag{4}$$

Taking the cohomology, this sequence realizes the length filtration of the Steenrod algebra $\mathcal{A}$ [14, Proposition 4.3]. That is, we have isomorphisms

$$H^*(D(n)) \cong \mathcal{A}/G_n, \quad \text{where } G_n \text{ is the span of } \mathrm{Sq}^I,\ I \text{ is admissible and } l(I) > n. \tag{5}$$

We note that G_n happens to be a left $\mathcal{A}$–ideal, so that this isomorphism is as $\mathcal{A}$ modules. It happens that this cohomological property characterizes the sequence of spectra $D(n)$ [5, Corollary 1.4.1]. Of course, as an immediate consequence of (5), we see that $\operatorname{hocolim} D(n) \cong H\mathbb{Z}/2$.

On the other hand, the spectrum $\mathrm{BO}(1)_+^{\times n}$ admits a natural (left) $\mathrm{Gl}_n(\mathbb{Z}/2)$–action. Thus the Steinberg idempotent $e_n \in \mathbb{Z}/2[\mathrm{Gl}_n(\mathbb{Z}/2)]$ [14, Definition 2.2] and its conjugate e_n' [14, the sentence above Proposition 2.6] give rise to a splitting of $\mathrm{BO}(1)_+^{\times n}$ and we have $M(n) \simeq e_n\mathrm{BO}(1)_+^{\times n} \simeq e_n'\mathrm{BO}(1)_+^{\times n}$ [14, Theorem 5.1]. Moreover, through the Becker–Gottlieb transfer map, this splitting gives rise to a splitting of $M(n)$ off $\mathrm{BO}(n)_+$. We will review this splitting in more details in Section 2. The spectra $M(n)$'s and $D(n)$'s are related by the cofibration sequences [14]

$$\cdots \to \Sigma^{n-1} M(n) \to D(n-1) \to D(n) \to \Sigma^n M(n) \to \cdots. \tag{6}$$

Thus one can say that MO can be constructed with $\mathrm{BO}(n)_+$'s as building blocks, whereas $H\mathbb{Z}/2$ can be constructed with $M(n)$'s as building blocks. Furthermore, $H\mathbb{Z}/2$ and $M(n)$'s split off MO and $\mathrm{BO}(n)_+$'s, respectively. It is then natural to ask whether one can split intermediate stages as well. The purpose of this paper is to answer affirmatively to this question, and discuss some consequences, including an answer to the question on generalization of the exact sequence (3). We have the following, the main results of this paper.

Theorem 1.1 *For each n, the spectrum $D(n)$ splits off $\Sigma^n \mathrm{MTO}(n)$.*

An immediate consequence of Theorem 1.1 is the following.

Corollary 1.2 *$H_*(\Omega^\infty \Sigma^{-n} D(n))$ splits off $H_*(\Omega^\infty \mathrm{MTO}(n))$ as a Hopf algebra.*

Thus the "correct way to extend" the exact sequence (3) is just the following standard fact.

Proposition 1.3 (Kuhn and Priddy [8]) *The sequence of Hopf algebras*

$$\cdots \to H_*(\Omega_0^\infty M(n)) \to H_*(\Omega_0^\infty M(n-1)) \to \cdots$$
$$\cdots \to H_*(\Omega_0^\infty M(2)) \to H_*(Q_0 B\mathbb{Z}/2_+) \to H_*(Q_0 S^0) \to \mathbb{Z}/2$$

is exact. Furthermore, the image of $H_*(\Omega_0^\infty M(n)) \to H_*(\Omega_0^\infty M(n-1))$ is isomorphic to $H_*(\Omega_0^\infty \Sigma^{-n+1} D(n-1))$.

As $D(0) \cong S^0$, $\Sigma^{-1} D(1) \cong \mathrm{MTO}(1)$, and $M(1) \cong \mathrm{BO}(1)_+$, combined with the $n = 2$ case of Theorem 1.1, we recover Theorems A and B of [16]. Of course, the cohomology being dual of homology, the exact sequences above give some information on certain characteristic classes. More precisely, recall from [6; 16] (with correction from Randal-Williams, via personal communication):

Definition 1.4 A universally defined characteristic class in $H^*(\Omega^\infty \mathrm{MTO}(n))$ is an element in the subalgebra generated by the image of

$$H^*(\mathrm{BO}(n)) \xrightarrow{\sigma^{\infty *}} H^*(Q\mathrm{BO}(n)_+) \xrightarrow{(\Omega^\infty \omega_{O(n)})^*} H^*(\Omega^\infty \mathrm{MTO}(n)).$$

We denote by $\mu_{i_1,\dots,i_n} = (\Omega^\infty \omega_{O(n)})^*(\sigma^{\infty *}(\sigma_1^{i_1},\dots,\sigma_n^{i_n}))$, where

$$H^*(\mathrm{BO}(n)) \cong \mathbb{Z}/2[\sigma_1,\dots,\sigma_n]$$

and $\sigma^{\infty *}$ denotes the cohomology suspension.

We note that in the definitions in [16] and [6], only basepoint components of the infinite loop spaces was considered. However, this has the effect of missing out nontrivial 0–dimensional classes as also confirmed by Randal-Williams (personal communication). Therefore, we have removed the restriction to the basepoint component in our definition. We note that [6, Theorem 1.9] remains valid as is stated.[2]

In [6], we used the summand $\mathrm{BSO}(2n+1)_+$ that split off $\mathrm{MTO}(2n)$ to show that some of these classes remain algebraically independent. Here we use the splitting of $D(n)$ off $\mathrm{MTO}(n)$ to show that there are "linear" relations corresponding to elements of $H^*(M(n))$, and that in the case of dimension 2, these relations together with the ones derived from the action of top Steenrod squares are the only relations. More precisely, we will show:

Theorem 1.5 (i) *In $H^*(\Omega^\infty \mathrm{MTO}(n))$, we have relations*

$$(\Omega^\infty \omega_{O(n)})^*(\sigma^{\infty *}(x)) = 0 \quad \text{for } x \in H^*(M(n)) \subset H^*(\mathrm{BO}(n)).$$

(ii) *For $n=2$, the only relations among the $\mu_{i,j}$ are those above, and $\mu_{2i,2j} = \mu_{i,j}^2$.*

(iii) *Again for $n=2$, the subalgebra of universally defined characteristic classes in $H^*(\Omega^\infty \mathrm{MTO}(2))$ is the polynomial algebra generated by $\nu_{i,j}$'s with i and j odd, where $\nu_{i,j}$ is defined in [6], tensored with the boolean algebra $\mathbb{Z}/2[\mu_{0,0}]/(\mu_{0,0}^2 - \mu_{0,0})$.*

We will give a more precise description of the inclusion $H^*(M(n)) \subset H^*(\mathrm{BO}(n))$ in Proposition 5.7.

The paper is organized as follows. In Section 2 we recall the splitting related to the Steinberg idempotents and construct a map from $D(n)$ to $\Sigma^n \mathrm{MTO}(n)$ for each n. In Section 3, we recall relevant results from [8] and construct a map going the other way around. In Section 4 we study the composition and show that we indeed have a splitting. In Section 5 we discuss the consequences in homology of infinite loop spaces.

[2] As a matter of fact, it was assumed implicitly that the sequence I was nonzero, due to the obvious relation $\mu_{0,0} = 0$ with the "old" definition. This relation holds no longer. One can easily adapt the proof of [6, Theorem 1.9] to the "new" definition.

Most of the current paper is independent of the results from the previous one, except for Theorem 1.5(ii), (iii) and the contents of Section 4.2. Thus, the current paper can be read separately from [6]. A word is due on the way some of proofs are written. In some places, the reader familiar with works we quote may find that our proofs are somewhat going backward. For example, we deduce Proposition 5.1 from Theorem 3.7, but as a matter of fact in [8, Section 5], a large part of the latter was proved as a main ingredient of the proof of the former. This is our deliberate choice; we preferred referring the readers to statements that are ready available to be quoted, rather than letting them look for details of proofs, or reproducing them ourselves.

Acknowledgements Kashiwabara thanks Andrew Baker, Masaki Kameko, Nick Kuhn, Bob Oliver, Stewart Priddy, Lionel Schwartz and Steve Wilson for helpful conversations. A special thanks is due to Oscar Randal-Williams for helpful discussions. Zare is grateful to Institut Fourier for its hospitality and support for a visit during October 2014. The authors thank Haynes Miller and Geoffrey Powell for helpful conversations. The authors also thank the referees for their constructive critics of earlier versions. Kashiwabara was supported in part by grant ANR-08–BLAN-0248 and ANR-16-CE40-0003 ChroK. Zare has been supported in part by IPM grant 93550117.

2 Some splitting derived from Steinberg idempotents

In this section we recall from [14] and [17] the splitting related to Steinberg idempotents.

Let X be a spectrum, $e \in [X, X]$ an idempotent, i.e. a map such that $e \circ e = e \in [X, X]$. Note that $[X, X]$ has a natural ring structure where the multiplication is given by the composition, and e is an idempotent in terms of ring theory. Denote by eX the homotopy colimit $X \xrightarrow{e} X \xrightarrow{e} \cdots$. Then we have a splitting

$$X \simeq eX \vee (1-e)X.$$

Furthermore, if we still denote by e the induced map in (co)homology, we get

$$H_*(eX) \cong eH_*(X), \quad H^*(eX) \cong H^*(X)e.$$

We are particularly interested in the case of idempotents arising from a group action on spectra. That is, let G be a group acting on the spectrum X from the left. There are several different notions of group action on spectra, here we can take any of them: all we need is a group homomorphism $G \to \mathrm{Aut}(X)$ where $\mathrm{Aut}(X)$ is the group consisting of invertible elements in $[X, X]$. This group homomorphism extends to a ring homomorphism $\mathbb{Z}_{(2)}[G] \to [X, X]$, thus sending an idempotent to an idempotent. We see that an idempotent in the group ring $\mathbb{Z}_{(2)}[G]$ gives rise to a splitting of spectra on which G acts. Actually the theory of lifting idempotents allows us to settle for something less, which is one of the reasons why completion is crucial in the theory of splitting, but we will not need this for our purpose.

Now, let $G = \mathrm{Gl}_n(\mathbb{Z}/2)$. Its group-ring $\mathbb{Z}_{(2)}[\mathrm{Gl}_n(\mathbb{Z}/2)]$ contains well-known Steinberg idempotents e_n and e'_n defined by

$$(7)\qquad e_n = \frac{1}{q_n} \sum_{g \in B_n} g \sum_{\sigma \in \Sigma_n} (-1)^{\mathrm{sgn}(\sigma)} \sigma, \quad e'_n = \frac{1}{q_n} \sum_{\sigma \in \Sigma_n} (-1)^{\mathrm{sgn}(\sigma)} \sigma \sum_{g \in B_n} g,$$

where B_n denotes the subgroup consisting of upper triangular matrices, Σ_n denotes the subgroup of permutation matrices, and q_n is the index of B_n in G.

Remark 2.1 (i) Traditionally we consider the above elements as idempotents modulo 2, and use the lifting theory. However, as was noticed in [14, proof of Proposition 2.6] (see also [8, page 462]), e_n and e'_n actually are conjugate idempotents, and they can even be defined in $\mathbb{Z}_{(2)}[\mathrm{Gl}_n(\mathbb{Z}/2)]$. Let's note that working with spectra completed at 2 has some advantages, e.g. we get a better control over maps among spectra [13, Corollary 1.4(b)]. However, as far as our current work is concerned, localization is sufficient.

(ii) We use the additive structure in $[X, X]$ to extend the G–action on X. Thus even in the case when G acts on the space X (via maps of spaces, not just maps of spectra), the idempotents are not necessarily maps of spaces. However, in this case they can be realized as self maps of the space ΣX. In other words, the spectrum ΣeX is a suspension spectrum.

Write Δ_n for $O(1)^n$. The identification of $O(1)$ with $\mathbb{Z}/2$ gives a natural action of $\mathrm{Gl}_n(\mathbb{Z}/2)$ on $B\Delta_n$, thus on $B\Delta_{n+}$, and we have:

Definition 2.2 We define the spectra $M(n)$ by

$$M(n) \cong e_n B\Delta_{n+}.$$

Remark 2.3 Originally $M(n)$ was defined as $\Sigma^{-n} D(n)/D(n-1)$, but in terms of [14, Theorem A] this is equivalent, and in recent literature we encounter this definition more often.

Now, results in representation theory imply that for any $\mathbb{Z}/2[\mathrm{Gl}_n(\mathbb{Z}/2)]$–module W, we have an isomorphism $We_n' \cong We_n$ induced by $\sum_{\sigma\in\Sigma_n}\sigma$ [14, Proposition 2.6(b)]. On the other hand, the composition

$$B\Delta_{n+} \xrightarrow{Bi} \mathrm{BO}(n)_+ \xrightarrow{\mathrm{Tr}_{Bi}} B\Delta_{n+}$$

induces $\sum_{\sigma\in\Sigma_n}\sigma$ in $H^*(B\Delta_n)$; that is, the composition

$$B\Delta_{n+} \xrightarrow{e_n} B\Delta_{n+} \to \mathrm{BO}(n)_+ \to B\Delta_{n+} \tag{8}$$

induces in the cohomology e_n'. Therefore

$$e_n B\Delta_{n+} \to B\Delta_{n+} \to \mathrm{BO}(n)_+ \to B\Delta_{n+} \to e_n' B\Delta_{n+}$$

induces an isomorphism in mod 2 cohomology. In other words:

Theorem 2.4 [14, Theorem C] *$M(n)$ splits off* $\mathrm{BO}(n)_+$.

Of course, cohomology of a space is related to that of Thom spectra of bundles over it via Thom isomorphisms, so we can "Thomify" all of the above. More precisely, let ρ_n be the reduced regular representation of Δ_n and $\gamma = \rho_1^n$ its canonical representation. The canonical representation is the direct sum of n distinct projections, while the regular representation is the direct sum of all possible 1–forms. As these 1–forms are tensor products of projections, we get an isomorphism of representations

$$\bigoplus_{i>0} \Lambda^i(\gamma) \cong \rho_n,$$

where $\Lambda^i(-)$ is the i^{th} exterior power functor. Therefore, if we define a representation $\bar\rho_n$ of $O(n)$ by

$$\bar\rho_n = \bigoplus_{i>0} \Lambda^i(\gamma_n),$$

it restricts to ρ_n over $\Delta_n \subset O(n)$. Now, if k denotes an integer, $k\rho_n$ is invariant under the action of $\mathrm{Gl}_n(\mathbb{Z}/2)$; thus if $g \in \mathrm{Gl}_n(\mathbb{Z}/2)$, we have $g^*(k\rho_n) = k\rho_n$, giving rise to a Thomified map $B\Delta_n^{k\rho_n} = B\Delta_n^{g^*(k\rho_n)} \xrightarrow{\mathrm{Th}(g)} B\Delta_n^{k\rho_n}$. Here, and throughout the paper, given a (virtual) vector bundle $\xi \to X$, we shall write X^ξ for its Thom spectrum. This furnishes the Thom spectrum $B\Delta_n^{k\rho_n}$ with a $\mathrm{Gl}_n(\mathbb{Z}/2)$–action. When k is negative, slightly more careful arguments are needed, but this is taken care of by [17]. Thus we can split it using the Steinberg idempotents e_n and e_n'. Then we get a sequence of maps

$$B\Delta_n{}^{k\rho_n} \xrightarrow{e_n} B\Delta_n^{k\rho_n} \to \mathrm{BO}(n)^{k\bar\rho_n} \to B\Delta_n^{k\rho_n},$$

where the last map is the twisted Becker–Gottlieb transfer [6, Theorem 1.1(1)]. As everything in sight is compatible with the Thom isomorphism, the effect of these maps in the cohomology can be deduced from those in the sequence (8). Noting that e_n' is also a sum of Thomified maps, we see that this composition induces e_n' in cohomology. Thus, as in Theorem 2.4:

Theorem 2.5 *$e_n B\Delta_n^{k\rho_n}$ splits off $\mathrm{BO}(n)^{k\bar{\rho}_n}$.*

The spectra $e_n B\Delta_n^{k\rho_n}$'s are studied notably in [17] where it is called $M(n)_k$; when $k=0$, we recover Theorem 2.4. The case $k=-1$ also interests us for the following result, which is implicit in [17]:

Theorem 2.6 $$e_n B\Delta_n^{-\rho_n} \cong \Sigma^{-n} D(n).$$

Proof This seems to be well known, but as we haven't found it spelled out in literature, for the sake of reference we give a proof here. It suffices to note that $\mathbf{R}(n)e_n$ in [17, Theorem 4.1.1(1)] is same as $\mathbf{M}(n)_{-1}$ in [17, Proposition 4.1.6], which is the cohomology of $M(n)_{-1}$ (cf. [17, page 386], whereas by Theorem 5.8 and Lemma 5.6 of [14] it is isomorphic to the cohomology of $\Sigma^{-n}D(n)$. □

Combining the theorems above shows that $\Sigma^{-n}D(n)$ splits off $\mathrm{BO}(n)^{-\bar{\rho}_n}$. As the inclusion of the representation $\gamma_n \subset \bar{\rho}_n$ induces a map of Thom spectra

$$\mathrm{BO}(n)^{-\bar{\rho}_n} \to \mathrm{BO}(n)^{-\gamma_n} = \mathrm{MTO}(n),$$

we get a map $\beta_n\colon \Sigma^{-n}D(n) \to \mathrm{MTO}(n)$. Or, equivalently, we can construct the map as the composition

$$\Sigma^{-n}D(n) \to B\Delta_n^{-\rho_n} \to B\Delta_n^{-\gamma} \to \mathrm{BO}(n)^{-\gamma_n} = \mathrm{MTO}(n).$$

We will denote the resulting map by β_n. Here, the map $B\Delta_n^{-\rho_n} \to B\Delta_n^{-\gamma}$ is induced by the inclusion of bundles $\gamma \subset \rho_n$ and the map $B\Delta_n^{-\gamma} \to \mathrm{BO}(n)^{-\gamma_n}$ is the twisted Becker–Gottlieb transfer [6, Theorem 1.1(1)], noting that $\gamma_n|_{\Delta_n} = \gamma$.

3 Maps from $\mathrm{MTO}(n)$ to $\Sigma^{-n}D(n)$

3.1 Exact sequences of spectra and the Whitehead conjecture

In this section we use results from [8] to construct maps from $\Sigma^n\mathrm{MTO}(n)$'s to $D(n)$'s. We start by fixing terminology.

Definition 3.1 (i) A filtered spectrum (X, F_*X, ι_*) is a sequence of spectra F_*X

$$F_0X \xrightarrow{\iota_0} F_1X \xrightarrow{\iota_1} \cdots \xrightarrow{\iota_{n-1}} F_nX \xrightarrow{\iota_n} F_{n+1}X \xrightarrow{\iota_{n+1}} \cdots \tag{9}$$

with a homotopy equivalence hocolim $F_nX \simeq X$. Usually ι_* is clear from the context, and X is determined by F_*X's, so we simply refer to it as F_*X. To distinguish with individual spectra, we also write $(F_*X, * \geq 0)$

(ii) A map of filtered spectra f_* from F_*X to F_*Y is a collection of maps $f_n \colon F_nX \to F_nY$ that makes the squares

$$\begin{tikzcd} F_nX \arrow[r] \arrow[d] & F_{n+1}X \arrow[d] \\ F_nY \arrow[r] & F_{n+1}Y \end{tikzcd}$$

commutative.

Note that we don't require any condition that would be a counterpart of the injectivity on ι_n's here.

Definition 3.2 (i) By a chain complex of spectra (C_n, d_n) we understand a sequence of spectra C_n with maps $d_{n-1} \colon C_n \to C_{n-1}$ such that the composition $C_{n+1} \to C_n \to C_{n-1}$ is null for all n. By a map f of chain complexes of spectra $(C_n, d_n^C) \to (C_n', d_n^{C'})$ we mean a collection of maps $f_n \colon C_n \to C'_n$ such that $f_n \circ d_n^C = d_n^{C'} \circ f_{n+1}$. Furthermore, if we have a map $\epsilon \colon C_0 \to E_{-1}$ with $\epsilon \circ d_0 = 0$, we say that the complex is augmented over E_{-1}.

(ii) Let F_*X be a filtered spectrum. Define its *associated graded* complex $\mathrm{Gr}_\bullet(F_*X)$ by $\mathrm{Gr}_0(F_*X) = F_0X$ and $\mathrm{Gr}_i(F_*X) = \Sigma^{-i}\,\mathrm{cofib}(F_{i-1}X \to F_iX)$. Then we can compose the canonical maps $\mathrm{Gr}_i(F_*(X)) \to \Sigma^{-i}F_{i-1}X \to \mathrm{Gr}_{i-1}(F_*(X))$ to define a map $d_{i-1}^{\mathrm{Gr}_\bullet(F_*X)}$. As a matter of the fact, the composition $d_{i-1}^{\mathrm{Gr}_\bullet(F_*X)} \circ d_i^{\mathrm{Gr}_\bullet(F_*X)}$ factors through the composition $F_iX \to \mathrm{Gr}_i F_*X \to \Sigma F_{i-1}X$, which is trivial.

Remark 3.3 Our notion of complex is more general than that in [7]. The complexes dealt with in [7] are the ones that arise as associated graded complexes of filtered spectra.

Example 3.4 (i) Let $F_nX = D(n)$. Then the associated graded complex $\mathrm{Gr}_\bullet(F_*X)$ is

$$\cdots \to M(n+1) \xrightarrow{\delta_n} M(n) \to \cdots \to M(0)$$

considered in [8, Corollary 1.2].

(ii) Let $F_nY = \Sigma^n\mathrm{MTO}(n)$. Then the associated graded complex $\mathrm{Gr}_\bullet(F_*X)$ is given by $(\mathrm{BO}(n)_+, \mathrm{tr})$, where tr is the Becker–Gottlieb transfer associated to the inclusion

$O(n-1) \subset O(n)$, as the Becker–Gottlieb transfer $\mathrm{BO}(n)_+ \to \mathrm{BO}(n-1)_+$ factors as $\mathrm{BO}(n)_+ \xrightarrow{\tau} \mathrm{MTO}(n-1) \xrightarrow{\omega_{O(n-1)}} \mathrm{BO}(n-1)_+$ [6, Proposition 2.3]. We can also see that $(\mathrm{BO}(n)_+, \mathrm{tr})$ is a complex directly as follows: $O(n) \subset O(n) \times O(2) \subset O(n+2)$; thus $O(2) \subset N_{O(n+2)}(O(n))$. So the composition of the transfer associated to $O(n) \subset O(n+1)$ and that associated to $O(n+1) \subset O(n+2)$, which is the transfer associated to $O(n) \subset O(n+2)$, is trivial by [9, Chapter 4, Lemma 2.12]. Moreover, the complex of free spectra $(\mathrm{BO}(n)_+, \mathrm{tr})$ is augmented over $H\mathbb{Z}/2$ since the composition

$$\mathrm{BO}(1)_+ \to \mathrm{BO}(0)_+ \to H\mathbb{Z}/2$$

is trivial. This is just another way of saying that the transfer in $\mathbb{Z}/2$–cohomology $H^*(\mathrm{BO}(0)_+; \mathbb{Z}/2) \to H^*(\mathrm{BO}(1)_+; \mathbb{Z}/2)$ is trivial.

Definition 3.5 [8] (i) A fibration sequence of spectra $F \to X \xrightarrow{f} Y$ is called exact if there exists a map of spaces (i.e. not a map between their suspension spectra) $g \colon \Omega^\infty Y \to \Omega^\infty X$ such that $\Omega^\infty f \circ g \simeq \mathrm{id}$.

(ii) A chain complex of spectra $\cdots \to X_n \to \cdots X_1 \to X_0 \to E_{-1}$ augmented over E_{-1} is called exact if for each $n \geq 0$, $E_n \to X_n \to E_{n-1}$ is exact, where E_n is inductively defined as the fiber of the map $X_n \to E_{n-1}$. Note that by the exactness of $E_{n-2} \to X_{n-2} \to E_{n-3}$, $[X_n, E_{n-2}]$ injects to $[X_n, X_{n-2}]$, so the triviality of the composition $X_n \to X_{n-1} \to X_{n-2}$ implies that of the composition $X_n \to X_{n-1} \to E_{n-2}$.

(iii) A spectrum is said to be projective if it is a summand of a suspension spectrum.

The category of spectra being a triangulated category instead of an abelian category, we have some complication here. The notion of exactness with three terms is more or less a counterpart of a split short exactness in abelian categories. The use of this seemingly too strong condition is motivated by the following fact. By definition, an exact sequence of spectra yields an exact sequence of abelian groups upon applying $[Y, -]$ for a suspension spectrum Y, or a spectrum which is a summand of a suspension spectrum. Thus one can regard suspension spectra as free objects, summands of suspension spectra as projective objects, and carry out homological algebra in the category of spectra. This idea was developed further in [7]. For example, we get the following:

Proposition 3.6 *Let $(P_\bullet, d_\bullet)$ be a chain complex of projective R–modules with an augmentation $P_0 \to A$, and $(A_\bullet, d_\bullet)$ be a projective resolution of A. Then we get a chain map from $(P_\bullet, d_\bullet)$ to $(A_\bullet, d_\bullet)$.*

Proof This is just [7, Proposition 2.11] applied to $\mathrm{id} \colon A \to A$. □

Note that the proof of [7, Proposition 2.11] is still valid with our broader notion of complexes. However, for readers who would rather not go through the proof, we also remark that we will be using this later only when $(P_\bullet, d_\bullet)$ is of the form $(\mathrm{Gr}_n(X), d_n)$ for a filtered spectrum $F_n(X)$, which is also a complex in the sense of [7].

Now, we are ready to quote from [8]:

Theorem 3.7 (mod 2 Whitehead conjecture [8, Corollary 1.2]) (i) *The sequence of Example 3.4(i),*

$$\cdots \xrightarrow{\delta_{k+1}} M(k+1) \xrightarrow{\delta_k} M(k) \to \cdots \to M(1) \xrightarrow{\delta_0} M(0) \xrightarrow{\epsilon} H\mathbb{Z}/2, \tag{10}$$

is exact.

(ii) *Denote by E_k the fiber of the map $\Sigma^{-k} D(k) \to \Sigma^{-k} H\mathbb{Z}/2$. Then the above sequence can be obtained splicing together short exact sequences $E_k \to M(k) \to E_{k-1}$.*

Remark 3.8 It is easy to see that our definition of E_k agrees with that in [8].

3.2 Maps into $(D(n), n \geq 0)$

With the above preparation, we are ready to prove the following.

Theorem 3.9 *Let (X, F_*X, ι) be a filtered spectrum such that*

(i) *$H_*(\iota_n)$ is injective for all n, and*

(ii) *$\mathrm{Gr}_n(F_*X)$ is a suspension spectrum.*

Then any map of spectra $F_0(X) \to S^0$ extends to a map of filtered spectra $F_*(X)$ to $D(*)$.

Proof First note that condition (i) implies that in the associated graded complex, the differential induces trivial map in cohomology. In particular, one can augment it by any map from $F_0(X) \to H\mathbb{Z}/2$. Let's do so by using the composition of the given map $F_0(X) \to S^0$ and the augmentation in the $(M(n), \delta_n)$, $S^0 \to H\mathbb{Z}/2$. Since $\mathrm{Gr}_0(F_*X) = F_0X$, this yields a map $\mathrm{Gr}_0(F_*X) \to H\mathbb{Z}/2$. By Theorem 3.7, the augmented complex $(M(n), \delta_n)$ is a projective resolution of $H\mathbb{Z}/2$, so we can apply Proposition 3.6 to obtain a map of complex of spectra f from $(\mathrm{Gr}_n(X), d_n)$ to $(M(n), \delta_n)$. From the proof of [7, Proposition 2.11], we see that we can choose f_0 to be the prescribed map in the statement of the theorem.

Thus we have found maps f_n making the square

$$
\begin{tikzcd}
\mathrm{Gr}_n(X) \arrow[r, "d_{n-1}"] \arrow[d, "f_n"] & \mathrm{Gr}_{n-1}(X) \arrow[d, "f_{n-1}"] \\
M(n) \arrow[r, "d_{n-1}"] & M(n-1)
\end{tikzcd}
\tag{11}
$$

commutative. Next we will show that there exists a map $\alpha_n \colon \Sigma^{-n} F_n(X) \to \Sigma^{-n} D(n)$ which makes the diagram

$$
\begin{tikzcd}
\Sigma^{-n} F_n(X) \arrow[r] \arrow[d, "\alpha_n"] & \mathrm{Gr}_n(X) \arrow[d, "f_n"] \\
\Sigma^{-n} D(n) \arrow[r] & M(n)
\end{tikzcd}
\tag{12}
$$

commutative for each n. We proceed by induction on n. The case $n = 0$ is trivial. Suppose that we have constructed such α_{n-1}. Consider the diagram

$$
\begin{tikzcd}
\mathrm{Gr}_n(X) \arrow[r] \arrow[d, "f_n"] & \Sigma^{1-n} F_{n-1}(X) \arrow[d, "\alpha_{n-1}"] \\
M(n) \arrow[r] & \Sigma^{1-n} D(n-1)
\end{tikzcd}
\tag{13}
$$

By the definition of associated graded complex, the fiber of the top row is $\Sigma^{-n} F_n(X)$ whereas by the cofibration (6), that of the bottom row is $\Sigma^{-n} D(n)$. Thus if we can show the commutativity of the diagram (13), then we can define the map α_n making the diagram (12) commute. Note that the two horizontal maps induce trivial maps in cohomology, which implies that the two compositions from the top left corner to the bottom right corner factor through E_{n-1} where E_i is the same as in Theorem 3.7. Thus it suffices to show that the lifts in $[\mathrm{Gr}_n(X), E_{n-1}]$ of the two maps agree. However, by Theorem 3.7(ii), $[\mathrm{Gr}_n(X), E_{n-1}]$ injects to $[\mathrm{Gr}_n(X), M(n-1)]$. Thus it suffices to show that the two maps agree after composition with the map $\Sigma^{1-n} D(n-1) \to M(n-1)$. Now, consider the diagram

$$
\begin{tikzcd}
\mathrm{Gr}_n(X) \arrow[r] \arrow[d, "f_n"] & \Sigma^{1-n} F_{n-1}(X) \arrow[r] \arrow[d, "\alpha_{n-1}"] & \mathrm{Gr}_{n-1}(X) \arrow[d, "f_{n-1}"] \\
M(n) \arrow[r] & \Sigma^{1-n} D(n-1) \arrow[r] & M(n-1)
\end{tikzcd}
$$

The right square is commutative by the inductive hypothesis. But we chose our maps f_n so that the big square commutes.

The proof is complete now, noting that, by considering the cofibers of the rows in the diagram (12), we see that the family $\{\alpha_n\}$ forms a map of filtered spectra. ◻

Corollary 3.10 *There exists a map of filtered spectra*

$$\alpha_* \colon (\mathrm{MO}, \Sigma^n \mathrm{MTO}(n)) \to (H\mathbb{Z}/2, D(n))$$

which extends the identity $\mathrm{MTO}(0) = S^0 \to S^0$.

Proof It follows from Thom isomorphism that the map

$$\iota_{n-1} \colon \Sigma^{n-1}\mathrm{MTO}(n-1) \to \Sigma^n \mathrm{MTO}(n)$$

induces a monomorphism in $\mathbb{Z}/2$–homology. Moreover, for the associated graded spectrum of $(Y, F_*Y) = (\mathrm{MO}, \Sigma^n\mathrm{MTO}(n))$ we have $\mathrm{Gr}_n(\mathrm{MO}) = \mathrm{BO}(n)_+$ (Example 3.4(ii)). We also have $F_0 Y = \mathrm{MTO}(0) = S^0$ and we take the identity $S^0 \to S^0$ as our map $F(0) \to D(0) = S^0$. The result now follows from Theorem 3.9. ◻

4 The splitting

4.1 Proof of Theorem 1.1

We have constructed the maps β_n in Section 2, and the maps α_n in Section 3. All that remains is to show that the composition $\alpha_n \circ \beta_n$ induces an isomorphism in 2–local cohomology. As $D(n)$ is of finite type, it is enough to show that it induces an isomorphism in mod 2 cohomology. Since a map of spectra induces a map of modules over Steenrod algebra in cohomology, and $H^*(D(n))$ is generated by the bottom class as a module over Steenrod algebra (5), it suffices to show that

$$H^{-n}(\alpha_n \circ \beta_n) = H^{-n}(\beta_n) \circ H^{-n}(\alpha_n)$$

is an isomorphism. Since α_0 is just the equivalence $\mathrm{MTS}(0) \cong S^0 \cong D(0)$, $H^0(\alpha^0)$ is an isomorphism. As the family $\{\alpha_n\}$ forms a map of filtered spectra, we see that $H^{-n}(\alpha_n)$ is an isomorphism for all $n \geq 0$.

Unfortunately we have been unable to prove the fact that the family of maps going the other way, $\{\beta_n\}$, forms a map of filtered spectra.[3] So we honestly compute $H^{-n}(\beta_n)$ for all n. We have

$$H^*(\mathrm{BO}(n)) \cong \mathbb{Z}/2[\sigma_1, \dots, \sigma_n] \subset H^*(B\Delta_n) \cong \mathbb{Z}/2[x_1, \dots, x_n],$$

[3]The claim we made in earlier versions available online on arXiv is erroneous: one of the errors is the fact that the fiber of j_{-2} has positive-dimensional cells if $n > 1$.

where σ_i denotes the i^{th} elementary symmetric polynomial in x_j's. Of course, the identification is made through Bi^* where $i\colon \Delta_n \cong O(1)^n \subset O(n)$ is the standard inclusion. Thus, the map $B\Delta_n^{-\rho_n} \to \mathrm{MTO}(n)$ induces an inclusion

$$H^*(\mathrm{MTO}(n)) \cong \mathbb{Z}/2[\sigma_1,\dots,\sigma_n]\cdot(\sigma_n)^{-1} \subset H^*B(\Delta_n^{-\rho_n}) \cong \mathbb{Z}/2[x_1,\dots,x_n]\cdot e(\rho_n)^{-1}$$

where

$$e(\rho) = \prod_{\epsilon_i \in 0,1,\, \prod_i \epsilon_i \neq 0} \Sigma \epsilon_i x_i.$$

Here and later, for a ring R and $a \in R$ nondivisor of 0, we denote $R \cdot a^{-1}$ the free R–module generated by a^{-1} in an appropriate localization of R. Now, we see that the only nontrivial element of $H^{-n}(\mathrm{MTO}(n))$, σ_n^{-1}, maps to $x_1^{-1}\cdots x_n^{-1} \in H^{-n}(B\Delta_n^{-\rho_n})$. But, this class is invariant under the e_n–action, so it survives in $H^{-n}(\Sigma^{-n}D(n))$ by [14, the first sentence of Remark 5.12]. Thus $H^{-n}(\beta_n)$ is also an isomorphism for all n. This concludes the proof of Theorem 1.1.

4.2 Further refinements

We have shown in [6, Theorem 1.1.A] that $\mathrm{BSO}(2n+1)_+$ splits off $\mathrm{MTO}(2n)$. More precisely, we show that the composition $Bf_{2n} \circ \omega_{O(2n)} \circ \mathrm{Tr}_{Bf_{2n}}$ is a homotopy equivalence, where $f_{2n}\colon O(2n) \to \mathrm{SO}(2n+1)$ is given by $X \mapsto (\det X)(X \oplus 1)$, $\omega_{O(2n)}$ is the map of Thom spectra induced by the embedding of $-\gamma_n$ in 0, and $\mathrm{Tr}_{Bf_{2n}}$ is the associated Becker–Schultz–Mann–Miller–Miller transfer $\mathrm{BSO}(2n+1)_+ \to \mathrm{MTO}(2n)$ [10, Section 2]; see also [1, Section 4]. One may ask how this splitting interacts with the splitting of the current paper. We show that they are complementary.

Corollary 4.1 $\Sigma^{-2n}D(2n) \vee \mathrm{BSO}(2n+1)_+$ *splits off* $\mathrm{MTO}(2n)$. *When* $n=1$, *we have a homotopy equivalence* $\mathrm{MTO}(2) \cong \Sigma^{-2}D(2) \vee \mathrm{BSO}(3)_+$.

Proof Consider the composition

$$H^*(\mathrm{BSO}(2n+1)) \oplus H^*(\Sigma^{-2n}D(2n)) \xrightarrow{(\alpha_{2n} \vee Bf_{2n} \circ \omega_{O(2n)})^* \circ (\beta_{2n} \vee \mathrm{Tr}_{Bf_{2n}})^*} H^*(\mathrm{BSO}(2n+1)) \oplus H^*(\Sigma^{-2n}D(2n)).$$

The components $H^*(\mathrm{BSO}(2n+1)) \to H^*(\mathrm{BSO}(2n+1))$ and $H^*(\Sigma^{-2n}D(2n)) \to H^*(\Sigma^{-2n}D(2n))$ are automorphisms by [6, Theorem 1.1.A] and Theorem 1.1, respectively. Consider now the component $H^*(\Sigma^{-2n}D(2n)) \to H^*(\mathrm{BSO}(2n+1))$. This is trivial since the source is generated over the Steenrod algebra by a negative-degree element, and the target is concentrated in nonnegative degrees by (5). Thus the map $(\alpha_{2n} \vee Bf_{2n} \circ \omega_{O(2n)})^* \circ (\beta_{2n} \vee \mathrm{Tr}_{Bf_{2n}})^*$ is an automorphism. This proves the splitting for general n. When $n = 1$, it suffices to compare the cohomology of both sides, or, alternatively, to compare the fibrations $\mathrm{MTO}(2) \to \mathrm{BO}(2)_+ \to \mathrm{MTO}(1)$ and $\Sigma^{-2} D(2) \to M(2) \to D(1)$. Noting that $\mathrm{BO}(2)_+ \cong M(2) \vee \mathrm{BSO}(3)_+$ (cf. [15, Theorem C]), we see that $(\alpha_2 \vee Bf_2 \circ \omega_{O(2)})^*$ induces mod 2 cohomology equivalence, which implies 2–local homotopy equivalence as everything is of finite type. □

5 Homology of the associated infinite loop spaces

In this section, we discuss the consequences of our splitting theorem to the homology of associated infinite loop spaces.

5.1 Exact sequences

We start with the following refinement of Proposition 1.3.

Proposition 5.1 *The sequence of Hopf algebras*

$$\cdots \to H_*(\Omega_0^\infty M(n)) \to H_*(\Omega_0^\infty M(n-1)) \to \cdots$$
$$\cdots \to H_*(\Omega_0^\infty M(2)) \to H_*(Q_0 B\mathbb{Z}/2_+) \to H_*(Q_0 S^0) \to \mathbb{Z}/2$$

is exact. It gives rise to an exact sequence of graded vector spaces after taking the module of indecomposables. Moreover, the image of $H_*(\Omega_0^\infty M(n)) \to H_*(\Omega_0^\infty M(n-1))$ is isomorphic to $H_*(\Omega_0^\infty \Sigma^{1-n} D(n-1))$.

Proof Suppose we have a short exact sequence of spectra $F \to X \to Y$. By the definition of the exactness, Definition 3.5, we see that the map $H_*(\Omega^\infty X) \to H_*(\Omega^\infty Y)$ is surjective. Thus by standard arguments (see e.g. [16, Section 2.6])

$$H_*(\Omega^\infty F) \to H_*(\Omega^\infty X) \to H_*(\Omega^\infty Y)$$

is short exact. Furthermore, it is clear that this short exact sequence splits as

$$H_*(\Omega_0^\infty F) \otimes k[\pi_0(F)] \to H_*(\Omega_0^\infty X) \otimes k[\pi_0(X)] \to H_*(\Omega_0^\infty Y) \otimes k[\pi_0(Y)],$$

where $k = \mathbb{Z}/2$. Noting that both in abelian categories and in the category of spectra, an exact sequence can be decomposed into a series of short exact sequences, we see that an exact sequence of spectra leads to an exact sequence of Hopf algebras by applying the functor $H_*(\Omega^\infty)$ or $H_*(\Omega_0^\infty)$.

Now, note that the $\mathrm{Gl}_n(\mathbb{Z}/2)$–action on $B\Delta_{n+}$ extends that on $B\Delta_n$. Thus it is easy to see from (7) that we have $e'_n B\Delta_n = e'_n B\Delta_{n+}$ for $n > 1$. As a matter of fact, $\mathrm{Gl}_n(\mathbb{Z}/2)$ acts trivially on $S^0 \subset B\Delta_{n+}$, so $q_n e'_n$ restricted to S^0 is the signed sum of 1's and (-1)'s which is zero. Thus for $n \geq 2$, $M(n)$ is a summand of $B\Delta_n$ (and not just a summand of $B\Delta_{n+}$), so $\Omega^\infty M(n)$ splits off $QB\Delta_n$ as infinite loop spaces. Of course, this also implies that $M(n)$ is connected for $n \geq 2$, so $\Omega_0^\infty M(n) = \Omega^\infty M(n)$. Therefore $H_*(\Omega_0^\infty M(n))$ splits off $H_*(QB\Delta_n)$ as Hopf algebras; in particular, the former is isomorphic to a Hopf subalgebra of the latter, which is a polynomial algebra. It is known that any Hopf subalgebra of a polynomial algebra is polynomial by the structure theorem of Hopf algebras over $\mathbb{Z}/2$ ([2, Theorem 6.1] or [11, Theorem 7.11]). So $H_*(\Omega_0^\infty M(n))$ is also a polynomial algebra. Thus everything in the exact sequence is polynomial. As any surjective map of algebras to a polynomial algebra splits, we see that a short exact sequence of Hopf algebras involving only polynomial algebras remain exact after passing to the modules of indecomposables. Noting that an exact sequence of polynomial Hopf algebras can be obtained by splicing together short exact sequences of polynomial Hopf algebras, we can say the same about an exact sequence of Hopf algebras, not necessarily short exact.

It remains to identify the image of each map. But this follows from Theorem 3.7(ii) and the fact that the map $E_n \to \Sigma^{-n} D(n)$ induces homotopy equivalence $\Omega_0^\infty E_n \to \Omega_0^\infty \Sigma^{-n} D(n)$. □

Remark 5.2 By the comments in the first paragraph of the above proof we also have an exact sequence of Hopf algebras even if we don't restrict to the base point components; that is we also have an exact sequence of Hopf algebras

$$\begin{aligned}\cdots \to H_*(\Omega^\infty M(n)) \to H_*(\Omega^\infty M(n-1)) \to \cdots \\ \cdots \to H_*(\Omega^\infty M(2)) \to H_*(QB\mathbb{Z}/2_+) \to H_*(QS^0) \to \mathbb{Z}/2.\end{aligned}$$

An immediate consequence of Proposition 5.1 is:

Corollary 5.3 *$H^*(\Omega_0^\infty \mathrm{MTO}(2))$ is a polynomial algebra.*

Proof By Corollary 4.1 we have $\Omega_0^\infty \mathrm{MTO}(2) \cong Q_0\mathrm{BSO}(3)_+ \times \Omega^\infty E_2$, noting that $\pi_0(E_2) = 0$ since it is a direct summand of $\pi_0(M(2))$. The short exact sequence above, dualized, implies that $H^*(\Omega^\infty E_2)$ injects to $H^*(\Omega^\infty M(3))$. Since $M(3)$ is a stable summand of $\mathrm{BO}(3)$, we see that $H^*(\Omega^\infty E_2)$ injects to $H^*(Q_0\mathrm{BO}(3))$ which is polynomial [19, Theorem 3.11]. Since $H^*(\Omega^\infty E_2)$ is a connected Hopf algebra, as in the above, by the structure theorem of Hopf algebras over $\mathbb{Z}/2$, this implies that $H^*(\Omega^\infty E_2)$ itself is a polynomial algebra. Now the corollary follows as the other tensor factor $H^*(Q_0\mathrm{BO}(3))$ is polynomial again by [19, Theorem 3.11]. □

5.2 Relations among μ–classes

We now prove Theorem 1.5 as an application. We start with the following definitions. For a $\mathbb{Z}/2$–algebra R, denote by $Q(R)$ its module of indecomposables, i.e. $I(R)/(I(R)^2)$ where $I(R)$ denotes the augmentation ideal. We will write often QR instead of $Q(R)$ to avoid heavy notations.

Lemma 5.4 *Let X be a pointed space, $u_X\colon X \to QX$ be the unit map, and*

$$\sigma_*^\infty\colon QH_*(QX) \to H_*(X)$$

the homology suspension. Write $W(QH_*(QX))$ for the image of $\widetilde{H}_*(X)$ in $QH_*(QX)$ by the composition of $H_*(u_X)$ and the projection $\widetilde{H}_*(QX) \to QH_*(QX)$, and write $F(QH_*(QX)) = \mathrm{Ker}(\sigma_*^\infty)$.[4] Then we have

$$QH_*(QX) \cong W(QH_*(QX)) \oplus F(QH_*(QX)).$$

This direct sum decomposition is natural with respect to maps of spaces (and not map of suspension spectra). We will refer to it as the WF decomposition.

Proof The direct sum decomposition is an immediate consequence of the standard fact that the homology suspension surjects to $\widetilde{H}_*(X)$ (e.g. [6, Lemmas 4.4 and 4.5]). Since σ_*^∞ is natural with respect to maps of spectra, and u_X is natural with respect to maps of spaces, the decomposition is natural with respect to maps of spaces. □

We will extend this to slightly wider category of infinite loop spaces including the $\Omega^\infty M(n)$'s.

Lemma 5.5 *Let X be a space on which a group G acts, $e \in \mathbb{Z}_{(2)}[G]$ an idempotent. Denote by $\pi\colon X \to eX$ the projection and by $i\colon eX \to X$ the section associated to the splitting of X by e such that $e = i \circ \pi$. Then one can decompose $QH_*(\Omega^\infty eX)$ as*

$$QH_*(\Omega^\infty eX) \cong W(QH_*(\Omega^\infty eX)) \oplus F(QH_*(\Omega^\infty eX))$$

so that the direct sum decomposition is compatible with that of $H_*(QX)$ via $H_*(\Omega^\infty \pi)$ as well as $H_*(\Omega^\infty i)$.

[4] The notation is voluntarily reminiscent of what we used in earlier versions. W here corresponds to W_1, F to F_2.

Proof This is equivalent to saying that $QH_*(\Omega^\infty e)$ respects the WF–decomposition. Let $g \in G$. Then g acts on X via a map of spaces, so $QH_*(\Omega^\infty g)$ respects the WF–decomposition. On the other hand, for $x \in H_*(QX)$,

$$H_*(\Omega^\infty(g_1+g_2))(x) = \Sigma H_*(\Omega^\infty(g_1))(x')H_*(\Omega^\infty(g_2))(x''),$$
$$H_*(\Delta_{QX})(x) = \Sigma x' \times x'',$$

but we have $H_*(\Delta_{QX})(x) = 1 \otimes x + x \otimes 1$ modulo $I \otimes I$ where I is the augmentation ideal of $H_*(QX)$. Thus,

$$QH_*(\Omega^\infty(g_1+g_2)) = QH_*(\Omega^\infty(g_1)) + QH_*(\Omega^\infty(g_2)),$$

so $QH_*(\Omega^\infty e)$ also respects the WF decomposition. □

As noted above, maps of spectra don't necessarily respect the WF decomposition. However, as the summand F is defined in terms of stable information only, some maps of spectra have nice behavior with respect to this decomposition. For example, we can prove:

Lemma 5.6 *The map*

$$QH_*(\Omega^\infty\delta_{n-1})\colon QH_*(\Omega^\infty M(n)) \to QH_*(\Omega^\infty M(n-1))$$

induces an inclusion

$$W(QH_*(\Omega^\infty M(n))) \to F(QH_*(\Omega^\infty M(n-1))).$$

Proof The long exact sequence for the homology of the cofibration (6) implies that $H_*(\delta_k)$ is trivial for all k. Thus by naturality of the homology suspension, we see that the image of $QH_*(\Omega^\infty\delta_{n-1})$ is included in $F(QH_*(\Omega^\infty M(n-1)))$. By Remark 5.2 we have $\mathrm{Ker}(QH_*(\Omega^\infty\delta_{n-1})) = \mathrm{Im}(QH_*(\Omega^\infty\delta_n))$, but as before this is included in $F(QH_*(\Omega^\infty M(n)))$. So the restriction of $QH_*(\Omega^\infty\delta_{n-1})$ to $W(QH_*(\Omega^\infty M(n)))$ is injective. □

Now we are ready to prove Theorem 1.5. The inclusion $H^*(M(n)) \subset H^*(\mathrm{BO}(n))$ is given by $H^*(f_n)$, and this is determined uniquely by its compatibility with $H^*(\alpha_n)$, which in turn is determined uniquely by the fact that $H^{-n}(\mathrm{MTO}(n))$ contains only one nontrivial element, and the fact that $H^*(D(n))$ is generated by the bottom class as a module over the Steenrod algebra (5). The cofibration sequence (2) implies that

$(\Omega^\infty\omega_{O(n)})^*(\sigma^{\infty *}(x))=0$ if $\sigma^{\infty *}(x)\in H^*(Q(\mathrm{BO}(n)_+))$ belongs to the image of $H^*(\Omega^\infty\mathrm{MTO}(n-1))$. Now, Theorem 3.9 implies that we have a commutative diagram

$$\begin{tikzcd}
\mathrm{BO}(n)_+ \arrow[r, "f_n"] \arrow[d] & M(n) \arrow[d] \arrow[dr, "\delta_{n-1}"] & \\
\mathrm{MTO}(n-1) \arrow[r, "\alpha_{n-1}"] & \Sigma^{1-n}D(n-1) \arrow[r] & M(n-1)
\end{tikzcd}$$

Thus we get

$$\begin{tikzcd}
H^*(Q(\mathrm{BO}(n)_+)) & H^*(\Omega^\infty M(n)) \arrow[l, "H^*(\Omega^\infty f_n)"'] & \\
H^*(\Omega^\infty\mathrm{MTO}(n-1)) \arrow[u] & H^*(\Omega^\infty(\Sigma^{1-n}D(n-1)) \arrow[l] \arrow[u] & H^*(\Omega^\infty M(n-1)) \arrow[l] \arrow[ul, "H^*(\Omega^\infty\delta^{n-1})"']
\end{tikzcd}$$

On the other hand, dualizing Lemma 5.6, we see that the dual of $F\big(QH_*(\Omega^\infty M(n-1))\big)$ surjects to the dual of $W(QH_*(\Omega^\infty M(n)))$, which is precisely the image of $\sigma^{\infty *}_{M(n)}$. Thus we have inclusions

$$\mathrm{Im}(\sigma^{\infty *}_{M(n)})\subset\mathrm{Im}(PH^*(\Omega^\infty\delta^{n-1}))\subset\mathrm{Im}(H^*(\Omega^\infty\delta^{n-1})).$$

Therefore by the commutativity of the diagram above, we see that the image of the composition

$$H^*(M(n))\xrightarrow{\sigma^{\infty *}}H^*(\Omega^\infty M(n))\to H^*(Q(\mathrm{BO}(n)_+))$$

is contained in the image of $H^*(\Omega^\infty\mathrm{MTO}(n-1))$. This concludes the proof of (i).

Now, notice that the splitting $\mathrm{BO}(2)_+\simeq\mathrm{BSO}(3)_+\vee M(2)$ combined with part (i) shows that the only nontrivially characteristic classes may arise from the restriction of $(\Omega^\infty\omega_{O(2)})^*\circ\sigma^{\infty *}$ to the $H^*\mathrm{BSO}(3)$ summand of $H^*\mathrm{BO}(2)$, which was studied in [6, Theorem 1.9]. Noting that [6, Remark 4.7] allows us to talk about μ–classes and ν–classes interchangeably, we get (ii) and (iii).

To conclude, we give some explicit examples of those relations. First of all, we have [12, Corollary 3.11].

Proposition 5.7 *The image of $H^*(M(n))$ in $H^*(\mathrm{BO}(n))$ is the free-module over $H^*(B\Delta_n)^{\mathrm{Gl}_n(\mathbb{Z}/2)}$ generated by a basis of $A(n-2)\,\mathrm{Sq}^{2^{n-1},\dots,2,1}(x_1^{-1}\cdots x_n^{-1})$, where $A(k)$ is the subalgebra of the Steenrod algebra generated by $\mathrm{Sq}^1,\mathrm{Sq}^2,\dots,\mathrm{Sq}^{2^k}$. Here we identify $H^*(\mathrm{BO}(n))$ with its image in*

$$H^*(B\Delta_n)\subset H^*(B\Delta_n)^{-\gamma_n}\cong H^*(B\Delta_n)\cdot(x_1\cdots x_n)^{-1}$$

via Bi^* where $i: \Delta_n \subset O(n)$. In terms of cohomology classes, we identify $H^*(\mathrm{BO}(n))$ with the subalgebra of $H^*(B\Delta_n)$ generated by the elementary symmetric polynomials $\sigma_1 = \sigma_1(x_1, \dots, x_n), \dots, \sigma_n = \sigma_n(x_1, \dots, x_n)$.

The action of the Steenrod algebra on $H^*(B\Delta_n) \cdot (x_1 \cdots x_n)^{-1}$ is determined by the action of the total Steenrod square (see e.g. [18]) $\mathrm{Sq}^T(x_i) = x_i + x_i^2$ for $1 \le i \le n$, and the Cartan formula $\mathrm{Sq}^T(yz) = \mathrm{Sq}^T(y)\,\mathrm{Sq}^T(z)$ for any y, z. Thus

$$\mathrm{Sq}^T(x_i^{-1}) = x_i^{-1}(1+x_i)^{-1} = x_i^{-1}(1 + x_i + x_i^2 + x_i^3 + \cdots).$$

When $n = 2$, $A(0)$ is just the exterior algebra generated by Sq^1, that is, a graded vector space spanned by 1 and Sq^1. Furthermore, by the above, we see that

$$\mathrm{Sq}^{2,1}(x_1^{-1}x_2^{-1}) = x_1 + x_2 = \sigma_1, \mathrm{Sq}^1\,\mathrm{Sq}^{2,1}(x_1^{-1}x_2^{-1}) = \mathrm{Sq}^1(\sigma^1) = x_1^2 + x_2^2 = \sigma_1^2.$$

Since the Dickson invariant algebra $H^*(B\Delta_n)^{\mathrm{Gl}_n(\mathbb{Z}/2)}$ is generated by

$$w_2 = x_1^2 + x_1x_2 + x_2^2 = \sigma_1^2 + \sigma_2, \quad w_3 = x_1x_2(x_1 + x_2) = \sigma_1\sigma_2$$

[12, Theorem A1], we derive:

Corollary 5.8 *The set*

$$\{(\sigma_1^2 + \sigma_2)^i (\sigma_1\sigma_2)^j \sigma_1^\epsilon \mid i \ge 0,\ j \ge 0, \epsilon \in \{1, 2\}\}$$

forms a basis of the image of $H^*(M(2))$ in $H^*(\mathrm{BO}(2))$.

Combined with Theorem 1.5, we get a table of these relations in low dimensions,

$$\begin{aligned}
\mu_{1,0} &= 0 \quad (i = 0, j = 0, \epsilon = 1),\\
\mu_{3,0} + \mu_{1,1} &= 0 \quad (i = 1, j = 0, \epsilon = 1),\\
\mu_{2,1} &= 0 \quad (i = 0, j = 1, \epsilon = 1),\\
\mu_{5,0} + \mu_{3,1} + \mu_{1,2} &= 0 \quad (i = 2, j = 0, \epsilon = 1),\\
\mu_{3,1} &= 0 \quad (i = 0, j = 1, \epsilon = 2),\\
\mu_{4,1} + \mu_{2,2} &= 0 \quad (i = 1, j = 1, \epsilon = 1).
\end{aligned}$$

Here we have omitted the relations that follow from lower degree relations and the general relation $\mu_{2i,2j} = \mu_{i,j}^2$. For example, setting $\epsilon = 2, i = 1$ and $j = 0$ gives $\mu_{4,0} + \mu_{2,1} = 0$; however, we have already listed $\mu_{2,1} = 0$, and we can deduce $\mu_{4,0} = 0$ from $\mu_{4,0} = \mu_{1,0}^4$ and $\mu_{1,0} = 0$.

References

[1] **J C Becker**, **R E Schultz**, *The real semicharacteristic of a fibered manifold*, Quart. J. Math. Oxford Ser. 33 (1982) 385–403 MR Zbl

[2] **A Borel**, *Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts*, Ann. of Math. 57 (1953) 115–207 MR Zbl

[3] **S Galatius**, **O Randal-Williams**, *Stable moduli spaces of high-dimensional manifolds*, Acta Math. 212 (2014) 257–377 MR Zbl

[4] **S Galatius**, **U Tillmann**, **I Madsen**, **M Weiss**, *The homotopy type of the cobordism category*, Acta Math. 202 (2009) 195–239 MR Zbl

[5] **D J Hunter**, **N J Kuhn**, *Characterizations of spectra with $\mathcal{U}$–injective cohomology which satisfy the Brown–Gitler property*, Trans. Amer. Math. Soc. 352 (2000) 1171–1190 MR Zbl

[6] **T Kashiwabara**, **H Zare**, *Splitting Madsen–Tillmann spectra, I: Twisted transfer maps*, Bull. Belg. Math. Soc. Simon Stevin 25 (2018) 263–304 MR Zbl

[7] **N J Kuhn**, *Spacelike resolutions of spectra*, from "Proceedings of the northwestern homotopy theory conference" (H R Miller, S B Priddy, editors), Contemp. Math. 19, Amer. Math. Soc., Providence, RI (1983) 153–165 MR Zbl

[8] **N J Kuhn**, **S B Priddy**, *The transfer and Whitehead's conjecture*, Math. Proc. Cambridge Philos. Soc. 98 (1985) 459–480 MR Zbl

[9] **L G Lewis, Jr**, **J P May**, **M Steinberger**, **J E McClure**, *Equivariant stable homotopy theory*, Lecture Notes in Math. 1213, Springer (1986) MR Zbl

[10] **B M Mann**, **E Y Miller**, **H R Miller**, *S^1–equivariant function spaces and characteristic classes*, Trans. Amer. Math. Soc. 295 (1986) 233–256 MR Zbl

[11] **J W Milnor**, **J C Moore**, *On the structure of Hopf algebras*, Ann. of Math. 81 (1965) 211–264 MR Zbl

[12] **S A Mitchell**, *Finite complexes with $A(n)$–free cohomology*, Topology 24 (1985) 227–246 MR Zbl

[13] **S A Mitchell**, *Splitting $B(\mathbb{Z}/p)^n$ and BT^n via modular representation theory*, Math. Z. 189 (1985) 1–9 MR Zbl

[14] **S A Mitchell**, **S B Priddy**, *Stable splittings derived from the Steinberg module*, Topology 22 (1983) 285–298 MR Zbl

[15] **S A Mitchell**, **S B Priddy**, *Symmetric product spectra and splittings of classifying spaces*, Amer. J. Math. 106 (1984) 219–232 MR Zbl

[16] **O Randal-Williams**, *The homology of the stable nonorientable mapping class group*, Algebr. Geom. Topol. 8 (2008) 1811–1832 MR Zbl

[17] **S Takayasu**, *On stable summands of Thom spectra of $B(\mathbb{Z}/2)^n$ associated to Steinberg modules*, J. Math. Kyoto Univ. 39 (1999) 377–398 MR Zbl

[18] **G Walker**, **R M W Wood**, *Polynomials and the* mod 2 *Steenrod algebra, II: Representations of* $\mathrm{GL}(n, \mathbb{F}_2)$, London Mathematical Society Lecture Note Series 442, Cambridge Univ. Press (2018) MR Zbl

[19] **R J Wellington**, *The unstable Adams spectral sequence for free iterated loop spaces*, Mem. Amer. Math. Soc. 258, Amer. Math. Soc., Providence, RI (1982) MR Zbl

TK: *Laboratoire de Mathématiques, Institut Fourier, Université Grenoble Alpes*
Grenoble, France

HZ: *School of Mathematics, Statistics and Computer Science, College of Science*
University of Tehran
Tehran, Iran

HZ: *School of Mathematics, Institute for Research in Fundamental Sciences (IPM)*
Tehran, Iran

takuji.kashiwabara@univ-grenoble-alpes.fr, hadi.zare@ut.ac.ir

Received: 8 December 2015 Revised: 20 December 2021

Algebraic & Geometric Topology **23**:5 (2023) 1959–2008
DOI: 10.2140/agt.2023.23.1959
Published: 25 July 2023

Free and based path groupoids

Andrés Ángel
Hellen Colman

We give an explicit description of the free path and loop groupoids in the Morita bicategory of translation topological groupoids. We prove that the free path groupoid of a discrete group acting properly on a topological space X is a translation groupoid given by the same group acting on the topological path space X^I. We give a detailed description of based path and loop groupoids and show that both are equivalent to topological spaces. We also establish the notion of homotopy and fibration in this context.

18B40, 55P35, 58E40; 55R91, 58D19

1 Introduction

Our aim is to give an explicit description of the path object in the bicategory of translation topological groupoids. Our main application will be in the setting of orbifolds as groupoids.

We adopt the model developed by Moerdijk and Pronk [9] to describe orbifolds in terms of groupoids. Essentially an orbifold is a Morita equivalence class of groupoids of a certain type, which we will call orbifold groupoids.

In this spirit, the right notion of morphism between orbifold groupoids is that of a generalized map. These generalized maps arise as morphisms in the bicategory of topological groupoids, functors and natural transformations when inverting the essential equivalences; see Pronk [13].

All orbifolds can be represented by a groupoid given by a certain type of action of a group G on a topological space X. This representation $G \ltimes X$ is called translation groupoid. In particular we will be interested in *developable* orbifolds defined by a translation groupoid given by a discrete group acting properly on a space.

For these orbifolds, we use their groupoid characterization to obtain a description of the generalized maps from the interval to the orbifold as a translation groupoid. We prove that the free path groupoid of the translation groupoid $G \ltimes X$ is the translation groupoid $G \ltimes X^I$. In fact we describe three different approaches resulting in three characterizations of the path groupoid: as a colimit of G–paths, as a groupoid of multiple G–paths and as a translation groupoid $G \ltimes X^I$. We prove that the three groupoids are equivalent.

We show that this construction of the path groupoid is functorial and invariant under Morita equivalence.

The pullback along the diagonal of this model gives us as a particular case, the free loop groupoid which coincides with the descriptions given by Lupercio and Uribe [7], Adem, Leida and Ruan [1] and Noohi [10] in various contexts.

Moreover, we use this model to calculate the based groupoid of paths between two points. We prove that this groupoid is actually equivalent to a topological space.

Using our description of the path groupoid, we provide an explicit characterization for a homotopy between two generalized maps, as well as a definition of orbifold fibrations. We prove that the evaluation map is both a groupoid homotopy equivalence and a groupoid fibration.

Organization

In Section 2 we present some basic definitions and constructions for topological groupoids. We define translation groupoids and introduce the bicategory of translation groupoids resulting from inverting the essential equivalences. Section 3 introduces the model for orbifolds as groupoids that gives the setting for the construction of the path groupoid in the next section. Section 4 is devoted to the construction of the free path groupoid. We give here an explicit equivalence between all models for the path groupoid. We prove that this construction is functorial and invariant under Morita equivalence. Section 5 provides a detailed description of the based path and loop groupoids and describes some examples. Section 6 concerns the characterization of the homotopy between generalized maps. In Section 7 we provide a definition of groupoid fibration and prove that the evaluation morphism is a groupoid fibration.

2 Context

2.1 Topological groupoids

A *topological groupoid* $\mathcal{G}$ is a groupoid object in the category Top of topological spaces and continuous maps. Our notation for groupoids is that G_0 is the space of objects and G_1 is the space of arrows, with source and target maps $s, t\colon G_1 \to G_0$, multiplication $m\colon G_1 \times_{G_0} G_1 \to G_1$, inversion $i\colon G_1 \to G_1$, and object inclusion $u\colon G_0 \hookrightarrow G_1$.

The set of arrows from x to y is denoted by $G(x,y) = \{g \in G_1 \mid s(g) = x \text{ and } t(g) = y\}$. The set of arrows from x to itself, $G(x,x)$, is a group called the *isotropy* group of $\mathcal{G}$ at x and denoted by G_x.

A *strict morphism* $\phi\colon \mathcal{K} \to \mathcal{G}$ of groupoids is a functor given by two continuous maps $\phi\colon K_1 \to G_1$ and $\phi\colon K_0 \to G_0$ that together commute with all the structure maps of the groupoids $\mathcal{K}$ and $\mathcal{G}$.

A *natural transformation* $T\colon \phi \Rightarrow \psi$ between two morphisms $\phi, \psi\colon \mathcal{K} \to \mathcal{G}$ is a continuous map $T\colon K_0 \to G_1$ with $T(x)\colon \phi(x) \to \psi(x)$ such that for any arrow $h\colon x \to y$ in K_1, the identity $\psi(h)T(x) = T(y)\phi(h)$ holds. Since we are in a topological groupoid and inversion is continuous, we also have a natural transformation $T^{-1}\colon \psi \Rightarrow \phi$ and write $\phi \sim_T \psi$.

Topological groupoids, strict morphisms and natural transformations form a 2–category, which we denote by TopG.

A strict morphism $\epsilon\colon \mathcal{K} \to \mathcal{G}$ of topological groupoids is an *essential equivalence* if:

(i) ϵ is essentially surjective in the sense that

$$s\pi_1\colon G_1 \times^t_{G_0} K_0 \to G_0$$

is an open surjection where $G_1 \times^t_{G_0} K_0$ is the pullback along the target $t\colon G_1 \to G_0$.

(ii) ϵ is fully faithful in the sense that K_1 is the pullback of topological spaces

$$\begin{tikzcd}
K_1 \arrow[r, "\epsilon"] \arrow[d, "{(s,t)}"'] & G_1 \arrow[d, "{(s,t)}"] \\
K_0 \times K_0 \arrow[r, "\epsilon \times \epsilon"] & G_0 \times G_0
\end{tikzcd}$$

Note that if there exists a functor $\delta\colon \mathcal{G} \to \mathcal{K}$ with natural transformations $\eta\colon \mathrm{id}_{\mathcal{G}} \Rightarrow \epsilon \circ \delta$ and $\nu\colon \delta \circ \epsilon \Rightarrow \mathrm{id}_{\mathcal{K}}$ in TopG, the functor ϵ is essentially surjective — indeed, $s\pi_1$ has a section defined by $(\eta_x, \delta(x))\colon G_0 \to G_1 \times^t_{G_0} K_0$, which implies that it is open and surjective — and ϵ is fully faithful because the map $K_1 \to \mathcal{K}_0 \times \mathcal{K}_0 \times_{\mathcal{G}_0 \times \mathcal{G}_0} \mathcal{G}_1$ has an inverse defined by $(x, y, h) \to \nu_y \circ \delta(h) \circ \nu_x^{-1}$.

An essential equivalence $\epsilon\colon \mathcal{K} \to \mathcal{G}$ does not generally have an inverse functor $\delta\colon \mathcal{G} \to \mathcal{K}$ such that $\epsilon \circ \delta \sim_T \mathrm{id}_{\mathcal{G}}$ and $\delta \circ \epsilon \sim_{T'} \mathrm{id}_{\mathcal{K}}$ in TopG. The functor δ exists by the axiom of choice but in general it is not *continuous*.

Definition 2.1 Let $\psi\colon \mathcal{K} \to \mathcal{G}$ and $\phi\colon \mathcal{L} \to \mathcal{G}$ be strict morphisms. The *groupoid pullback* $\mathcal{P} = \mathcal{K} \times_{\mathcal{G}} \mathcal{L}$ is the topological groupoid whose space of objects is

$$P_0 = K_0 \times^t_{G_0} G_1 \times^s_{G_0} L_0$$

and space of arrows is $P_1 = K_1 \times^t_{G_0} G_1 \times^s_{G_0} L_1$. Source and target maps are given by $s(k, g, l) = (s(k), \psi(k)^{-1} g \phi(l), s(l))$ and $t(k, g, l) = (t(k), g, t(l))$. There is a square of morphisms and a natural transformation T that makes the diagram

$$\begin{tikzcd}
\mathcal{K} \times_{\mathcal{G}} \mathcal{L} \arrow[r, "\pi_1"] \arrow[d, "\pi_2"'] \arrow[dr, phantom, "\sim_T"] & \mathcal{K} \arrow[d, "\psi"] \\
\mathcal{L} \arrow[r, "\phi"'] & \mathcal{G}
\end{tikzcd}$$

commutative and is universal with this property.

Definition 2.2 The groupoids $\mathcal{K}$ and $\mathcal{G}$ are *Morita equivalent* if there exists a groupoid $\mathcal{L}$ and a span

$$\mathcal{K} \xleftarrow{\sigma} \mathcal{L} \xrightarrow{\epsilon} \mathcal{G},$$

where ϵ and σ are essential equivalences. We write $\mathcal{G} \sim_M \mathcal{K}$.

The proof that a Morita equivalence is an equivalence relation is based on the groupoid pullback defined above.

A *generalized map* (ϵ, ϕ) from $\mathcal{K}$ to $\mathcal{G}$ is a span $\mathcal{K} \xleftarrow{\epsilon} \mathcal{J} \xrightarrow{\phi} \mathcal{G}$ such that ϵ is an essential equivalence. Two generalized maps $\mathcal{K} \xleftarrow{\epsilon} \mathcal{J} \xrightarrow{\phi} \mathcal{G}$ and $\mathcal{K} \xleftarrow{\epsilon'} \mathcal{J}' \xrightarrow{\phi'} \mathcal{G}$ are *equivalent* if there exists a diagram

$$\begin{tikzcd}
 & \mathcal{J} \arrow[dl, "\epsilon"'] \arrow[dr, "\phi"] & \\
\mathcal{K} \arrow[r, phantom, "\sim_T"] & \mathcal{L} \arrow[u, "u"'] \arrow[d, "v"] \arrow[r, phantom, "\sim_{T'}"] & \mathcal{G} \\
 & \mathcal{J}' \arrow[ul, "\epsilon'"] \arrow[ur, "\phi'"'] &
\end{tikzcd}$$

which is commutative up to natural transformations and where $\mathcal{L}$ is a topological groupoid, and u and v are essential equivalences.

2.2 The Morita bicategory of topological groupoids MTopG

Consider the class of arrows E given by the essential equivalences in the 2–category TopG. It was proven by Pronk in [13; 14] that E satisfies the conditions to admit a bicalculus of fractions. The bicategory of fractions $\mathsf{TopG}(E^{-1})$ obtained by formally inverting the essential equivalences is what we call the *Morita bicategory of topological groupoids* and we denote by MTopG.

The explicit description of the bicategory MTopG is as follows:

- Objects are topological groupoids $\mathcal{G}$.
- A 1–morphism from $\mathcal{K}$ to $\mathcal{G}$ is a *generalized map*
$$\mathcal{K} \xleftarrow{\epsilon} \mathcal{J} \xrightarrow{\phi} \mathcal{G}$$
such that ϵ is an essential equivalence.
- A 2–morphism from $\mathcal{K} \xleftarrow{\epsilon} \mathcal{J} \xrightarrow{\phi} \mathcal{G}$ to $\mathcal{K} \xleftarrow{\epsilon'} \mathcal{J}' \xrightarrow{\phi'} \mathcal{G}$ is given by a class of diagrams
$$\begin{tikzcd} & \mathcal{J} \arrow[dl, "\epsilon"'] \arrow[dr, "\phi"] & \\ \mathcal{K} & \sim_T \ \mathcal{L} \ \sim_{T'} \arrow[u, "u"'] \arrow[d, "v"] & \mathcal{G} \\ & \mathcal{J}' \arrow[ul, "\epsilon'"] \arrow[ur, "\phi'"'] & \end{tikzcd}$$
where $\mathcal{L}$ is a topological groupoid, and u and v are essential equivalences.

The horizontal composition of generalized maps $\mathcal{K} \xleftarrow{\epsilon} \mathcal{J} \xrightarrow{\phi} \mathcal{G}$ and $\mathcal{G} \xleftarrow{\zeta} \mathcal{J}' \xrightarrow{\psi} \mathcal{L}$ is given by the diagram
$$\begin{tikzcd} & & \mathcal{J}' \times_{\mathcal{G}} \mathcal{J} \arrow[dl] \arrow[dr] & & \\ & \mathcal{J} \arrow[dl, "\epsilon"'] \arrow[dr, "\phi"] & & \mathcal{J}' \arrow[dl, "\zeta"'] \arrow[dr, "\psi"] & \\ \mathcal{K} & & \mathcal{G} & & \mathcal{L} \end{tikzcd}$$
where $\mathcal{J}' \times_{\mathcal{G}} \mathcal{J}$ is the groupoid pullback. Note that this composition is associative only up to a 2–morphism.

2.3 Translation groupoids

Let G be a topological group with a continuous left action on a topological space X. Then the *translation groupoid* $G \ltimes X$ is defined by:

- The space of objects is X itself, and the space of arrows is the Cartesian product $G \times X$.
- The source $s\colon G \times X \to X$ is the second projection, and the target $t\colon G \times X \to X$ is given by the action. Then (g, x) is an arrow $x \to gx$.
- The other structure maps are defined by the unit $u(x) = (e, x)$, where e is the identity element in G, and $(h, gx) \circ (g, x) = (h \star g, x)$ where $\star$ is the group multiplication.

Example 2.3 These examples will appear later on in our applications.

(1) **Unit groupoid** Consider the groupoid $e \ltimes X$ given by the action of the trivial group e on the topological space X. This is a topological groupoid whose arrows are all units. In this way, any topological space can be considered as a groupoid.

(2) **Multiplication groupoid** Let H be a subgroup of a topological group G. Consider the translation groupoid $H \ltimes G$ where H acts by multiplication on G.

(3) **Conjugation groupoid** Let H be a subgroup of a topological group G. Consider the translation groupoid $H \ltimes G$ where H acts by conjugation on G.

(4) **Point groupoid** Let G be a topological group. Let $\bullet$ be a point. Consider the groupoid $G \ltimes \bullet$ where G acts trivially on the point. This is a topological groupoid with exactly one object, $\bullet$, and G is the space of arrows in which the maps s and t coincide. We call $G \ltimes \bullet$ the point groupoid associated to G. In this way any group can be considered as a groupoid.

We will denote by $\mathbf{1}$ the *trivial groupoid* with one object and one arrow; that is, $\mathbf{1} = e \ltimes \bullet$, the unit groupoid over a point or a point groupoid associated to the trivial group.

An *equivariant map* $G \ltimes X \to K \ltimes Y$ between translation groupoids consists of a pair $\varphi \ltimes f$, where $\varphi\colon G \to K$ is a group homomorphism and $f\colon X \to Y$ satisfies $f(gx) = \varphi(g) f(x)$ for $g \in G$ and $x \in X$.

Translation groupoids, equivariant maps and natural transformations form a 2–category that we denote by TrG.

2.4 The Morita bicategory of translation groupoids MTrG

We construct now a subbicategory MTrG of the Morita bicategory of topological groupoids MTopG where the objects are strictly the translation groupoids and the maps are equivariant ones.

Proposition 2.4 [14] *Let $\psi\colon G \ltimes X \to L \ltimes Z$ and $\phi\colon H \ltimes Y \to L \ltimes Z$ be equivariant maps. The fiber product $\mathcal{K}$*

$$\begin{array}{ccc} \mathcal{K} & \xrightarrow{\pi_1 \ltimes f} & G \ltimes X \\ {\scriptstyle \pi_2 \ltimes g}\downarrow & & \downarrow{\scriptstyle \psi} \\ H \ltimes Y & \xrightarrow{\phi} & L \ltimes Z \end{array}$$

is again a translation groupoid. Moreover, its structure group is $G \times H$, $\mathcal{K} = (G \times H) \ltimes P$ and the first components of the equivariant maps $\pi_1 \ltimes f$ and $\pi_2 \ltimes g$ are the group projections $\pi_1\colon G \times H \to G$ and $\pi_2\colon G \times H \to H$.

An *equivariant essential equivalence* is an equivariant map $\xi \ltimes \epsilon$ which is an essential equivalence.

Consider the bicategory whose

- objects are translation groupoids $G \ltimes X$;
- 1–morphisms from $G \ltimes X$ to $K \ltimes Y$ are *equivariant generalized maps*

  $$G \ltimes X \xleftarrow{\xi \ltimes \epsilon} L \ltimes Z \xrightarrow{\varphi \ltimes f} K \ltimes Y$$

  such that $\xi \ltimes \epsilon$ is an equivariant essential equivalence;
- a 2–morphism $\Rightarrow$ from the equivariant generalized map $G \ltimes X \xleftarrow{\xi \ltimes \epsilon} L \ltimes Z \xrightarrow{\varphi \ltimes f} K \ltimes Y$ to $G \ltimes X \xleftarrow{\xi' \ltimes \epsilon'} L' \ltimes Z' \xrightarrow{\varphi' \ltimes f'} K \ltimes Y$ is given by a class of diagrams

  $$\begin{tikzcd} & L \ltimes Z \arrow[dl, "\xi \ltimes \epsilon"'] \arrow[dr, "\varphi \ltimes f"] & \\ G \ltimes X \arrow[r, phantom, "\sim_T"] & R \ltimes U \arrow[u, "u"'] \arrow[d, "v"] \arrow[r, phantom, "\sim_{T'}"] & K \ltimes Y \\ & L' \ltimes Z' \arrow[ul, "\xi' \ltimes \epsilon'"] \arrow[ur, "\varphi' \ltimes f'"'] & \end{tikzcd}$$

  where $R \ltimes U$ is a translation groupoid, and u and v are equivariant essential equivalences.

Translation groupoids, equivariant generalized maps and diagrams as above form the *Morita bicategory of translation groupoids*, which we denote by MTrG.

3 Orbifolds as groupoids

We recall now the description of orbifolds as groupoids due to Moerdijk and Pronk [9; 13]. Orbifolds were first introduced by Satake [16] as a generalization of a manifold defined in terms of local quotients. The groupoid approach provides a global language to reformulate the notion of orbifold.

A groupoid $\mathcal{G}$ is *proper* if $(s,t)\colon G_1 \to G_0 \times G_0$ is a proper map and it is a *foliation* groupoid if each isotropy group is discrete.

Definition 3.1 An *orbifold* groupoid is a proper foliation groupoid.

Given an orbifold groupoid $\mathcal{G}$, its orbit space $|\mathcal{G}|$ is a locally compact Hausdorff space. Given an arbitrary locally compact Hausdorff space X we can equip it with an orbifold structure as follows:

Definition 3.2 An *orbifold structure* on a locally compact Hausdorff space X is given by an orbifold groupoid $\mathcal{G}$ and a homeomorphism $h\colon |\mathcal{G}| \to X$.

If $\epsilon\colon \mathcal{H} \to \mathcal{G}$ is an essential equivalence and $|\epsilon|\colon |\mathcal{H}| \to |\mathcal{G}|$ is the induced homeomorphism between orbit spaces, we say that the composition $h \circ |\epsilon|\colon |\mathcal{H}| \to X$ defines an *equivalent* orbifold structure.

Definition 3.3 An *orbifold* $\mathcal{X}$ is a space X equipped with an equivalence class of orbifold structures. A specific such structure, given by $\mathcal{G}$ and $h\colon |\mathcal{G}| \to X$, is a *presentation* of the orbifold $\mathcal{X}$.

If two groupoids are Morita equivalent, then they define the same orbifold. Therefore any structure or invariant for orbifolds, if defined through groupoids, should be invariant under Morita equivalence.

Definition 3.4 An *orbifold map* $f\colon \mathcal{Y} \to \mathcal{X}$ is given by an equivalence class of generalized maps (ϵ, ϕ) from $\mathcal{K}$ to $\mathcal{G}$ between presentations of the orbifolds such that the following diagram commutes:

$$\begin{array}{ccc} |\mathcal{K}| & \xrightarrow{|\phi||\epsilon|^{-1}} & |\mathcal{G}| \\ \downarrow & & \downarrow \\ Y & \longrightarrow & X \end{array}$$

A specific such generalized map (ϵ, ϕ) is called a *presentation* of the orbifold map f.

We can obtain an orbifold by considering the action of a compact group G acting on a space X with finite stabilizers. All orbifolds can be described in this way [12].

The orbifold $\mathcal{X}$ is *developable* if it is presented by a groupoid Morita equivalent to a translation groupoid $G \ltimes X$ with G a discrete group acting properly on X.

4 Path groupoid

From now on, we will focus on developable orbifolds and G will be a discrete group acting properly on X. In this context, we will show that in the bicategory of topological groupoids any path in $G \ltimes X$

$$I \leftarrow \mathcal{I} \rightarrow G \ltimes X$$

is equivalent to a strict map

$$I \rightarrow G \ltimes X$$

where I is the unit groupoid $e \ltimes I$, $I = [0, 1]$ and $\mathcal{I}$ is any topological groupoid.

4.1 Generalized paths

A *path* in the groupoid $G \ltimes X$ in the Morita bicategory of topological groupoids is a generalized map (δ, β) from the unit groupoid I to $G \ltimes X$. That is, a span

$$I \xleftarrow{\delta} \mathcal{I} \xrightarrow{\beta} G \ltimes X.$$

Since $I \xleftarrow{\delta} \mathcal{I}$ is an essential equivalence, we can use groupoid atlases [15; 17] to see that the equivalence class $[I \xleftarrow{\delta} \mathcal{I} \xrightarrow{\beta} G \ltimes X]$ has a representative of the form

$$I \xleftarrow{\epsilon} I_{S_n} \xrightarrow{\alpha} G \ltimes X,$$

where I_{S_n} is the groupoid associated to a subdivision

$$S_n = \{0 = r_0 \leq r_1 < \cdots < r_{n-1} \leq r_n = 1\}$$

of the interval $I = [0, 1]$ as explained below.

The space of objects of the groupoid I_{S_n} is the disjoint union

$$\bigsqcup_{i=1}^{n} I_i,$$

where I_i is a small open neighborhood of $[r_{i-1}, r_i]$ and (r, i) denotes an element r in the connected component I_i.

The space of arrows of I_{S_n} is given by the disjoint union

$$\bigsqcup_{i=1}^{n} I_i \bigsqcup_{i=1}^{n-1} (\tilde{I}_i \sqcup \tilde{I}_i),$$

where $\bigsqcup_{i=1}^{n} I_i$ is the set of unit arrows, $\tilde{I}_i = I_i \cap I_{i+1}$ and another copy $\tilde{I}_i$ was added for inverse arrows. For each point r_i in the subdivision S_n, $\tilde{I}_i$ is an open neighborhood of r_i. Two arrows were added for each point (r,i) in the interval $\tilde{I}_i$: $\tilde{r}_i$ and its inverse arrow such that the source of $\tilde{r}_i$ is (r,i) and its target is $(r,i+1)$.

Definition 4.1 A *generalized path* in the groupoid $G \ltimes X$ is a generalized map $I \xleftarrow{\epsilon} I_{S_n} \xrightarrow{\alpha} G \ltimes X$ such that:

(1) $\epsilon\colon I_{S_n} \to I$ on objects is the inclusion in each connected component, $\epsilon(r,i) = r$ and on arrows it sends all arrows to identity arrows, $\epsilon(\tilde{r}_i) = \mathrm{id}_r$.

(2) $\alpha\colon I_{S_n} \to G \ltimes X$ on objects is given by a map $\alpha_i : I_i \to X$ in each connected component and on arrows is given by $\alpha(\tilde{r}_i) = (k_i, \alpha_i(r))$ satisfying the condition $k_i \alpha_i(r) = \alpha_{i+1}(r)$ for all $r \in \tilde{I}_i$.

We denote by $\mathrm{Map}(I_{S_n}, G \ltimes X)$ this space of maps from I_{S_n} to $G \ltimes X$ with the compact open topology.

4.1.1 Equivalence of generalized paths We will establish now an equivalence relation between the generalized maps defining our generalized paths which will allow us to give a groupoid structure to the space of generalized paths.

Definition 4.2 Two generalized paths $I \xleftarrow{\epsilon} I_{S_m} \xrightarrow{\alpha} G \ltimes X$ and $I \xleftarrow{\epsilon'} I_{S_{m'}} \xrightarrow{\beta} G \ltimes X$ are equivalent if there exist a subdivision S_n and essential equivalences u and v such that the following diagram commutes up to natural transformations:

$$\begin{tikzcd}
 & I_{S_m} \arrow[dl, "\epsilon"'] \arrow[dr, "\alpha"] & \\
I \arrow[r, phantom, "\sim"] & I_{S_n} \arrow[u, "u"'] \arrow[d, "v"] \arrow[r, phantom, "\sim"] & G \ltimes X \\
 & I_{S_{m'}} \arrow[ul, "\epsilon'"] \arrow[ur, "\beta"'] &
\end{tikzcd}$$

Since G is discrete, the condition $\alpha u \sim \beta v$ guarantees the existence of a natural transformation $T\colon \bigsqcup_{i=1}^n I_i \to G \times X$ such that $T(r,i) = (g_i, \alpha_i(r))$ with $\beta_i(r) = g_i \alpha_i(t)$. By naturality of the transformation, the diagram

$$\begin{tikzcd} \alpha_i(r,i) \arrow[r, "g_i"] \arrow[d, "k_i"'] & \beta_i(r,i) \arrow[d, "k_i'"] \\ \alpha_{i+1}(r,i) \arrow[r, "g_{i+1}"] & \beta_{i+1}(r,i) \end{tikzcd}$$

commutes for all $r \in \tilde{I}_i$. Therefore $k_i' = g_{i+1} k_i g_i{}^{-1}$ for all $i = 1, \dots, n-1$.

Remark 4.3 Two generalized paths are equivalent if there exists a common subdivision S_n and $g_i \in G$ such that $\beta_i(r) = g_i \alpha_i(r)$ for all $i = 1, \dots, n$ and $k_i' = g_{i+1} k_i g_i{}^{-1}$ for all $i = 1, \dots, n-1$.

Then, we have a translation groupoid $G^n \ltimes \mathrm{Map}(I_{S_n}, G \ltimes X)$ given by this action of G^n on the space $\mathrm{Map}(I_{S_n}, G \ltimes X)$. Source and target are given by

$$s((g_1, \dots, g_n), (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})) = (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$$

and

$$\begin{aligned} &t((g_1, \dots, g_n), (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})) \\ &\qquad\qquad = (g_1\alpha_1, \dots, g_n\alpha_n, g_2 k_1 g_1^{-1}, \dots, g_n k_{n-1} g_{n-1}^{-1}). \end{aligned}$$

4.1.2 Colimit construction In order to account for all possible subdivisions, we will consider the colimit of the groupoids $G^n \ltimes \mathrm{Map}(I_{S_n}, G \ltimes X)$ over a partially ordered set that we describe next.

We define the category $\mathcal{C}_I$ as the category with objects the ordered tuples

$$S_n = \{0 = r_0 \le r_1 \le \cdots \le r_n = 1\}$$

with an open cover of $I = [0,1]$ given by connected intervals $\{I_i \mid 1 \le i \le n\}$. We require that:

(1) $[r_{i-1}, r_i] \subseteq I_i$ and $I_i \cap \{r_0, r_1, \dots, r_n\} = \{r_{i-1}, r_i\}$, which is one point if $r_{i-1} = r_i$ and two points if $r_{i-1} < r_i$.

(2) (a) If $r_{k-2} < r_{k-1} = r_k = \cdots = r_l < r_{l+1}$ then we require that

$$I_k = I_{k+1} = \cdots = I_l \subseteq I_{k-1} \cap I_{l+1}.$$

(b) If $0 = r_0 = r_1 = \cdots = r_k < r_{k+1}$ then we require that

$$I_1 = I_2 = \cdots = I_k \subseteq I_{k+1}.$$

(c) If $r_{k-1} < r_k = r_{k+1} = \cdots = r_n = 1$ then we require that

$$I_{k+1} = I_{k+2} = \cdots = I_n \subseteq I_k.$$

We have a morphism from $(\{r_0 \le r_1 \le \cdots \le r_n\}, \{I_i\})$ to $(\{t_0 \le t_1 \le \cdots \le t_m\}, \{\tilde{I}_j\})$ if:

(I) $\{r_0, r_1, \ldots, r_n\} \supseteq \{t_0, t_1, \ldots, t_m\}$.

(II) The multiplicity of repeated elements decreases; ie for every i,

$$\bigl|\{j \mid r_j = r_i\}\bigr| \ge \bigl|\{j \mid t_j = r_i\}\bigr|.$$

(III) The open cover $\{I_i\}$ is a refinement of the open cover $\{\tilde{I}_j\}$ in the following way:

(a) For each closed interval $[r_{i-1}, r_i]$ with nonempty interior there is a unique $[t_{j-1}, t_j]$ with $[r_{i-1}, r_i] \subseteq [t_{j-1}, t_j]$ and we have

$$\begin{tikzcd} I_i \arrow[r, "\subseteq"] & \tilde{I}_j \\ {[r_{i-1}, r_i]} \arrow[u, "\subseteq"] \arrow[r, "\subseteq"] & {[t_{j-1}, t_j]} \arrow[u, "\subseteq"'] \end{tikzcd}$$

(b) If there is a repeated element in the $\{t_0 \le t_1 \le \cdots \le t_m\}$, $t_{j-1} = t_j$, it is also a repeated element of $\{r_0 \le r_1 \le \cdots \le r_n\}$, $r_{i-1} = r_i$. We require $I_i \subseteq \tilde{I}_j$.

The morphisms are generated (as a category) by the set of morphisms:

(1) Eliminating a point from the subdivision $\{0 = r_0 \le r_1 \le \cdots \le r_i \le \cdots \le r_n = 1\}$:

$$d_i \colon (\{r_0 \le \cdots \le r_i \le \cdots \le r_n\}, \{I_i\}) \to (\{r_0 \le \cdots \le \hat{r}_i \le \cdots \le r_n\}, \{\tilde{I}_j\}),$$

where d_i drops the i^{th} element and concatenates the consecutive intervals I_i and I_{i+1}, ie $\tilde{I}_j = I_j$ for $j = 0, \ldots, i-1$, $\tilde{I}_i = I_i \cup I_{i+1}$ and $\tilde{I}_j = I_j$ for $j = i+1, \ldots, n$.

(2) Enlarging the intervals without changing the points of the subdivision given by $\{0 = r_0 \le r_1 \le \cdots \le r_n = 1\}$:

$$u \colon (\{r_0 \le \cdots \le r_n\}, \{I_i\}) \to (\{r_0 \le \cdots \le r_n\}, \{\tilde{I}_i\})$$

when $I_i \subseteq \tilde{I}_i$.

We call $\mathcal{C}_I$ the category of subdivisions of I which is a cofiltered category, which boils down to the fact that for two subdivisions there is a common refinement.

For every morphism, there is a continuous map given by concatenation and inclusion

$$\coprod_i I_i \to \coprod_j \tilde{I}_j.$$

To the morphism $d_i\colon S_n \to S_{n-1}$, we assign the functor $d_{i*}\colon I_{S_n} \to I_{S_{n-1}}$ that on objects concatenates $I_i \cup I_{i+1}$ and on morphisms sends $\tilde{r}_i$ and its inverse arrow $\tilde{r}_i'$ to the identity arrow on (r,i). Similarly, for $u\colon S_n \to S_n$, there is a functor $u_*\colon I_{S_n} \to I_{S_n}$ given by inclusion at the level of objects and morphisms. This gives a functor from $\mathcal{C}_I \to \mathsf{Gpd}$. We can obtain a contravariant functor ψ from $\mathcal{C}_I^{\mathrm{op}}$ to topological spaces that on objects sends S_n to $\mathrm{Map}(I_{S_n}, G \ltimes X)$ and on morphisms sends $d_i\colon S_n \to S_{n-1}$ to the morphism $d_i^*\colon \mathrm{Map}(I_{S_{n-1}}, G \ltimes X) \to \mathrm{Map}(I_{S_n}, G \ltimes X)$ given by taking $\alpha \in \mathrm{Map}(I_{S_{n-1}}, G \ltimes X)$ represented by $(\alpha_1, \dots, \alpha_{n-1}, k_1, \dots, k_{n-2})$ and sending it to $(\alpha_1, \dots, \alpha_i|_{I_i}, \alpha_i|_{I_{i+1}}, \dots, \alpha_{n-1}, k_1, \dots, k_{i-1}, \mathrm{id}, k_i, k_{i+1}, \dots, k_{n-2})$, ie taking $\alpha_i\colon I_i \cup I_{i+1} \to X$ to the restrictions to I_i and I_{i+1}. Similarly,

$$u^*\colon \mathrm{Map}(I_{S_n}, G \ltimes X) \to \mathrm{Map}(I_{S_n}, G \ltimes X)$$

is just restriction of all the paths: taking $\alpha \in \mathrm{Map}(I_{S_n}, G \ltimes X)$ represented by $(\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$ and sending it to $(\alpha_1|_{I_1}, \dots, \alpha_n|_{I_n}, k_1, \dots, k_{n-1})$.

We have an action of G^n on the space $\mathrm{Map}(I_{S_n}, G \ltimes X)$ given by

$$(g_1, \dots, g_n) \cdot (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1}) \\ = (g_1\alpha_1, \dots, g_n\alpha_n, g_2 k_1 g_1^{-1}, \dots, g_n k_{n-1} g_{n-1}^{-1}).$$

The map $d_i^*\colon \mathrm{Map}(I_{S_{n-1}}, G \ltimes X) \to \mathrm{Map}(I_{S_n}, G \ltimes X)$ is equivariant with respect to the map $\sigma_i\colon G^{n-1} \to G^n$ given by $\sigma_i(g_1, \dots, g_{n-1}) = (g_1, \dots, g_i, g_i, g_{i+1}, \dots, g_{n-1})$. This means that

$$\sigma_i(g_1, \dots, g_{n-1}) \cdot d_i^*(\alpha_1, \dots, \alpha_{n-1}, k_1, \dots, k_{n-2}) \\ = d_i^*((g_1, \dots, g_{n-1}) \cdot (\alpha_1, \dots, \alpha_{n-1}, k_1, \dots, k_{n-2}))$$

This is because $(g_1, \dots, g_i, g_i, g_{i+1}, \dots, g_{n-1})$ acting on

$$(\alpha_1, \dots, \alpha_i|_{I_i}, \alpha_i|_{I_{i+1}}, \dots, \alpha_{n-1}, k_1, \dots, k_{i-1}, \mathrm{id}, k_i, k_{i+1}, \dots, k_{n-2})$$

is equal in the first part to

$$(g_1\alpha_1, \dots, g_i\alpha_i|_{I_i}, g_i\alpha_i|_{I_{i+1}}, \dots, g_{n-1}\alpha_{n-1})$$

and in the second part to

$$(g_2k_1g_1^{-1},\dots,g_ik_{i-1}g_{i-1}^{-1},g_i\,\mathrm{id}\,g_i^{-1},g_{i+1}k_ig_i^{-1},\dots,g_{n-1}k_{n-2}g_{n-2}^{-1}),$$

which is

$$(g_2k_1g_1^{-1},\dots,g_ik_{i-1}g_{i-1}^{-1},\mathrm{id},g_{i+1}k_ig_i^{-1},\dots,g_{n-1}k_{n-2}g_{n-2}^{-1}).$$

This is precisely

$$d_i^*((g_1,\dots,g_{n-1})\cdot(\alpha_1,\dots,\alpha_{n-1},k_1,\dots,k_{n-2})).$$

Similarly the map $u^*\colon \mathrm{Map}(I_{S_n},G\ltimes X)\to\mathrm{Map}(I_{S_n},G\ltimes X)$ is equivariant with respect to the identity map $G^n\to G^n$.

Therefore we have a contravariant functor from $\mathcal{C}_I$ to the category of translation groupoids that on objects sends S_n to $G^n\ltimes\mathrm{Map}(I_{S_n},G\ltimes X)$ and on morphisms sends $d_i\colon S_n\to S_{n-1}$ to the functor (d_i^*,σ_i) and $u\colon S_n\to S_n$ to the functor (u^*,id); formally we have a (covariant) functor $\Phi\colon\mathcal{C}_I^{\mathrm{op}}\to\mathsf{TrG}$.

We consider now the (filtered) colimit of Φ,

$$P=\operatorname*{colim}_{\mathcal{C}_I^{\mathrm{op}}}\Phi$$

given by an object $P\in\mathsf{TrG}$ together with morphisms from $\mathrm{Map}(I_{S_n},G\ltimes X)$ for each S_n such that for each morphism the following diagrams commute:

For d_i:

$$\begin{tikzcd}
G^n\ltimes\mathrm{Map}(I_{S_n},G\ltimes X) \arrow[d, "\sigma_i\ltimes d_i^*"] \arrow[dr] & \\
& P \\
G^{n-1}\ltimes\mathrm{Map}(I_{S_{n-1}},G\ltimes X) \arrow[ur] &
\end{tikzcd}$$

For u:

$$\begin{tikzcd}
G^n\ltimes\mathrm{Map}(I_{S_n},G\ltimes X) \arrow[d, "\mathrm{id}\ltimes u^*"] \arrow[dr] & \\
& P \\
G^n\ltimes\mathrm{Map}(I_{S_n},G\ltimes X) \arrow[ur] &
\end{tikzcd}$$

Moreover, $P = \operatorname{colim} \Phi$ has the following universal property. Given another translation groupoid W with functors from $G^n \ltimes \operatorname{Map}(I_{S_n}, G \ltimes X)$ that are compatible, such functors factor uniquely through the colimit P as shown in the diagrams

$$\begin{tikzcd}
G^n \ltimes \operatorname{Map}(I_{S_n}, G \ltimes X) \arrow[dd, "\sigma_i \ltimes d_i^*"] \arrow[dr] \arrow[drr, bend left] & & \\
& P \arrow[r, dashed] & W \\
G^{n-1} \ltimes \operatorname{Map}(I_{S_{n-1}}, G \ltimes X) \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}$$

$$\begin{tikzcd}
G^n \ltimes \operatorname{Map}(I_{S_n}, G \ltimes X) \arrow[dd, "\mathrm{id} \ltimes u^*"] \arrow[dr] \arrow[drr, bend left] & & \\
& P \arrow[r, dashed] & W \\
G^{n} \ltimes \operatorname{Map}(I_{S_{n}}, G \ltimes X) \arrow[ur] \arrow[urr, bend right] & &
\end{tikzcd}$$

Definition 4.4 The *path groupoid* $P(G \ltimes X)$ of the translation groupoid $G \ltimes X$ is

$$P(G \ltimes X) = \operatorname*{colim}_{\mathcal{C}_I^{\mathrm{op}}} \Phi,$$

where $\Phi : \mathcal{C}_I^{\mathrm{op}} \to \mathsf{TrG}$ is as above.

We are ready now to give an explicit construction of the groupoid $P = P(G \ltimes X)$ by using the constructions of colimits in the category of topological spaces Top and in the category of groups Grp.

The colimit of the contravariant functor $\psi : \mathcal{C}_I^{\mathrm{op}} \to \mathsf{Top}$ is a topological space $M = \operatorname{colim} \psi$ such that

$$M = \Big(\coprod_{\mathcal{C}_I} \operatorname{Map}(I_{S_n}, G \ltimes X) \Big) \Big/ \sim,$$

where $\sim$ is the equivalence relation generated by $\alpha \sim d_i^*(\alpha)$ for all S_n and $d_i : S_n \to S_{n-1}$ and $\alpha \sim u^*(\alpha)$ for all S_n and $u : S_n \to S_n$.

This topological space $M = \operatorname{colim} \psi$ will be the space of objects of the path groupoid P. To construct the space of arrows of the path groupoid, we consider now a colimit in the category of groups.

Consider the functor $\varphi\colon \mathcal{C}_I^{\mathrm{op}} \to \mathsf{Grp}$ which sends S_n to G^n and on morphisms sends $u\colon S_n \to S_n$ to the identity $G^n \to G^n$ and $d_i : S_n \to S_{n-1}$ to the morphism $\sigma_i : G^{n-1} \to G^n$ given by $\sigma_i(g_1, \dots, g_{n-1}) = (g_1, \dots, g_i, g_i, g_{i+1}, \dots, g_{n-1})$.

The colimit of φ is a group $H = \operatorname{colim} \varphi$ such that

$$H = \Big(\coprod_{\mathcal{C}_I} G^n\Big)\Big/{\sim}$$

where $\sim$ is generated by $(g_1, \dots, g_{n-1}) \sim (g_1, \dots, g_i, g_i, g_{i+1}, \dots, g_{n-1})$. This group H is discrete and acts on the topological space M constructed above.

We can describe now explicitly the object and arrow spaces of the path groupoid $P = P(G \ltimes X)$ in TrG:

$$P_0 = M = \operatorname{colim} \psi = \coprod_{\mathcal{C}_I} \mathrm{Map}(I_{S_n}, G \ltimes X)/{\sim}$$

and

$$P_1 = H \times M = \operatorname{colim} \varphi \times \operatorname{colim} \psi = \bigg(\Big(\coprod_{\mathcal{C}_I} G^n\Big)\Big/{\sim}\bigg) \times \bigg(\Big(\coprod_{\mathcal{C}_I} \mathrm{Map}(I_{S_n}, G \ltimes X)\Big)\Big/{\sim}\bigg),$$

which we endow with the inductive topology.

Remark 4.5 Let G be a discrete group acting on X. The path groupoid of $G \ltimes X$ is the translation groupoid

$$P = P(G \ltimes X) = H \ltimes M.$$

We will show that this path groupoid $P = \operatorname{colim} \Phi$ described above is actually equivalent to the translation groupoid $G \ltimes X^I$. In order to give an explicit characterization of the equivalence of categories, we will introduce some auxiliary groupoids which in turn will relate to the idea introduced in [4] of multiple G–paths.

4.2 Multiple G–paths

We will provide now another description of the path groupoid in terms of equivariant generalized maps. We will see that for each generalized path (ϵ, α), its equivalence class $[I \xleftarrow{\epsilon} I_{S_n} \xrightarrow{\alpha} G \ltimes X]$ contains a representative in MTrG of the form

$$I \xleftarrow{\delta} G \ltimes Y \xrightarrow{\phi} G \ltimes X,$$

where $G \ltimes Y$ is a translation groupoid.

Given a generalized path $I \xleftarrow{\epsilon} I_{S_n} \xrightarrow{\alpha} G \ltimes X$, we will construct a space $Y = Y_\alpha$ such that $G \ltimes Y$ is Morita equivalent to I_{S_n}, and maps $\delta\colon G \ltimes Y \to I$ and $\phi\colon G \ltimes Y \to G \ltimes X$ such that (δ, ϕ) is 2–isomorphic to the given G–path (ϵ, α).

4.2.1 Construction of $G \ltimes Y_\alpha$ Let $\alpha = (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$. Consider the product space

$$G \times (I_{S_n})_0 = \{(g, (r, i)) \mid g \in G, (r, i) \in I_i\}$$

and the identifications, for all $r \in \tilde{I}_i$,

$$(g, (r, i+1)) \sim (k_i^{-1} g, (r, i)),$$

where $\alpha(\tilde{r}_i) = (k_i, \alpha_i(r))$.

Now Y_α is defined as the quotient space

$$\{[(g, (r, i))] \mid (g, (r, i)) \in G \times (I_{S_n})_0 \text{ and } (g, (r, i+1)) \sim (k_i^{-1} g, (r, i)) \text{ for all } r \in \tilde{I}_i\}.$$

Observe that the space Y_α depends on α in the sense that it is given by the subdivision S_n and the group elements $k_1, \dots, k_{n-1}$, but it is independent of the actual pieces of the path $\alpha_1, \dots, \alpha_n$.

The action of G on Y_α is given by the multiplication in the group $h\big([g, (r, i)]\big) = [gh^{-1}, (r, i)]$.

We can consider then the translation groupoid $G \ltimes Y_\alpha$ where the source and target are given by the maps $s\big(h, [g, (r, i)]\big) = [g, (r, i)]$ and $t\big(h, [g, (r, i)]\big) = [gh^{-1}, (r, i)]$.

4.2.2 Morita equivalence $I_{S_n} \sim_M G \ltimes Y_\alpha$ We will show now that the translation groupoid constructed above is Morita equivalent to the groupoid I_{S_n}. Let

$$\nu\colon I_{S_n} \to G \ltimes Y_\alpha$$

be the morphism defined by $\nu((r, i)) = [e, (r, i)]$ on objects and $\nu(\tilde{r}_i) = \big(k_i, [e, (r, i)]\big)$ on arrows for all $r \in \tilde{I}_i$. The open map ν is essentially surjective since

$$s\pi_1\colon G \times (I_{S_n})_0 \to Y_\alpha = (G \times (I_{S_n})_0)/\sim$$

is the quotient projection. It is also fully faithful since $(I_{S_n})_1$ is given by the pullback of the maps

$$\begin{tikzcd} & G \times Y_\alpha \arrow[d, "{(s,t)}"] \\ (I_{S_n})_0 \times (I_{S_n})_0 \arrow[r, "\nu \times \nu"] & Y_\alpha \times Y_\alpha \end{tikzcd}$$

Therefore given a groupoid I_{S_n}, we can construct another groupoid Y_α for each set of elements $k_1, \dots, k_{n-1}$ such that I_{S_n} is Morita equivalent to $G \ltimes Y_\alpha$.

4.2.3 The 2–isomorphism $(\epsilon, \alpha) \Rightarrow (\delta, \phi)$ We will define now the maps δ and ϕ to obtain the generalized map $I \xleftarrow{\delta} G \ltimes Y_\alpha \xrightarrow{\phi} G \ltimes X$ being 2–isomorphic to the given generalized path (ϵ, α).

We define $\phi\big([g, (r, i)]\big) = g^{-1}\alpha_i(r)$ on objects and $\phi(h, [g, (r, i)]) = (h, g^{-1}\alpha_i(r))$ on arrows. Moreover, the morphism ϕ is G–equivariant in the ordinary sense (the group homomorphism is the identity).

The essential equivalence $\delta\colon G \ltimes Y_\alpha \to I$ is given by projection on both objects and arrows, $\delta\big(h, [g, (r, i)]\big) = r$. Both morphisms ϕ and δ are well defined and δ is open, surjective on objects and fully faithful.

The diagram

$$\begin{tikzcd} & I_{S_n} \arrow[dl, "\epsilon"'] \arrow[dr, "\alpha"] \arrow[dd, "\nu"] & \\ I & & G \ltimes X \\ & G \ltimes Y_\alpha \arrow[ul, "\delta"] \arrow[ur, "\phi"'] & \end{tikzcd}$$

is commutative since $\phi\nu((r, i)) = \phi\big([e, (r, i)]\big) = \alpha_i(r_i)$ and $\phi\nu(\tilde{r}_i) = \phi\big(k_i, [e, (r, i)]\big) = (k_i, \alpha_i(r))$ for all $r \in \tilde{I}_i$.

Thus there is a 2–isomorphism between the generalized map $I \xleftarrow{\delta} G \ltimes Y_\alpha \xrightarrow{\phi} G \ltimes X$ and the generalized path $I \xleftarrow{\epsilon} I_{S_n} \xrightarrow{\alpha} G \ltimes X$.

Observe that the identifications we have made in the quotient to obtain the space Y_α determine a gluing of the segments I_i at the different levels of $G \times (I_{S_n})_0$ to obtain copies of the entire interval $I = [0, 1]$. This gluing is determined by the group elements $k_1, \dots, k_{n-1}$.

To define the map ϕ from the groupoid $G \ltimes Y_\alpha$ associated to the generalized path α, we are concatenating the different pieces α_i in these different levels by multiplying by the correct group element to obtain an honest path in X.

4.2.4 The homeomorphism $\gamma: Y_\alpha \to G \times I$ For each map $\alpha\colon I_{S_n} \to G \ltimes X$, let us show now that the space Y_α we just constructed is G–equivariantly homeomorphic to the space $G \times I$, where the action on the latter is determined by the action of G on Y_α given by $h[g,(r,i)] = [gh^{-1},(r,i)]$. We have that the action on $G \times I$ is given by

$$G \times (G \times I) \to G \times I, \quad (h,(g,r)) = (gh^{-1}, r).$$

We define the homeomorphism $\gamma\colon Y_\alpha \to G \times I$ as

$$\gamma\big([g,(r,i)]\big) = ((k_{i-1}\cdots k_1)^{-1} g, r)$$

for $i = 1,\dots,n$. The morphism γ depends only on S_n and $k_1,\dots,k_{n-1}$ and is independent on the actual paths $\alpha_1,\dots,\alpha_n$. The inverse morphism $\gamma^{-1}\colon G \times I \to Y_\alpha$ is given by

$$\gamma^{-1}(h,r) = [k_{i-1}\cdots k_1 h,(r,i)]$$

if $r \in I_i$. Moreover, the homeomorphism γ is G–equivariant by construction.

Definition 4.6 A *multiple G–path* in the groupoid $G \ltimes X$ is a generalized map

$$I \leftarrow G \ltimes (G \times I) \xrightarrow{\sigma} G \ltimes X,$$

where σ is a G–equivariant map in the ordinary sense.

4.2.5 Equivalence of multiple G–paths Given two multiple G–paths

$$I \leftarrow G \ltimes (G \times I) \xrightarrow{\sigma} G \ltimes X \quad \text{and} \quad I \leftarrow G \ltimes (G \times I) \xrightarrow{\tau} G \ltimes X,$$

they are equivalent if there exists a subdivision S_n and $k_1,\dots,k_{n-1} \in G$ such that the diagram

$$\begin{tikzcd}
 & G \ltimes (G \times I) \arrow[dl, "p"'] \arrow[dr, "\sigma"] & \\
I & I_{S_n} \arrow[u, "\nu"'] \arrow[d, "\eta"] & G \times X \\
 & G \ltimes (G \times I) \arrow[ul, "p"] \arrow[ur, "\tau"'] &
\end{tikzcd}$$

commutes up to natural transformations, where $\nu = \nu_{k_1,\dots,k_{n-1}}$ and $\eta = \eta_{k_1,\dots,k_{n-1}}$.

Since p is an essential equivalence, we have that $\nu \sim \eta$ and then $\sigma\nu \sim \tau\nu$. That means that there exists a natural transformation $T\colon (I_{S_n})_0 \to G \times X$ such that $T(r,i)$ is an arrow between $\sigma\nu(r,i) = \sigma((k_{i-1}\cdots k_1)^{-1}, r)$ and $\tau((k_{i-1}\cdots k_1)^{-1}, r)$. Therefore we have that the multiple G–paths are equivalent if there exists a subdivision S_n, $k_1,\dots,k_{n-1} \in G$ and $g_1,\dots,g_n \in G$ such that

$$g_i\sigma((k_{i-1}\cdots k_1)^{-1}, r) = \tau((k_{i-1}\cdots k_1)^{-1}, r) \quad \text{if } r \in I_i.$$

Since σ is equivariant,

$$g_i(k_{i-1}\cdots k_1)\sigma(e,r)=(k_{i-1}\cdots k_1)\tau(e,r)\quad\text{if } r\in I_i.$$

For $i=1$ this means that there exists $g_1\in G$ such that $\tau(e,r)=g_1\sigma(e,r)$. Since the interval $e\times I$ is connected, we have that $g_i=(k_{i-1}\cdots k_1)g_1(k_{i-1}\cdots k_1)^{-1}$ for all $i=1,\dots,n$. In other words, all other g_i for $i=2,\dots,n$ are determined by g_1. Once that we have a group element $g_1\in G$ that makes $\tau(e,r)=g_1\sigma(e,r)$ in the first piece of the interval, $r\in[0,r_1]$, then all the other pieces of the interval coming from the subdivision S_n will also coincide since, for all $r\in I_i$,

$$g_i(k_{i-1}\cdots k_1)\sigma(e,r)=(k_{i-1}\cdots k_1)\tau(e,r)$$

and

$$g_i=(k_{i-1}\cdots k_1)g_1(k_{i-1}\cdots k_1)^{-1}.$$

Then

$$(k_{i-1}\cdots k_1)g_1(k_{i-1}\cdots k_1)^{-1}(k_{i-1}\cdots k_1)\sigma(e,r)=(k_{i-1}\cdots k_1)\tau(e,r),$$

which implies that $g_1\sigma(e,r)=\tau(e,r)$ for all $r\in I$.

Proposition 4.7 *Two multiple G–paths σ and τ are equivalent if there exists $g\in G$ such that*

$$g\sigma(e,r)=\tau(e,r).$$

We have the group G acting now on the space of equivariant maps $G\mathrm{Map}(G\times I,X)$. Let $P'=G\ltimes G\mathrm{Map}(G\times I,X)$ be the *multiple G–path groupoid.*

Since $\sigma(g,r)=g\sigma(e,r)$, we observe that a multiple G–path is determined by the honest path $\beta\colon I\to X$ given by $\beta(r)=\sigma(e,r)$. Conversely, any path $\beta\colon I\to X$ can be made into a multiple G–path by putting $\sigma(g,r)=g\beta(r)$. Consider the translation groupoid of honest paths, given by the obvious action of G on X. Let $P''=G\ltimes X^I$, where $X^I=\mathrm{Map}(I,X)$.

We will prove next that all three characterizations of the path groupoid, as generalized paths, as multiple G–paths and as honest paths are equivalent.

4.3 Equivalence of the different models for path groupoids

Recall the definition of the path groupoid and the other two characterizations introduced in the previous section:

(1) The groupoid $P = \operatorname{colim}\varphi \ltimes \operatorname{colim}\psi$, where $M = \operatorname{colim}\psi$ is the space of classes of generalized paths.

(2) The groupoid $P' = G \ltimes G\mathrm{Map}(G \times I, X)$, where $G\mathrm{Map}(G \times I, X)$ is the space of G–equivariant maps.

(3) The groupoid $P'' = G \ltimes X^I$, where X^I is the free path space.

4.3.1 The equivalence of categories $\chi\colon \operatorname{colim}\varphi \ltimes \operatorname{colim}\psi \to G \ltimes G\mathrm{Map}(G \times I, X)$

Recall that $M = \operatorname{colim}\psi$ is the space of classes of generalized paths, ie

$$M = \Bigl(\coprod_{\mathcal{C}_I} \mathrm{Map}(I_{S_n}, G \ltimes X)\Bigr)\Big/{\sim},$$

where $\sim$ is the equivalence relation generated by $\alpha \sim d_i^*(\alpha)$ for all S_n and $d_i\colon S_n \to S_{n-1}$ and $\alpha \sim u^*(\alpha)$ for all S_n and $u\colon S_n \to S_n$. We will use the same notation,

$$(\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1}),$$

to denote the elements in M.

The idea is to complete each piece α_i of the generalized path

$$\alpha = (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$$

to have the entire branch σ_i of a multiple G–path σ.

The functor $\chi\colon \operatorname{colim}\varphi \ltimes \operatorname{colim}\psi \to G \ltimes G\mathrm{Map}(G \times I, X)$ Given a generalized path $\alpha = (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$ for the subdivision S_n of the interval I, we can define (as in the previous section)

(1) a space $Y_\alpha = \{[(g, (r, i))] \mid (g, (r, i)) \in G \times (I_{S_n})_0\}$ with the relation

$$(g, (r, i+1)) \sim (k_i^{-1} g, (r, i))$$

for all $r \in \tilde{I}_i$,

(2) a homeomorphism $\gamma_\alpha\colon G \ltimes Y_\alpha \to G \ltimes (G \times I)$,

(3) an essential equivalence $\nu_\alpha\colon I_{S_n} \to G \ltimes Y_\alpha$, and

(4) a generalized map $I \xleftarrow{\delta} G \ltimes Y_\alpha \xrightarrow{\phi_\alpha} G \ltimes X$ such that $(\epsilon, \alpha) \Rightarrow (\delta, \phi_\alpha)$.

We define $\chi\colon \operatorname{colim}\varphi \ltimes M \to G \ltimes G\mathrm{Map}(G \times I, X)$ as $\chi(\alpha) = \phi_\alpha {\gamma_\alpha}^{-1}$ on objects and $\chi(g_1, \dots, g_n, \alpha) = (g_1, \phi_\alpha {\gamma_\alpha}^{-1})$ on arrows.

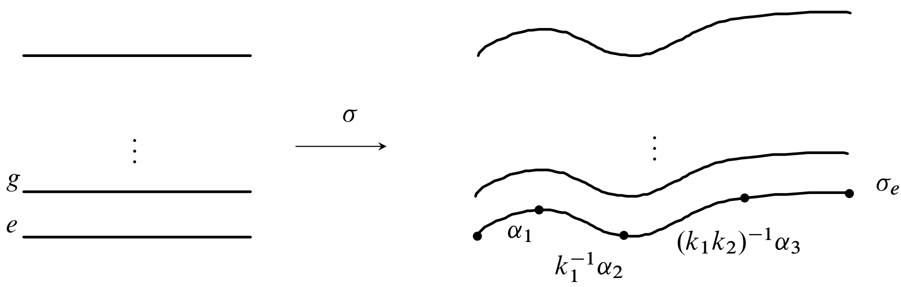


Figure 1

Then $\chi(\alpha)(g,r) = \phi_\alpha {\gamma_\alpha}^{-1}(g,r) = \phi_\alpha[k_{i-1}\cdots k_1 g,(r,i)] = (k_{i-1}\cdots k_1 g)^{-1}\alpha_i(r)$ if $r \in I_i$. We are sending each generalized path $\alpha = (\alpha_1,\cdots,\alpha_n,k_1,\cdots,k_{n-1})$ into the multiple G–path σ given by

$$\sigma(g,r) = g^{-1}(k_{i-1}\cdots k_1)^{-1}\alpha_i(r) \quad \text{if } r \in I_i.$$

In particular, we have that the branch σ_e corresponding to the interval $e \times I$ is given by the concatenation (see also Figure 1)

$$\sigma(e,r) = \alpha_1(r) * {k_1}^{-1}\alpha_2(r) * (k_2k_1)^{-1}\alpha_3(r) * \cdots * (k_{n-1}\cdots k_1)^{-1}\alpha_n(r).$$

On arrows, we send $((g_1,\cdots,g_n),\alpha_1,\cdots,\alpha_n,k_1,\cdots,k_{n-1}) \in \operatorname{colim}\varphi \times \operatorname{colim}\psi$ into the arrow (g_1,σ_α), where σ_α is defined as before.

We will show next that χ is an equivariant map between translation groupoids where the group homomorphism is given by the projection on the first coordinate.

Let $\alpha' = (g_1\alpha_1,\ldots,g_n\alpha_n, g_2k_1{g_1}^{-1},\ldots,g_nk_{n-1}{g_{n-1}}^{-1})$. We have that

$$\chi(\alpha') = g_1\chi((\alpha_1,\ldots,\alpha_n,k_1,\cdots,k_{n-1}))$$

since

$$\begin{aligned}
&\sigma_{\alpha'}(e,r)\\
&\quad= g_1\alpha_1(r)*(g_2k_1{g_1}^{-1})^{-1}g_2\alpha_2(r)*\cdots*(g_nk_{n-1}{g_{n-1}}^{-1}\cdots g_2k_1{g_1}^{-1})^{-1}g_n\alpha_n(r)\\
&\quad= g_1(\alpha_1(r)*{k_1}^{-1}\alpha_2(r)*(k_2k_1)^{-1}\alpha_3(r)*\cdots*(k_{n-1}\cdots k_1)^{-1}\alpha_n(r) = g_1\sigma_\alpha(e,r).
\end{aligned}$$

The functor $\chi^{-1}\colon G \ltimes G\mathrm{Map}(G\times I, X) \to \operatorname{colim}\varphi \ltimes \operatorname{colim}\psi$ Consider the continuous functor given by $\chi^{-1}(\sigma) = \sigma|_{e\times I}\circ i_e$ on objects and $\chi^{-1}((g,\sigma)) = (g,\sigma|_{e\times I}\circ i_e)$ on arrows, where $i_e\colon I \to e\times I$ sends $r \in I$ to $(e,r) \in e\times I$. Recall that by our notation convention the right side means in both cases the class in the colimit. Note that the generalized path $\sigma|_{e\times I}\circ i_e$ corresponds to a subdivision S_1 with only one subinterval; that is, $\sigma|_{e\times I}\circ i_e$ is an honest path.

The functors χ and χ^{-1} are inverse up to natural transformation The composition $\chi \circ \chi^{-1}\colon G \ltimes G\mathrm{Map}(G \times I, X) \to G \ltimes G\mathrm{Map}(G \times I, X)$ is the identity map since the generalized map α_σ associated to σ has only one piece. On objects,

$$\chi \circ \chi^{-1}(\sigma) = \chi(\alpha_\sigma) = \sigma_{\alpha_\sigma}$$

such that $\sigma_{\alpha_\sigma}(g, r) = g^{-1}\sigma(e, r) = \sigma(g, r)$, so $\sigma_{\alpha_\sigma} = \sigma$. On arrows,

$$\chi \circ \chi^{-1}(g, \sigma) = \chi(g, \sigma_{\alpha_\sigma}) = \chi(g, \sigma) = (g, \sigma).$$

We will prove next that the composition in the other direction is equivalent by a natural transformation to the identity. We have that

$$\chi^{-1} \circ \chi\colon \mathrm{colim}\,\varphi \ltimes \mathrm{colim}\,\psi \to \mathrm{colim}\,\varphi \ltimes \mathrm{colim}\,\psi$$

sends each generalized path class $\alpha = (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$ to the generalized path α_{σ_α}, where

$$\alpha_{\sigma_\alpha}(r) = \sigma_\alpha(e, r) = \alpha_1(r) * {k_1}^{-1}\alpha_2(r) * (k_2k_1)^{-1}\alpha_3(r) * \dots * (k_{n-1}\cdots k_1)^{-1}\alpha_n(r),$$

and each arrow $((g_1, \dots, g_n), \alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1}) \in \mathrm{colim}\,\varphi \times \mathrm{colim}\,\psi$ to the arrow $(g_1, \alpha_{\sigma_\alpha})$.

There is a natural transformation $T\colon \mathrm{colim}\,\psi \to \mathrm{colim}\,\varphi \times \mathrm{colim}\,\psi$ given by

$$T(\alpha) = ((\mathrm{id}, {k_1}^{-1}, (k_2k_1)^{-1}, \dots, (k_{n-1}\cdots k_1)^{-1}), (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1}))$$

which is an arrow between α and α_{σ_α} since

$$\begin{aligned}
&(\mathrm{id}, {k_1}^{-1}, (k_2k_1)^{-1}, \dots, (k_{n-1}\cdots k_1)^{-1})(\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1}) \\
&= \big((\mathrm{id}\,\alpha_1, {k_1}^{-1}\alpha_2, \dots, (k_{n-1}\cdots k_1)^{-1}\alpha_n), ({k_1}^{-1}k_1, \dots, (k_{n-1}\cdots k_1)^{-1}k_{n-1}(k_{n-2}\cdots k_1))\big) \\
&\qquad \big((\alpha_1, {k_1}^{-1}\alpha_2, \dots, (k_{n-1}\cdots k_1)^{-1}\alpha_n), (\mathrm{id}, \dots, \mathrm{id})\big).
\end{aligned}$$

This generalized path is equal to the concatenation of the n pieces

$$\alpha_1(r) * {k_1}^{-1}\alpha_2(r) * (k_2k_1)^{-1}\alpha_3(r) * \dots * (k_{n-1}\cdots k_1)^{-1}\alpha_n(r)$$

since the connecting arrows are all identities. Moreover, T satisfies the naturality condition and is continuous by the universal property of the colimit.

Therefore χ is an equivalence of categories between the groupoid of generalized paths and the groupoid of multiple G–paths. We will see next that the groupoid of multiple G–paths is just the free path space X^I together with G acting on it.

4.3.2 The isomorphism of categories $\xi\colon G \ltimes G\mathrm{Map}(G \times I, X) \to G \ltimes X^I$ To construct this isomorphism we will use the fact that a multiple G–path σ is determined by its value at the branch σ_e corresponding to the interval $e \times I$, since σ is equivariant.

We define $\xi(\sigma) = \sigma i_e \in X^I$ on objects and $\xi(g, \sigma) = (g, \sigma i_e)$ on arrows. Conversely, $\xi^{-1}(\beta) = \sigma_\beta$ where $\sigma_\beta(g, r) = g^{-1}\beta(r)$. The functor

$$\xi\colon G \ltimes G\mathrm{Map}(G \times I, X) \to G \ltimes X^I$$

is an isomorphism of categories since it has a strict inverse functor, $\xi \circ \xi^{-1} = \mathrm{id}_{G \ltimes X^I}$ and $\xi^{-1} \circ \xi = \mathrm{id}_{G \ltimes G\mathrm{Map}(G \times I, X)}$, satisfying that the restriction ξ and the action ξ^{-1} are both continuous.

Theorem 4.8 *All models for the path groupoid of $G \ltimes X$ are equivalent*;

$$P(G \ltimes X) = \operatorname{colim} \varphi \ltimes \operatorname{colim} \psi \sim G \ltimes G\mathrm{Map}(G \times I, X) = G \ltimes X^I.$$

Remark 4.9 We can also prove that any generalized map is equivalent to a strict map in the context of translation groupoids, without using groupoids atlases. It was proven by Pronk and Scull in [14] that any generalized map

$$I \xleftarrow{\delta} \mathcal{I} \xrightarrow{\beta} G \ltimes X$$

between translation groupoids is equivalent to a generalized map

$$I \xleftarrow{\epsilon} G \ltimes Y \xrightarrow{\beta'} G \ltimes X,$$

where the middle groupoid is a translation groupoid. In the same paper, they proved that any essential equivalence between translation groupoids has to be of some prescribed form. In our case, this implies that the essential equivalence $I \xleftarrow{\epsilon} G \ltimes Y$ satisfies $e = G/K$ and $I = Y/K$ where K acts freely on Y. Hence $K = G$ and G acts freely on Y. Since G acts also properly on Y we have that $Y = G \times I$. Then any generalized map (δ, β) is equivalent to a generalized map

$$I \leftarrow G \ltimes (G \times I) \xrightarrow{\beta''} G \ltimes X.$$

Now, applying our isomorphism $\xi\colon P' \to P''$ to the right leg of the span, we obtain $\alpha = \xi(\beta'') \in X^I$ which gives the equivalent strict map $I \xrightarrow{\alpha} G \ltimes X$.

4.4 Functoriality and Morita invariance of the path groupoid

In this section we will see that the path groupoid is functorial and that the path groupoid is well defined up to Morita equivalence.

4.4.1 Functoriality We will show that an equivariant map between translation groupoids induces an equivariant map between the path groupoids.

For a strict equivariant map $\varphi \ltimes f\colon G \ltimes X \to H \ltimes Y$, we have an induced map $\varphi_* \ltimes f_*\colon G \ltimes X^I \to H \ltimes Y^I$ defined by $f_*(\alpha) = f \circ \alpha$ for all $\alpha \in X^I$ and $\varphi_* = \varphi$. We construct now an equivariant map $P(\varphi \ltimes f)\colon P(G \ltimes X) \to P(H \ltimes Y)$ between the colimit constructions.

For every n we have induced maps

$$(\varphi \ltimes f)_*\colon \mathrm{Map}(I_{S_n}, G \ltimes X) \to \mathrm{Map}(I_{S_n}, H \ltimes Y)$$

in terms of the description $\mathrm{Map}(I_{S_n}, G \ltimes X) = G^n \ltimes (X^I)^n \times_{X^{n-1}} G^{n-1}$; this map corresponds just to $\varphi^n \ltimes (f_*^{n-1} \times \varphi^n)$. By taking the colimit we obtain an equivariant map $P(\varphi \ltimes f)\colon P(G \ltimes X) \to P(H \ltimes Y)$.

Similarly, we have an induced map

$$\varphi_* \ltimes f_*\colon G \ltimes G\mathrm{Map}(G \times I, X) \to H \ltimes H\mathrm{Map}(H \times I, Y)$$

between the multiple G–path groupoids. We consider an equivariant map $(G \times I) \xrightarrow{\sigma} X$ and define $f_*(\sigma)\colon H \times I \to Y$ by $f_*(\sigma)(h,r) = h^{-1} f(\sigma(e,r))$.

In any of the three models the functoriality is easy to check and we have the following theorem.

Theorem 4.10 *The path groupoid of $G \ltimes X$ is functorial for equivariant maps.*

Moreover, the equivalence of the three models for the path groupoid is natural.

Theorem 4.11 *For a strict equivariant map $\varphi \ltimes f\colon G \ltimes X \to H \ltimes Y$ the following diagram is commutative*:

$$\begin{tikzcd}
P(G \ltimes X) \arrow[r, "P(\varphi\ltimes f)"] \arrow[d, "\chi"] & P(H \ltimes Y) \arrow[d, "\chi"] \\
G \ltimes G\mathrm{Map}(G \times I, X) \arrow[r, "\varphi_*\ltimes f_*"] \arrow[d, "\xi"] & H \ltimes H\mathrm{Map}(H \times I, Y) \arrow[d, "\xi"] \\
G \ltimes X^I \arrow[r, "\varphi_*\ltimes f_*"] & H \ltimes Y^I
\end{tikzcd}$$

4.4.2 Morita invariance We will start by proving that an essential equivalence $G \times X \to H \times Y$ induces an essential equivalence between the path groupoids,

$$P(G \ltimes X) \to P(H \ltimes Y).$$

This will give that for a given Morita equivalence

$$G \ltimes X \xleftarrow{\sigma} G' \ltimes X' \xrightarrow{\epsilon} H \ltimes Y,$$

where ϵ and σ are essential equivalences, we have induced essential equivalences

$$P(G \ltimes X) \xleftarrow{P(\sigma)} P(G' \ltimes X') \xrightarrow{P(\epsilon)} P(H \ltimes Y).$$

Proposition 4.12 *If $\varphi \ltimes f : G \ltimes X \to H \ltimes Y$ is an essential equivalence, then $P(\varphi \ltimes f)$ is an essential equivalence.*

Proof (1) $P(\varphi \ltimes f)$ is fully faithful. We will show that $P(G \ltimes X)_1$ is the pullback of topological spaces

$$\begin{tikzcd}
P(G \ltimes X)_1 \arrow[r, "\epsilon"] \arrow[d, "{(s,t)}"'] & P(H \ltimes Y)_1 \arrow[d, "{(s,t)}"] \\
P(G \ltimes X)_0 \times P(G \ltimes X)_0 \arrow[r, "\epsilon\times\epsilon"] & P(H \ltimes Y)_0 \times P(H \ltimes Y)_0
\end{tikzcd}$$

Specifically we have to prove that the natural map ξ from $P(G \ltimes X)_1$ to the fibered product $P(G \ltimes X)_0 \times P(G \ltimes X)_0 \times_{P(H\ltimes Y)_0 \times P(H\ltimes Y)_0)} P(H \ltimes Y)_1$ is a homeomorphism.

Let us define the inverse map ξ^{-1}. For $\alpha, \beta \in P(G \ltimes X)_0$ and an element $\theta \in P(H \ltimes Y)_1$ with $s(\theta) = t(\theta) = \alpha$, we can assume that there is a subdivision of the interval such that α and β are represented both by elements of $\mathrm{Map}(I_{S_n}, G \ltimes X)$ and θ by an element of H^n. Therefore we have $\alpha = (\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1})$ and $\beta = (\beta_1, \dots, \beta_n, k'_1, \dots, k'_{n-1})$ such that $f(\beta_i(r)) = h_i f(\alpha_i(r))$ for all $i = 1, \dots, n$ and $k'_i = h_{i+1} k_i {h_i}^{-1}$ for all $i = 1, \dots, n-1$.

But then by fixing r and using that $\varphi \ltimes f$ is an essential equivalence, we have a fibered product of topological spaces

$$\begin{tikzcd}
(G \ltimes X)_1 \arrow[r, "\epsilon"] \arrow[d, "{(s,t)}"'] & (H \ltimes Y)_1 \arrow[d, "{(s,t)}"] \\
(G \ltimes X)_0 \times (G \ltimes X)_0 \arrow[r, "\epsilon\times\epsilon"] & (H \ltimes Y)_0 \times (H \ltimes Y)_0
\end{tikzcd}$$

and therefore for every $r \in I_i$ there is $g_i^r \in G$ such that $\phi(g_i^r) = h_i$. Since G is discrete and the dependence on r is continuous, the n–tuple $(g_1^r, \dots, g_n^r)$ actually does not depend on r and represents an element of $P(G \ltimes X)_1$.

(2) $P(\varphi \ltimes f)$ is essentially surjective. We will show that

$$s\pi_1\colon P(H \ltimes Y)_1 \times^t_{P(H\ltimes Y)_0} P(G \ltimes X)_0 \to P(H \ltimes Y)_0$$

is an open surjection.

For étale groupoids, the condition that the morphism

$$s\pi_1\colon (H \ltimes Y)_1 \times^t_{(H\ltimes Y)_0} (G \ltimes X)_0 \to (H \ltimes Y)_0$$

is an open surjection implies that it has local sections. We will use these local sections to construct local sections of $s\pi_1\colon P(H \ltimes Y)_1 \times^t_{P(H\ltimes Y)_0} P(G \ltimes X)_0 \to P(H \ltimes Y)_0$.

Let $\{U_\alpha\}_{\alpha\in\Delta}$ be a cover of Y and $s_\alpha\colon U_\alpha \to (H \ltimes Y)_1 \times^t_{(H\ltimes Y)_0} (G \ltimes X)_0$ the local sections. Take $\gamma \in P(H \ltimes Y)_0$ and suppose that γ is represented by an element of $\mathrm{Map}(I_{S_n}, H \ltimes Y)$; therefore $\gamma = (\gamma_1, \dots, \gamma_n, k_1, \dots, k_{n-1})$ with $\gamma_i\colon I_i \to Y$.

Given the subdivision $\{0 = r_0 \le r_1 \le \cdots \le r_n = 1\}$ associated to γ and with an open cover of $[0,1]$ given by connected intervals $\{I_i \mid 1 \le i \le n\}$, we want to construct a refinement of the subdivision

$$\{0 = r_0 \le s^1_0 \le \cdots \le s^1_{m_1} = r_1 \le \cdots \le r_{i-1} = s^i_0 \le \cdots \le s^i_{m_i} = r_i \le \cdots \le r_n = 1\}$$

along with connected intervals I^i_j with the property that $\gamma_i(I^i_j) \subseteq U_{\alpha^i_j}$ for some α^i_j. To construct the subdivision, first for $[r_{i-1}, r_i]$ with nonempty interior, we consider the covering $\{I_i \cap U_\alpha\}_{\alpha\in\Delta}$. By compactness of the interval $[r_{i-1}, r_i]$ we can find a partition $r_{i-1} = s^i_0 < \cdots < s^i_{m_i} = r_i$ such that each $\gamma_i([s^i_{j-1}, s^i_j])$ is contained in some $U_{\alpha^i_j}$. Let I^i_j be an open connected neighborhood of $[s^i_{j-1}, s^i_j]$ small enough such that $I^i_j \cap \{s^i_0, s^i_1, \dots, s^i_{m_i}\} = \{s^i_{j-1}, s^i_j\}$.

For the repeated elements $r_{i-1} = r_i$, it is a matter of just shrinking the interval I_i to get $\gamma_i(I_i) \subseteq U_{\alpha_i}$ for some α_i and to obtain an object of the category of subdivisions.

With the local sections $s_{\alpha^i_j}$ we obtain maps $\pi_2 s_{\alpha^i_j}(\gamma_i(r))\colon I^i_j \to X$ and functions $\pi_1\pi_1 s_{\alpha^i_j}(\gamma_i(r))\colon I^i_j \to H$, since the intervals are connected and H is a discrete group, actually these functions are constant and we have elements $h^i_j \in H$ with $s\pi_1(f(\pi_2 s_{\alpha^i_j}(\gamma_i(r))), h^i_j) = \gamma_i(r)$, ie

$$f(\pi_2 s_{\alpha^i_j}(\gamma_i(r))) = h^i_j \gamma_i(r)$$

for $r \in I^i_j$.

Note that $(h^i_j)^{-1} f(\pi_2 s_{\alpha^i_j}(\gamma_i(s^i_j))) = (h^i_{j+1})^{-1} f(\pi_2 s_{\alpha^i_{j+1}}(\gamma_i(s^i_j)))$ (both are $\gamma_i(s^i_j)$) and therefore

$$h^i_{j+1}(h^i_j)^{-1} f(\pi_2 s_{\alpha^i_j}(\gamma_i(s^i_j))) = f(\pi_2 s_{\alpha^i_{j+1}}(\gamma_i(s^i_j))).$$

Since f is full and faithful, there is a $g^i_j \in G$ with $\phi(g^i_j) = h^i_{j+1}(h^i_j)^{-1}$ such that

$$g^i_j \pi_2 s_{\alpha^i_j}(\gamma_i(s^i_j)) = \pi_2 s_{\alpha^i_{j+1}}(\gamma_i(s^i_j)).$$

Similarly at the intersection points of two consecutive paths we have $k_i\gamma_i(r) = \gamma_{i+1}(r)$ for all $r \in \tilde{I}^i_j$ and therefore

$$k_i(h^i_{m_i})^{-1} f(\pi_2 s_{\alpha^i_{m_i}}(\gamma_i(r_i)) = k_i\gamma_i(r_i) = \gamma_{i+1}(r_i) = (h^{i+1}_0)^{-1} f(\pi_2 s_{\alpha^{i+1}_0}(\gamma_{i+1}(r_i)).$$

Then

$$h^{i+1}_0 k_i(h^i_{m_i})^{-1} f(\pi_2 s_{\alpha^i_{m_i}}(\gamma_i(r_i)) = f(\pi_2 s_{\alpha^{i+1}_0}(\gamma_{i+1}(r_i)),$$

and since f is full and faithful, we have elements $g^i \in G$ with $\phi(g^i) = h^{i+1}_0 k_i(h^i_{m_i})^{-1}$ such that

$$g^i \pi_2 s_{\alpha^i_{m_i}}(\gamma_i(r_i)) = \pi_2 s_{\alpha^{i+1}_0}(\gamma_{i+1}(r_i)).$$

Therefore we have a generalized path

$$\big((\pi_2 s_{\alpha^i_j}(\gamma_i(r)))_{i,j}, g^1_1, g^1_2, \dots, g^1_{m_1}, g^1, g^2_1, \dots, g^n_{m_n}\big)$$

and elements $(h^1_1, h^1_2, \dots, h^1_{m_1-1}, \dots, h^n_{m_n-1})$ of H. By construction,

$$(h^1_1, h^1_2, \dots, h^1_{m_1-1}, \dots, h^n_{m_n-1})\big((\pi_2 s_{\alpha^i_j}(\gamma_i(r)))_{i,j}, g^1_1, g^1_2, \dots, g^1_{m_1}, g^1, g^2_1, \dots, g^n_{m_n}\big)$$

is

$$(\gamma_1|_{I^1_1}, \gamma_1|_{I^1_2} \dots, \gamma_1|_{I^1_{m_1}}, \dots, \gamma_n|_{I^n_{mn}}, \mathrm{id}, \mathrm{id}, \dots, k_1, \mathrm{id}, \dots, k_n).$$

In the colimit this represents the same element as $(\gamma_1, \dots, \gamma_n, k_1, \dots, k_{n-1})$. Therefore we have constructed local sections on the set

$$\big\{(\gamma_1, \dots, \gamma_n, k_1, \dots, k_{n-1}) \in \mathrm{Map}(I_{S_n}, H \ltimes Y) \mid \gamma_i([s^i_{j-1}, s^i_j]) \subseteq U_{\alpha^i_j}\big\},$$

which is an open set in the compact open topology of $\mathrm{Map}(I_{S_n}, H \ltimes Y)$. □

Thus, we have proved that the path groupoid functor sends essential equivalences to essential equivalences and therefore the path groupoid is invariant under Morita equivalence.

Theorem 4.13 *If $G \ltimes X \sim_M H \ltimes Y$, then $P(G \ltimes X) \sim_M P(H \ltimes Y)$.*

4.5 The free loop groupoid $L(G \ltimes X)$

We use the model of the path groupoid $P'' = G \ltimes X^I$ to define the loop groupoid as the following pullback along the diagonal:

$$\begin{tikzcd} & G \ltimes X^I \arrow[d, "\mathrm{ev}"] \\ \Delta\colon G \ltimes X \arrow[r] & (G \times G) \ltimes (X \times X) \end{tikzcd} \tag{1}$$

Definition 4.14 The *free loop groupoid* $L(G \ltimes X)$ of a translation groupoid $G \ltimes X$ is

$$L(G \ltimes X) = (G \times G) \ltimes L_0,$$

where

$$L_0 = \{(\beta, h, l) \in X^I \times G \times G \,|\, \beta(0) = hl^{-1}\beta(1)\}$$

and the group $G \times G$ acts on L_0 by $(a,b)(\beta,h,l) = (a\beta, bha^{-1}, bla^{-1})$.

Figure 2 depicts an arrow $(a,b) \in G \times G$ from (β,h,l) to $(a\beta, bha^{-1}, bla^{-1})$.

We will show that this groupoid $(G \times G) \ltimes L_0$ is Morita equivalent to the translation groupoid $G \ltimes L$ where

$$L = \{(\alpha, g) \in X^I \times G \,|\, \alpha(0) = g\alpha(1)\}$$

and the action is given by $(\alpha, g) \sim (k\alpha, kgk^{-1})$. Figure 3 depicts an arrow $(k, (\alpha, g))$ between (α, g) and $(k\alpha, kgk^{-1})$.

Proposition 4.15 *If $G \ltimes X$ is a topological groupoid, then the loop groupoid $L(G \ltimes X)$ is Morita equivalent to $G \ltimes L$, where L and the action are defined above.*

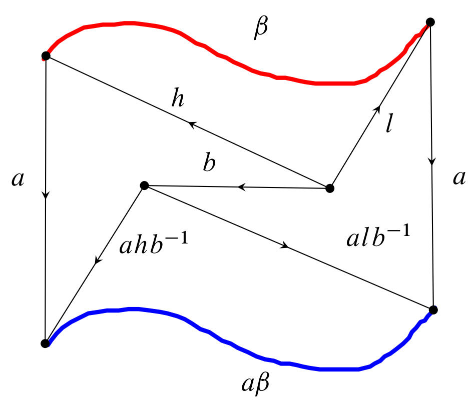


Figure 2

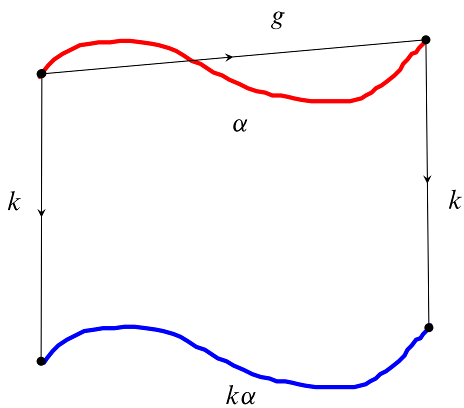


Figure 3

Proof We define an equivariant map $\phi \ltimes \epsilon \colon (G \times G) \ltimes L_0 \to G \ltimes L$ by $\phi((a,b)) = a$ and $\epsilon((\beta, h, l)) = (\beta, l^{-1}h)$. This map is an essential equivalence since the map

$$s\pi_1 \colon G \times L_0 \to L$$

given by $s\pi_1(k, (\beta, h, l)) = (k^{-1}\beta, k^{-1}l^{-1}hk)$ is an open surjection and $G \times G \times L_0$ is given by the pullback of the maps

$$\begin{tikzcd} & G \times L \arrow[d, "{(s,t)}"] \\ L_0 \times L_0 \arrow[r, "\epsilon \times \epsilon"] & L \times L \end{tikzcd}$$

□

Remark 4.16 We can use our description for the free loop groupoid in the special case of the point groupoid. We obtain that $L(G \ltimes \bullet) = (G \times G) \ltimes (G \times G)$ with the action $(a,b) \cdot (h,l) = (bha^{-1}, bla^{-1})$. This groupoid is equivalent to G acting on itself by conjugation by using the second characterization of the loop groupoid as $G \ltimes L$ with $L = \{(\beta, g) \in X^I \times G \mid \beta(0) = g\beta(1)\}$. In this way, we recover a result of Lupercio and Uribe in [7]. Observe that $L(G \ltimes \bullet) = G \ltimes G$, whereas $P(G \ltimes \bullet) = G \ltimes \bullet$.

5 Based path and loop groupoids

Now that we have defined the free path groupoid of a translation groupoid and have given several equivalent models, we can give an explicit characterization of the various groupoids resulting from fixing certain points. These based groupoids of paths will be of great significance to the groupoid Lusternik–Schnirelmann theory defined in [3] and further studied in [2].

5.1 The groupoid $\Omega_{x,y}$ of paths from x to y

The groupoid of paths from x to y, $\Omega_{x,y}$, is defined as a pullback of the evaluation map $\mathrm{ev}\colon P(G \ltimes X) \to (G \times G) \ltimes (X \times X)$ and the constant map $x \times y\colon \mathbf{1} \to (G \times G) \ltimes (X \times X)$, where $\mathbf{1}$ is the trivial groupoid with one object and one arrow, ie $\mathbf{1} = e \ltimes \bullet$, and $(x \times y)(\bullet) = (x, y)$. That is,

$$\begin{tikzcd} \Omega_{x,y} \arrow[r] \arrow[d] & P(G \ltimes X) \arrow[d, "\mathrm{ev}"] \\ \mathbf{1} \arrow[r, "x \times y"] & (G \times G) \ltimes (X \times X) \end{tikzcd}$$

Note that by the definition of groupoid pullback, we have that if we take the model of the path groupoid of generalized paths, $P = \operatorname{colim} \phi \ltimes \operatorname{colim} \psi$, then the object space of the pullback is

$$\{((\alpha_1, \dots, \alpha_n, k_1, \dots, k_{n-1}), h, l) \in \operatorname{colim} \psi \times (G \times G) \mid \alpha_1(0) = hx \text{ and } \alpha_n(1) = ly\},$$

ie the objects of $\Omega_{x,y}$ are sequences of paths and arrows $(h, \alpha_1, k_1, \dots, k_n, \alpha_n, l)$ where $s(k_i) = \alpha_{i+1}(r_i)$ for $i = 1, \dots, n-1$, $t(k_i) = \alpha_i(r_i)$ for $i = 0, \dots, n$, $s(h) = x$ and $s(l) = y$; which are precisely the Haefliger G–paths [6] when restricted to the closed intervals in the subdivision. Note that the sequences in Haefliger paths start and end with *arrows* and not with *paths* like our generalized paths in the free path groupoid defined in Section 4. We recover the original sequence in the Haefliger G–paths when we fix the endpoints x and y in our free generalized paths.

For an equivalent characterization of the groupoid of paths from x to y, we can consider our simplest model for the path groupoid $P'' = G \ltimes X^I$. In this case, we describe the space of objects as $(\Omega_{x,y})_0 = \{(\beta, h, l) \in X^I \times (G \times G) \mid \beta(0) = hx \text{ and } \beta(1) = ly\}$.

These are paths that start at any point in the orbit of x and end at any point in the orbit of y. The space of arrows is the Cartesian product $G \times (\Omega_{x,y})_0$ where the action is given by $g(\beta, h, l) = (g\beta, gh, gl)$; see Figure 4.

Since $(\beta, h, l) \sim (g\beta, gh, gl)$ for all $g \in G$, we can consider $g = h^{-1}$ and we have that all classes $[(\beta, h, l)]$ have a representative of the form (α, k) with $\alpha = h^{-1}\beta$ and $k = h^{-1}l$. Then we can consider the space of objects

$$P_{x,y} = \{(\alpha, k) \in X^I \times G \mid \alpha(0) = x \text{ and } \alpha(1) = ky\}.$$

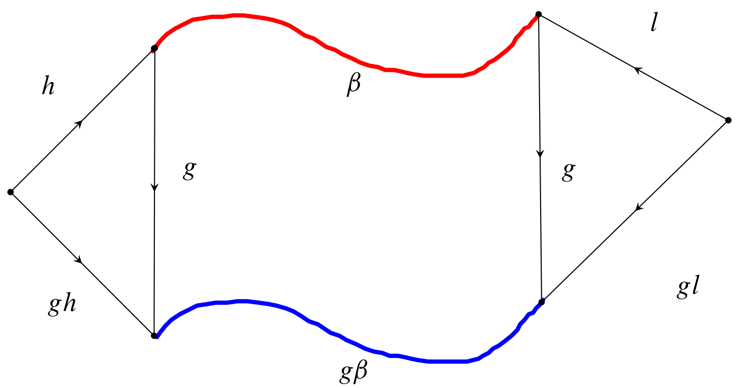


Figure 4

Observe that

$$(\alpha,k)\sim(h^{-1}\beta,e,h^{-1}l)\sim(gh^{-1}\beta,ge,gh^{-1}l)=(g\alpha,g,gk)\sim(e\alpha,e,k)\sim(\alpha,k),$$

so the action is trivial on the space of objects $P_{x,y}$.

Therefore the *groupoid of paths between x and y* is the translation groupoid $\Omega_{x,y}=G\ltimes(\Omega_{x,y})_0$ which is equivalent to the topological space $P_{x,y}$.

5.2 The groupoid Ω_x of based loops

Similarly, we define the based loop groupoid as the groupoid pullback,

$$\begin{tikzcd}
\Omega_x \arrow[r] \arrow[d] & P(G\ltimes X) \arrow[d, "\mathrm{ev}"] \\
\mathbf{1} \arrow[r, "x\times x"] & G\times G\ltimes(X\times X)
\end{tikzcd}$$

where $x\times x$ is the constant map with $(x\times x)(\bullet)=(x,x)$.

That is, the *based loop groupoid* is the translation groupoid $\Omega_x=G\ltimes(\Omega_x)_0$ where the object space is

$$(\Omega_x)_0=\{(\beta,h,l)\in X^I\times(G\times G)\mid\beta(0)=hx\text{ and }\beta(1)=lx\},$$

ie the space of paths that begin and end at (possibly different) points in the orbit of x. The action is given by $g(\beta,h,l)=(g\beta,gh,gl)$; see Figure 5.

Again, the groupoid Ω_x is equivalent to the topological space

$$P_{x,x}=\{(\alpha,k)\mid\alpha(0)=x\text{ and }\alpha(1)=kx\}.$$

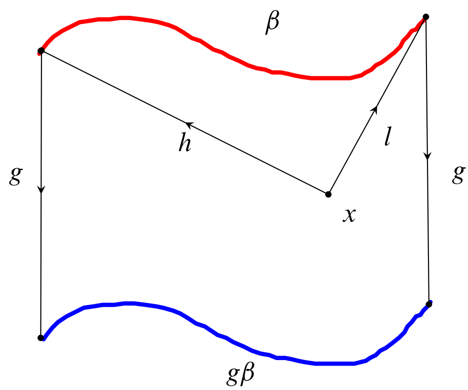


Figure 5

Alternatively, the based loop groupoid Ω_x can be obtained as the groupoid pullback

$$\begin{tikzcd} \Omega_x \arrow[r] \arrow[d] & L(G \ltimes X) \arrow[d, "\mathrm{ev}_0"] \\ \mathbf{1} \arrow[r, "x"] & G \ltimes X \end{tikzcd}$$

where $L(G \ltimes X)$ is the free loop groupoid.

5.3 The groupoid P_x of paths from x

We define the x–based path groupoid as the groupoid pullback

$$\begin{tikzcd} P_x \arrow[r] \arrow[d] & P(G \ltimes X) \arrow[d, "\mathrm{ev}"] \\ \mathbf{1} \times (G \ltimes X) \arrow[r, "(x,\mathrm{id})"] & (G \times G) \ltimes (X \times X) \end{tikzcd}$$

where $(x,\mathrm{id})\colon \mathbf{1} \times (G \ltimes X) \to (G \times G) \ltimes (X \times X)$ is given by $(x,\mathrm{id})(\bullet, z) = (x, z)$. Then the object space of the pullback P_x is

$$\begin{aligned} (P_x)_0 &= \{(\beta, (h,l), (\bullet, z)) \in X^I \times G \times G \times \mathbf{1} \times X \mid \beta(0) = hx \text{ and } \beta(1) = lz\} \\ &= \{(\beta, (h,l), z) \mid \beta(0) = hx \text{ and } \beta(1) = lz\}. \end{aligned}$$

The group $G \times G$ acts on $(P_x)_0$ by $(g,k)(\beta, (h,l), z) = (g\beta, (gh, glk^{-1}), kz)$; see Figure 6.

The *x–based path groupoid* is the translation groupoid $P_x = (G \times G) \ltimes (P_x)_0$.

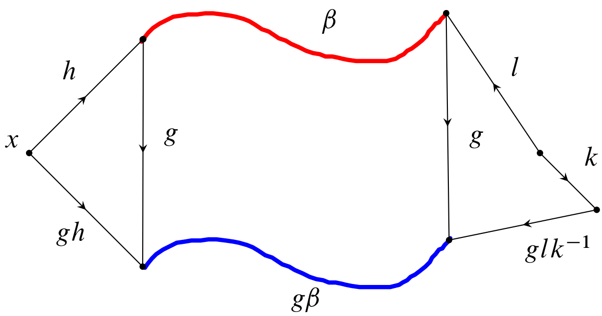


Figure 6

Observing that the equivalence class of each $(\beta,(h,l),z)\in(P_x)_0$ contains an element of the form $(\alpha,g,w)\in X^I\times G\times X$ we have that the based path groupoid P_x is equivalent to $G\ltimes P$ where $P=\{(\alpha,g,w)\mid \alpha(0)=x \text{ and } \alpha(1)=gw\}$ and the action is given by $k(\alpha,g,w)=(\alpha,gk^{-1},kw)$. Figure 7 depicts an arrow $(k,(\alpha,g,w))\in G\times P$ between (α,g,w) and (α,gk^{-1},kw).

The x–based path groupoid P_x is not in general equivalent to a topological space.

Given points $x,y\in X$, our various path groupoids are related by

$$\begin{tikzcd}
 & \Omega_{x,y} \arrow[rr] \arrow[d] & & \mathbf{1_y} \arrow[d] \\
\Omega_x \arrow[r] & P_x \arrow[r] \arrow[d] & G\ltimes X^I \arrow[r, "\mathrm{ev}_1"] \arrow[d, "\mathrm{ev}_0"] & G\ltimes X \\
 & \mathbf{1_x} \arrow[r] & G\ltimes X &
\end{tikzcd}$$

where $\mathbf{1_x}=e\ltimes x$ and $\mathbf{1_y}=e\ltimes y$ and all diagrams are commutative up to a natural transformation.

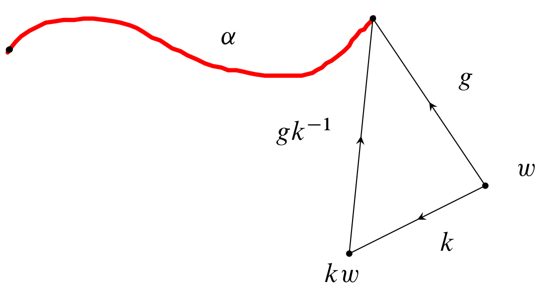


Figure 7

5.4 Examples

We will illustrate in this section the concepts described in the previous sections by calculating various path groupoids in some particular cases.

5.4.1 Topological spaces The free path groupoid of the topological space X is $P(e \ltimes X) = e \ltimes X^I = X^I$ and the free loop groupoid is $L(e \ltimes X) = e \ltimes L$ where $L = \{\alpha \in X^I \mid \alpha(0) = \alpha(1)\}$. In this way we recover the classical free path and loop spaces of a topological space. Likewise, the based path and loop groupoids also coincide with the classical ones for topological spaces.

5.4.2 Groups For a point groupoid $G \ltimes \bullet$ we have shown before that the path groupoid is itself and the loop groupoid is $(G \times G) \ltimes (G \times G)$ with the action $(a, b) \cdot (h, l) = (bha^{-1}, bla^{-1})$, which is equivalent to G acting on itself by conjugation; that is, $L(G \ltimes \bullet) = G \ltimes G$ and $P(G \ltimes \bullet) = G \ltimes \bullet$. The based loop groupoid is the unit groupoid G, as a discrete topological space. The based path groupoid of paths emanating from $\bullet$ is $G \ltimes G$.

5.4.3 Free actions If G acts freely on a topological space X, we observe that the groupoid $G \ltimes X$ and the topological space X/G are Morita equivalent. Then, we have that $P(G \ltimes X) = P(e \ltimes X/G) = e \ltimes (X/G)^I = (X/G)^I$ and the free loop groupoid is $L(G \ltimes X) = L(X/G)$ where $L(X/G)$ is the free loop space of the topological space X/G. In the same way, we have that the based groupoids coincide with the ones of the topological space X/G.

5.4.4 Orbifolds We proved that for developable orbifolds $G \ltimes X$, the free path groupoid is $P(G \ltimes X) = G \ltimes X^I$ and the free loop groupoid is $L(G \ltimes X) = G \ltimes L$ where $L = \{(\alpha, g) \in X^I \times G \mid \alpha(0) = g\alpha(1)\}$. Also, the groupoid of paths between x and y is the topological space $P_{x,y} = \{(\alpha, k) \in X^I \times G \mid \alpha(0) = x \text{ and } \alpha(1) = ky\}$, the groupoid of based loops is the topological space $P_{x,x} = \{(\alpha, k) \mid \alpha(0) = x \text{ and } \alpha(1) = kx\}$ and the groupoid of based paths from x is the translation groupoid $P_x = (G \times G) \ltimes (P_x)_0$.

6 Homotopy

We will define in this section a notion of homotopy based on the explicit description of the path groupoid $P(G \ltimes X)$ given in the previous section. This will provide a concrete alternative to the more abstract presentation given by Noohi in [10; 11] for stacks.

6.1 Natural transformations for translation groupoids

The equivariant maps $\varphi \ltimes f\colon K \ltimes Z \to G \ltimes X$ and $\psi \ltimes g\colon K \ltimes Z \to G \ltimes X$ are equivalent by a natural transformation if there exists a K–map $\gamma\colon Z \to G$ such that $\gamma(z)f(z) = g(z)$ for all $z \in Z$ where both Z and G are K–spaces considering the action of K on G,

$$K \times G \to G, \quad (k, g) \mapsto \psi(k)g\varphi(k)^{-1}.$$

Therefore $\varphi \ltimes f \sim \psi \ltimes g$ if there exists $\gamma\colon Z \to G$ such that

(1) $\gamma(z)f(z) = g(z)$ for all $z \in Z$, and

(2) $\gamma(kz) = \psi(k)\gamma(z)\varphi(k)^{-1}$ for all $k \in K$.

If Z is connected, then γ is a constant map since G is discrete. Then $\varphi \ltimes f \sim \psi \ltimes g$ if there exists $h \in G$ such that $hf(z) = g(z)$ for all $z \in Z$ and $h = \psi(k)h\varphi(k)^{-1}$ for all $k \in K$. Then $g = hf$ and $\psi = h\varphi h^{-1}$. In other words, $\psi(k)$ is conjugated to $\varphi(k)$ for all $k \in K$.

In addition, if G is abelian, then $\varphi \ltimes f \sim \psi \ltimes g$ if $g = hf$ for some $h \in G$ and $\varphi = \psi$.

If $X = Z = \bullet$, then $\varphi \ltimes \bullet \sim \psi \ltimes \bullet$ if and only if φ and ψ are conjugate, $\varphi = h^{-1}\psi h$. In particular, when the group acting is abelian we have that two maps between point groupoids are equivalent by a natural transformation only if they are equal.

We give now a characterization of 2–isomorphism for strict maps. Namely, if two strict maps are 2–isomorphic then when composed with an essential equivalence they are equivalent by a natural transformation, and if two strict maps are equivalent by a natural transformation then they are 2–isomorphic as generalized maps.

Proposition 6.1 *If f and g are equivalent by a natural transformation, then $f \Rightarrow g$ as generalized equivariant maps.*

Proof Just consider the essential equivalences η and ν as identity maps and the following diagram is commutative up to natural transformations since $f \sim g$:

$$\begin{tikzcd}
 & G \ltimes X \arrow[dl, "\mathrm{id}"'] \arrow[dr, "f"] & \\
G \ltimes X \arrow[r, phantom, "\sim"] & G \ltimes X \arrow[u, "\mathrm{id}"'] \arrow[d, "\mathrm{id}"] \arrow[r, phantom, "\sim"] & H \ltimes Y \\
 & G \ltimes X \arrow[ul, "\mathrm{id}"] \arrow[ur, "g"'] &
\end{tikzcd}$$

□

Proposition 6.2 *If two strict maps $f\colon G\ltimes X\to H\ltimes Y$ and $g\colon G\ltimes X\to H\ltimes Y$ are 2–isomorphic, then there exists an essential equivalence $\eta\colon\mathcal{L}\to G\ltimes X$ such that $f\eta\sim g\eta$.*

Proof We have that there exist essential equivalences η,ν such that the diagram

$$\begin{tikzcd} & G\ltimes X \arrow[dl, "\mathrm{id}"'] \arrow[dr, "f"] & \\ G\ltimes X & \mathcal{L} \arrow[u, "\eta"'] \arrow[d, "\nu"] \arrow[l, phantom, "\sim"] \arrow[r, phantom, "\sim"] & H\ltimes Y \\ & G\ltimes X \arrow[ul, "\mathrm{id}"] \arrow[ur, "g"'] & \end{tikzcd}$$

commutes up to natural transformation. That is, $\eta\sim\nu$ and $f\eta\sim g\nu$. Therefore, $f\eta\sim g\eta$. □

Proposition 6.3 *If $(\epsilon,f)\Rightarrow(\sigma,g)$, then there exist essential equivalences ν and η such that $f\nu\Rightarrow g\eta$.*

Proof By definition of 2–isomorphism, there are essential equivalences ν and η such that $f\nu\sim g\eta$. The result follows from Proposition 6.1. □

Proposition 6.4 *If $f\Rightarrow g$, then $(\epsilon,f)\Rightarrow(\sigma,g)$ for all essential equivalences ϵ and σ with $\epsilon\sim\sigma$.*

6.2 Diagonal map

We will consider the pullback of the unique morphism $G\ltimes X\xrightarrow{c}\mathbf{1}$ with itself, where $\mathbf{1}$ is the terminal object in MTopG. This pullback defines the product and then by the universal property we obtain the definition of the diagonal map. Then, the path groupoid will be a factorization of that diagonal.

Definition 6.5 [5] An object T in a bicategory $\boldsymbol{B}$ is terminal if the category $\boldsymbol{B}[C,T]$ is equivalent to the terminal category for every object C in $\boldsymbol{B}$. A terminal object is unique up to equivalence when it exists.

The trivial groupoid $\mathbf{1}=e\ltimes\bullet$ is the terminal object in the bicategory of translation groupoids MTrG since the category MTrG$[G\ltimes X,\mathbf{1}]$ is equivalent to the category $\mathbf{1}$.

Indeed, the objects in the category $\mathsf{MTrG}[G \ltimes X, \mathbf{1}]$ are generalized maps and the arrows are classes of diagrams. We can see that all objects are related by an arrow, ie $\mathsf{MTrG}[G \ltimes X, \mathbf{1}]$ is the pair groupoid. Given two generalized maps,

$$G \ltimes X \xleftarrow{\epsilon'} G' \ltimes X' \xrightarrow{c'} \mathbf{1} \quad \text{and} \quad G \ltimes X \xleftarrow{\epsilon''} G'' \ltimes X'' \xrightarrow{c''} \mathbf{1},$$

we can see that they are equivalent, ie

$$\begin{tikzcd}
 & G' \ltimes X' \arrow[dl, "\epsilon'"'] \arrow[dr] & \\
G \ltimes X \arrow[r, phantom, "\sim"] & P \arrow[u] \arrow[d] \arrow[r, phantom, "\sim"] & \mathbf{1} \\
 & G'' \ltimes X'' \arrow[ul, "\epsilon''"] \arrow[ur] &
\end{tikzcd}$$

by considering P as the pullback of ϵ' and ϵ''. In particular, the strict constant map $G \ltimes X \xrightarrow{c} \mathbf{1}$ is the (unique up to 2–isomorphism) map to the terminal object.

Let us now consider the pullback of this constant map with itself which defines the product

$$\begin{tikzcd}
G \times G \ltimes (X \times X) \arrow[r] \arrow[d] & G \ltimes X \arrow[d, "c"] \\
G \ltimes X \arrow[r, "c"'] & \mathbf{1}
\end{tikzcd}$$

The product $(G \times G) \ltimes (X \times X)$ of the object $G \ltimes X$ with itself is unique up to equivalence.

By the universal property of the pullback, there exists a map Δ that makes the two triangles commutative up to natural transformation

$$\begin{tikzcd}
G \ltimes X \arrow[dr, dashed, "\Delta"] \arrow[drr, bend left, "\mathrm{id}"] \arrow[ddr, bend right, "\mathrm{id}"'] & & \\
 & (G \times G) \ltimes (X \times X) \arrow[r, "p_1"] \arrow[d, "p_2"] & G \ltimes X \arrow[d, "c"] \\
 & G \ltimes X \arrow[r, "c"] & \mathbf{1}
\end{tikzcd}$$

The map $\Delta\colon G \ltimes X \to (G \times G) \ltimes (X \times X)$ is the *diagonal map*. Its explicit definition on objects is $\Delta(x) = (x, x)$ and on arrows, $\Delta(g, x) = (g, g, x, x)$. The diagonal map is defined up to 2–isomorphism.

Remark 6.6 The diagonal defined in [1] is 2–isomorphic to this one.

Definition 6.7 The *evaluation map* $\mathrm{ev}\colon G \ltimes X^I \to (G \times G) \ltimes (X \times X)$ is given by $\mathrm{ev}(g,\alpha) = (g, g, \alpha(0), \alpha(1))$.

We have that the diagonal map factors through the path groupoid as expected.

Proposition 6.8 *There is a factorization of the diagonal map* Δ

$$\begin{tikzcd} G \ltimes X \arrow[rr, "\Delta"] \arrow[dr, "k"'] & & (G \times G) \ltimes (X \times X) \\ & G \ltimes X^I \arrow[ur, "e"'] & \end{tikzcd}$$

where k *and* e *are generalized maps.*

Proof Let k be the functor $G \ltimes X \to G \ltimes X^I$ given by $x \rightsquigarrow \alpha_x$ on objects, and $(g,x) \rightsquigarrow (g,\alpha_x)$, where $\alpha_x\colon I \to X$ is a constant path at $x \in X$, and let e be the evaluation map, $e = \mathrm{ev}$. Then, we have that the composition $e \circ c$ is equivalent by a natural transformation to the diagonal Δ. □

6.3 Homotopic maps

We will now give an explicit characterization of the homotopy between generalized maps.

Definition 6.9 Two generalized maps,

$$K \ltimes Y \xleftarrow{\sigma} K' \ltimes Y' \xrightarrow{f} G \ltimes X \quad \text{and} \quad K \ltimes Y \xleftarrow{\tau} K'' \ltimes Y'' \xrightarrow{g} G \ltimes X,$$

are *homotopic* if there is a generalized map $K \ltimes Y \xleftarrow{\epsilon} \widetilde{K} \ltimes \widetilde{Y} \xrightarrow{H} G \ltimes X$ such that the following diagram commutes up to 2–isomorphism:

$$\begin{tikzcd} G \ltimes X & & G \ltimes X^I \arrow[ll, "\mathrm{ev}_0"'] \arrow[rr, "\mathrm{ev}_1"] & & G \ltimes X \\ & K' \ltimes Y' \arrow[ul, "f"] \arrow[dr, "\sigma"'] & \widetilde{K} \ltimes \widetilde{Y} \arrow[u, "H"'] \arrow[d, "\epsilon"] & K'' \ltimes Y'' \arrow[ur, "g"'] \arrow[dl, "\tau"] & \\ & & K \ltimes Y & & \end{tikzcd}$$

This means that the generalized map (σ, f) is isomorphic to the generalized map $(\epsilon, \mathrm{ev}_0 \circ H)$ and (τ, g) is isomorphic to $(\epsilon, \mathrm{ev}_1 \circ H)$.

That is, (σ, f) is homotopic to (τ, g) if there exists (ϵ, H) and two commutative diagrams up to natural transformations,

$$
\begin{tikzcd}
 & \widetilde{K} \ltimes \widetilde{Y} \arrow[dl, "\epsilon"'] \arrow[dr, "\mathrm{ev}_0\, H"] & \\
\mathcal{K} \ltimes Y \arrow[r, phantom, "\sim"] & \mathcal{L}_0 \arrow[u, "u_0"'] \arrow[d, "v_0"] \arrow[r, phantom, "\sim"] & G \ltimes X \\
 & K' \ltimes Y' \arrow[ul, "\sigma"] \arrow[ur, "f"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \widetilde{K} \ltimes \widetilde{Y} \arrow[dl, "\epsilon"'] \arrow[dr, "\mathrm{ev}_1\, H"] & \\
\mathcal{K} \ltimes Y \arrow[r, phantom, "\sim"] & \mathcal{L}_1 \arrow[u, "u_1"'] \arrow[d, "v_1"] \arrow[r, phantom, "\sim"] & G \ltimes X \\
 & K'' \ltimes Y'' \arrow[ul, "\tau"] \arrow[ur, "g"'] &
\end{tikzcd}
$$

where $\mathcal{L}_i$ is a translation groupoid, and u_i and v_i are equivariant essential equivalences for $i = 0, 1$. We will denote this homotopy between equivariant generalized maps by $\simeq$.

Remark 6.10 $(\sigma, f) \simeq (\tau, g)$ if there exists (ϵ, H) and essential equivalences u_0, u_1, v_0 and v_1 such that

$$f v_0 \sim \mathrm{ev}_0\, H u_0 \quad \text{and} \quad g v_1 \sim \mathrm{ev}_1\, H u_1$$

with $\epsilon u_0 \sim \sigma v_0$ and $\epsilon u_1 \sim \tau v_1$.

Proposition 6.11 *If* $(\sigma, f) \Rightarrow (\tau, g)$, *then* $(\sigma, f) \simeq (\tau, g)$.

Proof Consider $H = i_X \circ f$ where i_X is the inclusion of X in X^I given by $i_X(x) = \alpha_x$ with α_x being the constant map $\alpha_x(t) = x$ for all $t \in I$. Then the following diagram is commutative up to 2–isomorphism:

$$
\begin{tikzcd}
G \ltimes X & & G \ltimes X^I \arrow[ll, "\mathrm{ev}_0"'] \arrow[rr, "\mathrm{ev}_1"] & & G \ltimes X \\
 & K' \ltimes Y' \arrow[ul, "f"] \arrow[dr, "\sigma"'] & K' \ltimes Y' \arrow[u, "H"'] \arrow[d, "\sigma"] & K'' \ltimes Y'' \arrow[ur, "g"'] \arrow[dl, "\tau"] & \\
 & & K \ltimes Y & &
\end{tikzcd}
$$

The first triangle is an equality and the second is commutative since $(\sigma, f) \Rightarrow (\tau, g)$. $\square$

Remark 6.12 Let f and g be strict maps. Following the characterization for isomorphism of strict maps given in Proposition 6.2 and the definition of groupoid homotopy, we have that $f \simeq g$ if there exists a generalized map (ϵ, H) and essential equivalences η and ν such that $f\epsilon\eta \sim \mathrm{ev}_0\, H\eta$ and $g\epsilon\nu \sim \mathrm{ev}_1\, H\nu$.

Proposition 6.13 *Let f and g be strict maps.*

1. *If f and g are ψ–equivariantly homotopic maps, then $f \simeq g$ as generalized equivariant maps.*
2. *If f and g are equivalent by a natural transformation, then $f \simeq g$ as generalized equivariant maps.*

Proof (1) Let $H : Y \to X^I$ be the ψ–equivariant homotopy, ie $H_t(ky) = \psi(k) H_t(y)$. Then the following diagram is commutative:

$$\begin{tikzcd} G \ltimes X & G \ltimes X^I \arrow[l, "\mathrm{ev}_0"'] \arrow[r, "\mathrm{ev}_1"] & G \ltimes X \\ & K \ltimes Y \arrow[ul, "f"] \arrow[u, "H"'] \arrow[ur, "g"'] & \end{tikzcd}$$

(2) This follows from Propositions 6.1 and 6.11. □

Therefore our definition of homotopy generalizes both the notion of natural transformation and the notion of equivariant homotopy.

Proposition 6.14 *If $(\epsilon, f) \simeq (\sigma, g)$ then there exist essential equivalences a and b such that $fa \simeq gb$ as strict maps.*

Proof Since we have a homotopy between generalized maps, we know that there exists (δ, H) and essential equivalences u_0, v_0, u_1 and v_1 such that

$$f v_0 \sim \mathrm{ev}_0\, H u_0, \quad g v_1 \sim \mathrm{ev}_1\, H u_1, \quad \delta u_0 \sim \epsilon v_0, \quad \delta u_1 \sim \sigma v_1.$$

Take $a = v_0(u_0)^{-1}\delta^{-1}$ and $b = v_1(u_1)^{-1}\delta^{-1}$. Then fa and gb are homotopic. □

Proposition 6.15 *The path groupoid $G \ltimes X^I$ is homotopy equivalent to the groupoid $G \ltimes X$. The evaluation $e_1 : G \ltimes X^I \to G \ltimes X$ is a homotopy equivalence.*

Proof Consider the map $H : G \ltimes X^I \to G \ltimes (X^I)^I$ such that $H(\alpha) = \lambda$ with

$$\lambda : I \to X^I, \quad \lambda(t) = \alpha(r + t - rt).$$

We have the commutative diagram

$$\begin{tikzcd} G \ltimes X^I & G \ltimes (X^I)^I \arrow[l, "\mathrm{ev}_0"'] \arrow[r, "\mathrm{ev}_1"] & G \ltimes X^I \\ & G \ltimes X^I \arrow[ul, "\mathrm{id}"] \arrow[u, "H"'] \arrow[ur, "i \circ e_1"'] & \end{tikzcd}$$

showing that $i \circ e_1$ is homotopic to the identity map. □

7 Fibrations

We recall the definition of fibration for topological spaces given as a dualization of the notion of cofibration.

Definition 7.1 [8; 18] A map $p\colon E \to B$ is a fibration if for all spaces U with $\mathrm{ev}_0 \circ K = p \circ k$ in the diagram

$$\begin{tikzcd}
U \arrow[dr, dashed, "\widetilde{K}"] \arrow[drr, bend left, "K"] \arrow[ddr, bend right, "k"'] & & \\
& E^I \arrow[r, "p_*"] \arrow[d, "\mathrm{ev}_0"] & B^I \arrow[d, "\mathrm{ev}_0"] \\
& E \arrow[r, "p"] & B
\end{tikzcd}$$

there exists $\widetilde{K}$ that makes the diagram commute.

We want to introduce a notion of fibration for generalized maps. First, let us note that a strict equivariant map $\varphi \ltimes f : G \ltimes X \to H \ltimes Y$ induces a map $\varphi_* \ltimes f_* : G \ltimes X^I \to H \ltimes Y^I$ by $f_*(\alpha) = f \circ \alpha$ for all $\alpha \in X^I$ and $\varphi_* = \varphi$. We proved in Proposition 4.12 that if $\epsilon\colon G \ltimes X \to H \ltimes Y$ is an essential equivalence, then $\epsilon_*\colon G \ltimes X^I \to H \ltimes Y^I$ is an essential equivalence as well.

Then every generalized map $G \ltimes X \xleftarrow{\epsilon} G' \ltimes X' \xrightarrow{f} H \ltimes Y$ induces a generalized map $G \ltimes X^I \xleftarrow{\epsilon_*} G' \ltimes X'^I \xrightarrow{f_*} H \ltimes Y^I$ between the path groupoids.

Definition 7.2 A generalized map $G \ltimes X \xleftarrow{\epsilon} G' \ltimes X' \xrightarrow{f} H \ltimes Y$ is a *groupoid fibration* if for all translation groupoids $L \ltimes U$ with $\mathrm{ev}_0 \circ (\Omega, K) \Rightarrow (\omega, k) \circ (\epsilon, f)$ in the diagram

$$\begin{tikzcd}
L \ltimes U & & & & \\
& \widetilde{\mathcal{L}} \arrow[ul, dashed, "\widetilde{\Omega}"'] \arrow[dr, dashed, "\widetilde{K}"] & \mathcal{L} \arrow[ull, bend right, "\Omega"'] \arrow[drr, bend left, "K"] & & \\
& \ell \arrow[uul, bend left, "\omega"] \arrow[dr, bend right, "k"'] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] & G' \ltimes X'^I \arrow[l, "\epsilon_*"'] \arrow[r, "f_*"] \arrow[d, "\mathrm{ev}_0"] & H \ltimes Y^I \arrow[d, "\mathrm{ev}_0"] \\
& & G \ltimes X & G' \ltimes X' \arrow[l, "\epsilon"'] \arrow[r, "f"] & H \ltimes Y
\end{tikzcd}$$

there exists $(\widetilde{\Omega}, \widetilde{K})$ that makes the diagram commute up to 2–isomorphism.

Since a 2–isomorphism between strict maps induces a 2–isomorphism between the induced maps between their path groupoids, being a fibration is a property invariant under 2–isomorphism.

Proposition 7.3 *Consider 2–isomorphic maps*

$$f\colon G \ltimes X \to H \ltimes Y \quad \text{and} \quad g\colon G \ltimes X \to H \ltimes Y,$$

$f \Rightarrow g$*. Then* f *is a fibration if and only if* g *is a fibration.*

We will see that for (ϵ, f) to be a groupoid fibration it is necessary and sufficient that the right leg of the span is a groupoid fibration (considered as a generalized map with identity as a left leg).

Proposition 7.4 *A generalized map* $G \ltimes X \xleftarrow{\epsilon} G' \ltimes X' \xrightarrow{f} H \ltimes Y$ *is a groupoid fibration if and only if* $f\colon G' \ltimes X' \to H \ltimes Y$ *is a groupoid fibration.*

Proof If the generalized map (ϵ, f) is a groupoid fibration, then there exists $(\tau', \widetilde{H}')$ that makes the diagram

$$\begin{tikzcd}
L \ltimes U & & \mathcal{L} \arrow[ll, "\sigma"'] \arrow[ddrr, "H"] & & \\
& \widetilde{\mathcal{L}} \arrow[ul, "\tau'"] \arrow[dr, "\widetilde{H}'"] & P \arrow[ull, dashed, "\tau"'] \arrow[dr, dashed, "\widetilde{H}"] & & \\
& & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] & G' \ltimes X'^I \arrow[l, "\epsilon_*"] \arrow[r, "f_*"] \arrow[d, "\mathrm{ev}_0"] & H \ltimes Y^I \arrow[d, "\mathrm{ev}_0"] \\
& \ell \arrow[uuul, "\sigma_0"] \arrow[r, "\epsilon H_0"] \arrow[rr, bend right, "H_0"'] & G \ltimes X & G' \ltimes X' \arrow[l, "\epsilon"] \arrow[r, "f"] & H \ltimes Y
\end{tikzcd}$$

commute up to 2–isomorphism.

Let P be the pullback

$$\begin{tikzcd}
P \arrow[r, "\widetilde{H}''"] \arrow[d, "\epsilon'_*"] & G' \ltimes X'^I \arrow[d, "\epsilon_*"] \\
\widetilde{\mathcal{L}} \arrow[r, "\widetilde{H}'"] & G \ltimes X^I
\end{tikzcd}$$

Take $\tau = \tau' \epsilon'_*$ and $\widetilde{H} = \widetilde{H}''$. Then $f\colon G' \ltimes X' \to H \ltimes Y$ is a groupoid fibration.

Conversely, if f is a groupoid fibration then we have this commutative diagram

$$
\begin{tikzcd}
L \ltimes U & & \mathcal{L} \arrow[ll, "\sigma"'] \arrow[ddrr, "H"] & & \\
& & P \arrow[ull, "\tau"'] \arrow[d, dashed, "\widetilde{H}"] \arrow[dr, "\widetilde{H}'"] & & \\
& \ell \arrow[uul, "\sigma_0"] \arrow[dr, "H_0"] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] & G' \ltimes X'^I \arrow[l, "\epsilon_*"'] \arrow[r, "f_*"] \arrow[d, "\mathrm{ev}_0"] & H \ltimes Y^I \arrow[d, "\mathrm{ev}_0"] \\
& \ell' \arrow[uuul, bend left, "\sigma_0'"] \arrow[rr, bend right, "H_0'"'] & G \ltimes X & G' \ltimes X' \arrow[l, "\epsilon"'] \arrow[r, "f"] & H \ltimes Y
\end{tikzcd}
$$

where $(\sigma_0', H_0') = \epsilon \circ (\sigma_0, H_0)$. Now take $\widetilde{H} = \epsilon_* \widetilde{H}'$ and $\tau = \tau'$. Therefore, (ϵ, f) is a fibration. $\square$

Then, the test to decide if a generalized map is a groupoid fibration amounts to check the definition of groupoid fibration with a strict map. Moreover, we know that any generalized map $L \ltimes U \leftarrow \ell \rightarrow G \ltimes X$ is equivalent to a generalized map of the form $L \ltimes U \leftarrow L' \ltimes U' \rightarrow G \ltimes X$, where L' may be chosen as $L \times G$ and the group homomorphisms are the appropriate projections onto L and G [14].

The groupoid fibration definition specializes to the following:

Definition 7.5 A strict map $f\colon G \ltimes X \rightarrow H \ltimes Y$ is a groupoid fibration if for all translation groupoids $L \ltimes U$ with $\mathrm{ev}_0 \circ (\Omega, K) \Rightarrow f \circ (\omega, k)$ in the diagram

$$
\begin{tikzcd}
L \ltimes U & & & \\
& \widetilde{\mathcal{L}} \arrow[ul, dashed, "\widetilde{\Omega}"'] \arrow[dr, dashed, "\widetilde{K}"] & L'' \ltimes U'' \arrow[ull, bend right, "\Omega"'] \arrow[dr, bend left, "K"] & \\
& L' \ltimes U' \arrow[uul, bend left, "\omega"] \arrow[dr, bend right, "k"] & G \ltimes X^I \arrow[r, "f_*"] \arrow[d, "\mathrm{ev}_0"] & H \ltimes Y^I \arrow[d, "\mathrm{ev}_0"] \\
& & G \ltimes X \arrow[r, "f"] & H \ltimes Y
\end{tikzcd}
$$

there exists $(\widetilde{\Omega}, \widetilde{K})$ that makes the diagram commute up to 2–isomorphism.

In other words, f is a groupoid fibration if for all commutative diagrams

$$\begin{tikzcd}
 & & (L'')\ltimes U'' \arrow[dll, "\Omega"'] \arrow[drr, "\mathrm{ev}_0\circ K"] & & \\
L\ltimes U & \sim & {} \arrow[u, "\eta"'] \arrow[d, "\nu"] & \sim & H\ltimes Y \\
 & & (L\times G)\ltimes U' \arrow[ull, "\omega"] \arrow[urr, "f\circ k"'] & &
\end{tikzcd}$$

there exists $(\widetilde{\Omega}, \widetilde{K})$ such that the following diagrams commute:

$$\begin{tikzcd}
 & & \widetilde{\mathcal{L}} \arrow[dll, "\widetilde{\Omega}"'] \arrow[drr, "\mathrm{ev}_0\circ \widetilde{K}"] & & \\
L'' & \sim & {} \arrow[u, "\eta'"'] \arrow[d, "\nu'"] & \sim & G\ltimes X \\
 & & (L\times G)\ltimes U' \arrow[ull, "\omega"] \arrow[urr, "k"'] & &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & & \widetilde{\mathcal{L}} \arrow[dll, "\widetilde{\Omega}"'] \arrow[drr, "f_*\circ \widetilde{K}"] & & \\
L\ltimes U & \sim & {} \arrow[u, "\eta''"'] \arrow[d, "\nu''"] & \sim & H\ltimes Y^I \\
 & & (L\times H)\ltimes U'' \arrow[ull, "\Omega"] \arrow[urr, "K"'] & &
\end{tikzcd}$$

Proposition 7.6 *The evaluation map* $\mathrm{ev}_0\colon G\ltimes X^I \to G\ltimes X$ *is a groupoid fibration.*

Proof For all translation groupoids $L\ltimes U$ making the following diagram commutative up to 2–isomorphism, we will construct the required generalized map $(\widetilde{\Omega}, \widetilde{K})$:

$$\begin{tikzcd}
L\ltimes U & & & \\
 & \widetilde{\mathcal{L}} \arrow[ul, dashed, "\widetilde{\Omega}"'] \arrow[dr, dashed, "\widetilde{K}"] & L''\ltimes U'' \arrow[ull, bend right, "\Omega"'] \arrow[dr, bend left, "K"] & \\
 & L'\ltimes U' \arrow[uul, bend left, "\omega"] \arrow[dr, bend right, "k"'] & G\ltimes (X^I)^I \arrow[r] \arrow[d] & G\ltimes X^I \arrow[d, "\mathrm{ev}_0"] \\
 & & G\ltimes X^I \arrow[r, "\mathrm{ev}_0"] & G\ltimes X
\end{tikzcd}$$

Since there is a 2–isomorphism between the generalized maps $(\Omega, \mathrm{ev}_0 K)$ and $(\omega, \mathrm{ev}_0 k)$, we know that there exists a groupoid $\widetilde{\mathcal{L}}$ and essential equivalences ν and η such that the following diagram commutes up to natural transformations:

$$\begin{tikzcd}
 & & L''\ltimes U'' \arrow[dll, "\Omega"'] \arrow[drr, "\mathrm{ev}_0\circ K"] & & \\
L\ltimes U & \sim & \widetilde{\mathcal{L}} \arrow[u, "\eta"'] \arrow[d, "\nu"] & \sim & G\ltimes X \\
 & & L'\ltimes U' \arrow[ull, "\omega"] \arrow[urr, "\mathrm{ev}_0\circ k"'] & &
\end{tikzcd}$$

We take $\widetilde{\Omega} = \Omega\eta$ and will construct a map $\widetilde{K}\colon \widetilde{\mathcal{L}} \to G \ltimes (X^I)^I$ such that the following diagram commutes up to natural transformations:

$$\begin{tikzcd}
L \ltimes U & & & \\
& \widetilde{\mathcal{L}} \arrow[ul, "\widetilde{\Omega}"'] \arrow[r, "\eta"] \arrow[d, "\nu"] \arrow[dr, dashed, "\widetilde{K}"] & L'' \ltimes U'' \arrow[ull, bend right, "\Omega"'] \arrow[dr, bend left, "K"] & \\
& L' \ltimes U' \arrow[uul, bend left, "\omega"] \arrow[dr, bend right, "k"'] & G \ltimes (X^I)^I \arrow[r] \arrow[d] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] \\
& & G \ltimes X^I \arrow[r, "ev_0"] & G \ltimes X
\end{tikzcd}$$

Consider the groupoid pullback

$$\begin{tikzcd}
P \arrow[r, "\pi_1"] \arrow[d, "\pi_2"] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] \\
G \ltimes X^I \arrow[r, "\mathrm{ev}_0"] & G \ltimes X
\end{tikzcd}$$

where P is the translation groupoid

$$P = (G \times G) \ltimes (X^I \times_X X^I \times_X G)$$

with $X^I \times_X X^I \times_X G = \{(\alpha_1, \alpha_2, k) \mid k\alpha_1(0) = \alpha_2(0)\}$. We observe that in fact P is equivalent to $\mathrm{Map}(I_{S_2}, G \ltimes X)$. To show this equivalence, we construct first a functor $\overline{K}\colon P \to G \ltimes (X^I \times_X X^I)$, where $X^I \times_X X^I = X^{I\vee I}$ is the pullback of the diagram

$$\begin{tikzcd}
& X^I \arrow[d, "\mathrm{ev}_0"] \\
X^I \arrow[r, "\mathrm{ev}_0"] & X
\end{tikzcd}$$

given by $\overline{K}(\alpha_1, \alpha_2, k)) = (k\alpha_1, \alpha_2)$ on objects and $\overline{K}(g_1, g_2) = g_2$ on morphisms.

Since $(g_1, g_2) \cdot (\alpha_1, \alpha_2, k) = (g_1\alpha_1, g_2\alpha_2, g_2 k g_1^{-1})$ and

$$\overline{K}(g_1\alpha_1, g_2\alpha_2, g_2 k g_1^{-1}) = (g_2 k g_1^{-1} g_1\alpha_1, g_2\alpha_2) = (g_2 k\alpha_1, g_2\alpha_2) = g_2(k\alpha_1, \alpha_2),$$

we can see that this is just a special case of the equivalences of the path groupoid models from Section 4,

$$\mathrm{Map}(I_{S_2}, G \ltimes X) \cong P \sim G \ltimes X^{I\vee I} \cong G \ltimes X^I.$$

We observe that the diagram of functors

$$\begin{tikzcd}
G \ltimes X^I & G \ltimes X^{I\vee I} \arrow[l, "j_1^*"'] \arrow[r, "j_2^*"] & G \ltimes X^I \\
& P \arrow[ul, "\pi_1"] \arrow[u, "\overline{K}"'] \arrow[ur, "\pi_2"'] &
\end{tikzcd}$$

commutes up to natural transformations since the right-hand side commutes on the nose and the left-hand side commutes up to a natural transformation. Here $j_1\colon I \to I \vee I$ and $j_2\colon I \to I \vee I$ are the natural maps for the coproduct of pointed spaces

$$\begin{tikzcd}
I \arrow[r, "i_1"] \arrow[dr, "j_1"'] & I \times I \arrow[d, "\pi"] & I \arrow[l, "i_2"'] \arrow[dl, "j_2"] \\
& I \vee I &
\end{tikzcd}$$

where $i_1(t) = (t, 0)$, $i_2(s) = (0, s)$ and $\pi\colon I \times I \to I \vee I$ is a deformation retract. Therefore, we have the commutative diagram

$$\begin{tikzcd}
P \arrow[dr, dashed, "\overline{K}"] \arrow[drrr, bend left, "\pi_1"] \arrow[ddrr, bend right, "\pi_2"'] & & & \\
& G \ltimes X^{I\vee I} \arrow[dr, dashed, "\pi^*"] \arrow[drr, bend left, "j_1^*"] \arrow[dr, bend right, "j_2^*"'] & & \\
& & G \ltimes X^{I\times I} \arrow[r, "i_1^*"] \arrow[d, "i_2^*"] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] \\
& & G \ltimes X^I \arrow[r, "\mathrm{ev}_0"] & G \ltimes X
\end{tikzcd} \tag{2}$$

Now, by the universal property of the groupoid pullback, there exists a functor $\xi\colon \mathcal{L} \to P$ such that the diagram

$$\begin{tikzcd}
\mathcal{L} \arrow[dr, "\xi"] \arrow[drr, bend left, "K\eta"] \arrow[ddr, bend right, "k\nu"'] & & \\
& P \arrow[r] \arrow[d] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] \\
& G \ltimes X^I \arrow[r, "\mathrm{ev}_0"] & X
\end{tikzcd} \tag{3}$$

commutes up to natural transformation.

Now, we put together diagrams (2) and (3) to obtain

$$\begin{tikzcd}
\widetilde{\mathcal{L}} \arrow[r, "\eta"] \arrow[d, "\nu"] \arrow[dr, dashed, "\xi"] & L'' \ltimes U'' \arrow[rrrddd, bend left, "K"] & & & \\
L' \ltimes U' \arrow[dddrrr, bend right, "k"] & P \arrow[rrrdd, bend left, "\pi_1"] \arrow[dr, dashed, "\overline{K}"] \arrow[dddrr, bend right, "\pi_2"] & & & \\
& & G \ltimes X^{I \vee I} \arrow[rrd, bend left, "j_1^*"] \arrow[dr, dashed, "\pi^*"] \arrow[ddr, bend right, "j_2^*"] & & \\
& & & G \ltimes X^{I \times I} \arrow[r, "i_1^*"] \arrow[d, "i_2^*"] & G \ltimes X^I \arrow[d, "\mathrm{ev}_0"] \\
& & & G \ltimes X^I \arrow[r, "\mathrm{ev}_0"] & G \ltimes X
\end{tikzcd}$$

and define $\widetilde{K} = \pi^* \circ \overline{K} \circ \xi$. □

Acknowledgements

Ángel acknowledges and thanks the financial support provided by the Max Planck Institute for Mathematics and the grant P12.160422.004/01 -FAPA ANDRES ANGEL from Vicedecanatura de Investigaciones de la Facultad de Ciencias de la Universidad de los Andes, Colombia.

Colman is partially supported by the Simons Foundation (#278333). Collaboration between the coauthors was possible thanks to this financial support.

Part of this paper was written during the visit of the second author to the Universidad de los Andes and during the visit of the first author to Wright College. They gratefully acknowledge the host institutions for their hospitality.

Both authors thank Gustavo Granja and Behrang Noohi for their helpful insights on an earlier version of this manuscript. Special thanks to Dorette Pronk for valuable discussions.

Both authors would like to thank the referee for the careful reading and for giving constructive comments which helped improved the quality of the paper.

References

[1] **A Adem**, **J Leida**, **Y Ruan**, *Orbifolds and stringy topology*, Cambridge Tracts in Mathematics 171, Cambridge Univ. Press (2007) MR Zbl

[2] **A Ángel**, **H Colman**, *G–category versus orbifold category*, Topol. Methods Nonlinear Anal. 61 (2023) 179–197 MR Zbl

[3] **H Colman**, *The Lusternik–Schnirelmann category of a Lie groupoid*, Trans. Amer. Math. Soc. 362 (2010) 5529–5567 MR Zbl

[4] **H Colman**, *On the 1–homotopy type of Lie groupoids*, Appl. Categ. Structures 19 (2011) 393–423 MR Zbl

[5] **N Gambino**, **A Joyal**, *On operads, bimodules and analytic functors*, Mem. Amer. Math. Soc. 1184, Amer. Math. Soc., Providence, RI (2017) MR Zbl

[6] **K Guruprasad**, **A Haefliger**, *Closed geodesics on orbifolds*, Topology 45 (2006) 611–641 MR Zbl

[7] **E Lupercio**, **B Uribe**, *Loop groupoids, gerbes, and twisted sectors on orbifolds*, from "Orbifolds in mathematics and physics" (A Adem, J Morava, Y Ruan, editors), Contemp. Math. 310, Amer. Math. Soc., Providence, RI (2002) 163–184 MR Zbl

[8] **J P May**, *A concise course in algebraic topology*, Univ. Chicago Press (1999) MR Zbl

[9] **I Moerdijk**, **D A Pronk**, *Orbifolds, sheaves and groupoids*, K–Theory 12 (1997) 3–21 MR Zbl

[10] **B Noohi**, *Mapping stacks of topological stacks*, J. Reine Angew. Math. 646 (2010) 117–133 MR Zbl

[11] **B Noohi**, *Fibrations of topological stacks*, Adv. Math. 252 (2014) 612–640 MR Zbl

[12] **J Pardon**, *Enough vector bundles on orbispaces*, Compos. Math. 158 (2022) 2046–2081 MR Zbl

[13] **D A Pronk**, *Etendues and stacks as bicategories of fractions*, Compositio Math. 102 (1996) 243–303 MR Zbl

[14] **D Pronk**, **L Scull**, *Translation groupoids and orbifold cohomology*, Canad. J. Math. 62 (2010) 614–645 MR Zbl

[15] **D Pronk**, **L Scull**, **M Tommasini**, *Atlases for ineffective orbifolds*, preprint (2017) arXiv 1606.04439v2

[16] **I Satake**, *On a generalization of the notion of manifold*, Proc. Nat. Acad. Sci. U.S.A. 42 (1956) 359–363 MR Zbl

[17] **A Sibih**, *Orbifold atlas groupoids*, master's thesis, Dalhousie University (2013) Available at `https://dalspace.library.dal.ca/handle/10222/21458`

[18] **J Strom**, *Modern classical homotopy theory*, Graduate Studies in Math. 127, Amer. Math. Soc., Providence, RI (2011) MR Zbl

Departamento de Matemáticas, Universidad de los Andes
Bogota, Colombia

Department of Mathematics, Wright College
Chicago, IL, United States

ja.angel908@uniandes.edu.co, hcolman@ccc.edu

https://www.hellencolman.com/math

Received: 2 September 2017 Revised: 16 November 2021

Algebraic & Geometric Topology **23**:5 (2023) 2009–2028
DOI: 10.2140/agt.2023.23.2009
Published: 25 July 2023

Discrete real specializations of sesquilinear representations of the braid groups

NANCY SCHERICH

Using Salem numbers, this paper gives real specializations of sesquilinear representations of the braid groups that make the images discrete groups. This method is applied to the Burau, Jones and Lawrence–Krammer representations, and some details on the commensurability of the target groups are given.

20C08, 57M99

1 Introduction

Representations of the braid groups have attracted attention because of their wide variety of applications from discrete geometry to quantum computing. This paper takes the point of view that one should ask structural questions about the image of a braid group representation, in particular whether the image is discrete for specializations of the parameter. Venkataramana in [15] also followed this pursuit for discrete specializations of the Burau representation but with a different approach toward arithmeticity.

Since the Jones representations are used in modeling quantum computations, much work has been done to understand specializations at roots of unity, as explored by Funar and Kohno in [7], Freedman, Larsen and Wang in [6], and many others. However, there seems to be a lack of exploration of the *real* specializations of these representations. This paper takes a more general approach to find discrete real specializations of any sesquilinear group representation, and show how this can be applied to representations of the braid groups. The main theorem follows.

Theorem 1.1 *Let* $\rho_t : G \to \mathrm{GL}_m(\mathbb{Z}[t, t^{-1}])$ *be a group representation with parameter* t*. Suppose there exists a matrix* J_t *such that*:

(1) *For all* M *in the image of* ρ_t, $M^* J_t M = J_t$, *where* $M^*(t) = M^{\mathsf{T}}(1/t)$.

(2) $J_t = (J_{1/t})^\intercal$.

(3) $J_t \in \mathrm{GL}_m(\mathbb{Q}(t))$, *where no entry of* J_t *has denominator with* 1 *as a root.*

(4) J_t *is positive definite for* t *in a neighborhood* η *of* 1 *in* $\mathbb{C}$.

Then there exist infinitely many Salem numbers s *such that the specialization representation* ρ_s *at* $t = s$ *is discrete.*

Further applying a classification theorem of hermitian forms from Scharlau [13] proves the following commensurability result of the target groups.

Corollary 1.2 *For* $\rho_t : G \to \mathrm{SL}_{2m+1}(\mathbb{Z}[t, t^{-t}])$ *as in Theorem 1.1, there exist infinitely many Salem numbers* s *such that for infinitely many integers* n *and* k *the specializations* ρ_{s^k} *at* $t = s^k$ *and* ρ_{s^n} *at* $t = s^n$ *map into commensurable lattices.*

Squier showed in [14] that the reduced Burau representation is sesquilinear and satisfies the criteria for Theorem 1.1. Example 3.2 gives explicit Salem numbers such that specializing the reduced Burau representation to these numbers is discrete. Using the Burau representation as motivation, Section 2 introduces the main tools of discrete generalized unitary groups and Salem numbers. In Section 3 we apply Theorem 1.1 to the Jones and Lawrence–Krammer representations of the braid group, and we suspect it also applies to all of the BMW representations. Lastly, Section 4 discusses the lattice structure and commensurability of the target groups for the Salem number specializations.

Acknowledgements

The results of this paper formed part of the author's dissertation. The author would like to thank her advisor, Darren Long, for endless insight, guidance and support on this project, Vaughan Jones for helpful conversations, and Warren Scherich for moral support. This paper was funded in part by the NSF grant DMS-1463740.

2 Discrete representations using Salem numbers

2.1 Motivation from Squier and the Burau representation

The (reduced) Burau representation $\rho_{n,t} \colon B_{n+1} \to \mathrm{GL}_n(\mathbb{Z}[t, t^{-1}])$ is an irreducible representation of the braid group. These representations depend on n, where $n+1$

is the number of braid strands, and are parametrized by a variable t. Squier showed in [14] that there is a nondegenerate n–dimensional matrix $J_{n,t}$ satisfying the equation

$$M^* J_{n,t} M = J_{n,t}, \tag{2-1}$$

for all M in the image of $\rho_{n,t}$. Here M^* is the transpose of M after replacing t with $1/t$ in the entries of M, $M^*(t) := M(1/t)^\intercal$. $J_{n,t}$ is sesquilinear with respect to $*$, $J^*_{n,t} = J_{n,t}$, and letting $t = x^2$, $J_{n,t}$ is given by the tridiagonal matrix

$$J_{n,t} = \begin{bmatrix} x + x^{-1} & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & & -1 \\ & & -1 & x + x^{-1} \end{bmatrix}.$$

If t is a unit complex number, (2-1) agrees with the usual unitary relation $(\overline{M})^\intercal M = \mathrm{Id}$. Representations that satisfy (2-1) are called *sesquilinear*, and are said to map into a *generalized unitary group*. This terminology will be made precise in the next section.

These generalized unitary groups are the key to finding discrete specializations. The method described here is to show that carefully chosen specializations of the parameter t make the entire generalized unitary group discrete, thus making the image of the representation discrete.

2.2 Unitary groups

In general, unitary groups are matrix groups which respect a form, or inner product. These notions heavily rely on the ring of coefficients and an involution of that ring. Let R be a ring and ϕ an order 2 automorphism of R. For a matrix M defined over R, let $M^* = (M^\phi)^\intercal$, where M^ϕ is the matrix obtained by applying ϕ to the entries of M. For the Burau representation in Section 2.1, ϕ is the map given by $t \mapsto 1/t$, and $R = \mathbb{Z}[t, t^{-1}]$.

Definition 2.1 For a matrix J such that $J^* = J$, the *generalized unitary group* is

$$U_m(J, \phi, R) := \{M \in \mathrm{GL}_m(R) \mid M^* J M = J\}.$$

Here, J is called a *sesquilinear form* with respect to ϕ. For example, in this notation the familiar unitary group U_m can be written as $U_m(\mathrm{Id}, -, \mathbb{C})$, where "–" is complex conjugation. A representation is called *sesquilinear* if its image is contained in a generalized unitary group.

2.2.1 Creating discrete unitary groups The Burau representation can be written as $\rho_{n,t} \colon B_{n+1} \to U_n(J_{n,t}, \phi, \mathbb{Z}[t, t^{-1}])$. With the goal of parameter specialization in mind, the relevant choice for the coefficient ring is a number ring. Discreteness of the unitary group is a delicate relationship between the form J and the algebraic structure of the number ring. More precisely, let L be a totally real algebraic field extension of $\mathbb{Q}$ and let K be a degree 2 field extension of L, with $L, K \subseteq \mathbb{C}$. Let ϕ be the order 2 generator of $\mathrm{Gal}(K/L)$, and let $\mathcal{O}_K$ and $\mathcal{O}_L$ denote the rings of integers of K and L, respectively:

$$\begin{tikzcd}
\mathcal{O}_K \arrow[loop above, "\phi"] \arrow[d, dash] & K \arrow[loop above, "\phi"] \arrow[d, dash, "2"] & K^\sigma \subseteq \mathbb{C} \arrow[loop above, "\phi^\sigma"] \\
\mathcal{O}_L & L & L^\sigma \subseteq \mathbb{R}
\end{tikzcd}$$

Let σ be a complex place of K, which in this setting is a field homomorphism $\sigma \colon K \to \mathbb{C}$ different from ϕ and the identity map. We write $X^\sigma = \sigma(X)$ for any X in K. The algebraic structure is passed along by σ, meaning $\mathcal{O}_{K^\sigma} = (\mathcal{O}_K)^\sigma$ is the ring of integers for K^σ and $\phi^\sigma = \sigma\phi\sigma^{-1}$ is an involution on K^σ.

Let J be a matrix over $\mathcal{O}_K$ that is sesquilinear with respect to ϕ. Then J^σ is sesquilinear with respect to ϕ^σ. So, in particular,

$$U_m(J^\sigma, \phi^\sigma, \mathcal{O}_{K^\sigma}) = \{M \in \mathrm{GL}_m(\mathcal{O}_{K^\sigma}) \mid (M^{\phi^\sigma})^{\mathsf{T}} J^\sigma M = J^\sigma\}.$$

Since σ is a homomorphism, we can see that $(U_m(J, \phi, \mathcal{O}_K))^\sigma = U_m(J^\sigma, \phi^\sigma, \mathcal{O}_{K^\sigma})$ by applying σ to the equation $J = M^* J M$.

The following results outline compatibility requirements between J and $\mathcal{O}_K$, which show that $U_m(J, \phi, \mathcal{O}_K)$ is a discrete subgroup of $\mathrm{GL}_m(\mathbb{R})$ under the standard euclidean topology.

Proposition 2.2 *$U_m(J^\sigma, \phi^\sigma, \mathcal{O}_{K^\sigma})$ is a bounded group when J^σ is positive definite and ϕ^σ is complex conjugation.*

Proof Because J^σ is positive definite, by Sylvester's law of inertia and the Gram–Schmidt process, there exists a matrix $Q \in \mathrm{GL}_m(\mathbb{C})$ such that $J^\sigma = Q^* \,\mathrm{Id}\, Q$. This implies that $Q U_m(J^\sigma, \phi^\sigma, \mathcal{O}_{K^\sigma}) Q^{-1} \subseteq U_m(\mathrm{Id}, \phi^\sigma, \mathbb{C})$, which is a subgroup of the compact group U_m. $\square$

Theorem 2.3 *$U_m(J,\phi,\mathcal{O}_K)$ is discrete if, for every complex place σ of K, J^σ is positive definite and ϕ^σ is complex conjugacy.*

Proof Assume that $\{M_n\}$ converges to the identity in $U_m(J,\phi,\mathcal{O}_K)$. To show $\{M_n\}$ is eventually constant, we will show that for n large there are only finitely many possibilities for the entries $(M_n)_{ij}$.

By assumption, for each σ the group $U_m(J^\sigma,\phi^\sigma,O_{K^\sigma})$ is bounded by Proposition 2.2. Also, for every M_n, $M_n^\sigma \in U_m(J^\sigma,\phi^\sigma,O_{K^\sigma})$. Thus, there exists a B such that for large n, for all i, j, and for all σ, we have that $|(M_n^\sigma)_{ij}| < B$.

For every $M \in U_m(J,\phi,\mathcal{O}_K)$, the equation $M^*JM = J$ can be rearranged to $JMJ^{-1} = ((M^\phi)^\mathsf{T})^{-1}$, showing that M and $((M^\phi)^\mathsf{T})^{-1}$ are simultaneously conjugate. Thus $\{M_n^\phi\}$ also converges to the identity. Convergent sequences are bounded, so for large enough n, $|(M_n)_{ij}| < B$ and $|(M_n)_{ij}^\phi| < B$ for every ij–entry.

L is a totally real degree 2 subfield of K, and ϕ generates $\mathrm{Gal}(K/L)$. So K has one nonidentity real embedding ϕ, and all other embeddings are complex. Thus we have shown above that for large n there is a uniform bound B for each entry $(M_n)_{ij}$ and each Galois conjugate of $(M_n)_{ij}$. There are only finitely many algebraic integers α such that $\deg(\alpha) \le \deg(K/\mathbb{Q})$, and with the property that α and all of the Galois conjugates of α have absolute values bounded above by B. So there are only finitely many possible entries for $(M_n)_{ij}$, which implies the sequence $\{M_n\}$ is eventually constant. □

Corollary 2.4 *If $\rho\colon G \to U_m(J,\phi,\mathcal{O}_K)$ is a representation of a group G such that for every nonidentity place σ of K, J^σ is positive definite and ϕ^σ is complex conjugacy, then ρ is a discrete representation.*

At first glance, the requirements for Corollary 2.4 seem very specific and perhaps it is doubtful that any such a representation could exist. However, as described in Section 2.1, Squier showed that the Burau representation maps into a generalized unitary group over $\mathbb{Z}[t,t^{-1}]$, so the next task is to find values of t such that the form and coefficient ring satisfy the specific hypothesis of Corollary 2.4.

2.3 Salem numbers

Salem numbers are the key ingredient to the application of Corollary 2.4, which requires a real algebraic number field with tight control and understanding of each of its complex embeddings.

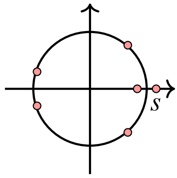


Figure 1: A schematic picture of an order 6 Salem number.

Definition 2.5 A Salem number s is a real algebraic unit greater than 1, with one real Galois conjugate $1/s$, and all of the complex Galois conjugates of s have absolute value equal to 1.

For example, the largest real root of Lehmer's polynomial, called Lehmer's number,

$$x^{10} + x^9 - x^7 - x^6 - x^5 - x^4 - x^3 + x + 1,$$

is a Salem number. Trivial Salem numbers of degree 2 are solutions to $s^2 - ns + 1$ for $n \in \mathbb{N}$ and $n > 2$. It is well known that there are infinitely many Salem numbers of arbitrarily large absolute value and degree. In particular, if s is a Salem number, then s^m is also a Salem number for every positive integer m. One geometric consequence of the property that powers of Salem numbers are Salem numbers is that by taking powers, one can control the spatial configuration of the complex Galois conjugates of a Salem number, as described in Lemma 2.6.

Lemma 2.6 *For any Salem number s, and for any interval containing 1 on the complex unit circle, there exist infinitely many integers m such that every complex Galois conjugate of s^m lies in the interval.*

Proof Let $e^{i\theta_1}, \dots, e^{i\theta_k}$ be all the Galois conjugates of the Salem number s with positive imaginary part. Suppose that $\prod_{j=1}^{k} (e^{i\theta_j})^{m_j} = 1$. Let ϕ be the automorphism of the Galois closure of s with the property that $\phi(e^{i\theta_1}) = s$. Since ϕ must permute the

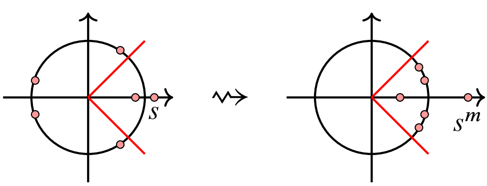


Figure 2: A schematic picture of Lemma 2.6.

Galois conjugates of s, for $j \neq 1$, $\phi(e^{i\theta_j})$ is again on the complex unit circle. Thus,

$$1=\phi\left(\prod_{j=1}^{k}(e^{i\theta_j})^{m_j}\right)=s^{m_1}\prod_{j=2}^{k}\phi(e^{i\theta_j})^{m_j}, \quad \text{which implies} \quad \prod_{j=2}^{k}\phi(e^{i\theta_j})^{m_j}=1/s^{m_1}.$$

Since each $\phi(e^{i\theta_j})$ is a unit complex number, it must be the case that each $m_j = 0$. This shows that the point $p = (e^{i\theta_1}, \dots, e^{i\theta_k})$ satisfies the criteria for Kronecker's theorem. In particular, the set $\overline{\{p^m \mid \in \mathbb{Z}\}}$ is dense in the torus T^k. □

Fixing an arbitrary Salem number s, let $K = \mathbb{Q}(s)$, $L = \mathbb{Q}(s + 1/s)$ and $\mathcal{O}_K$ be the ring of integers of K:

$$\begin{array}{c} \mathbb{Q}(s) = K \\ \big|\,{\scriptstyle 2} \\ \mathbb{Q}(s+1/s) = L \\ \big| \\ \mathbb{Q} \end{array}$$

Since s and $1/s$ are real and all other Galois conjugates of s are complex, K has exactly two real embeddings. For a complex embedding σ of K, $(s+1/s)^\sigma = 2\,\mathrm{Re}(s^\sigma)$, which is real. This shows that all embeddings of L are real, and that L is a totally real subfield of K. Since s is a root of $X^2 - (s + 1/s)X + 1$, K is degree 2 over L.

The Galois group of K/L is generated by ϕ, which maps s to $1/s$. (This exactly matches the involution $t \mapsto 1/t$ needed in the sesquilinear condition for the Burau representation.) On the complex unit circle, inversion is the same as complex conjugation. So for the complex embeddings σ of K, ϕ^σ is complex conjugacy. Notice for a sesquilinear matrix J_t over $\mathcal{O}_K$ with a parameter t, specializing $t = s$ leaves J_s^σ hermitian.

Theorem 1.1 *Let $\rho_t\colon G \to \mathrm{GL}_m(\mathbb{Z}[t, t^{-1}])$ be a representation of a group G. Suppose there exists a matrix J_t such that:*

(1) $M^* J_t M = J_t$ *for all* M *in the image of* ρ_t.

(2) $J_t = (J_{1/t})^\mathsf{T}$.

(3) $J_t \in \mathrm{GL}_m(\mathbb{Q}(t))$, *where no entry of* J_t *has denominator with* 1 *as a root.*

(4) J_t *is positive definite for* t *in a neighborhood* η *of* 1 *in* $\mathbb{C}$.

Then there exist infinitely many Salem numbers s such that the specialization ρ_s at $t = s$ is discrete.

Proof The neighborhood η can be chosen so that no entry of J_t has a denominator with a root in η. By Lemma 2.6, there are infinitely many Salem numbers with the property that all the complex Galois conjugates lie in η. Let s be one such Salem number. Specializing t to s gives $\rho_s\colon G \to U_m(J_s, \phi, O_{\mathbb{Q}(s)})$, where ϕ is the usual map given by $s \mapsto 1/s$.

Let σ be a complex place of $\mathbb{Q}(s)$ which is given by $s \mapsto z$ for z a complex Galois conjugate of s. Then $J_s^\sigma = J_z$, and since $z \in \eta$, J_z is positive definite. By Corollary 2.4, the specialization ρ_s at $t = s$ is discrete. □

Remark 2.7 If the representations in Theorem 1.1 all have determinant 1, then the image is more than just discrete, and in fact is a subgroup of a lattice. See Section 4 for more details.

3 Applications to braid group representations

3.1 The Burau representation

Proposition 3.1 *There are infinitely many Salem numbers s such that the Burau representation specialized to $t = s$ is discrete.*

Proof The specialization of $\rho_{n,1}$ at $t = 1$ collapses to an irreducible representation of the symmetric group. As a representation of a finite group, $\rho_{n,1}$ fixes a positive definite form which is unique up to scaling, by Lemma 3.7, which is proved later. At $t = 1$, $J_{n,1}$ is positive definite, and the signature of $J_{n,t}$ can only change at zeroes of its determinant.

An inductive computation shows that $\det(J_{n,t}) = (t^{2n+2} - 1)/(t^n(t^2 - 1))$ for $t \neq 1$, and the zeroes of $\det(J_{n,t})$ occur at $(n+1)^{\text{th}}$ roots of unity. Thus, $J_{n,t}$ remains positive definite for unit complex values of t with argument less than $2\pi/(n + 1)$. This shows the Burau representation satisfies the criteria of Theorem 1.1. □

Example 3.2 The Burau representation $\rho_{4,t}$ of B_4 is discrete when specializing t to the following Salem numbers:

- Lehmer's number raised to the powers 16, 32 and 47,
- the largest real root of $1-x^4-x^5-x^6+x^{10}$ raised to the powers 17, 23, and 43.

3.2 The Jones representations

The *Hecke algebra* (of type A_n), denoted by $H_n(q)$, is the complex algebra generated by invertible elements $g_1, \dots, g_{n-1}$ with relations

$$\begin{aligned} g_i g_{i+1} g_i &= g_{i+1} g_i g_{i+1} && \text{for all } i < n, \\ g_i g_j &= g_j g_i && \text{for } |i-j| > 1, \\ g_i^2 &= (1-q) g_i + q && \text{for all } i < n. \end{aligned} \tag{3-1}$$

Here, q is viewed as a complex parameter. $H_n(q)$ is a quotient of $\mathbb{C}[B_n]$ by the relation (3-1). The *Jones representations* of B_n are defined by precomposing a representation of $H_n(q)$ by the quotient map from $\mathbb{C}[B_n]$. The Jones representations have matrix entries in $\mathbb{Z}[t, t^{-1}]$.

3.2.1 The Jones representations are sesquilinear The Hecke algebras have a natural automorphism, denoted here by ϕ, which sends q to $1/q$. Taking q to be a unit complex number, this automorphism becomes complex conjugacy. The Jones representations are known to be sesquilinear with respect to ϕ for various complex specializations of q and with many different types of proofs, as in [2; 3; 9; 17]. To be overtly clear that all of the criteria of Theorem 1.1 are satisfied by the Jones representations, we provide a simple proof of sesquilinearity here that is very similar to [3, Proposition 3.7] by Brunat and Marin.

The irreducible representations of $H_n(q)$ are parametrized by the Young diagrams. (For a more detailed discussion of Young diagrams see [18], and for a construction of the Jones representations see [16].) Every Young diagram contains sub-Young diagrams, obtained by removing boxes in a way that retains the weakly decreasing row length condition. If λ is a Young diagram with n boxes, then we will call the sub-Young diagrams found by removing one box from λ the $(n-1)$*–subdiagrams of* λ. A Young diagram is completely determined by any two of its $(n-1)$–subdiagrams. These $(n-1)$–subdiagrams also determine representations of the Hecke algebras in a powerful way. The following theorem, originally due to Curtis, Iwahori and Kilmoyer in [5] and popularized by Jones in [8], states concretely the relationship between Young diagrams and the representations of the Hecke algebras.

Theorem 3.3 *Up to equivalence, the finite-dimensional irreducible representations of $H_n(q)$, for generic q, are in one to one correspondence with the Young diagrams of n boxes. Moreover, if ρ is a representation corresponding to Young diagram λ, then ρ*

restricted to $H_{n-1}(q)$ is equivalent to the representation $\bigoplus_{i=1}^k \rho_{\lambda_i}$, where $\lambda_1,\dots,\lambda_k$ are all of the $(n-1)$–subdiagrams of λ and each ρ_{λ_i} is an irreducible representation of $H_{n-1}(q)$ corresponding to λ_i.

Here equivalence means the existence of an intertwining isomorphism, made precise by the following definition.

Definition 3.4 The representations $\varphi\colon G\to \mathrm{GL}(V)$ and $\psi\colon G\to\mathrm{GL}(W)$ are *equivalent* if there exists a linear isomorphism $T\colon V\to W$ such that $T\varphi(g)(v)=\psi(g)T(v)$ for all $g\in G$ and $v\in V$, or that the following diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\varphi(g)} & V \\ \downarrow{\scriptstyle T} & & \downarrow{\scriptstyle T} \\ W & \xrightarrow{\psi(g)} & W \end{array}$$

Choosing bases for V and W, the equivalence T gives the matrix equation

$$[T][\varphi(g)][T]^{-1}=[\psi(g)].$$

At the level of matrices, representations are equivalent exactly when they are simultaneously conjugate. In the context of Theorem 3.3, the restriction of ρ to $H_{n-1}(q)$ is equivalent to the representation $\bigoplus_{i=1}^k \rho_{\lambda_i}$, which means there is a change of basis such that the restriction of ρ is block diagonal. These restriction rules are combinatorially depicted in the Young lattice of Young diagrams; see [18].

A representation is sesquilinear if there exists an invertible matrix J such that for every M in the image of the representation the following equation is satisfied:

$$M^*JM=J. \tag{3-2}$$

Rearranging this equation, we see that $M=J^{-1}((M^{\phi})^{\mathsf{T}})^{-1}J$, showing that M and $((M^{\phi})^{\mathsf{T}})^{-1}$ are simultaneously conjugate and the conjugating matrix J is the sesquilinear form. Changing views slightly, consider the following definition.

Definition 3.5 For $\varphi\colon G\to\mathrm{GL}(V)$ a complex linear representation, $\tilde\varphi\colon G\to\mathrm{GL}(V^*)$ is called the *ϕ–twisted contragredient representation* of φ and is given by

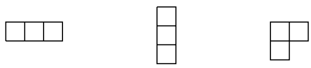
$$\tilde\varphi(g)f(v)=f(\varphi(g^{-1})^{\phi}v)$$

for every $g\in G$, $v\in V$ and $f\in V^*$.

If a basis for V is chosen, then as matrices, $[\tilde{\varphi}(g)] = ([\varphi(g)^{\phi}]^{\mathsf{T}})^{-1}$. So another way to view a sesquilinear representation is one that is *equivalent to its ϕ–twisted contragredient*.

Lemma 3.6 *Every finite-dimensional irreducible representation of the Hecke algebra is equivalent to its ϕ–twisted contragredient representation, when q is a generic complex number.*

Proof We can establish this result for $n = 3$. There are three nonequivalent irreducible representations of $H_3(q)$ corresponding to the following Young diagrams:



Up to equivalence, the first two representations are 1–dimensional, given by $g_i \mapsto q$ and $g_i \mapsto -1$, and are in fact equal to their ϕ–twisted contragredient representations. The third representation is known to be the Burau representation for B_3. As described earlier, Squier showed that the Burau representations are sesquilinear and are therefore equivalent to their ϕ–twisted contragredient.

Inductively moving forward, let $\rho\colon H_n(q) \to \mathrm{GL}(V)$ be a finite-dimensional irreducible representation and $\tilde{\rho}$ be the ϕ–twisted contragredient representation of ρ. Up to equivalence, ρ corresponds to a Young diagram λ. To show that ρ and $\tilde{\rho}$ are equivalent, it suffices to show that both representations correspond to the same λ. A Young diagram is completely characterized by its list of $(n-1)$–subdiagrams, which correspond to the restriction of the representation to $H_{n-1}(q)$. So it is enough to show that the restrictions of ρ and $\tilde{\rho}$ correspond to the same list of $(n-1)$–subdiagrams.

Writing $\rho| = \rho|_{H_{n-1}(q)}$, by Theorem 3.3 there is an equivalence T such that

$$T\rho|(h)T^{-1} = \bigoplus_{i=1}^{k} \rho_{\lambda_i}(h) \quad \text{for every } h \in H_{n-1}(q),$$

where each λ_i is an $(n-1)$–subdiagram of λ, k is the number of $(n-1)$–subdiagrams of λ, and ρ_{λ_i} is an irreducible representation of $H_{n-1}(q)$ corresponding to λ_i. Choosing a basis for V, the matrix for $[T\rho\,|\,(h)T^{-1}]$ is block diagonal. Taking the ϕ–twisted contragredient of a block diagonal matrix preserves the block decomposition, which gives

$$([T^{\phi}]^{\mathsf{T}})^{-1}[\tilde{\rho}\,|\,(h)][T^{\phi}]^{\mathsf{T}} = \bigoplus_{i=1}^{k}[\tilde{\rho}_{\lambda_i}(h)] \quad \text{for every } h \in H_{n-1}(q).$$

This equation shows that $\tilde{\rho}|$ is equivalent to $\bigoplus \tilde{\rho}_{\lambda_i}$. Since each ρ_{λ_i} is an irreducible representation of $H_{n-1}(q)$, we can inductively assume that ρ_{λ_i} is equivalent to $\tilde{\rho}_{\lambda_i}$, for all $i \le k$. Therefore, ρ_{λ_i} and $\tilde{\rho}_{\lambda_i}$ correspond to the same Young diagram λ_i. Thus the restrictions of ρ and $\tilde{\rho}$ correspond to the same list of $(n-1)$–subdiagrams. □

Lemma 3.7 *If an absolutely irreducible matrix representation has an invertible matrix J satisfying $M^* J M = J$ for all M in the representation, then J is unique up to scaling.*

Proof Suppose there were two such matrices J_1 and J_2. Then (3-2) gives, for all matrices M in the representation,

$$J_1 M J_1^{-1} = ((M^{\phi})^{\mathsf{T}})^{-1} = J_2 M J_2^{-1} \implies (J_1^{-1} J_2)^{-1} M (J_1^{-1} J_2) = M.$$

This shows that $J_1^{-1} J_2$ is in the centralizer of the entire irreducible representation. Schur's lemma gives that $J_1^{-1} J_2 = \alpha \cdot \mathrm{Id}$ for some scalar α, and finally $J_2 = \alpha J_1$. □

Proposition 3.8 *If ρ is an irreducible Jones representation of B_n and q is a generic unit complex number close to* 1*, then there exists a nondegenerate, positive definite, sesquilinear matrix J with entries in $\mathbb{Q}(q)$ such that for all M in the image of ρ, $(M^{\phi})^{\mathsf{T}} J M = J$.*

Proof Let ρ be a finite-dimensional irreducible representation of $H_n(q)$ over V. By Lemma 3.6, ρ is equivalent to its ϕ–twisted contragredient representation $\tilde{\rho}$ by an equivalence T. Choose a basis for V and its dual basis for V^*, and let $\mathcal{T}$ be the matrix for T with respect to these bases. We will use this matrix $\mathcal{T}$ to find the desired matrix J. Let the superscript $*$ denote the ϕ–twisted transpose of a matrix to ease computation. For all $g \in H_n(q)$, we get the matrix equations

$$\begin{aligned} \mathcal{T}[\rho(g)]\mathcal{T}^{-1} = [\tilde{\rho}(g)] = ([\rho(g)]^{-1})^* &\implies (\mathcal{T}^{-1})^*[\rho(g)]^*\mathcal{T}^* = [\rho(g)]^{-1} \\ &\implies \mathcal{T}^*[\rho(g)](\mathcal{T}^*)^{-1} = ([\rho(g)]^{-1})^*. \end{aligned} \tag{3-3}$$

This shows that $\mathcal{T}$ and $\mathcal{T}^*$ are two possible forms for ρ. By Lemma 3.7, $\mathcal{T} = \alpha\mathcal{T}^*$ for some $\alpha \in \mathbb{C}$. Applying $*$ again gives $\mathcal{T} = \alpha\alpha^*\mathcal{T}$ and $\alpha\alpha^* = 1$.

Define $J = \beta\mathcal{T} + \beta^*\mathcal{T}^* = (\alpha\beta + \beta^*)\mathcal{T}^*$ where β is as follows. (Here β is needed to ensure that J is invertible.) If $\alpha \ne -1$, let $\beta = 1$, which gives that $\det J = \det((\alpha+1)\mathcal{T})$, which is nonzero. If $\alpha = -1$, let $\beta \in \mathbb{C}$ be such that $\beta^* \ne \beta$. Then

$$\det J = \det[(\alpha\beta + \beta^*)\mathcal{T}^*] = \det[(-\beta + \beta^*)\mathcal{T}]$$

is nonzero. So, in both cases, J is invertible.

Next note that J is sesquilinear, that is, $J^* = (\beta\mathcal{T} + \beta^*\mathcal{T}^*)^* = \beta^*\mathcal{T}^* + \beta\mathcal{T} = J$. If M is a matrix in the image of ρ, rearranging the right-hand equation of (3-3) gives $M^*\mathcal{T}^*M = \mathcal{T}$. So inserting J gives

$$M^*JM = M^*(\alpha\beta + \beta^*)\mathcal{T}^*M = (\alpha\beta + \beta^*)M^*\mathcal{T}^*M = (\alpha\beta + \beta^*)\mathcal{T} = J.$$

To show that the entries of J are in $\mathbb{Q}(q)$ we will proceed by induction on n, as in the proof of Lemma 3.6. As a base case with $n = 3$, Squier's form for the Burau representation has entries in $\mathbb{Q}(q)$. Let ρ be an irreducible Jones representation of B_n with $\rho|_{B_{n-1}} = \bigoplus_{i=1}^k \rho_i$, where each ρ_i is an irreducible Jones representation of B_{n-1}. We can inductively assume each ρ_i is sesquilinear with form J_i whose coefficients are in $\mathbb{Q}(q)$. Thus, there exist some scalars α_i such that $J = [\alpha_i J_i]$, the block diagonal matrix, and J is the sesquilinear form for ρ.

It remains to show that J is positive definite. Taking $q = 1$, ρ is an irreducible representation of the symmetric group Σ_n. As a linear representation of a finite group, V admits an inner product that is invariant under the action of Σ_n, given by a positive definite nondegenerate matrix $\widehat{J}$. Lemma 3.7 guarantees that $\widehat{J}$ is unique up to scaling. Since $J|_{q=1}$ is also a form for this representation, it must be that $\widehat{J}$ is a multiple of $J|_{q=1}$, which gives that J is positive definite for $q = 1$. Since J is Hermitian for unit complex q it has real eigenvalues, and continuity of the determinant map finally gives that either J or $-J$ is positive definite for q close to 1. □

Corollary 3.9 *For each irreducible Jones representation, there are infinitely many Salem numbers s such that specializing $q = s^m$, for some m, is a discrete representation.*

3.3 The Lawrence–Krammer and BMW representations

The BMW algebras are a 2–parameter family of algebras, denoted by $C_n(l, m)$, with $n - 1$ generators and parameters l and m. The BMW representations of the braid group come from representations of the BMW algebras [1; 12]. Similar to how the Burau representation is one irreducible summand of the Jones representations, Zinno proved in [19] that the Lawrence–Krammer representation is one summand of the BMW representations. To make sense of the $*$ operation, the relevant involution for the BMW algebra is given by $l \mapsto 1/l$, $m \mapsto m$ and $\alpha \mapsto 1/\alpha$, where $m = \alpha + 1/\alpha$. Budney proved that the Lawrence–Krammer representation is sesquilinear [4], and Brunat and Marin give a more general proof that all the BMW representations are sesquilinear [3]; see also [10]. It is also known that the sesquilinear forms $J_{l,m}$ are positive definite for a neighborhood of $(1,1)$ in the unit complex sphere in $\mathbb{C}^2$ [17]; see [2, Theorem 1.2] for a concise restatement. It is suspected that the forms $J_{l,m}$ have coefficients in $\mathbb{Q}(m,l)$, and this is known to be true for the Lawrence–Krammer representation.

Corollary 3.10 *For the Lawrence–Krammer representation there are infinitely many Salem numbers s such that specializing $l = s^{k_1}$ and $\alpha = s^{k_2}$, for some k_1 and k_2, is a discrete representation.*

Example 3.11 Let ρ be the Lawrence–Krammer representation of B_4 given on page 272 of [1]. Taking the Salem number $S = 1/2 + 1/\sqrt{2} + 1/(2\sqrt{-1+2\sqrt{2}})$, specializing $\alpha = S^{15}$ and $l = S^3$ makes ρ a discrete representation.

4 Commensurability

Ideally, we would like to find *real* specializations so that the Jones representations have images that are not just discrete, but are arithmetic groups or lattices in $\mathrm{GL}_n(\mathbb{R})$. A first step in this direction is to further study the unitary groups coming from Salem number specializations, and consider when the images are subsets of lattices. Specializing to two different powers of the same Salem number can give *commensurable* unitary groups, but the defining sesquilinear forms might be very different.

Recall the notation of K, L, $\mathcal{O}_K$ and ϕ from Section 2.3. In general, fixing a number ring $\mathcal{O}_K$ and dimension m, the group $U_m(J,\phi,\mathcal{O}_K)$ is determined by the form J. Notice that $U_m(J,\phi,\mathcal{O}_K) = U_m(\lambda J,\phi,\mathcal{O}_K)$ for every $\lambda \in L$, and that the form J is not completely unique. This motivates that following definition.

Definition 4.1 Matrices J and H are equivalent over K if $Q^* J Q = \lambda H$ for some $Q \in \mathrm{GL}_m(K)$ and $\lambda \in \mathrm{Fix}(\phi)$.

It would be nice if equivalent forms gave rise to *equal* unitary groups, but this is not true in general. However, in the careful scenario that the unitary group is a lattice in $\mathrm{SL}_m(\mathbb{R})$, changing the form by equivalence yields "the same" lattice, up to commensurability in the following sense.

Definition 4.2 Two groups G_1 and G_2 are *commensurable* if there are finite-index subgroups $H_1 \subseteq G_1$ and $H_2 \subseteq G_2$ such that H_1 is isomorphic to H_2.

Definition 4.3 A *lattice* in a semisimple Lie group is a discrete subgroup with finite covolume.

For our purposes, we will take $\mathrm{SL}_m(\mathbb{R})$ or $\mathrm{PSL}_m(\mathbb{R})$ as the semisimple Lie group.

Proposition 4.4 *Assume that $\mathrm{SU}_m(J_1,\phi,\mathcal{O}_K)$ and $\mathrm{SU}_m(J_2,\phi,\mathcal{O}_K)$ are lattices in $\mathrm{SL}_m(\mathbb{R})$. If J_1 and J_2 are equivalent over K, then $\mathrm{SU}_m(J_1,\phi,\mathcal{O}_K)$ is commensurable to $\mathrm{SU}_m(J_2,\phi,\mathcal{O}_K)$*

Proof Let $\lambda J_1 = Q^* J_2 Q$ for some $Q \in \mathrm{GL}_m(K)$ and $\lambda \in \mathrm{Fix}(\phi)$. For notational clarity, write $\mathrm{SU}(J_i,\mathcal{O}_K) = \mathrm{SU}_m(J_i,\phi,\mathcal{O}_K)$.

Since scalar multiplication commutes with matrix multiplication, $M^*JM = J$ if and only if $M^*\lambda JM = \lambda J$. So scaling the form preserves the unitary group, and without loss of generality we may assume $\lambda = 1$.

It is clear that $M^*JM = J$ if and only if $(Q^*M^*Q^{*-1})(Q^*JQ)(Q^{-1}MQ) = Q^*JQ$, which seems like it implies that $\mathrm{SU}(Q^*J_1Q,\mathcal{O}_K) = Q^{-1}\,\mathrm{SU}(J_1,\mathcal{O}_K)Q$. However, since Q has coefficients in K, $Q^{-1}MQ$ may not have coefficients in $\mathcal{O}_K$, so we can only conclude that $Q^{-1}\,\mathrm{SU}(J,\mathcal{O}_K)Q \subseteq \mathrm{SU}(Q^*JQ, K)$. To avoid this, we need to pass to a finite-index subgroup.

Since K is the ring of fractions of $\mathcal{O}_K$, there exists $\gamma \in \mathcal{O}_K$ such that $\gamma Q \in M_m(\mathcal{O}_K)$. As a ring of integers of an algebraic extension, $\mathcal{O}_K$ is a Dedekind domain and every quotient is finite. So $\mathcal{O}_K/\langle\gamma^2\rangle$ is finite and $\mathrm{SU}(J_1,\mathcal{O}_K/\langle\gamma^2\rangle)$ is finite. The kernel N of the quotient map $\mathrm{SU}(J_1,\mathcal{O}_K) \to \mathrm{SU}(J_1,\mathcal{O}_K/\langle\gamma^2\rangle)$ has finite index in $\mathrm{SU}(J_1,\mathcal{O}_K)$.

Any element B in the kernel has the form $B = \mathrm{Id} + \gamma^2 A$ for some matrix $A \in M_m(\mathcal{O}_K)$. Substituting Q^*J_2Q for J_1 in the equation $B^*J_1B = J_1$ gives that QBQ^{-1} fixes the form J_2. Because Q has coefficients over K, QBQ^{-1} has coefficients in K and not necessarily in $\mathcal{O}_K$. However, since $QBQ^{-1} = \mathrm{Id} + (\gamma Q)A(\gamma Q^{-1})$, and both A and γQ are integral, QBQ^{-1} is also integral. Thus $QBQ^{-1} \in \mathrm{SU}(J_2,\mathcal{O}_K)$.

Since $\mathrm{SU}(J_1,\mathcal{O}_K)$ is a lattice and N is a finite-index subgroup, N is also a lattice in $\mathrm{SL}(\mathbb{R})$ with finite covolume. Thus QNQ^{-1} has finite covolume in $\mathrm{SL}(\mathbb{R})$ and is therefore a lattice. So QNQ^{-1} is a sublattice of $\mathrm{SU}(J_2,\mathcal{O}_K)$ and must have finite index by Margulis's theorem for lattices.

This shows that N is a finite-index subgroup of $\mathrm{SU}(J_1,\mathcal{O}_K)$ and QNQ^{-1} has finite index in $\mathrm{SU}(J_2,\mathcal{O}_K)$. □

So how does this lattice information apply to the Jones representations? Firstly, after rescaling and reparametrization the Jones representations can be made to have determinant ± 1, allowing the image to land in $\mathrm{PSU}(J, \phi, \mathcal{O}_K)$ instead of just $U(J, \phi, \mathcal{O}_K)$. Secondly, an arithmetic group theory result of Harish and Chandra that is formalized in our setting in Chapter 6 of Morris [11], states that $\mathrm{SU}_m(J, \phi, \mathcal{O}_K)$ is a lattice in $\mathrm{SL}_m(\mathbb{R})$ under the exact Salem number circumstances as required by Theorem 1.1. So Corollary 3.9 can be restated using this new vocabulary.

Corollary 4.5 *For each irreducible Jones representation, after a change of parameter, there are infinitely many Salem numbers s such that specializing q to a power of s maps into a lattice in $\mathrm{PSL}_m(\mathbb{R})$.*

Proof Let ρ_q be an irreducible Jones representation of dimension m. The images of the braid generators under ρ_q have determinant $\pm q^k$ for some $k \in \mathbb{N}$. After a change of variable $q = y^m$ and scaling the generators by $1/y^{m-k}$, this adjusted representation $\tilde{\rho}_y$ maps into $\mathrm{PSU}_m(J^y, \mathbb{Z}[y^{\pm 1}])$.

The subgroup B_n^{even} of even braids (the preimage of the alternating group under the standard projection to S_n) is a noncentral normal subgroup of B_n of finite index. The restriction $\tilde{\rho}_y|$ maps B_n^{even} into $\mathrm{SU}_m(J_y, \mathbb{Z}[y^{\pm 1}])$, and by Theorem 1.1 there exist infinitely many Salem numbers s such that the specialization $\rho_s|$ at $y = s$ is discrete. Further, by the results in Chapter 6 of [11] described above, these specializations make $\mathrm{SU}_m(J_s, \mathcal{O}_K)$ lattices in $\mathrm{SL}_m(\mathbb{R})$. Finite-index arguments imply $\mathrm{PSU}_m(J_s, \mathcal{O}_K)$ is a lattice in $\mathrm{PSL}_m(\mathbb{R})$. □

Since the goal is to obtain commensurable lattices as images of our Jones representations, and it is more natural to think of lattices in $\mathrm{SL}_m(\mathbb{R})$ instead of in $\mathrm{PSL}_m(\mathbb{R})$, one may simply pass to the finite-index subgroup B_n^{even} and continue to think only about lattices in $\mathrm{SL}_m(\mathbb{R})$. To apply Proposition 4.4 requires equivalent defining forms. In general, it is difficult to determine when two forms are equivalent. The following theorem gives a complete classification of the sesquilinear forms in a very specific algebraic setting that applies to the Salem number field scenario.

Theorem 4.6 (Scharlau [13, Chapter 10]) *If L is a global field and $K = L(\sqrt{\delta})$, sesquilinear forms over K/L are classified by dimension, determinant class and the signatures for those orderings of L for which δ is negative.*

This classification relies on the *determinant class*, which is defined here. Recall for a Salem number s the following tower of fields:

$$\begin{array}{ccc} \mathbb{Q}(s) = K & K^\times & \\ \Big| \, 2 & \Big\downarrow \text{Norm} & \\ \mathbb{Q}(s+1/s) = L & L^\times & (L^\times)^2 \subseteq \operatorname{Norm}(K^\times) \\ \Big| & & \\ \mathbb{Q} & & \end{array}$$

The Galois group of K/L is generated by ϕ, which maps s to $1/s$. There is a multiplicative group homomorphism $\operatorname{Norm}\colon K^\times \to L^\times$ given by $\operatorname{Norm}(\alpha) = \alpha\alpha^\phi$, where $K^\times = K - \{0\}$. Notice for $\beta \in L$ we have $\operatorname{Norm}(\beta) = \beta\beta^\phi = \beta^2$, and so $(L^\times)^2 \subseteq \operatorname{Norm}(K)$.

Definition 4.7 The *determinant class* of a sesquilinear form H over K/L is the coset of $\det(H)$ in $K^\times/\operatorname{Norm}(K^\times)$:

$$[\det(H)] = \det(H)\operatorname{Norm}(K).$$

Taking $\delta = (s-1/s)^2$, K can be rewritten as $K = L(\sqrt{\delta})$. Thus we can restate Scharlau's classification in the specific context of Salem numbers.

Theorem 4.8 (Scharlau restated) *Sesquilinear forms over K/L are classified by dimension, determinant class and the signatures for those orderings of L for which $(s-1/s)^2$ is negative.*

In odd dimensions it is very simple to show that *all* sesquilinear forms have the same determinant class, up to scaling. However, for even dimensions, the situation is very unclear.

Proposition 4.9 *For every odd-dimensional invertible sesquilinear matrices H and J over K, $[\det(H)] = [\det(\lambda J)]$ for $\lambda \in L$.*

Proof Let H and J be sesquilinear matrices over K of dimension $2k+1$. Since H and J are Hermitian, they are both diagonalizable with diagonal entries fixed by ϕ. So, the determinants of both H and J are elements in L. Let d_H and d_J denote the nonzero determinants of H and J. Thus,

$$d_H = \frac{d_H}{d_J} d_J \overset{\operatorname{mod}(L^\times)^2}{\equiv} \left(\frac{d_H}{d_J}\right)^{2k+1} d_J = \det\left(\frac{d_H}{d_J} J\right).$$

Since $(L^\times)^2 \subseteq \mathrm{Norm}(K)$, we have that H and λJ have the same determinant class for $\lambda = d_H/d_J \in L$. $\square$

As a result, to determine whether two forms of the same odd dimension are equivalent, it suffices to only check that they have the same signatures.

Theorem 4.10 *For J_t a sesquilinear form that is positive definite for t in a neighborhood η of 1, there are infinitely many Salem numbers s and integers n and m such that, in all odd dimensions, $\mathrm{SU}_{2k+1}(J_{s^n}, \phi, \mathcal{O}_K)$ and $\mathrm{SU}_{2k+1}(J_{s^m}, \phi, \mathcal{O}_K)$ are commensurable, discrete groups.*

Proof By Lemma 2.6 there are infinitely many Salem numbers s and integers n and m such that every complex Galois conjugate of s^m and s^n lies in η. Fix one such Salem number s, and K, L and δ as above.

By Theorem 4.6, sesquilinear forms are completely classified by dimension, determinant class, and the signatures for the places of L for which $(s-1/s)^2$ is negative. By Proposition 4.9, J_{s^n} and λJ_{s^m} have the same determinant class for λ in L, namely $\lambda = \det J_{s^n}/\det J_{s^m}$.

Let σ be a complex placement of L. Then $\sigma(s^m)$ is a complex Galois conjugates of s^m, and similarly for $\sigma(s^n)$ and s^n. Since n and m were chosen so that all of the complex Galois conjugates of s^m and s^n have arguments in η, $J_{\sigma(s^m)}$ and $J_{\sigma(s^n)}$ are positive definite. Moreover, $\det J_{s^n}/\det J_{s^m}$ and $\sigma(\det J_{s^n}/\det J_{s^m})$ are both positive, making $\lambda > 0$. So regardless of whether $\sigma((s-1/s)^2)$ is positive or negative, the forms $J_{\sigma(s^i)}$ have the same signature.

Therefore J_{s^n} is equivalent to λJ_{s^m}, and so $\mathrm{SU}(J_{s^n}, \phi, \mathcal{O}_K)$ is commensurable to $\mathrm{SU}(J_{s^m}, \phi, \mathcal{O}_K)$. The groups are discrete by Theorem 2.3. $\square$

Corollary 4.11 *Let $\rho_t\colon G \to \mathrm{SL}_{2k+1}(\mathbb{Z}[t, t^{-t}])$ be a group representation with a parameter t. Suppose there exists a matrix J_t such that:*

(1) *For all M in the image of ρ_t, $M^* J_t M = J_t$, where $M^*(t) = M^\mathsf{T}(1/t)$.*

(2) $J_t = (J_{1/t})^\mathsf{T}$.

(3) *$J_t \in \mathrm{GL}_m(\mathbb{Q}(t))$, where no entry of J_t has denominator with 1 as a root.*

(4) *J_t is positive definite for t in an neighborhood η of 1.*

Then there exist infinitely many Salem numbers s such that for infinitely many integers n and m the specializations ρ_{s^m} at $t = s^m$ and ρ_{s^n} at $t = s^n$ map into commensurable lattices of $\mathrm{SL}_{2k+1}(\mathbb{R})$.

Example 4.12 The reduced Burau representation of B_4 is 3–dimensional and, after the appropriate rescaling to have determinant 1, satisfies Corollary 4.11. So certain powers of the specializations in Example 3.2 map into commensurable lattices in $\mathrm{SL}_3(\mathbb{R})$.

References

[1] **J S Birman**, **H Wenzl**, *Braids, link polynomials and a new algebra*, Trans. Amer. Math. Soc. 313 (1989) 249–273 MR Zbl

[2] **O Brunat**, **K Magaard**, **I Marin**, *Image of the braid groups inside the finite Iwahori–Hecke algebras*, J. Reine Angew. Math. 733 (2017) 161–182 MR Zbl

[3] **O Brunat**, **I Marin**, *Image of the braid groups inside the finite Temperley–Lieb algebras*, Math. Z. 277 (2014) 651–664 MR Zbl

[4] **R D Budney**, *On the image of the Lawrence–Krammer representation*, J. Knot Theory Ramifications 14 (2005) 773–789 MR Zbl

[5] **C W Curtis**, **N Iwahori**, **R Kilmoyer**, *Hecke algebras and characters of parabolic type of finite groups with (B, N)–pairs*, Inst. Hautes Études Sci. Publ. Math. 40 (1971) 81–116 MR Zbl

[6] **M H Freedman**, **M J Larsen**, **Z Wang**, *The two-eigenvalue problem and density of Jones representation of braid groups*, Comm. Math. Phys. 228 (2002) 177–199 MR Zbl

[7] **L Funar**, **T Kohno**, *On Burau's representations at roots of unity*, Geom. Dedicata 169 (2014) 145–163 MR Zbl

[8] **V F R Jones**, *Hecke algebra representations of braid groups and link polynomials*, Ann. of Math. (2) 126 (1987) 335–388 MR Zbl

[9] **I Marin**, *Caractères de rigidité du groupe de Grothendieck–Teichmüller*, Compos. Math. 142 (2006) 657–678 MR Zbl

[10] **I Marin**, *Braids inside the Birman–Wenzl–Murakami algebra*, Algebr. Geom. Topol. 10 (2010) 1865–1886 MR Zbl

[11] **D W Morris**, *Introduction to arithmetic groups*, Deductive, Lethbridge, AB (2015) MR Zbl

[12] **J Murakami**, *The Kauffman polynomial of links and representation theory*, Osaka J. Math. 24 (1987) 745–758 MR Zbl

[13] **W Scharlau**, *Quadratic and Hermitian forms*, Grundl. Math. Wissen. 270, Springer (1985) MR Zbl

[14] **C C Squier**, *The Burau representation is unitary*, Proc. Amer. Math. Soc. 90 (1984) 199–202 MR Zbl

[15] **T N Venkataramana**, *Image of the Burau representation at d^{th} roots of unity*, Ann. of Math. (2) 179 (2014) 1041–1083 MR Zbl

[16] **H Wenzl**, *Hecke algebras of type A_n and subfactors*, Invent. Math. 92 (1988) 349–383 MR Zbl

[17] **H Wenzl**, *Quantum groups and subfactors of type B, C, and D*, Comm. Math. Phys. 133 (1990) 383–432 MR Zbl

[18] **Y Zhao**, *Young tableaux and the representations of the symmetric group*, Harvard College Math. Rev. 2 (2008) 33–45

[19] **M G Zinno**, *On Krammer's representation of the braid group*, Math. Ann. 321 (2001) 197–211 MR Zbl

ICERM, Brown University
Providence, RI, United States

nancy.scherich@gmail.com

Received: 13 October 2019 Revised: 3 December 2021

Algebraic & Geometric Topology **23**:5 (2023) 2029–2106
DOI: 10.2140/agt.2023.23.2029
Published: 25 July 2023

A model for configuration spaces of points

RICARDO CAMPOS
THOMAS WILLWACHER

We construct a real combinatorial model for the configuration spaces of points of compact smooth oriented manifolds without boundary. We use these models to show that the real homotopy type of configuration spaces of a simply connected such manifold only depends on the real homotopy type of the manifold.

Moreover, we show that for framed D–dimensional manifolds these models capture a natural right homotopy action of the little D–disks operad.

55P62, 81Q30; 16E45, 18M60, 18N40, 53C15

1 Introduction

Given a smooth manifold M, we study the configuration space of n nonoverlapping points on M,

$$\mathrm{Conf}_n(M) = \{(m_1, \dots, m_n) \in M^n \mid m_i \neq m_j \text{ for } i \neq j\}.$$

These spaces are classical objects in topology, whose homological and homotopical properties have been subject to intensive study over the decades. One of the first important results dates back to 1978 when Cohen and Taylor [15] constructed a spectral sequence converging to the cohomology $H^\bullet(\mathrm{Conf}_n(M))$. A different spectral sequence was constructed by Bendersky and Gitler [4] and both spectral sequences have been shown to coincide from the E^2 term on by Felix and Thomas [19]. In the particular case of smooth compact projective complex manifolds, it was shown by Totaro [44] that the Cohen–Taylor spectral sequence collapses after the second page, and Kriz [30] showed that for those manifolds the E^2 page is actually a model of $\mathrm{Conf}_n(M)$ in the sense of rational homotopy theory.

In this paper, we aim to understand the rational homotopy type of configuration spaces. Classical rational homotopy theory à la Sullivan [42] states that we can understand topological spaces via algebraic *models* which are differential graded commutative $\mathbb{K}$–algebras (dgca), where $\mathbb{K}$ is a field of characteristic zero. This roughly amounts to capturing the nontorsion part of the homotopy groups of such spaces. Usually, the field $\mathbb{K}$ is taken to be the rational numbers, but due to the transcendental methods we use, we take the base field $\mathbb{K} = \mathbb{R}$ to be the real numbers and we will therefore refer to the *real* homotopy type of configuration spaces.

Our first main result is the construction of a differential graded commutative $\mathbb{R}$–algebra model ${}^*\mathsf{Graphs}_M$ for $\mathrm{Conf}_\bullet(M)$, in the case when M is a D–dimensional compact smooth oriented manifold without boundary, with $D \geq 2$. Our model depends on M only through the following data:

- The cohomology $V = H^\bullet(M)$ as a vector space with a nondegenerate pairing of degree $D = \dim(M)$.
- The partition function Z_M of the "universal" perturbative AKSZ topological field theory on M. This is a Maurer–Cartan element in a certain graph complex only depending on V.

In particular, this shows that the latter perturbative invariants Z_M — special cases of which have been studied in the literature, eg by Bonechi, Cattaneo and Mnev [6] — contain at least as much information as the real homotopy type of $\mathrm{Conf}_\bullet(M)$. Furthermore, the real homotopy type of M is encoded in the tree-level components of Z_M. The higher loop order pieces of Z_M "indicate" (in a vague sense) the failure of the homotopy type of $\mathrm{Conf}_\bullet(M)$ to depend only on M. Finally, the real cohomology of $\mathrm{Conf}_\bullet(M)$ can be computed just from the tree level knowledge; see Section 7.

Now suppose that M is furthermore framed, ie the frame bundle of M is trivialized. Then the totality of spaces $\mathrm{Conf}_\bullet(M)$ carries additional algebraic structure, in that it can be endowed with a homotopy right action of the little D–disks operad as follows. First we consider the natural compactification $\mathsf{FM}_M(n)$ of $\mathrm{Conf}_n(M)$ introduced by Axelrod and Singer [2]; see also Sinha [41]. It is naturally acted upon from the right by the Fulton–MacPherson–Axelrod–Singer variant of the little disks operad FM_D introduced by Getzler and Jones in [21] by "insertion" of configurations of points.

The right E_D–module structure on configuration spaces has been receiving much interest in the last decade, since it has been realized that the homotopy theory of these right modules captures much of the homotopy theory of the underlying manifolds. For example, by the Goodwillie–Weiss embedding calculus the derived mapping spaces ("Ext's") of those right E_D–modules capture (under good technical conditions) the homotopy type of the embedding spaces of the underlying manifolds; see Boavida de Brito, Goodwillie and Weiss [9; 10; 22]. Dually, the factorization homology ("Tor's") of E_D–algebras has been widely studied and captures interesting properties of both the manifold and the E_D–algebra; see Ayala and Francis [3]. However, in order to use these tools in concrete situations it is important to have models for $\mathrm{Conf}_\bullet(M)$ (as a right Hopf E_D–module) that are computationally accessible, ie combinatorial. In this paper we provide such models.

Concretely, our second main result is that our model ${}^*\mathsf{Graphs}_M$ above combinatorially captures this action of the little D–disks operad as well, in the sense that it is a right Hopf operadic comodule over the Kontsevich Hopf cooperad ${}^*\mathsf{Graphs}_D$, modeling the topological little D–disks operad, and the combinatorially defined action models the topological action of E_D on $\mathrm{Conf}_\bullet(M)$.

In fact, one can consider the following "hierarchy" of invariants of a manifold M:

(1) The real (or rational) homotopy type of M.

(2) The real (or rational) homotopy types of $\mathsf{FM}_M(m)$ for $m = 1, 2, \ldots$.

(3) The real (or rational) homotopy type of FM_M considered as a right FM_D–module, for parallelized M. (For nonparallelizable M one may consider similarly the homotopy type of the FM–module of framed configuration spaces of points FFM_M.)

The relative strength of this invariants has been unknown. In particular, it is a long standing open problem if for simply connected M the rational homotopy type of $\mathrm{Conf}_\bullet(M)$ depends only on the rational homotopy type of M; see Félix, Halperin and Thomas [18, Problem 8, page 518] — see also Levitt [34] for a stronger conjecture disproved by Longoni and Salvatore in [37].

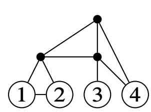
In our model the above hierarchy is nicely encoded in the loop order filtration on a certain graph complex GC_M, in which item (1) is encoded by the tree level piece of Z_M along with the cohomology of item (2), while the full Z_M encodes item (3).

Our third main result states that for a simply connected smooth closed framed manifold M, these invariants are of equal strength. We show furthermore that without the framed assumption, item (1) is still equally strong as item (2); thus establishing [18, Problem 8, page 518] under the assumption of smoothness.

Finally, if we consider a nonparallelized manifold there is still a way to make sense of the insertion of points at the boundary, but the price to pay is that one has to consider configurations of *framed points* in M. The resulting framed configuration spaces $\mathrm{Conf}^{\mathrm{fr}}_\bullet(M)$ then come equipped with a natural right action of the framed little disks operad E^{fr}_D. In Section 9 we present $\mathsf{BVGraphs}_M$, a natural modification of Graphs_M that encompasses the data of the frames and we show that if we consider a two-dimensional orientable manifold Σ, $\mathsf{BVGraphs}_\Sigma$ models this additional right action. In the framed case we restrict ourselves to the 2–dimensional setting.

Outline and statement of the main result

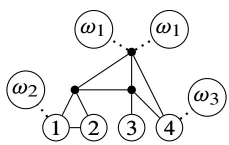
Let us summarize the construction and state the main result here. First recall from [27] the Kontsevich dg cooperad ${}^*\mathsf{Graphs}_D$. Elements of ${}^*\mathsf{Graphs}_D(r)$ consist of linear combinations of graphs with r numbered and an arbitrary number of unidentifiable vertices, like the following:



The precise definition of ${}^*\mathsf{Graphs}_D$ will be recalled in Section 3. The graphs contributing to ${}^*\mathsf{Graphs}_D$ may be interpreted as the nonvaccuum Feynman diagrams of the perturbative AKSZ σ–models on $\mathbb{R}^D$; see Alexandrov, Schwarz, Zaboronsky and Kontsevich [1].

Kontsevich constructs an explicit map ${}^*\mathsf{Graphs}_D \to \Omega_{\mathrm{PA}}(\mathsf{FM}_D)$ to the dgca of PA forms on the compactified configuration spaces FM_D. This map is compatible with the (co)operadic compositions, in the sense described in Section 3 below.

Now fix a smooth compact manifold M of dimension D, of which we pick an algebraic realization, so that we can talk about PA forms $\Omega_{\mathrm{PA}}(M)$. Then we consider a collection of dg commutative algebras ${}^*\mathsf{Graphs}_M(r)$. Elements of ${}^*\mathsf{Graphs}_M(r)$ are linear combinations of graphs, but with additional decorations of each vertex in the symmetric algebra $S(\tilde{H}(M))$ generated by the reduced cohomology $\tilde{H}(M)$. The following graph is an example, where we fixed some basis $\{\omega_j\}$ of $\tilde{H}(M)$:



These graphs may be interpreted as the nonvaccuum Feynman diagrams of the perturbative AKSZ σ–model on M. We equip the spaces ${}^*\mathsf{Graphs}_M(r)$ with a nontrivial differential built using the partition function Z_M of those field theories. This partition function can be considered as a special Maurer–Cartan element of a certain graph complex GC_M. Algebraically, the spaces ${}^*\mathsf{Graphs}_M(r)$ assemble into a right dg Hopf cooperadic comodule over the Hopf cooperad ${}^*\mathsf{Graphs}_D$.

By mimicking the Kontsevich construction, we construct, for a parallelized manifold M, a map of dg Hopf collections[1]

$${}^*\mathsf{Graphs}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M),$$

compatible with the (co)operadic (co)module structure, where we consider FM_M as equipped with the right FM_D–action. If M is not parallelized, we do not have an FM_D–action on FM_M. Nevertheless we may consider a (quasi-isomorphic) dg Hopf collection

$${}^*\mathsf{Graphs}_M^{\not\varnothing} \subset {}^*\mathsf{Graphs}_M$$

that still comes with a map of dg Hopf collections

$${}^*\mathsf{Graphs}_M^{\not\varnothing} \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M).$$

[1] A (dg) Hopf collection $\mathcal{C}$ for us is a sequence $\mathcal{C}(r)$ of dg commutative algebras, with actions of the symmetric groups S_r. A (dg) Hopf cooperad is a cooperad in dg commutative algebras.

Our first main result is the following.

Theorem 1 *The map* ${}^*\mathsf{Graphs}_M^{\not\varnothing} \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ *is a quasi-isomorphism of dg Hopf collections. In the parallelized case the map* ${}^*\mathsf{Graphs}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ *is a quasi-isomorphism of dg Hopf collections*, *compatible with the* (*co*)*operadic* (*co*)*module structures.*

This result provides us with explicit combinatorial dgca models for configuration spaces of points, compatible with the right E_D–action on these configuration spaces in the parallelizable setting. An extension to the nonparallelized case is provided in Section 9, albeit only in dimension $D = 2$.

We note that our model ${}^*\mathsf{Graphs}_M$ depends on M only through the partition function $Z_M \in \mathrm{GC}_M$. The tree part of this partition function encodes the real homotopy type of M. The loop parts encode invariants of M. Now, simple degree counting arguments may be used to severely restrict the possible graphs occurring in M. In particular, one finds that if $H^1(M,\mathbb{R})$ vanishes, then for $D \geq 4$ there are no contributions to Z_M of positive loop order, and one hence arrives at the following result.

Corollary 2 (Theorem 63) *Let M be an orientable compact manifold without boundary of dimension $D \geq 4$, such that $H^1(M,\mathbb{R}) = 0$. Then the* (*naive*)[2] *real homotopy type of* $\mathrm{Conf}_\bullet(M)$ *depends only on the* (*naive*) *real homotopy type of M.*

For $D = 2$ the analogous statement is empty, as there is only one connected manifold satisfying the assumption. If we replace the condition $H^1(M,\mathbb{R}) = 0$ by the stronger condition of simple connectivity, the statement is also true in dimension 3, but for the trivial reason that by the Poincaré conjecture there is only one simply connected manifold M in dimension 3. Hence the above result also solves the real version of the long standing question in algebraic topology of whether for simply connected M the rational homotopy type of the configuration space of points on M is determined by the rational homotopy type of M; see [18, Problem 8, page 518]

Remark 3 Our result also shows that the "perturbative AKSZ"-invariant Z_M is at least as strong as the invariant of M given by the totality of the real homotopy types of the configuration spaces of M, considered as right E_D–modules. The latter "invariant"

[2]We call the naive real homotopy type the quasi-isomorphism type of the dg commutative algebra of (PL or smooth) forms. Note that in the nonsimply connected case this definition is not the correct one; one should rather consider the real homotopy type of the universal cover with the action of the fundamental group. We do not consider this better notion here, and in this paper "real homotopy type" shall always refer to the naive real homotopy type.

is the data entering the factorization or "manifoldic" homology — see [3] and Markarian and Tanaka [38] — and the Goodwillie–Weiss calculus [22] (over the reals). Conversely, from the fact that the models ${}^*\mathsf{Graphs}_M$ encode the real homotopy type of configuration spaces, one may see that the expectation values of the perturbative AKSZ theories on M may be expressed through the factorization homology of M. However, we will leave the physical interpretation to forthcoming work and focus here on the algebraic-topological goal of providing models for configuration spaces.

Remark 4 After the first version of this article appeared on the arXiv, Idrissi [25] obtained results very similar to ours by showing that for simply connected closed oriented manifolds the Lambrechts–Stanley dg model [32] is actually a real model of $\mathrm{Conf}_n(M)$. We sketch in Appendix A how this latter statement can also be obtained as a consequence of our main results.

Plan of the paper

In Section 2 we introduce the spaces FM_M, the compactifications of configuration spaces of points on a smooth manifold M (with $D = \dim M$) and its semialgebraic realizations and adapt results in the literature to construct the propagator.

Starting with the framed case, in Section 3 we construct the first graph complex ${}^*\mathsf{Gra}_M$ (a Hopf ${}^*\mathsf{Gra}_D$–comodule) and construct the map into $\Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ which is not yet a quasi-isomorphism. In Section 4 we use operadic twisting to obtain the graph complex ${}^*\mathsf{Graphs}_M$ and in Section 5 we show that ${}^*\mathsf{Graphs}_M$ is indeed a model for the real homotopy type of FM_M as a right FM_D–module.

In Section 6 we construct a no-tadpole variant of the graph complex to deal with the case where M is not parallelized and show that it models the real homotopy type of the collection of topological spaces FM_M, concluding the proof of Theorem 1.

The next goal is to study the dependence of the homotopy type of the configuration spaces on the base manifold. In Sections 7 and 8 we study the partition function Z_M that gives rise to the differential in ${}^*\mathsf{Graphs}_M$ and we show that it is gauge equivalent to one vanishing on graphs containing ≤ 2–valent vertices. We conclude that in good conditions the real homotopy type of M can be recovered from the tree piece of the graph complex, thus proving Corollary 2.

Finally, in the last section we construct a graphical model of configuration spaces of framed points in 2–dimensions, together with the action of the framed little disks operad.

1.1 Notation and conventions

Throughout the text all algebraic objects (vector spaces, algebras, operads, etc) are differential graded (or just dg) and are defined over the field $\mathbb{R}$.

We use cohomological conventions, ie all differentials have degree $+1$. We use the language of operads and follow mostly the conventions of Loday and Vallette's textbook [36]. One notable exception is that we denote the k–fold operadic (de)suspension of an operad $\mathcal{P}$ by $\Lambda^k\mathcal{P}$.

Acknowledgements

We would like to thank Pascal Lambrechts for useful remarks and references, and Najib Idrissi, Nils Prigge and Victor Turchin for valuable discussions and for pointing out some mistakes in the original version. Both authors have been supported by the Swiss National Science Foundation, grant 200021_150012, by the NCCR SwissMAP funded by the Swiss National Science Foundation, and the ERC starting grant 678156 (GRAPHCPX).

2 Compactified configuration spaces

2.1 Semialgebraic manifolds

Given a compact semialgebraic set X one can consider its dgca of piecewise semialgebraic (PA) forms, $\Omega_{\mathrm{PA}}(X)$, which is quasi-isomorphic to Sullivan's dgca of piecewise polynomial forms [24; 29].

Dually, one can also consider its complex of semialgebraic chains, which we denote by $\mathrm{Chains}(X)$, which is also quasi-isomorphic to the usual complex of singular chains.

By the Nash–Tognoli theorem [43] — see also [5, Section 14] — any smooth compact manifold is diffeomorphic to a component of a nonsingular algebraic subset of $\mathbb{R}^N$ for some N. In particular, any such manifold can be realized as a smooth semialgebraic (ie Nash) submanifold of Euclidean space. Throughout this paper, whenever we consider a closed smooth manifold M we will consider implicitly a chosen such realization of M as a Nash submanifold of $\mathbb{R}^N$.

We refer to [5] for an introduction to real algebraic geometry. An overview is also contained in the introductory sections of [24].

Even though all manifolds considered in this paper will be smooth, it is not sufficient for our purposes to consider the de Rham complex. The main reason for this is that we would like to consider fiber integration over nonsmooth fiber bundles $E \to B$. Nonetheless, the relevant bundles will be SA (semialgebraic) bundles [24] and, for such bundles, there is a pushforward map $\Omega_{\mathrm{min}}(E) \to \Omega_{\mathrm{PA}}(B)$, where $\Omega_{\mathrm{min}}(M) \subset \Omega_{\mathrm{PA}}(M)$ is the (nonquasi-isomorphic) subalgebra of minimal forms.

While this pushforward cannot be naturally extended to the whole space of PA forms $\Omega_{\mathrm{PA}}(E)$, as described in Appendix C, we can consider a subalgebra of trivial forms $\Omega_{\mathrm{triv}}(E)$, sitting between $\Omega_{\mathrm{min}}(E)$ and $\Omega_{\mathrm{PA}}(E)$ and quasi-isomorphic to $\Omega_{\mathrm{PA}}(E)$, such that the pushforward extends naturally to a map $\Omega_{\mathrm{triv}}(E) \to \Omega_{\mathrm{PA}}(B)$.

2.2 Configuration spaces of points in $\mathbb{R}^D$

Let D be a positive integer. We will use the Fulton–MacPherson topological operad FM_D that was introduced by Getzler and Jones [21]. Its n–ary space $\mathsf{FM}_D(n)$ is a suitable compactification of the quotient of the configuration space

$$\mathrm{Conf}_n(\mathbb{R}^D)/(\mathbb{R}_{>0} \ltimes \mathbb{R}^D),$$

with the Lie group $\mathbb{R}_{>0} \ltimes \mathbb{R}^D$ acting by scaling and translations. For $n > 1$ the spaces $\mathsf{FM}_D(n)$ are $(Dn-D-1)$–dimensional manifolds with corners whose boundary strata represent sets of points getting infinitely close.

The first few terms are[3]

$$\mathsf{FM}_D(0) = \{*\}, \quad \mathsf{FM}_D(1) = \{*\}, \quad \mathsf{FM}_D(2) = S^{D-1}.$$

The operadic composition $\circ_i$ is given by inserting a configuration at the boundary stratum at the point labeled by i. A thorough study of these operads can be found in [33].

The operad FM_D can be related to a shifted version of the homotopy Lie operad via the operad morphism

$$\Lambda^{D-1} L_\infty \to \mathrm{Chains}(\mathsf{FM}_D), \tag{1}$$

given by sending the generator $\mu_n \in \Lambda^{D-1} L_\infty(n)$ to the fundamental chain of $\mathsf{FM}_D(n)$, ie the semialgebraic chain corresponding to $\mathsf{FM}_D(n)$ as a submanifold of itself.[4]

[3] We work with the unital version of the Fulton–MacPherson operad.

[4] Recall that due to our cohomological conventions these spaces live in nonpositive degree. In particular, the generator $\mu_n \in L_\infty$ has degree $2-n$.

2.3 Configuration spaces of points on a manifold

Let M be a closed smooth oriented manifold of dimension D. We denote by $\mathrm{Conf}_n(M)$, the configuration space of n points in M. Concretely, $\mathrm{Conf}_n(M) = M^n - \Delta$, where Δ is the fat (or long) diagonal $\Delta = \{(m_1, \dots, m_n) \in M^n \mid \exists i \neq j : m_i = m_j\}$.

The Fulton–MacPherson–Axelrod–Singer compactification of $\mathrm{Conf}_n(M)$ is a smooth manifold with corners $\mathsf{FM}_M(n)$ whose boundary strata correspond to nested groups of points becoming "infinitely close"; see [41] for more details and a precise definition. Since the inclusion $\mathrm{Conf}_n(M) \hookrightarrow \mathsf{FM}_M(n)$ is a homotopy equivalence, we work preferably with $\mathsf{FM}_M(n)$ as these spaces have a richer structure.

Convention 5 (semialgebraicity of $\mathsf{FM}_M(n)$) The choice of semialgebraic structure on $\mathsf{FM}_M(n)$ is done in a way compatible with the one from M as follows: Let us consider the chosen semialgebraic realization of the manifold M in $\mathbb{R}^N$ for some N.

For $1 \leq i \neq j \leq n$, let $\theta_{i,j} : \mathrm{Conf}_n(M) \to S^{N-1}$ be defined by

$$\theta_{i,j}((x_1, \dots, x_n)) = \frac{x_i - x_j}{\|x_i - x_j\|_{\mathbb{R}^N}}.$$

For $1 \leq i \neq j \neq k \leq n$, we define $d_{i,j,k} : \mathrm{Conf}_n(M) \to (0, +\infty)$ by

$$d_{i,j,k}((x_1, \dots, x_n)) = \frac{\|x_i - x_j\|}{\|x_i - x_k\|}.$$

Considering all possibilities of i, j and k, we have defined a natural embedding

$$\iota : \mathrm{Conf}_n(M) \to M^n \times (S^{N-1})^{n(n-1)} \times [0, +\infty]^{n(n-1)(n-2)}.$$

We define $\mathsf{FM}_M(n)$ as the closure $\iota(\mathrm{Conf}_n(M))$, thus inheriting a semialgebraic structure.

Remark 6 (SA bundles) For every $m > n$ there are various projection maps

$$\mathsf{FM}_M(m) \to \mathsf{FM}_M(n)$$

corresponding to forgetting $m - n$ of the points. These maps are not smooth fiber bundles, but they are SA (semialgebraic) bundles [24], which allows us to consider pushforwards (fiber integration) of forms along these maps.

The proof of this fact is a straightforward adaptation of the proof of the same fact for FM_D done in [33, Section 5.9]. In this case one starts instead by associating to a configuration in $\mathsf{FM}_M(n)$ a configuration of nested disks in M.

Convention 7 From here onward, we fix representatives of the cohomology of M, ie we fix an embedding

$$\iota\colon H^\bullet(M) \hookrightarrow \Omega^\bullet_{\mathrm{triv}}(M) \tag{2}$$

that yields a right inverse of the projection from closed forms to cohomology.

2.3.1 The diagonal class Since M is compact and oriented, the pairing

$$\int\colon H^\bullet(M) \otimes H^\bullet(M) \to \mathbb{R}, \quad (\omega, \nu) \mapsto \int_M \omega \wedge \nu$$

given by Poincaré duality is nondegenerate. We shall also consider a version of this pairing which is antisymmetric for odd D,

$$\langle \omega, \nu \rangle = (-1)^{D\ \deg(\nu)} \int_M \omega \wedge \nu.$$

The diagonal map $\Delta\colon M \to M \times M$ defines an element in $H_\bullet(M \times M)$ and its dual under Poincaré duality is called the diagonal class, which is also denoted by $\Delta \in H^\bullet(M \times M) = H^\bullet(M) \otimes H^\bullet(M)$.

If we pick a homogeneous basis $e_1, \dots, e_k$ of $H^\bullet(M)$, we have $\Delta = \sum_{i,j} g^{ij} e_i \otimes e_j$, where (g^{ij}) is the matrix inverse to the Poincaré duality pairing $\langle -, - \rangle$. Alternatively, this can also be written as $\Delta = \sum_i (-1)^{\deg(e_i)} e_i \otimes e_i^*$, where $\{e_i^*\}$ is the dual basis of $\{e_i\}$ with respect to $(-,-)$.

In $\mathsf{FM}_M(2)$, if we consider the case in which the two points come infinitely close to one another, we obtain a map $\partial \mathsf{FM}_M(2) \to M \cong \Delta \subset M \times M$ which is a sphere bundle (with S^{D-1} fibers). Notice that $\partial \mathsf{FM}_M(2)$ can be identified with $ST(M)$, the sphere tangent bundle of M.

The following proposition can essentially be found in the literature — see for instance [7, Section 3; 14; 13, Lemma 2] — we only have to apply minor modifications in order to work in the semialgebraic setting.

Proposition 8 *Let $p_1\colon \mathsf{FM}_M(2) \to M$ (resp. $p_2\colon \mathsf{FM}_M(2) \to M$) be the map that forgets the point labeled by* 2 *(resp.* 1*) from a configuration. Then there is a form $\phi_{12} \in \Omega^{D-1}_{\mathrm{triv}}(\mathsf{FM}_M(2))$ satisfying the following properties*:

(i) $d\phi_{12} = p_1^* \wedge p_2^*(\Delta) = \sum_{i,j} g^{ij} p_1^*(e_i) \wedge p_2^*(e_j) \in \Omega^D_{\mathrm{triv}}(\mathsf{FM}_M(2))$.

(ii) *The fiber integral of the restriction of ϕ_{12} to $\partial \mathsf{FM}_M(2)$ is equal to $(-1)^D$. (We then say that this restriction is a global angular form.) Additionally, if $D = 2$,* *the restriction of ϕ_{12} to every fiber of the circle bundle yields a round volume form of that circle, with respect to some metric.*

(iii) *The form ϕ_{12} is symmetric with respect to the $\mathbb{Z}_2$–action induced by swapping points 1 and 2 for D even and antisymmetric for D odd.*

(iv) *For any $\alpha \in H^\bullet(M)$,*
$$\int_2 \phi_{12}\, p_2^* \iota(\alpha) = 0,$$
where ι is as in (2) and the integral is along the fiber of p_1, ie one integrates out the second coordinate.

Notice that the form $\phi_{12} \in \Omega^{D-1}_{\mathrm{triv}}(\mathsf{FM}_M(2))$ is also called the propagator in the literature. More generally, we consider the forms $\phi_{ij} \in \Omega^{D-1}_{\mathrm{triv}}(\mathsf{FM}_M(n))$ to be $p_{ij}^*(\phi_{12})$, where $p_{ij} : \mathsf{FM}_M(n) \to \mathsf{FM}_M(2)$ is the projection map that remembers only the points labeled i and j.

Proof Let $\psi \in \Omega^{D-1}_{\mathrm{triv}}(\partial\mathsf{FM}_M(2))$ be a global angular form of the sphere bundle. Such a form always exists — see for example [8] where such construction is made in the smooth case — but the argument can be adapted to the semialgebraic case. It is also shown in [8] that for a circle bundle the global angular form can be chosen to restrict to the standard volume form on each fiber. Moreover, the differential of such a form is basic (it is the pullback of the Euler class of the sphere bundle). Let E be a collar neighborhood of $\partial\mathsf{FM}_M(2)$ inside $\mathsf{FM}_M(2)$. (See [40, Lemma VI.1.6] for the existence of a semialgebraic (even Nash) collar.) Let us extend the form ψ to E by pulling it back along the projection $E \to \partial\mathsf{FM}_M(2)$. We can consider a semialgebraic cutoff function $\rho\colon \mathsf{FM}_M(2) \to \mathbb{R}$ such that ρ is constant equal to zero outside of E and is constant equal to 1 in some open set U such that $\partial\mathsf{FM}_M(2) \subset U \subset E$. We can therefore consider the well-defined form $\rho\psi \in \Omega^{D-1}_{\mathrm{triv}}(\mathsf{FM}_M(2))$.

Since $d(\rho\psi)|_{\partial\mathsf{FM}_M(2)} = d\psi$ is basic, the form $d(\rho\psi) \in \Omega^{D}_{\mathrm{triv}}(\mathsf{FM}_M(2))$ induces a form in $\Omega^{D}_{\mathrm{triv}}(M \times M)$, still denoted by $d(\rho\psi)$. This form is clearly closed, but not necessarily exact, as $\rho\psi$ itself might not extend to the boundary.

Let $\omega \in H^\bullet(M \times M) \subset \Omega_{\mathrm{triv}}(M \times M)$. Then

$$\int_{M\times M} \omega\, d(\rho\psi) = \int_{\mathsf{FM}_M(2)} \omega\, d(\rho\psi) = (-1)^D \int_{\partial\mathsf{FM}_M(2)} \omega\rho\psi = \int_{\Delta\cong M} \omega. \tag{3}$$

It follows that the cohomology class of $d(\rho\psi)$ is the Poincaré dual of the diagonal $\Delta \cong M \subset M\times M$. Thus $p_1^* \wedge p_2^*(\Delta)$ and $d(\rho\psi)$ are cohomologous in $\Omega^{D}_{\mathrm{triv}}(M\times M)$. It follows that there exists a form $\beta \in \Omega^{D-1}_{\mathrm{triv}}(M \times M)$ such that $d\beta = p_1^* \wedge p_2^*(\Delta) - d(\rho\psi)$. We define the form $\phi_{12} \in \Omega^{D-1}_{\mathrm{triv}}(\mathsf{FM}_M(2))$ to be $\pi^*\beta + \rho\psi$, where

$$\pi\colon \mathsf{FM}_M(2) \to M \times M$$

is the projection. It is clear that ϕ_{12} satisfies property (i) and since the restriction of $\pi^*\beta$ to the boundary is a basic form and properties (ii) is preserved.

To ensure (iv) one can replace the ϕ_{12} constructed so far by

$$\phi_{12} - \int_3 \phi_{13} p_{23}^* \Delta - \int_3 \phi_{23} p_{13}^* \Delta + \int_{3,4} \phi_{34} (p_{13}^* \Delta)(p_{24}^* \Delta),$$

where p_{ij} is the forgetful map, forgetting all but points i and j from a configuration of points. We refer the reader to [13] where more details can be found. (The reference contains a construction of the propagator in the smooth setting, but the trick to ensure (iv) is verbatim identical in our semialgebraic setup.)

Finally, we can (anti)symmetrize ϕ_{12} to ensure it satisfies property (iii), while preserving the other properties. □

Remark 9 For M parallelizable, we can (and will) require a stronger version of property (ii). A parallelization is a choice of a trivialization $\partial \mathsf{FM}_M(2) \simeq M \times S^{D-1}$ and given such a parallelization, in the proof of the previous proposition we can take $\psi = \pi^*(\omega_{S^{D-1}}) \in \Omega^{D-1}_{\mathrm{triv}}(M \times S^{D-1})$, the pullback of the standard volume form of S^{D-1} via the projection $\pi\colon M \times S^{D-1} \to S^{D-1}$. By construction of ϕ_{12} the restriction of ϕ_{12} to $\partial \mathsf{FM}_M(2)$ has the form

$$\phi_{12}|_{\partial \mathsf{FM}_M(2)} = \psi + p^*\eta, \tag{4}$$

where $p\colon \partial \mathsf{FM}_M(2) \to M$ is the projection to the base and $\eta \in \Omega_{\mathrm{triv}}(M)$ is some form on the base. Note in particular that from the closedness of ψ and condition (i) above, it follows that

$$d\eta = \Delta_M, \tag{5}$$

where $\Delta_M \in \Omega^D_{\mathrm{triv}}(M)$ denotes the pullback of Δ along the diagonal map (ie the wedge product of its components).

3 The Cattaneo–Felder–Mnev graph complex and operad

Let n, N and D be positive integers and let V be an N–dimensional graded vector space with a nondegenerate pairing of degree $-D$; $\langle \cdot, \cdot \rangle\colon V \otimes V \to \mathbb{R}[-D]$. We

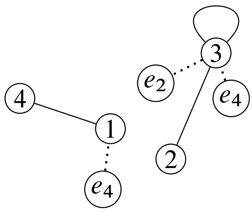

$$= \pm s^{41} e_4^1 s^{23} s^{33} e_2^3 e_4^3$$

Figure 1: An example of a graph describing an element in ${}^*\mathsf{Gra}_V(4)$.

require that for all homogeneous $x, y \in V$ of degrees k and l the pairing satisfies the (anti)symmetry condition $\langle x, y\rangle = (-1)^{kl+D}\langle y, x\rangle$. Moreover, we assume V to be "augmented" in the sense that we are also given a canonical decomposition $V = \mathbb{R} \oplus \overline{V}$. One should keep in mind the example of the Poincaré pairing on the cohomology of a connected N–manifold.

Let $e_2, e_3, \dots, e_N$ be a graded basis of $\overline{V}$ and for convenience of notation we define $e_1 = 1 \in \mathbb{R}$. We consider the free graded commutative algebra generated by symbols s^{ij} of degree $D-1$, where $1 \le i, j \le n$, $s^{ij} = (-1)^D s^{ji}$, and symbols $e_1^j, \dots, e_N^j$ for $j = 1, \dots, n$ of the same degrees as the elements of the basis $e_1, \dots, e_N$. We define a differential on it by the rules

$$de_\alpha^j = 0, \quad ds^{ij} = \sum_{\alpha,\beta} g^{\alpha\beta} e_\alpha^i e_\beta^j,$$

where g^{kl} is the inverse of the matrix describing the pairing on V. (So $\sum_{\alpha,\beta} g^{\alpha\beta} e_\alpha^i e_\beta^j$ is the "diagonal class".)

We define the dgca ${}^*\mathsf{Gra}_V(n)$ as the quotient of this algebra by the sub-dgca generated by elements of the form $e_1^j - 1$. Notice that there is a natural right action of the symmetric group $\mathbb{S}_n$ on ${}^*\mathsf{Gra}_V(n)$ by permuting the superscript indices (the i and j above) running from 1 to n.

Remark 10 All definitions are independent of the choice of graded basis of V and can be given in a basis-free way.

Remark 11 The notation ${}^*\mathsf{Gra}_V(n)$ stands for "predual graphs" as one may represent elements of ${}^*\mathsf{Gra}_V(n)$ as linear combinations of decorated directed graphs with n vertices and an ordering of the edges. The decorations are elements of V that may be attached to vertices; see Figure 1. Each such graph corresponds to monomial in ${}^*\mathsf{Gra}_V(n)$, an edge between vertices i and j corresponds to one occurrence of s^{ij} and a decoration by an element $e_\alpha \in V$ at vertex j corresponds to one occurrence of e_α^j.

Directions of the edges and their ordering might be ignored, keeping in mind that then a graph is only well defined up to a ± 1 prefactor. Notice that while both tadpoles and double edges are allowed, for (anti)symmetry reasons one has that $s^{ii} = 0$ if D is odd and $s^{ij}s^{ij} = 0$ if D is even.

3.1 Cooperadic comodule structure

Definition 12 Let D be a positive integer. For $n \geq 2$, the space ${}^*\mathsf{Gra}_D(n)$ is defined to be the free graded commutative algebra generated by symbols s^{ij} in degree $D-1$, for $i \neq j$, quotiented by the relations $s^{ij} = (-1)^D s^{ji}$. We set ${}^*\mathsf{Gra}_D(0) = {}^*\mathsf{Gra}_D(1) = \mathbb{R}$.

As before, the spaces ${}^*\mathsf{Gra}_D(n)$ can be seen as the span of undecorated graphs such that every edge has degree $D-1$.

Proposition 13 *The spaces ${}^*\mathsf{Gra}_D(n)$ form a cooperad in dg commutative algebras. The cooperadic structure is given by removal (contraction) of subgraphs; ie for $\Gamma \in {}^*\mathsf{Gra}(n)$, the component of $\Delta(\Gamma)$ in ${}^*\mathsf{Gra}_D(k) \otimes {}^*\mathsf{Gra}_D(i_1) \otimes \cdots \otimes {}^*\mathsf{Gra}_D(i_k)$ is*

$$\sum \pm \Gamma' \otimes \Gamma_1 \otimes \cdots \otimes \Gamma_k, \tag{6}$$

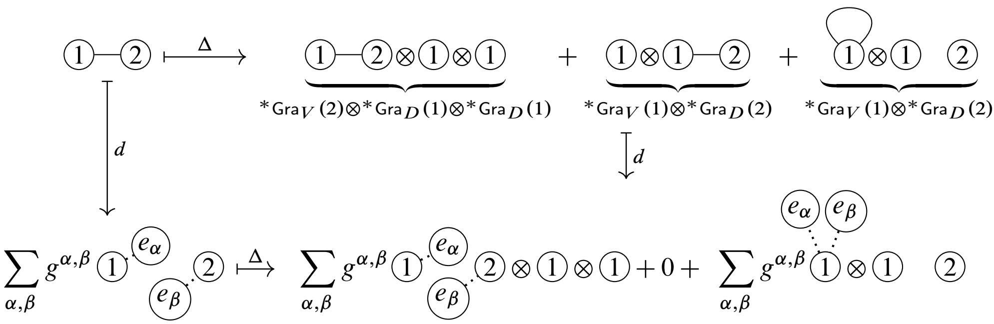
where the sum runs over all $(k+1)$–tuples $(\Gamma', \Gamma_1, \ldots, \Gamma_k)$ such that when each graph Γ_i is inserted at the vertex i of Γ', there is a way of reconnecting the loose edges such that one obtains Γ.

To obtain the appropriate signs one has to consider the full data of graphs with an ordering of oriented edges. In this situation the orientation of the edges of Γ is preserved and one uses the symmetry relations on Γ in such a way that for all $i = 1, \ldots, k$, the labels of the edges of the subgraph Γ_i come before the labels of the edges of the subgraph Γ_{i+1} and all of those come before the labels of the edges of the subgraphs Γ'. Notice that if one of the $i_j = 0$, the cooperadic cocomposition is given by adding a disconnected vertex to Γ' [20, Section 2.2.1]. The cooperad axioms are a straightforward verification.

Proposition 14 *The dg commutative algebras ${}^*\mathsf{Gra}_V(n)$ for $n = 1, 2, \ldots$ assemble to form a cooperadic right ${}^*\mathsf{Gra}_D$–comodule ${}^*\mathsf{Gra}_V$ in dg commutative algebras.*

Proof The cooperadic coactions are defined through formulas similar to (6), and proof of the associativity axiom is formally the same as the proof of the previous proposition.

To show that the differential respects the comodule structure it suffices to check this on generators of the commutative algebra. This is clear for decorations e^i_α and for tadpoles s^{ii}. For edges connecting different vertices let us do the verification for $s^{12} \in {}^*\mathsf{Gra}_V(2)$ for simplicity of notation. The only nontrivial commutative diagram to check is



where the vertical arrows correspond to the differential and the horizontal ones to the coaction. $\square$

3.2 Forms on (closed) manifolds

Let M be a closed smooth framed connected manifold of dimension D and let FM_M be the Fulton–MacPherson compactification of the spaces of configurations of points of M as described in Section 2. It is naturally an operadic right module over the operad FM_D, where the i^{th} composition of $c \in \mathsf{FM}_D(k)$ in a configuration $\bar{c} \in \mathsf{FM}_M(n)$ corresponds to the insertion of the configuration c at the i^{th} point of $\bar{c}$. The parallelization of the manifold ensures that this insertion can be made in a consistent way.

It follows that $\Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ is naturally equipped with a right cooperadic coaction of the cooperad (in dg commutative algebras) $\Omega_{\mathrm{PA}}(\mathsf{FM}_D)$ (mind Remark 15 below). The coaction is obtained from the restriction of forms to boundary strata where multiple points collide.

There is a map of ("almost") cooperads in dg commutative algebras

$$ {}^*\mathsf{Gra}_D \to \Omega_{\mathrm{PA}}(\mathsf{FM}_D) \tag{7}$$

given by associating to every edge the angle form relative to the two incident vertices [28; 33]. More explicitly, one considers the standard volume form

$$\phi_{12} \in \Omega_{\mathrm{PA}}^{D-1}(S^{D-1}) = \Omega_{\mathrm{PA}}^{D-1}(\mathsf{FM}_D(2)),$$

which plays the role of the propagator. The forms $\phi_{ij} \in \Omega_{\mathrm{PA}}^{D-1}(\mathsf{FM}_D(2))$ are defined by pulling back ϕ_{12} by the appropriate projection map. Finally, the map (7) above is obtained by extending the assignment $s^{ij} \mapsto \phi_{ij}$ to a map of dgcas.

Remark 15 The functor Ω_{PA} is not comonoidal since the canonical map

$$\Omega_{\mathrm{PA}}(A)\otimes\Omega_{\mathrm{PA}}(B)\to\Omega_{\mathrm{PA}}(A\times B)$$

goes "in the wrong direction"; therefore $\Omega_{\mathrm{PA}}(\mathsf{FM}_D)$ is not a cooperad. Nevertheless, by abuse of language throughout this paper we will refer to maps such as (7) as maps of cooperads (or cooperadic modules) if they satisfy a compatibility relation such as commutativity of the diagram

$$\begin{tikzcd}
{}^*\mathsf{Gra}_D(n) \arrow[rr] \arrow[dd] & & \Omega_{\mathrm{PA}}(\mathsf{FM}_D(n)) \arrow[d] \\
& & \Omega_{\mathrm{PA}}(\mathsf{FM}_D(n-k+1)\times\mathsf{FM}_D(k)) \\
{}^*\mathsf{Gra}_D(n-k+1)\otimes{}^*\mathsf{Gra}_D(k) \arrow[rr] & & \Omega_{\mathrm{PA}}(\mathsf{FM}_D(n-k+1))\otimes\Omega_{\mathrm{PA}}(\mathsf{FM}_D(k)) \arrow[u]
\end{tikzcd}$$

Since M is connected, its cohomology $H^\bullet(M)$ has a canonical augmentation given by the constant functions on M, and since M is closed, Poincaré duality gives us a pairing on $H^\bullet(M)$ of degree $-D$. We define, for any manifold M,

$${}^*\mathsf{Gra}_M := {}^*\mathsf{Gra}_{H^\bullet(M)}.$$

Let us denote by $\iota\colon H^\bullet(M)\hookrightarrow\Omega_{\mathrm{triv}}(M)$ the embedding from Convention 7; that is, for every $\omega\in H^\bullet(M)$, $\iota(\omega)$ is a representative of the class ω. Following Cattaneo and Mnev [13], we can define a map of dg commutative algebras (which a priori depends on various pieces of data)

$$(8)\qquad {}^*\mathsf{Gra}_M\to\Omega_{\mathrm{triv}}(\mathsf{FM}_M)\subset\Omega_{\mathrm{PA}}(\mathsf{FM}_M)$$

as follows: The map sends the generator s^{ij} for $i\neq j$ to ϕ_{ij}, where ϕ_{ij} is the form constructed in the discussion preceding the proof of Proposition 8 with the additional assumption from Remark 9. The map sends the decoration $\omega\in H^\bullet(M)$ on the j^{th} vertex to $\omega^j\in{}^*\mathsf{Gra}_D$ to $p_j^*(\iota(\omega))$, where $p_j\colon\mathsf{FM}_M\to M$ is the map that remembers only the point labeled by j. Finally the map sends s^{jj} to $p_j^*\eta$, where η is as in (4).

Lemma 16 *The map ${}^*\mathsf{Gra}_M\to\Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ is a map of dg Hopf collections, compatible with the cooperadic comodule structures along the map ${}^*\mathsf{Gra}_D\to\Omega_{\mathrm{PA}}(\mathsf{FM}_D)$, in the sense of Remark 15. In other words there is a map of 2–colored dg Hopf collections,*

$${}^*\mathsf{Gra}_M \circlearrowright {}^*\mathsf{Gra}_D\to\Omega_{\mathrm{PA}}(\mathsf{FM}_M)\circlearrowleft\Omega_{\mathrm{PA}}(\mathsf{FM}_D),$$

compatible with the (2–colored) cooperadic cocompositions.

Proof The compatibility with the differentials is clear for every generator of ${}^*\mathsf{Gra}_M$ except possibly s^{jj}, for which one uses (5). By definition the map consists of morphisms of commutative algebras; therefore it is enough to check the compatibility of the cocompositions on generators. For elements e^j_α this is clear. For the other generators we will sketch the verification for the case of $s^{12} \in {}^*\mathsf{Gra}_M(2)$ for simplicity of notation.

The composition map in $(\mathsf{FM}_M, \mathsf{FM}_D)$ is done by insertion at the boundary stratum. Since the cocomposition map $\Omega_{\mathrm{PA}}(\mathsf{FM}_M) \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M) \circ \Omega_{\mathrm{PA}}(\mathsf{FM}_D)$ is given by the pullback of the composition map we get, using (4),[5]

$$\phi_{12} \in \Omega_{\mathrm{PA}}(\mathsf{FM}_M(2)) \mapsto \underbrace{\phi_{12} \otimes 1 \otimes 1}_{\Omega_{\mathrm{PA}}(\mathsf{FM}_M(2)) \otimes \Omega_{\mathrm{PA}}(\mathsf{FM}_D(1)) \otimes \Omega_{\mathrm{PA}}(\mathsf{FM}_D(1))} + \underbrace{1 \otimes \phi_{12} + \eta \otimes 1}_{\Omega_{\mathrm{PA}}(\mathsf{FM}_M(1)) \otimes \Omega_{\mathrm{PA}}(\mathsf{FM}_D(2))}.$$

On the other hand, the corresponding cocomposition ${}^*\mathsf{Gra}_M \to {}^*\mathsf{Gra}_M \circ {}^*\mathsf{Gra}_D$ given by de-insertion sends s^{12} to

$$\underbrace{1 - 2 \otimes 1 \otimes 1}_{{}^*\mathsf{Gra}_M(2) \otimes {}^*\mathsf{Gra}_D(1) \otimes {}^*\mathsf{Gra}_D(1)} + \underbrace{1 \otimes 1 - 2}_{{}^*\mathsf{Gra}_M(1) \otimes {}^*\mathsf{Gra}_D(2)} + \underbrace{1 \otimes 1 \quad 2}_{{}^*\mathsf{Gra}_M(1) \otimes {}^*\mathsf{Gra}_D(2)};$$

therefore the cocomposition is respected by the map. $\square$

4 Twisting Gra_M and the comodule ${}^*\mathsf{Graphs}_M$

Let Gra_D and Gra_V be the duals of ${}^*\mathsf{Gra}_D$ and ${}^*\mathsf{Gra}_V$, respectively. Gra_V is an operadic right Gra_D–module in dg cocommutative coalgebras.

Recall that there is a map of operads $\Lambda^{D-1}\mathsf{Lie} \to \mathsf{Gra}_D$ given by mapping the generator $\mu \in \Lambda^{D-1}\mathsf{Lie}(2)$ to the single edge graph in $\mathsf{Gra}_D(2)$ [45]. This extends to a map from the canonical operadic right module

$$\Lambda^{D-1}\mathsf{Lie} \circlearrowleft \Lambda^{D-1}\mathsf{Lie} \to \mathsf{Gra}_M \circlearrowleft \mathsf{Gra}_D$$

sending the generator μ to $s^{12} \in \mathsf{Gra}_M(2)$. One can then apply the right module twisting procedure described in [45, Appendix I] to $\mathsf{Gra}_M \circlearrowleft \mathsf{Gra}_D$, thus obtaining the bimodule $\mathrm{Tw}\,\mathsf{Gra}_M \circlearrowleft \mathrm{Tw}\,\mathsf{Gra}_D$.

[5]Notice that on the second summand ϕ_{12} refers to the volume form of $S^{D-1} = \mathsf{FM}_D(2)$. We are using Remark 9 to ensure that this term is indeed of that form.

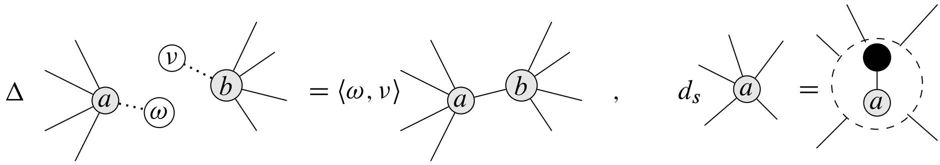


Figure 2: Internal vertices are depicted in black. Gray vertices are either internal or external vertices.

$\mathrm{Tw}\,\mathsf{Gra}_M$ can be described via a different kind of graphs. The space $\mathrm{Tw}\,\mathsf{Gra}_M(n)$ is spanned by graphs with n vertices labeled from 1 to n, called "external" vertices and k indistinguishable "internal" vertices. Both types of vertices can be decorated by elements of $(H^\bullet(M))^*$ (with $\bullet \geq 1$, see Remark 17 below), that can be identified with $H^{|D|-\bullet}(M)$ via the canonical pairing. The degree of the internal vertices is D, the degree of edges is $1-D$ and the degree of the decorations is the correspondent degree in $(H^\bullet(M))^*$, even if there is an identification with the cohomology. The differential in $\mathrm{Tw}\,\mathsf{Gra}_M$ can be split into 3 pieces: $d = \Delta + d_{\mathrm{ex}} + d_{\mathrm{in}} = \Delta + d_s$, where Δ is the differential coming from Gra_M, that connects decorations by making an edge, d_{ex} splits an internal vertex out of every external vertex and reconnecting incident edges in all possible ways and d_{in} splits similarly an internal vertex out of every internal vertex; see Figure 2.

Remark 17 Notice that due to ${}^*\mathsf{Gra}_M$ being given by a quotient by $e_1^j - 1$, if a certain vertex v of $\Gamma \in {}^*\mathsf{Gra}_M$ is decorated with the volume form on M, then we find as summands of $\Delta(\Gamma)$ all possibilities of connecting v to every other vertex in Γ.

The operad $\mathrm{Tw}\,\mathsf{Gra}_D$ is spanned by similar kinds of graphs, except that there are no decorations. We will therefore also refer to the vertices of $\mathrm{Tw}\,\mathsf{Gra}_D$ as internal and external.

The degrees of graphs in $\mathrm{Tw}\,\mathsf{Gra}_D$ are computed similarly, but the differential of $\mathrm{Tw}\,\mathsf{Gra}_D$ is different (since Gra_M is twisted as a Lie–module whereas Gra_D is twisted as an operad under Lie). Not only there is no Δ term, but also the splitting piece has an extra term subtracting all possible ways of adding a univalent internal vertex.

We are interested in a suboperad of $\mathrm{Tw}\,\mathsf{Gra}_D$, since $\mathrm{Tw}\,\mathsf{Gra}_D$ is in homologically "too big". The following result originates in [27; 33].

Proposition 18 [45] *The operad* $\mathrm{Tw}\,\mathsf{Gra}_D$ *has a suboperad that we call* Graphs_D *spanned by graphs* Γ *such that*:

- *All internal vertices of* Γ *are at least trivalent*,
- Γ *has no connected components consisting only of internal vertices.*

Moreover there is a cooperadic quasi-isomorphism

$$^*\mathsf{Graphs}_D \to \Omega_{\mathrm{PA}}(\mathsf{FM}_D),$$

extending the map (7).

This quasi-isomorphism is defined by integrating over all possible configurations of points corresponding to the internal vertices, a formula similar to the one from Lemma 19.

We will from now on interpret $\mathrm{Tw}\,\mathsf{Gra}_M$ as a right Graphs_D–module.

Let $^*\mathrm{Tw}\,\mathsf{Gra}_M$ be the cooperadic right $^*\mathsf{Graphs}_D$–comodule that is (pre)dual to $\mathrm{Tw}\,\mathsf{Gra}_M$.

The differential in $^*\mathrm{Tw}\,\mathsf{Gra}_M$ decomposes as $d = \delta_{\mathrm{cut}} + \delta_{\mathrm{contr}}$, where δ_{cut} is the piece originating from $^*\mathsf{Gra}_M$ that splits edges into "diagonal classes" and δ_{contr} contracts any edge adjacent to one or two internal vertices.

Lemma 19 *For* M *a closed compact framed connected manifold as above there is a natural map of right cooperadic comodules*

$$\omega_\bullet \colon {}^*\mathrm{Tw}\,\mathsf{Gra}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M),$$

extending the map $f\colon {}^*\mathsf{Gra}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ *from* (8).

Proof Let Γ be a graph in $^*\mathsf{Gra}_M(n+k)^{\mathbb{S}_k} \subset {}^*\mathrm{Tw}\,\mathsf{Gra}_M(n)$, ie Γ has n external and k internal vertices. Let us consider $f(\Gamma) \in \Omega_{\mathrm{PA}}(\mathsf{FM}_M(n+k))$, the image of Γ under the map (8). We define ω_Γ to be the integral of $f(\Gamma)$ over all configurations of the internal vertices. Concretely, if $\mathsf{FM}_M(n+k) \to \mathsf{FM}_M(n)$ denotes the map that forgets the last k points, then ω_Γ is given by the fiber[6] integral

$$\int_{\mathsf{FM}_M(n+k)\to\mathsf{FM}_M(n)} f(\Gamma).$$

The commutativity with the right cooperadic cocompositions is formally the same as why $^*\mathrm{Tw}\,\mathsf{Gra}_D \to \Omega_{\mathrm{PA}}(\mathsf{FM}_D)$ is a map of cooperads — see [33, Section 9.5] — together with the fact that the propagator on FM_M on clusters of points is given by

[6] Notice that here we make use of the fact that $f(\Gamma)$ is actually in $\Omega_{\mathrm{triv}}(\mathsf{FM}_M(n+k))$.

the corresponding propagator of FM_D. It remains to check the compatibility of the differentials.

Notice that ${}^*\mathrm{Tw}\,\mathsf{Gra}_M$ is a quasifree dgca generated by internally connected graphs, ie graphs that remain connected if we delete all external vertices. Since the map $\omega_\bullet$ is compatible with the products, it suffices to check the compatibility of the differentials on internally connected graphs. Let $\Gamma \in {}^*\mathrm{Tw}\,\mathsf{Gra}_M(n)$ be such a graph with k internal vertices.

If we denote by F the fiber of the map $\mathsf{FM}_M(n+k) \to \mathsf{FM}_M(n)$, we have, by Stokes' theorem,

$$d\omega_\Gamma = \int_F df(\Gamma) \pm \int_{\partial F} f(\Gamma).$$

If we compute $d\Gamma = \delta_{\mathrm{cut}}\Gamma + \delta_{\mathrm{contr}}\Gamma$, we retrieve

$$\omega_{\delta_{\mathrm{cut}}\Gamma} = \int_F f(d_{\mathrm{cut}}\Gamma) = \int_F df(\Gamma).$$

The boundary of the fiber decomposes into various pieces, namely

$$\partial F = \bigcup_{n<i,j\le n+k} \partial_{i,j} F \cup \bigcup_{\substack{a\le n \\ n<i\le n+k}} \partial_{a,i} F \cup \partial_{\ge 3} F,$$

where $\partial_{i,j} F$ is the boundary piece where points i and j (corresponding to internal vertices) collided, $\partial_{a,i} F$ is the boundary piece where point i (corresponding to an internal vertex) collided with point a (corresponding to an external vertex) and $\partial_{\ge 3} F$ is the boundary piece in which at least three points corresponding to internal vertices collided.

If points i and j are not connected by an edge in Γ, then $\int_{\partial_{i,j} F} f(\Gamma) = 0$. To see this, note that this integral has the form $\int_{\partial_{i,j} F} f(\Gamma) = \int_i f(\Gamma)|_{i=j} \int_{S^{D-1}} 1 = 0$. Here the integral vanishes by degree reasons since there is no top degree component of the form on the factor S^{D-1}. Here we used that the tangent bundle is trivialized. However, the same argument goes through without using this feature by using trivializations of the tangent bundle.

If points i and j are connected by an edge, then by property (ii) of Proposition 8 we have $\int_{\partial_{i,j} F} f(\Gamma) = \omega_{\Gamma/e}$, where Γ/e is the graph Γ with edge e contracted. An analogous argument for the boundary pieces $\partial_{a,i} F$ allows us to conclude that $\omega_{d\Gamma} = d\omega_\Gamma \pm \int_{\partial_{\ge 3} F} f(\Gamma)$.

The vanishing of $\int_{\partial_{\geq 3}F} f(\Gamma)$ results from Kontsevich's vanishing lemmas. Concretely, suppose there are $3 \leq l \leq k$ points colliding together. By integrating over the l points first we obtain an integral of the form $\int_{\mathsf{FM}_D(l)} \nu$, where ν is a product of $\phi_{i,j}$. If the dimension D is at least 3, this integral vanishes as in [26, Lemma 2.2], using property (iii) of Proposition 8.

To factor the integral we used the trivialization of the tangent bundle in this step. For later use we shall however remark that this is not necessary. More precisely, let the full subgraph on the "colliding" vertices be γ. Then by the same argument as in the proof of [26, Lemma 2.2], using property (iii) of Proposition 8, we may assume that all vertices of γ have valence ≥ 3. But then the inner integral describes a form of degree $\geq \frac{3}{2}l(D-1) - lD + D + 1 = \frac{1}{2}l(D-3) + D + 1 > D$ on M, and M is of dimension D. Hence this integral must vanish.

If $D = 2$, because of property (ii) of Proposition 8 we can use the Kontsevich vanishing lemma from [28, Section 6.6] to ensure the vanishing of the integral. □

4.1 The full graph complex and Graphs_M

The map constructed in Lemma 19 is not (in general) a quasi-isomorphism and the fundamental obstruction is the existence of graphs containing connected components of only internal vertices in ${}^*\mathrm{Tw}\,\mathsf{Gra}_M$. The desired complex ${}^*\mathsf{Graphs}_M$ will be a quotient of ${}^*\mathrm{Tw}\,\mathsf{Gra}_M$ through which the map $\omega_\bullet$ factors. A formal construction can be done making use of the full graph complex that we define as follows.

Definition 20 The full graph complex of M, ${}^*\mathrm{fGC}_{H^\bullet(M)}$ is defined to be the complex ${}^*\mathrm{Tw}\,\mathsf{Gra}_M(0)$ consisting of graphs with no external vertices. This vector space forms a differential graded commutative $\mathbb{R}$–algebra with product defined to be the disjoint union of graphs. We reserve the symbol $\mathrm{fGC}_{H^\bullet(M)} = ({}^*\mathrm{fGC}_{H^\bullet(M)})^*$ for the dual complex and the symbol $\mathrm{GC}_{H^\bullet(M)} \subset \mathrm{fGC}_{H^\bullet(M)}$ for the subcomplex of connected graphs.

The vector space ${}^*\mathrm{Tw}\,\mathsf{Gra}_M$ can be naturally regarded as a left module over the algebra ${}^*\mathrm{fGC}_{H^\bullet(M)}$, where the action is given by taking the disjoint union of graphs. Furthermore, we define the partition function

$$Z_M \colon {}^*\mathrm{fGC}_{H^\bullet(M)} \to \mathbb{R} \tag{9}$$

to be the map of dg commutative algebras obtained by restriction of the map $\omega_\bullet$ from Lemma 19.

There is a commutative diagram of dg commutative algebras and modules

$$\begin{tikzcd} {}^*\mathrm{fGC}_{H^\bullet(M)} \arrow[d, "Z_M"] & \circlearrowright & {}^*\mathrm{Tw}\,\mathsf{Gra}_M \arrow[d, "\omega_\bullet"] \\ \mathbb{R} & \circlearrowright & \Omega_{\mathrm{PA}}(\mathsf{FM}_M) \end{tikzcd}$$

Definition 21 The right ${}^*\mathsf{Graphs}_D$ cooperadic comodule ${}^*\mathsf{Graphs}_M$ is defined by

$${}^*\mathsf{Graphs}_M = \mathbb{R} \otimes_{Z_M} {}^*\,\mathrm{Tw}\,\mathsf{Gra}_M.$$

Remark 22 We pick as representatives for a basis of ${}^*\mathsf{Graphs}_M$ the set of graphs that contain no connected components without external vertices. With this convention it still makes sense to talk about the total number of vertices of a graph in ${}^*\mathsf{Graphs}_M$.

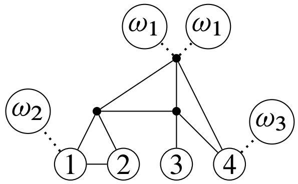
Notice that as a consequence, part of the differential of ${}^*\mathsf{Graphs}_M$ might reduce the number of vertices by more than 1 by "cutting away" a part of the graph that contains only internal vertices, which did not happen with ${}^*\mathrm{Tw}\,\mathsf{Gra}_M$.

Corollary 23 *The map ${}^*\mathrm{Tw}\,\mathsf{Gra}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ defined in Lemma 19 induces a map of cooperadic comodules ${}^*\mathsf{Graphs}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$, still denoted by $\omega_\bullet$.*

Remark 24 One may also endow $\mathrm{fGC}_{H^\bullet(M)}$ with the (free) product given by union of graphs. The differential is not a derivation with respect to this product, but it is a coderivation and it splits into a first-order and a second-order part, say $\delta_1 + \delta_2$. Concretely, the second-order part δ_2 replaces a pair of $H^\bullet(M)$–decorations in different connected components by an edge, while the first-order piece splits vertices and joins decorations in the same connected component. By Koszul duality, the commutator of the product and the operator δ_2 defines a Lie bracket of degree 1 on $\mathrm{fGC}_{H^\bullet(M)}$, which reduces to a Lie bracket on the connected piece $\mathrm{GC}_{H^\bullet(M)}$.

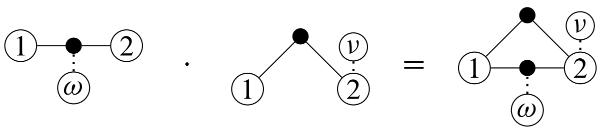
Now the partition function $Z_M \in \mathrm{fGC}_{H^\bullet(M)}$ is a map from the free graded commutative algebra ${}^*\mathrm{fGC}_{H^\bullet(M)}$ and hence completely characterized by the restriction to the generators, ie to the connected graphs, say $z_M \in \mathrm{GC}_{H^\bullet(M)}$. The closedness of Z_M then translates to the statement that the connected part z_M satisfies the Maurer–Cartan equation. See Section 7.1 for details.

To summarize, we constructed a cooperadic right Hopf comodule ${}^*\mathsf{Graphs}_M$. As a vector space, ${}^*\mathsf{Graphs}_M(n)$ is spanned by graphs with n labeled external vertices and an unspecified number of indistinguishable internal vertices that can be decorated by (possibly multiple) cohomology classes of degree at least 1, under the condition that there are no connected components without external vertices:

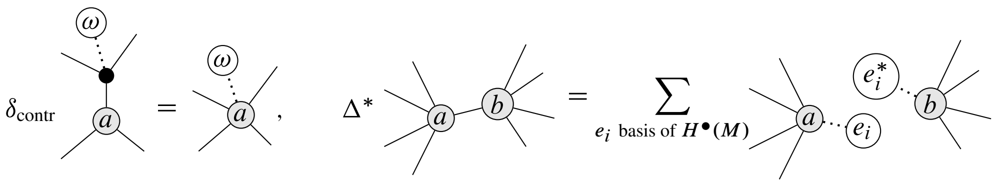


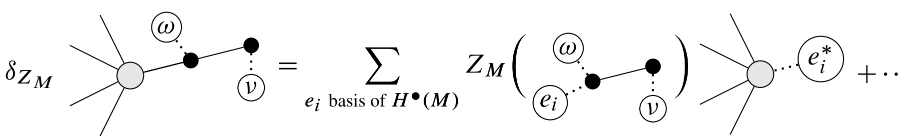
$\in {}^*\mathsf{Graphs}_M(4)$

There is a graded commutative algebra structure given by superposition of external vertices:



The differential δ splits as $\delta = \delta_{\mathrm{contr}} + \delta_{\mathrm{cut}}$, where δ_{contr} contracts edges adjacent to at least one internal vertex and δ_{cut} splits any edge into two decorations given by the diagonal class of M. Due to the constraint of not allowing connected components without external vertices, $\delta_{\mathrm{cut}} = \Delta^* + \delta_{Z_M}$ splits again into two pieces, Δ^* which does not create *forbidden* graphs and δ_{Z_M} that when creating such forbidden connected components transforms them into a scalar as prescribed by the partition function Z_M:





The cooperadic right comodule structure is obtained by collapsing subgraphs containing at least one external vertex into a single external vertex.

4.2 Historic remark

The above graph complexes can be seen as a version of the nonvacuum Feynman diagrams appearing in the perturbative expansion of topological field theories of AKSZ type, in the presence of zero modes. In this setting the field theories have been studied by Cattaneo and Felder [12] and Cattaneo and Mnev [13], whose names we hence attach to the above complexes of diagrams, though the above construction of ${}^*\mathsf{Graphs}_M$ does not appear in these works directly. Furthermore, it has been pointed out to us by A Goncharov that similar complexes have been known by experts before the works of the aforementioned authors. Finally, in the local case the construction is due to M Kontsevich [27].

5 Cohomology of the CFM (co)operad

The following theorem relates the right Graphs_D–module Graphs_M with the right FM_D–module FM_M.

Theorem 25 *The map $\omega_\bullet\colon {}^*\mathsf{Graphs}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ established in Corollary 23 is a quasi-isomorphism. Similarly, the composition map*

$$\mathsf{Chains}(\mathsf{FM}_M) \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)^* \xrightarrow{\omega_\bullet^*} \mathsf{Graphs}_M$$

is a quasi-isomorphism of right modules.

Note that there is in general no known explicit formula for the cohomology of the configuration spaces $\mathsf{FM}_M(n)$ on a manifold. However, two spectral sequences converging to the (co)homology are known, one by Cohen and Taylor [15] and one by Bendersky and Gitler [4]. Both spectral sequences have been shown to be isomorphic (via Poincaré duality) from the E^2 term on by Felix and Thomas [19]. The E^2 term is the cohomology of a relatively simple complex described below. It was shown by B Totaro [44] and I Kriz [30] that the spectral sequence abuts at the E^2 term for smooth projective varieties. However, it does not in general abut at the E^2 term; a counterexample was given in [19].

The strategy to prove Theorem 25 will be as follows. We will compare the double complex BG giving rise to the Bendersky–Gitler spectral sequence (its definition will be recalled below) to ${}^*\mathsf{Graphs}_M$. There is a complex $\widetilde{\mathrm{BG}}$ quasi-isomorphic to BG that comes with a natural map $f\colon \widetilde{\mathrm{BG}} \to \mathsf{Graphs}_M$. Our goal is to show that f is a quasi-isomorphism, and for that we set up another spectral sequence. The detailed proof is contained in Section 5.6.

5.1 The Bendersky–Gitler spectral sequence

Let us recall the definition of the Bendersky–Gitler spectral sequence. See also the exposition in [19].

Recall that the configuration space of n points in M is $\mathrm{Conf}_n(M) := M^n \setminus \Delta$, where $\Delta = \{(p_1, \dots, p_n) \mid \exists i \neq j : p_i = p_j\}$. By Poincaré–Lefschetz duality,

$$H_{-d}(\mathrm{Conf}_n(M)) \cong H^{n \dim(M) - d}(M^n, \Delta).$$

The relative cohomology $H^\bullet(M^n, \Delta)$ on the right is the cohomology of the complex

$$H^\bullet(M^n) \to H^\bullet(\Delta).$$

The left-hand side is the cohomology of $\Omega_{\mathrm{PA}}(M)^{\otimes n}$. The right-hand side may be computed as the cohomology of the Čech–de Rham complex corresponding to any covering of Δ. To obtain the Bendersky–Gitler double complex one takes the cover of the diagonal by the sets

$$U_{i,j} = \{p_i = p_j\} \subset \Delta.$$

The Bendersky–Gitler complex is the total complex of the double complex obtained using the natural quasi-isomorphism $\Omega_{\mathrm{PA}}(M)^{\otimes n} \to \Omega_{\mathrm{PA}}(M^n)$, ie

$$\mathrm{BG}(n) := \mathrm{Total}(\Omega_{\mathrm{PA}}(M)^{\otimes n} \to \text{Čech–de Rham}(\Delta)).$$

By the statements above and a simple spectral sequence argument, it follows that $H^\bullet(\mathrm{BG}(n)) \cong H(M^n, \Delta)$.

For what we will say below it is important to describe $\mathrm{BG}(n)$ in a more concise way. Elements of $\mathrm{BG}(n)$ can be seen as linear combinations of decorated graphs on n vertices, the decoration being one element of $\Omega_{\mathrm{PA}}(M)$ for each connected component of the graph. The degree of such a graph is computed as

$$(\text{degree}) = \#(\text{edges}) + \#(\text{total degree of decorations}) - n \cdot \dim(M).$$

The differential is composed of two parts, one of which comes from the de Rham differential and one of which comes from the Čech differential:

$$d_{\text{total}} = d_{\text{dR}} + \delta.$$

Concretely, δ adds an edge in all possible ways, and multiplies the decorations of the connected components the edge joins.

Remark 26 The original construction of the Bendersky–Gitler spectral sequence uses the de Rham complex of M, but since there is only semialgebraic data involved, namely intersections of sets $U_{i,j} \cong M^{n-1}$, we are allowed to replace differential forms by piecewise algebraic (PA) forms.

5.2 A general construction

Recall that the monoidal product of symmetric sequences $\circ$ is given by

$$(\mathcal{S}\circ\mathcal{S}')(n) = \bigoplus_{k=k_1+\cdots+k_n} \mathcal{S}(k)\otimes\mathcal{S}'(k_1)\otimes\cdots\otimes\mathcal{S}'(k_n)\otimes\mathbb{R}[\mathrm{Sh}(k_1,\dots,k_n)],$$

where $\mathrm{Sh}(k_1,\dots,k_n)$ are the $k_1,\dots,k_n$ shuffles. Let $\mathcal{C}$ be a cooperad, $\mathcal{M}$ be a cooperadic right $\mathcal{C}$–comodule with coaction $\Delta_{\mathcal{M}}\colon\mathcal{M}\to\mathcal{M}\circ\mathcal{C}$, and let A be some dg commutative algebra, which can be seen as a symmetric sequence concentrated in arity 1. Then the spaces

$$\mathcal{M}(n)\otimes A^{\otimes n} = (\mathcal{M}\circ A)(n)$$

assemble into another cooperadic right $\mathcal{C}$–comodule.

More formally, since A is a dg commutative algebra, for every symmetric sequence $\mathcal{S}$ there is a morphism

$$s\colon\mathcal{S}\circ A\to A\circ\mathcal{S}$$

given by the multiplication in A.

The coaction of $\mathcal{C}$ on $\mathcal{M}\circ A$ is given by the composition of the maps

$$\mathcal{M}\circ A\xrightarrow{\Delta_{\mathcal{M}}\circ\mathrm{id}_A}(\mathcal{M}\circ\mathcal{C})\circ A\cong\mathcal{M}\circ(\mathcal{C}\circ A)\xrightarrow{\mathrm{id}_{\mathcal{M}}\circ s}\mathcal{M}\circ(A\circ\mathcal{C})\cong(\mathcal{M}\circ A)\circ\mathcal{C}.$$

It is a straightforward verification to check that the axioms for cooperadic comodules hold.

5.3 The definition of $\widetilde{\mathrm{BG}}$

Let $\mathcal{C}$ be a coaugmented cooperad and $\mathcal{M}$ be a right $\mathcal{C}$–comodule. Applying the cobar construction to the cooperad $\mathcal{C}$ we obtain an operad $\Omega(\mathcal{C})$. Applying the cobar construction to the comodule $\mathcal{M}$ we obtain a right $\Omega(\mathcal{C})$–module $\Omega_{\Omega(\mathcal{C})}(\mathcal{M})$, also denoted just by $\Omega(\mathcal{M})$. As a symmetric sequence, $\Omega(\mathcal{M})=\mathcal{M}\circ\Omega(\mathcal{C})$ and the differential splits as $d=d_1+d_2+d_3$, where d_1 comes from the differential in $\mathcal{M}$, d_2 comes from the differential in $\Omega(\mathcal{C})$ and d_3 is induced by the comodule structure. Of course, if A is a dg commutative algebra, then replacing $\mathcal{M}$ by $\mathcal{M}\circ A$ we obtain a right $\Omega(\mathcal{C})$–module $\Omega(\mathcal{M}\circ A)$. We can now define

$$\widetilde{\mathrm{BG}} := \Omega_{\Lambda^{D-1}L_\infty}(s^{-D}\Lambda^{D}\mathsf{coComm}\circ\Omega_{\mathrm{PA}}(M)),$$

where on the right-hand side we consider $s^{-D}\Lambda^D\mathsf{coComm}$ as a right comodule over $\Lambda^D\mathsf{coComm}$ and then we use the construction from the previous section that gives us a $\Lambda^D\mathsf{coComm}$–right comodule structure on $s^{-D}\Lambda^D\mathsf{coComm}\circ\Omega(\mathcal{M})$. Notice that the operadic cobar construction is given by

$$\Omega(\Lambda^D\mathsf{coComm}) = \Omega((\Lambda^{D-1}\mathsf{Lie})^\vee) = \Lambda^{D-1}L_\infty.$$

Up to degrees, one can picture $\widetilde{\mathrm{BG}}$ as multiple ("commuting") L_∞ words, each labeled by a PA form on M. Besides the de Rham and the L_∞ differential, the cobar differential acts by merging two L_∞ words while multiplying the associated forms.

5.4 Some other general remarks and the definition of sBG

Let $\mathcal{P}$ be a Koszul operad, $\mathcal{P}^\vee$ the Koszul dual cooperad and $\mathcal{P}_\infty = \Omega(\mathcal{P}^\vee)$ the minimal cofibrant model for $\mathcal{P}$. There are bar and cobar construction functors between the categories of right $\mathcal{P}$–modules and right $\mathcal{P}^\vee$–comodules,

$$B_{\mathcal{P}^\vee}\colon \mathrm{Mod}\text{–}\mathcal{P} \leftrightarrow \mathrm{coMod}\text{–}\mathcal{P}^\vee : \Omega_{\mathcal{P}}.$$

Given a right $\mathcal{P}^\vee$–comodule $\mathcal{M}$ there are two ways to construct a right $\mathcal{P}_\infty$–module:

(1) Take the right $\mathcal{P}_\infty$–module $\Omega_{\mathcal{P}_\infty}(\mathcal{M})$.

(2) Take $\Omega_{\mathcal{P}}(\mathcal{M})$ and consider it as a right $\mathcal{P}_\infty$–module via the morphism of operads $p\colon \mathcal{P}_\infty \to \mathcal{P}$.

Lemma 27 *Let $\mathcal{P}$ be a Koszul operad with zero differential such that $\mathcal{P}(0) = 0$ and $\mathcal{P}(1) = \mathbb{R}$ and let $\mathcal{M}$ be a right $\mathcal{P}^\vee$–comodule. Then there is a canonical (surjective) quasi-isomorphism*

$$\pi\colon \Omega_{\mathcal{P}_\infty}(\mathcal{M}) \to \Omega_{\mathcal{P}}(\mathcal{M}).$$

Proof As symmetric sequences, $\Omega_{\mathcal{P}_\infty}(\mathcal{M}) = \mathcal{M}\circ\mathcal{P}_\infty$ and $\Omega_{\mathcal{P}}(\mathcal{M}) = \mathcal{M}\circ\mathcal{P}$. We define $\pi = \mathrm{id}_{\mathcal{M}}\circ p$. It is clear that each piece of the differential commutes with π. The remaining claim that π is a quasi-isomorphism follows from a spectral sequence argument.

Concretely, we consider a filtration $\mathcal{F}^p\Omega_{\mathcal{P}_\infty}(\mathcal{M})$ spanned by elements for which the sum of the degree in $\mathcal{M}$ with the weight in $\mathcal{P}_\infty$ (the amount of elements from $\mathcal{P}^\vee$ used) does not exceed p. On the first page of the spectral sequence given by this filtration we recover $\Omega_{\mathcal{P}}(\mathcal{M})$ and thus the result follows. □

Now let us give the definition of sBG,

$$\mathrm{sBG} = \Omega_{\Lambda^{D-1}\mathsf{Lie}}(s^{-D}\Lambda^D\mathsf{coComm}\circ\Omega_{\mathrm{PA}}(M)),$$

where on the right we consider $\Lambda^D \mathsf{coComm} = (\Lambda^{D-1}\mathsf{Lie})^\vee$ as a right comodule over itself and the algebra of differential forms $\Omega_{\mathrm{PA}}(M)$. Then, by the lemma above, we see that there is a canonical quasi-isomorphism

$$\widetilde{\mathrm{BG}} \to \mathrm{sBG}.$$

Similar to $\widetilde{\mathrm{BG}}$, one can picture sBG as connected components of Lie words, each labeled by a PA form on M. One can consider a basis of $\mathsf{Lie}(n)$ consisting of graphs on n vertices, with $n-1$ edges, such that there are no two edges (i,r) and (j,r) with r bigger than both i and j. Taking the degrees and differentials into account, we see that $\mathrm{sBG}(n)$ is precisely what in [19] is denoted by $\bar{E}(n,A)$, for $A = \Omega_{\mathrm{PA}}(M)$.

Furthermore it was shown in [19, Proposition 2.4] that there is a canonical quasi-isomorphism

$$\mathrm{BG} \to \mathrm{sBG}.$$

In particular one obtains:

Corollary 28 *The following symmetric sequences are isomorphic*:

$$H_\bullet(\mathrm{Conf}_\bullet(M)) \cong H(\mathrm{BG}) \cong H(\mathrm{sBG}) \cong H(\widetilde{\mathrm{BG}}).$$

5.5 The map $\widetilde{\mathrm{BG}} \to \mathsf{Graphs}_M$

The goal of this subsection is to construct the map of right $\Lambda^{D-1}L_\infty$–modules

$$\Phi\colon \widetilde{\mathrm{BG}} \to \mathsf{Graphs}_M.$$

Since $\widetilde{\mathrm{BG}} := \Omega_{\Lambda^{D-1}L_\infty}(s^{-D}\Lambda^D\mathsf{coComm} \circ \Omega_{\mathrm{PA}}(M))$ is quasifree as right $\Lambda^{D-1}L_\infty$–module, it suffices to define our map Φ on the module generators and verify that this map is compatible with the differential. Note that $s^{-D}\Lambda^D\mathsf{coComm}(n) = \mathbb{R}[nD]\mu_n$ is one-dimensional, generated by the n–fold coproduct μ_n. We define the map Φ on generators by setting, for $\alpha_1,\dots,\alpha_n \in \Omega_{\mathrm{PA}}(M)$ and $\Gamma \in {}^*\mathsf{Graphs}_M$,

$$(\Phi(\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n))(\Gamma) := \int_{\mathsf{FM}_M(n)} (\pi_1^*\alpha_1)\cdots(\pi_n^*\alpha_n)\omega_\Gamma. \tag{10}$$

Here $\pi_j\colon \mathsf{FM}_M(n) \to \mathsf{FM}_M(1) = M$ is the map that forgets the position of all points in the configuration except for the j^{th} point. Notice that the element $\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n$ has degree $-nD + |\alpha_1| + \cdots + |\alpha_n| = -(\dim(\mathsf{FM}_M(n)) - |\pi_1^*\alpha_1| - \cdots - |\pi_n^*\alpha_n|)$; therefore F preserves degrees.

A general element of $\widetilde{\mathrm{BG}}$ is a linear combination of elements obtained by acting with elements of the operad $\ell_j \in \Lambda^{D-1} L_\infty$ on generators,

$$x := (\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n) \circ (\ell_1, \dots, \ell_n).$$

For such elements x we have that $\Phi(x) = \Phi(\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n) \circ (\ell_1, \dots, \ell_n)$, using the right action of $\Lambda^{D-1} L_\infty$ on Graphs_M. This latter action factors through the right action of Graphs_D on Graphs_M via the maps

$$\Lambda^{D-1} L_\infty \xrightarrow{f} \mathrm{Chains}(\mathsf{FM}_D) \xrightarrow{\omega^*} \mathsf{Graphs}_D.$$

Denoting the cooperadic coaction on $\Gamma \in {}^*\mathsf{Graphs}_M$ by $\sum \Gamma' \otimes \gamma_1 \otimes \cdots \otimes \gamma_k$, with $\gamma_j \in {}^*\mathsf{Graphs}_D$, this implies that

$$\begin{aligned} \Phi(x)(\Gamma) &= (\Phi(\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n) \circ (\ell_1, \dots, \ell_n))(\Gamma) \\ &= \sum \pm \Phi(\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n)(\Gamma') \cdot \prod_j \int_{f(\ell_j)} \omega_{\gamma_j} \\ &= \sum \pm \int_{\mathsf{FM}_M(n)} (\pi_1^* \alpha_1) \cdots (\pi_n^* \alpha_n) \omega_{\Gamma'} \prod_j \int_{f(\ell_j)} \omega_{\gamma_j} \\ &= \int_{\mathrm{Fund}(\mathsf{FM}_M(n)) \circ (f(\ell_1), \cdots, f(\ell_n))} (\pi_{i_1}^* \alpha_1) \cdots (\pi_{i_n}^* \alpha_n) \omega_\Gamma. \end{aligned} \tag{11}$$

In the last line we are integrating over the fundamental chain of a boundary stratum of FM_M in which groups of points are infinitesimally close together. The indices $i_1, \dots, i_n$ shall be those of one (arbitrary) point in each such group. Furthermore, we used the compatibility of the map ω with the operadic FM_D–action on FM_M. Using the formula above we can show the following result.

Lemma 29 *The map $\Phi \colon \widetilde{\mathrm{BG}} \to \mathsf{Graphs}_M$ defined above is compatible with the differentials and is hence a map of right $\Lambda^{D-1} L_\infty$–modules. It furthermore factorizes through the adjoint ω^* of the map $\omega \colon {}^*\mathsf{Graphs}_M \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ of Corollary 23 as*

$$\widetilde{\mathrm{BG}} \xrightarrow{F} \Omega_{\mathrm{PA}}(\mathsf{FM}_M)^* \xrightarrow{\omega^*} \mathsf{Graphs}_M$$

with

$$\begin{aligned} &F((\mu_n \otimes \alpha_1 \otimes \cdots \otimes \alpha_n) \circ (\ell_1, \dots, \ell_n))(\omega) \\ &\qquad\qquad = \int_{\mathrm{Fund}(\mathsf{FM}_M(n)) \circ (f(\ell_1), \cdots, f(\ell_n))} (\pi_{i_1}^* \alpha_1) \cdots (\pi_{i_n}^* \alpha_n) \omega. \end{aligned}$$

Proof The factorization through ω^* is clear by (11).

It remains to check that the differentials are preserved by Φ. Note that the differential on $\widetilde{\mathrm{BG}}$ decomposes into three terms, $d = d_{\Omega_{\mathrm{PA}}(M)} + d_{\Lambda^{D-1}Ł_\infty} + d_{\mathrm{cobar}}$, stemming from the internal differentials on $\Omega_{\mathrm{PA}}(M)$ and $\Lambda^{D-1}Ł_\infty$, and the cobar construction respectively. Note that the second summand is zero on generators.

On the other hand we compute, applying Stokes' theorem,

$$\begin{aligned}
&(\Phi(\mu_n\otimes\alpha_1\otimes\cdots\otimes\alpha_n))(d\Gamma)\\
&\quad=\int_{\mathsf{FM}_M(n)}(\pi_1^*\alpha_1)\cdots(\pi_n^*\alpha_n)\omega_{d\Gamma}\\
&\quad=\int_{\mathsf{FM}_M(n)}(\pi_1^*\alpha_1)\cdots(\pi_n^*\alpha_n)d\omega_\Gamma\\
&\quad=\sum_{j=1}^n\pm\int_{\mathsf{FM}_M(n)}(\pi_1^*\alpha_1)\cdots(\pi_j^*d\alpha_j)\cdots(\pi_n^*\alpha_n)\omega_\Gamma\pm\int_{\partial\mathsf{FM}_M(n)}(\pi_1^*\alpha_1)\cdots(\pi_n^*\alpha_n)\omega_\Gamma.
\end{aligned}$$

The two terms exactly reproduce the differential on $\widetilde{\mathrm{BG}}$. The first term corresponds to the part from the internal differential on $\Omega_{\mathrm{PA}}(M)$. The second term (the boundary integral) produces the part of the differential from the cobar construction. More precisely, it is the sum over codimension 1 boundary strata corresponding to some subset of the n points colliding. But each such term is, using the computation (11) again, identified with an action of a generator of $\Lambda^{D-1}L_\infty$, so that all these terms together assemble to $\pm d_{\mathrm{cobar}}\Phi(\mu_n\otimes\alpha_1\otimes\cdots\otimes\alpha_n)$. □

5.6 The map $\widetilde{\mathrm{BG}}\to\mathsf{Graphs}_M$ is a quasi-isomorphism

In this section we will show the following proposition:

Proposition 30 *The map $\Phi\colon\widetilde{\mathrm{BG}}\to\mathsf{Graphs}_M$ constructed above is a quasi-isomorphism.*

There is a filtration on Graphs_M by the number of connected components in graphs. Concretely, let $\mathcal{F}^p\mathsf{Graphs}_M$ be the set of elements of Graphs_M which contain only graphs with p or fewer connected components. There is a similar filtration on $\widetilde{\mathrm{BG}}$ coming from the arity of elements of the generating symmetric sequence $s^{-D}\Lambda^D\mathsf{coComm}$. Concretely, elements of $\mathcal{F}^p\widetilde{\mathrm{BG}}$ are those elements of $\widetilde{\mathrm{BG}}$ that can be built without using any generators $\mu_{p+1},\mu_{p+2},\dots$ in $\Lambda^D\mathsf{coComm}$. The filtration is aritywise bounded, since the number of connected components in arity r is necessarily between 1 and r.

Lemma 31 *The map Φ from above is compatible with the filtration*, *ie*

$$\Phi(\mathcal{F}^p\widetilde{\mathrm{BG}})\subset\mathcal{F}^p\mathsf{Graphs}_M.$$

Proof The result is clear for generators of $\widetilde{\mathrm{BG}}$, since graphs with n vertices cannot have more than n connected components. In general Φ is compatible with the filtration since is a morphism of $\Lambda^{D-1}L_\infty$ right modules and the right action of $\Lambda^{D-1}L_\infty$ on Graphs_M is given by the insertion of connected graphs which cannot increase the number of connected components. □

It follows that Φ induces a morphism of the respective spectral sequences. We will show the following lemma:

Lemma 32 *The map Φ induces an isomorphism at the first pages of the associated spectral sequences.*

The statement of the lemma is equivalent to saying that the graded version of Φ,

$$\mathrm{gr}\,\Phi\colon \mathrm{gr}\,\widetilde{\mathrm{BG}} \to \mathrm{gr}\,\mathsf{Graphs}_M,$$

is a quasi-isomorphism.

One can compute the cohomology of $\mathrm{gr}\,\widetilde{\mathrm{BG}}$ explicitly.

Lemma 33 $H(\mathrm{gr}\,\widetilde{\mathrm{BG}}) = (s^{-D}\Lambda^D\mathsf{coComm} \circ H^\bullet(M)) \circ \Lambda^{D-1}\mathsf{Lie} =: \mathrm{sBG}_{H(M)}$.

Proof The differential on $\mathrm{gr}\,\widetilde{\mathrm{BG}}$ is precisely the one induced by the de Rham differential and the differential on $\Lambda^{D-1}L_\infty$. Therefore, by the Künneth formula,

$$\begin{aligned} H(\mathrm{gr}\,\widetilde{\mathrm{BG}}) &= H(s^{-D}\Lambda^D\mathsf{coComm}) \circ H(\Omega_{\mathrm{PA}}(M)) \circ H(\Lambda^{D-1}L_\infty) \\ &= (s^{-D}\Lambda^D\mathsf{coComm} \circ H^\bullet(M)) \circ \Lambda^{D-1}\mathsf{Lie}. \end{aligned}$$ □

Having fixed the embedding $H^\bullet(M) \hookrightarrow \Omega_{\mathrm{PA}}(M)$ and fixing any aritywise right inverse (as cochain complexes) of the projection $\Lambda^{D-1}L_\infty \to \Lambda^{D-1}\mathsf{Lie}$, from now on we interpret the space $\mathrm{sBG}_{H(M)}$ (with zero differential) as a subcomplex of $\mathrm{gr}\,\widetilde{\mathrm{BG}}$.

Proposition 34 *The map $\mathrm{gr}\,\Phi$ restricts to an injective map $\mathrm{sBG}_{H(M)} \to \mathrm{gr}\,\mathsf{Graphs}_M$ and the inclusion morphism $\Phi(\mathrm{sBG}_{H(M)}) \hookrightarrow \mathrm{gr}\,\mathsf{Graphs}_M$ is a quasi-isomorphism.*

The proof is by an argument similar to the one used by P Lambrechts and I Volic in [33, Lemma 8.3]. If we believe Proposition 34 for now, Lemma 32 follows as a corollary.

Proof of Proposition 30 As a consequence of Lemma 32, the map Φ induces a quasi-isomorphism at the level of the associated graded with respect to an (aritywise) bounded filtration, and therefore is a quasi-isomorphism itself. □

5.7 Proof of Proposition 34

Proposition 35 *The vector spaces* $\mathrm{sBG}_{H(M)}(n)$ *satisfy the recursion*

$$\mathrm{sBG}_{H(M)}(n) = \mathrm{sBG}_{H(M)}(n-1) \otimes H^\bullet(M) \oplus \mathrm{sBG}_{H(M)}(n-1)[D-1]^{\oplus n-1}. \tag{12}$$

Proof We have

$$\mathrm{sBG}_{H(M)}(n) = \bigoplus_{i_1+\cdots+i_k=n} H^\bullet(M)^{\otimes k}[kD] \otimes \Lambda^{D-1}\mathsf{Lie}(i_1) \otimes \cdots \otimes \Lambda^{D-1}\mathsf{Lie}(i_k) \otimes \mathbb{R}[\mathrm{Sh}(i_1,\ldots,i_k)].$$

Let us take an element of $\mathrm{sBG}_{H(M)}(n)$ and consider two different cases. If the input labeled by 1 corresponds to the unit $1 \in \Lambda^{D-1}\mathsf{Lie}(1)$, it is associated to an element of $H^\bullet(M)$ and by ignoring these we are left with a generic element of $\mathrm{sBG}_{H(M)}(n-1)$, thus giving us the first summand of (12).

Alternatively, if the vertex labeled by 1 corresponds to some Lie word in $\Lambda^{D-1}\mathsf{Lie}(i_j)$ with $j > 1$, the only possibility is that it came from the insertion of the generator $\mu_2 \in \Lambda^{D-1}\mathsf{Lie}(2)$ in some other Lie word. Since there are $n-1$ such choices and μ_2 has degree has degree $1-D$, we obtain the summand $\mathrm{sBG}_{H(M)}(n-1)[D-1]^{\oplus n-1}$. □

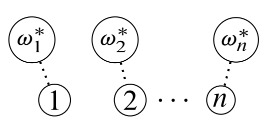
Lemma 36 *The map* $\mathrm{gr}\,\Phi$ *restricts to an isomorphism from* $\mathrm{sBG}_{H(M)}(n)$ *onto its image* $\Phi(\mathrm{sBG}_{H(M)}(n)) \subset \mathrm{gr}\,\mathsf{Graphs}_M(n)$.

Proof It suffices to show the injectivity of the map $\mathrm{gr}\,\Phi$ when restricted to $\mathrm{sBG}_{H(M)}(n)$.

Recall that

$$\mathrm{sBG}_{H(M)}(n) = \bigoplus_{i_1+\cdots+i_k=n} H^\bullet(M)^{\otimes k}[kD] \otimes \Lambda^{D-1}\mathsf{Lie}(i_1) \otimes \cdots \otimes \Lambda^{D-1}\mathsf{Lie}(i_k) \otimes \mathbb{R}[\mathrm{Sh}(i_1,\ldots,i_k)].$$

Let us start by considering the case in which the numbers $i_1,\ldots,i_n$ are all equal to 1. Let $\omega_1 \otimes \cdots \otimes \omega_n \in H^\bullet(M)^{\otimes n}[nD] \otimes \Lambda^{D-1}\mathsf{Lie}(1) \otimes \cdots \otimes \Lambda^{D-1}\mathsf{Lie}(1)$. The element $\Phi(\omega_1 \otimes \cdots \otimes \omega_n) \in \mathsf{Graphs}_M(n)$ is in principle a sum of many terms, but its projection into the subspace of $\mathsf{Graphs}_M(n)$ made only of graphs with no internal vertices, no more than one decoration per vertex, and precisely n connected components is simply the graph

$\pm$ 

where ω_i^* is dual to ω_i under the pairing on $H^\bullet(M)$. This implies in particular that Φ is injective when restricted to $H^\bullet(M)^{\otimes n}[nD] \otimes \Lambda^{D-1}\mathsf{Lie}(1) \otimes \cdots \otimes \Lambda^{D-1}\mathsf{Lie}(1)$.

The same idea can be adapted for the case of arbitrary i_j. The image of the elements of $\mathrm{sBG}_{H(M)}$ might be very complicated, but to conclude injectivity it is enough to see that the components on a "disconnected enough" subspace are different and by compatibility with the L_∞–action these components are just given by insertion of graphs representing L_∞ words.

Let $p \subset 2^{\{1,\dots,n\}}$ denote a partition of the numbers $1,\dots,n$. To every such p we can associate a subspace $V_p \subset \mathsf{Graphs}_M(n)$ spanned by graphs with no internal vertices and such that the vertices labeled by a and b are on the same connected component if and only if a and b are in the same element of the partition p.

Every partition p is determined the number of elements of the partition, which is a number $k \leq n$, the sizes of the partitions, $i_1,\dots,i_k$ such that $i_1+\cdots+i_k=n$ and an element of $\mathrm{Sh}(i_1,\dots,i_k)$ specifying which numbers are included in each element of the partition. This data defines a subspace W_p of $\mathrm{sBG}_{H(M)}(n)$ and the map Φ induces maps $\Phi_p\colon \overline{W}_p \to \overline{V}_p$, where $\overline{V}_p = \bigoplus_{p' \text{ coarser than } p} V_{p'}$ and similarly for W_p. It can shown by induction on the size of the partition p that the maps Φ_p are injective for every partition p, so in particular for p the discrete partition we obtain the injectivity of the full map.

This follows from the fact that a linear map $f\colon A\oplus B \to V$ is injective if its restriction to both A and B is injective and $f(A)\cap f(B)=0$ and in our case these two conditions can be verified just by looking at the component of $V_p \subset \overline{V}_p$. □

Corollary 37 *The family of graded vector spaces $\Phi(\mathrm{sBG}_{H(M)}) \subset \mathrm{gr}\,\mathsf{Graphs}_M$ satisfies the recursion*

$\Phi(\mathrm{sBG}_{H(M)}(0)) = \mathbb{R}$,

$\Phi(\mathrm{sBG}_{H(M)}(n)) = \Phi(\mathrm{sBG}_{H(M)}(n-1)) \otimes H^\bullet(M) \oplus \Phi(\mathrm{sBG}_{H(M)}(n-1))[D-1]^{\oplus n-1}$.

Proposition 34 will follow from showing that the inclusion $\Phi(\mathrm{sBG}_{H(M)}) \hookrightarrow \mathrm{gr}\,\mathsf{Graphs}_M$ is a quasi-isomorphism and for this we will use some additional filtrations.

The differential on $\mathrm{gr}\,\mathsf{Graphs}_M$ splits into the terms

$$\delta = \delta_s + \Delta + \Delta_1,$$

where δ_s is obtained by splitting vertices, Δ (the BV part of the differential) removes two decorations and creates an edge instead and Δ_1 connects a connected component of (possibly decorated) internal vertices to the given graph. Let us call the emv-degree (edges minus vertices) of a graph the number

$$\#(\text{edges}) - \#(\text{vertices}).$$

The differential can only increase or leave constant the emv degree. Hence we can put a filtration on $\operatorname{gr}\mathsf{Graphs}_M$ by emv degree. We will denote the associated graded by

$$\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M.$$

The induced differential on the associated graded ignores the Δ part of the differential.

Lemma 38 $$H(\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M) = \Phi(\mathrm{sBG}_{H(M)}).$$

Since in $\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M$ the Δ part of the differential is zero, all pieces of the differential increase the number of internal vertices by at least one. To show this lemma, we will put yet another filtration on $\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M$ by $\#(\text{internal vertices}) - \text{degree}$. Let us call the associated graded

$$\operatorname{gr}''\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M.$$

Notice that in $\operatorname{gr}''\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M$ we have $\Delta = 0$ and the only "surviving" pieces of Δ_1 replace any decoration by an internal vertex with the same decoration or connect a single internal vertex to another vertex of the graph. These pieces also appear in δ_s and it can be checked that they appear with opposite signs, thus canceling out.

Lemma 39 $$H(\operatorname{gr}''\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M) = \Phi(\mathrm{sBG}_{H(M)}).$$

Proof Let us write $V(n) = \operatorname{gr}''\operatorname{gr}'\operatorname{gr}\mathsf{Graphs}_M(n)$ for brevity. We will show that $H(V(n)) \cong \Phi(\mathrm{sBG}_{H(M)}(n))$ by induction on n. We can split

$$\begin{tikzcd} V(n) = & V_0 \arrow[loop above] & \oplus & V_1 \arrow[loop above] & \oplus & V_{\geq 2} \arrow[loop above] \arrow[ll, bend right] \end{tikzcd}$$

according to the valence of the external vertex 1 (where decorations are considered to increase the valence of the vertices). The arrows indicate how the differential maps the individual parts to each other. The complex V_0 is isomorphic to $V(n-1)$ and we can invoke the induction hypothesis. For the remainder we consider a spectral sequence whose first differential is $V_{\geq 2} \to V_1$. Concretely, we consider $(\mathcal{F}_k)_{k\in\mathbb{Z}}$, a descending filtration $V(n) \supset \cdots \supset \mathcal{F}_k \supset \mathcal{F}_{k+1} \supset \cdots \supset 0$, such that $\mathcal{F}_k$ is spanned by graphs of degree at least k in which the vertex 1 is not 1–valent and by graphs of degree at least $k+1$ in which the vertex 1 has valence 1. The map $V_{\geq 2} \to V_1$ is injective and its cokernel is generated by graphs of one of the following types:

(1) Vertex 1 has a decoration and no incoming edges.

(2) Vertex 1 has no decoration and is connected to some other external vertex.

In the first case we obtain a complex isomorphic to $V(n-1)$ for every choice of decoration, with a degree shift given by the decoration. In the second case, each choice of connecting external vertex yields a complex isomorphic to $V(n-1)$ with a degree shift given by the additional edge. This gives us the following expression of the first page of the spectral sequence:

$$\begin{aligned}E_1(V(n)) = H(\mathrm{gr}\, V(n)) &= V_0 \oplus V(n-1) \otimes \overline{H}^\bullet(M) \oplus V(n-1)[D-1]^{\oplus n-1} \\ &= V(n-1) \otimes H^\bullet(M) \oplus V(n-1)[D-1]^{\oplus n-1}.\end{aligned}$$

Under this identification, on the this page of the spectral sequence we obtain precisely the differential of $V(n-1)$. Notice that $V_1 \oplus V_{\geq 2}$ is a double complex concentrated on a double column and therefore the spectral sequence collapses at the second page E_2. From this observation we obtain the recursion

$$H(V(n)) = H(V(n-1)) \otimes H^\bullet(M) \oplus H(V(n-1))[D-1]^{\oplus n-1},$$

which is the same as the recursion for $\Phi(\mathrm{sBG}_{H(M)}(n))$, as shown in Corollary 37. To see that the inclusion $\Phi(\mathrm{sBG}_{H(M)}(n)) \to V(n)$ induces a quasi-isomorphism on the second page of the spectral sequence, we start by noticing that the result holds trivially on the 1–dimensional initial terms $\Phi(\mathrm{sBG}_{H(M)}(0))$ and $H(V(0))$, and therefore $\Phi(\mathrm{sBG}_{H(M)}(n))$ and $H(V(n))$ have the same dimension.

The second page of the inclusion map

$$\begin{aligned}&\Phi(\mathrm{sBG}_{H(M)}(n-1)) \otimes H^\bullet(M) \oplus \Phi(\mathrm{sBG}_{H(M)}(n-1))[D-1]^{\oplus n-1} \\ &\qquad \to H(V(n-1)) \otimes H^\bullet(M) \oplus H(V(n-1))[D-1]^{\oplus n-1}\end{aligned}$$

can be written as

$$\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix},$$

where $f_{12}\colon \Phi(\mathrm{sBG}_{H(M)}(n-1))[D-1]^{\oplus n-1} \to H(V(n-1)) \otimes H^\bullet(M)$ is actually the 0 map, since $\Phi(\mathrm{sBG}_{H(M)}(n-1))[D-1]^{\oplus n-1}$ corresponds to the image of elements in $H^\bullet(M)^{\otimes k}[kD] \otimes \Lambda^{D-1}\mathsf{Lie}(i_1) \otimes \cdots \otimes \Lambda^{D-1}\mathsf{Lie}(i_k)$ with $i_1 \geq 2$ and the vertex 1 cannot be the only labeled vertex in its connected component. The maps f_{11} and f_{22} are isomorphisms by induction, and therefore the second page of the inclusion map is an isomorphism, whence the result follows. □

Proof of Lemma 38 The E^1 term of the spectral sequence is a quotient complex, hence it abuts at that point. □

Proof of Theorem 25 We have shown that the composition

$$\widetilde{\mathrm{BG}} \xrightarrow{F} \Omega_{\mathrm{PA}}(\mathsf{FM}_M)^* \xrightarrow{\omega^*} \mathsf{Graphs}_M$$

is a quasi-isomorphism, but since the homology of $\Omega_{\mathrm{PA}}(\mathsf{FM}_M)^*$ is also isomorphic to the other two homologies which are finite-dimensional in each arity and degree, it follows that F and ω^* are quasi-isomorphisms themselves.

Consequentially, the map $\mathrm{Chains}(\mathsf{FM}_M) \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)^* \to \mathsf{Graphs}_M$ is a composition of quasi-isomorphisms, therefore is a quasi-isomorphism as well. □

Remark 40 For the proof of Theorem 25 we consider the functor Ω_{PA} of semialgebraic forms, but one could equally use any contravariant functor Ω landing in dgca's satisfying the following properties:

- Ω is quasi-isomorphic to the Sullivan functor A_{PL} of piecewise-linear de Rham forms.
- Ω admits pushforwards of the forgetful maps $\mathsf{FM}_M(n) \to \mathsf{FM}_M(n-k)$ satisfying the usual properties of fiber integrals, in particular Stokes' theorem.
- Ω is "almost" comonoidal, as in Remark 15.

6 The nonparallelizable case

Let M be a closed oriented connected manifold. In this section we show that even in absence of the parallelizability hypothesis, a slight variant of the collection of commutative algebra ${}^*\mathsf{Graphs}_M$ is still a model of FM_M.

In this respect it is not natural to consider graphs with tadpoles as the compatibility of the differential of the map from Lemma 16 depended on the vanishing of the Euler characteristic for those graphs. More precisely, the problem is that in the map of Lemma 16 a tadpole edge is sent to a form whose coboundary is the Euler class.

We define ${}^*\mathsf{Gra}_M^{\not\circlearrowleft} \subset {}^*\mathsf{Gra}_M$ to be the dg Hopf subcollection spanned by graphs without tadpoles.

Note that this subcollection is indeed closed under the product and differential. It furthermore retains a $\Lambda^{D-1}\mathsf{Lie}^*$–comodule structure from ${}^*\mathsf{Gra}_M$, but not the full

${}^*\mathsf{Gra}_D$–comodule structure, as the proof of Proposition 14 fails in the absence of tadpoles. Furthermore, the map (8) naturally restricts to a map of dg Hopf collections

$$ {}^*\mathsf{Gra}_M^{\not\circlearrowleft} \to \Omega_{\mathrm{triv}}(\mathsf{FM}_M) \subset \Omega_{\mathrm{PA}}(\mathsf{FM}_M), $$

which is well defined even if M has a nontrivial Euler class. The twisting construction of Section 4 and in particular the construction of the map ω of Corollary 23 also naturally yields a map

$$ \omega\colon {}^*\mathsf{Graphs}_M^{\not\circlearrowleft} \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M), \quad \Gamma \mapsto \omega_\Gamma, $$

where we denote by ${}^*\mathsf{Graphs}_M^{\not\circlearrowleft} \subset {}^*\mathsf{Graphs}_M$ the subcollection spanned by graphs without tadpoles.

To be clear, if M has nonvanishing Euler class then the map ω of Corollary 23 is not a priori not well-defined on Graphs_M because we would need to send a tadpole edge to a form whose coboundary is the Euler class. Furthermore, the partition function (9) is only well defined on the tadpole-free part ${}^*\mathrm{fGC}_{H^\bullet(M)}^{\not\circlearrowleft} \subset {}^*\mathrm{fGC}_{H^\bullet(M)}$. Hence one does not even get a well-defined (square-zero) differential on the graded collection ${}^*\mathsf{Graphs}_M$ from the partition function, one only has this on the tadpole-free part ${}^*\mathsf{Graphs}_M^{\not\circlearrowleft}$.

In particular, we note that the differential on ${}^*\mathsf{Graphs}_M^{\not\circlearrowleft}$ can indeed not produce tadpoles. The only term in the differential that is able to produce a tadpole is the edge contraction in the presence of a multiple edge. However, multiple edges are zero by symmetry reasons for even D while tadpoles are not present by symmetry reasons for odd D, hence no problem arises.

Also, if M is not parallelized, there is no consistent way of defining a right FM_D–action on FM_M. Nonetheless, disregarding the cooperadic coactions, the map

$$ {}^*\mathsf{Graphs}_M^{\not\circlearrowleft} \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M) $$

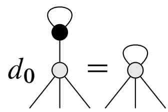
is well defined as a map of dgcas since the proof of Lemma 19 uses parallelizability condition only for the tadpoles and the coaction; see the remarks within that proof on using the trivialization of the tangent bundle.

Before proceeding, let us furthermore show that the exclusion of tadpoles has no effect on the homotopy type, provided ${}^*\mathsf{Graphs}_M$ is well defined. (See [45, Proposition 3.4] for similar results and arguments.)

Proposition 41 *Suppose that M is parallelizable* (*or at least has vanishing Euler class*), *so that the dg Hopf collection ${}^*\mathsf{Graphs}_M$ is well defined. Then the inclusion*

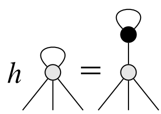
$^*\mathsf{Graphs}_M^{\not\varnothing} \to {}^*\mathsf{Graphs}_M$ *is a quasi-isomorphism of collections of dg commutative algebras.*

Proof sketch We consider a spectral sequence on $^*\mathsf{Graphs}_M$ whose associated graded has a differential contracting internal vertices with only an adjacent edge and a tadpole along the nontadpole edge:



Such a spectral sequence can be obtained by filtering first by the number of tadpoles and then by $l+$ degree, where l is the sum of lengths of maximal connected subgraphs consisting of 2–valent internal vertices and one internal vertex with just a tadpole at the end.

We can then set up a homotopy h that splits out an internal vertex with a tadpole:



We have $d_0 h + h d_0 = T\mathrm{id}$, where T is the number of tadpoles, whence it follows that $H({}^*\mathsf{Graphs}_M, d_0) = {}^*\mathsf{Graphs}_M^{\not\varnothing}$. □

Finally, one has the following version of Theorem 25 for nonparallelizable M.

Theorem 42 *Let M be a closed oriented manifold. The map*

$$\omega_\bullet \colon {}^*\mathsf{Graphs}_M^{\not\varnothing} \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)$$

is a quasi-isomorphism of symmetric sequences of dg commutative algebras. Similarly, the composition map

$$\mathrm{Chains}(\mathsf{FM}_M) \to \Omega_{\mathrm{PA}}(\mathsf{FM}_M)^* \xrightarrow{\omega_\bullet^*} \mathsf{Graphs}_M^{\not\varnothing} := ({}^*\mathsf{Graphs}_M^{\not\varnothing})^*$$

is a quasi-isomorphism.

Proof We follow the proof of Theorem 25. First we note that while in general one does not have a right FM_D–module structure on FM_M if M is not framed, the insertion of fundamental chains of FM_D at points in FM_M is independent of the framing so in fact it gives us a well-defined operadic action $\mathrm{Chains}(\mathsf{FM}_M) \circ \Lambda^{D-1} L_\infty \to \mathrm{Chains}(\mathsf{FM}_M)$. Similarly, as mentioned above, $\mathsf{Graphs}_M^{\not\varnothing}$ inherits a right $\Lambda^{D-1} L_\infty$–module structure from the one on $\mathsf{Gra}_M^{\not\circ} := ({}^*\mathsf{Gra}_M^{\not\circ})^*$. These structures suffice to define the map of right $\Lambda^{D-1}L_\infty$–modules

$$\Phi\colon \widetilde{\mathrm{BG}} \to \mathsf{Graphs}_M^{\not\circ}$$

as in Section 5.5 by formula (10) (respectively (11)). Furthermore, Lemma 29 does not make use of tadpoles and holds in this case as well.

Furthermore, the remaining arguments of Sections 5.6 and 5.7 leading to Theorem 25 are agnostic to the presence or absence of parallelizability of M or tadpoles in graphs, and hence also show Theorem 42. □

7 A simplification of ${}^*\mathsf{Graphs}_M$ and relations to the literature

7.1 An alternative construction of Graphs_M.

Recall that in Section 4 the space ${}^*\mathsf{Graphs}_M$ was constructed by identifying connected components without external vertices with real numbers via a "partition function", which is a map of commutative algebras $Z_M\colon {}^*\mathrm{fGC}_{H^\bullet(M)} \to \mathbb{R}$.

In this subsection and the next we present an alternative construction of Graphs_M that will allow us to understand better the relevance of the partition function Z_M in the homotopy type of Graphs_M.

Notice that ${}^*\mathrm{fGC}_{H^\bullet(M)}$ is a quasifree commutative algebra generated by its subspace of connected graphs ${}^*\mathrm{GC}_{H(M)}$. The differential d on ${}^*\mathrm{fGC}_{H^\bullet(M)}$ defines then a ΛL_∞ coalgebra structure on ${}^*\mathrm{GC}_{H^\bullet(M)}$. In fact, since the differential can increase the number of connected components by at most one, this is in fact a strict Lie coalgebra structure.

The dual Lie algebra structure is denoted by $\mathrm{GC}_{H^\bullet(M)} = ({}^*\mathrm{GC}_{H^\bullet(M)})^*$ and is represented by infinite sums of graphs decorated by $H_\bullet(M)$ (or dually by $H^\bullet(M)$, via the Poincaré pairing). The Lie bracket $[\Gamma, \Gamma']$ is given by summing over all possible ways of selecting a decoration in Γ and another decoration in Γ' and connecting them into an edge, with a factor given by their pairing. The differential acts by vertex splitting and joining decorations.

It follows that maps of dg commutative algebras ${}^*\mathrm{fGC}_{H^\bullet(M)} \to \mathbb{R}$ are identified with maps in the Lie algebra satisfying the Maurer–Cartan equation,

$$\mathsf{MC}(\mathrm{GC}_{H^\bullet(M)}) = \mathrm{Hom}_{\mathrm{dgca}}({}^*\mathrm{fGC}_{H^\bullet(M)}, \mathbb{R}).$$

We denote by $z_M \in \mathrm{GC}_{H^\bullet(M)}$ the Maurer–Cartan element corresponding to the partition function Z_M. If we consider the subrepresentation $S \subset \mathrm{Tw}\,\mathsf{Gra}_M$ given by graphs with no connected components consisting only of internal vertices, then Graphs_M is obtained by twisting S by the Maurer–Cartan element z_M, as recalled in the following section.

Analogously, we denote by $\mathrm{GC}_M := \mathrm{GC}^{z_M}_{H^\bullet(M)}$ the Lie algebra obtained by twisting with the Maurer–Cartan element z_M.

For later use let us also split the Maurer–Cartan element

$$z_M = \underbrace{\sum_{i,j=1}^{D} g^{ij}\, e_i \cdots \bullet \cdots e_j}_{=: z_0} + z'_M$$

into a part z_0 given by graphs with exactly one vertex and 2 or 1 decorations and a remainder $z'_M := z_M - z_0$. Note in particular that z_0 is determined solely by the nondegenerate pairing on $H(M)$. The element z_0 is itself a Maurer–Cartan element, and below we will consider the twisted dg Lie algebra

$$\mathrm{GC}'_{H(M)} := \mathrm{GC}^{z_0}_{H(M)},$$

and consider z'_M as a Maurer–Cartan element in $\mathrm{GC}'_{H(M)}$.

7.2 Twisting of modules

While the differential of Graphs_M can be very nonexplicit, expressing it as twist by a Maurer–Cartan element opens the door to simplifications of the model, as long as we have some control over the gauge equivalence class of the Maurer–Cartan element.

Indeed, let us pause for a moment to consider the following general situation. Suppose $\mathfrak{g}$ is a dg Lie algebra, acting on M, where M can be just a dg vector space, or a (co)operad or a (co)operadic (co)module, or a pair of a (co)operad and a (co)operadic (co)module. In any case we require the $\mathfrak{g}$–action to respect the given algebraic structure, in the sense that the action is by (co)derivations.

Suppose now that $m \in \mathfrak{g}$ is a Maurer–Cartan element, ie $dm + \frac{1}{2}[m,m] = 0$. Then we can form the twisted Lie algebra $\mathfrak{g}^m$ with the same Lie bracket, but differential $d_m = d + [m, -]$. We can furthermore form the twisted ($\mathfrak{g}^m$–)module M^m, which is the same space as M, carrying the same action and underlying algebraic structure (operad, operadic module, etc), but whose differential becomes

$$d_m = d + m\cdot,$$

where $m\cdot$ shall denote the action of m and we denote the original differential on M by d. Next suppose that $m' \in \mathfrak{g}$ is another Maurer–Cartan element. We say that m and m' are gauge equivalent if there is a Maurer–Cartan element $\hat{m} \in \mathfrak{g}[t, dt]$ whose restriction to $t = 0$ agrees with m, and whose restriction to $t = 1$ agrees with m'. More concretely,

$$\hat{m} = m_t + dt h_t,$$

where m_t can be understood as a family of Maurer–Cartan elements in $\mathfrak{g}$, connected by a family of infinitesimal homotopies (gauge transformations) h_t. The Maurer–Cartan equation for $\hat{m}$ translates into the two equations

$$dm_t + \tfrac{1}{2}[m_t, m_t] = 0, \qquad \frac{\partial m_t}{\partial t} + dh_t + [h_t, m_t] = 0.$$

Now suppose that $\mathfrak{g}$ is pro-nilpotent. Then we may form the exponential group $\mathrm{Exp}(\mathfrak{g})$, which is identified with the degree 0 subspace $\mathfrak{g}_0 \subset \mathfrak{g}$, with group product given by the Baker–Campbell–Hausdorff formula. We can integrate the flow of h_t into the element $H_t \in \mathrm{Exp}(\mathfrak{g})$, which acts on $x \in \mathfrak{g}$ by

$$H_t(x) = \exp(h_t) \cdot x = \alpha + \sum_{n \geq 0} \frac{\mathrm{ad}^n(h_t)}{(n+1)!}([h_t, x] - dh_t).$$

The action of H_t is compatible with the Lie bracket and has the property that, for every $x \in \mathfrak{g}$,

$$H_t(dx + [m, x]) = (d + [m_t, -])H_t(x).$$

In particular, the action of H_1 induces an isomorphism of dg Lie algebras,

$$H_1 \colon \mathfrak{g}^m \to \mathfrak{g}^{m'}.$$

Next suppose that also the action of $\mathfrak{g}$ on M is pro-nilpotent. Then, by a similar argument, the action of H_1 yields an isomorphism

$$H_1 \cdot \colon M^m \to M^{m'}. \tag{13}$$

Now let us relate these general statements to the objects of relevance in this paper. First consider $\mathfrak{g} = \mathrm{GC}_D$ to be the graph complex, but as a graded Lie algebra, ie considered with zero differential. The correct differential on the graph complex is then obtained by twisting with the Maurer–Cartan element [45]

$$m_0 = \bullet\!\!-\!\!\!-\!\!\bullet\,.$$

Furthermore, consider $M = {}^*\mathsf{Graphs}_D$, again with zero differential. There is a natural action of $\mathfrak{g}$ on M [17; 45]. The differential on ${}^*\mathsf{Graphs}_D = M^{m_0}$ is then reproduced by twisting with m_0.

Secondly, the above picture can be extended to include the (co)operadic right modules. First, GC_D acts on $\mathrm{GC}_{H(M)}$. We take

$$\mathfrak{g} = \mathrm{GC}_D \ltimes \mathrm{GC}_{H(M)},$$

where we consider again the first factor with trivial differential, and the second factor only with the part of the differential joining two decorations to an edge. The element m_0 from above is then a Maurer–Cartan element, and twisting by this Maurer–Cartan element reproduces the differential on the factors of $\mathfrak{g}$ considered above. Similarly, we may consider the Maurer–Cartan elements

$$m' := m_0 + z_0 \quad \text{or} \quad m_M := m_0 + z_M,$$

where z_0 and z_M are as above. Twisting then reproduces on the second factor either the differential on $\mathrm{GC}'_{H(M)}$, or that on GC_M.

Next consider for M the pair consisting of a cooperad and a comodule

$$({}^*\mathsf{Graphs}_D, {}^*\mathsf{Graphs}_M),$$

where the first factor we consider with the zero differential, and in the second we consider only the part that connects two decorations to an edge. Then twisting with the Maurer–Cartan element m_M reproduces the full differential on the factors.

Remark 43 An immediate consequence of the above way of constructing ${}^*\mathsf{Graphs}_M$ is that one has a large class of (co)derivations at hand. Namely, we have an action of $\mathfrak{g}^{z_M}$ on M^{z_M}. In particular, it was shown in [45] that the 0^{th} cohomology of GC_2 is the Grothendieck–Teichmüller algebra $\mathfrak{grt}_1$. Hence, overstretching the analogy a bit, we may consider the dg Lie algebra $\mathfrak{g}^{z_M}$, consisting of factors GC_D and GC_M, as a version of the Grothendieck–Teichmüller dg Lie algebra associated to the manifold M. Note however that this "definition" is a little provisional; a more invariant definition would be to define the M–Grothendieck–Teichmüller Lie algebra as the homotopy derivations of a real model of the pair $(\mathsf{FM}_D, \mathsf{FM}_M)$. It is yet an open question in how far the homotopy derivations in $\mathfrak{g}^{z_M}$ exhaust all homotopy derivations. For example, $\mathfrak{g}^{z_M}$ itself does not readily capture the (nonnilpotent) action of the Lie algebra $\mathfrak{o}(H(M))$ (of linear maps that preserve the pairing) on all objects involved.

Next, let us note that the right comodule ${}^*\mathsf{Graphs}_M$ is unaltered (up to isomorphism) if one replaces the Maurer–Cartan element z_M used in its definition by a gauge equivalent Maurer–Cartan element. Indeed, the action of $\mathrm{GC}_{H(M)}$ is nilpotent since the action of any element in $\mathrm{GC}_{H(M)}$ always kills at least on vertex. Hence given two gauge-equivalent Maurer–Cartan elements, an explicit isomorphism between the two versions of ${}^*\mathsf{Graphs}_M$ produced is given by (13).

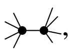
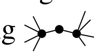
Finally, let us note that the above construction works equally well for the tadpole-free version ${}^*\mathsf{Graphs}_M^{\not\circlearrowleft}$ of ${}^*\mathsf{Graphs}_M$. In this case, one needs to work with the tadpole-free version of the graph complex GC_M. Also, in this case one does not have a right ${}^*\mathsf{Graphs}_D$ coaction.

7.3 Valence conditions

In this section we show that the Hopf comodule ${}^*\mathsf{Graphs}_M$ is quasi-isomorphic to (essentially) a quotient that can be identified with graphs containing only ≥ 3–valent internal vertices. For this, we would like that the Maurer–Cartan element (partition function) z'_M above vanished on the subspace spanned by graphs containing a ≤ 2–valent internal vertex. While this might not be the case in general, we show that z_M is gauge equivalent to a partition function satisfying this property.

Lemma 44 *The subspace $\mathrm{GC}^{\geq 3}_{H^\bullet(M)} \subset \mathrm{GC}'_{H^\bullet(M)}$ spanned by graphs having no 1 or 2–valent vertex is a dg Lie subalgebra.*

Proof $\mathrm{GC}^{\geq 3}_{H^\bullet(M)}$ is closed under the Lie bracket since it does not decrease the valence of vertices. It remains to check the stability under the differential.

Recall that the differential has three pieces, a first one that splits an internal vertex, a second one that joins decorations into an edge, and a third one arising from the twist by z_0. Joining decorations into an edge cannot decrease the valency on vertices and therefore preserves $\mathrm{GC}^{\geq 3}_{H^\bullet(M)}$. Univalent or bivalent vertices can a priori be created both by the second and third term in the differential. However, one easily checks that these ≤ 2–valent contributions cancel due to signs. For example, when computing the differential of the graph , bivalent vertices are created by vertex splitting . However, since there are two contributions corresponding to each of the two vertices, and they appear with opposite signs, they cancel out. For bivalent vertices carrying a decoration, or for a univalent vertex, the argument is similar. □

Let $\mathrm{GC}''_{H^\bullet(M)}$ be the subspace of $\mathrm{GC}'_{H^\bullet(M)}$ spanned by graphs that (i) do not contain any univalent vertices, and (ii) contain at least one ≥ 3–valent vertex. Notice that $\mathrm{GC}''_{H^\bullet(M)}$ is a sub-Lie algebra of $\mathrm{GC}_{H^\bullet(M)}$ since the Lie bracket cannot decrease any valences. Furthermore, we have the following easy result.

Lemma 45 *The Maurer–Cartan element $z'_M \in \mathrm{GC}'_{H^\bullet(M)}$ constructed above lives in the subspace $\mathrm{GC}''_{H^\bullet(M)}$.*

Proof First note that by definition z'_M contains no graphs with a single ≤ 2–valent vertex, as those graphs have been absorbed into z_0 above. Hence the only instance of a (connected) graph with a univalent vertex is a graph with an "antenna", ie an edge connected to a univalent vertex. However, to such graphs the configuration space integral formula associates weight 0, by property (iv) of Proposition 8 (or alternatively by a degree argument, since there are not enough form degrees depending on the position of the antenna vertex). Next, if the graph has no trivalent vertices, it is either a string, with some decorations at the ends, or a loop. In case of a string, the weight is zero again by (iv) of Proposition 8. Finally, the loops all have zero weight by degree reasons. □

The following proposition is essentially proven in [45, Proposition 3.4]. One uses essentially the dual argument of Theorem 49.

Proposition 46 *The inclusion map $\mathrm{GC}^{\geq 3}_{H^\bullet(M)} \hookrightarrow \mathrm{GC}''_{H^\bullet(M)}$ is a quasi-isomorphism of Lie algebras. Furthermore, endowing both sides with the descending complete filtrations by the number of nonbivalent vertices,*[7] *the map between the associated graded spaces is already a quasi-isomorphism.*

Due to this proposition we can apply the Goldman–Millson theorem [16] to conclude that any Maurer–Cartan element in $\mathrm{GC}''_{H^\bullet(M)}$ is gauge equivalent to a Maurer–Cartan element in the subspace $\mathrm{GC}^{\geq 3}_{H^\bullet(M)}$. In particular:

Corollary 47 *The Maurer–Cartan element z'_M is gauge equivalent to a Maurer–Cartan element in the subspace $\mathrm{GC}^{\geq 3}_{H^\bullet(M)}$.*

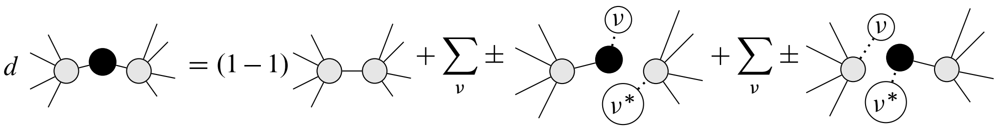
Next, we apply the remark of the previous subsection to conclude that we may use a ≥ 3–valent Maurer–Cartan element (say z^3_M) gauge equivalent to z'_M to construct ${}^*\mathsf{Graphs}_M$. For the sake of concreteness, let us temporarily (for this subsection) denote the version of ${}^*\mathsf{Graphs}_M$ constructed as before by $\mathsf{Graphs}_M^{z'_M}$, and the one constructed with z_3 instead by ${}^*\mathsf{Graphs}_M^{z_3}$, though this is an abuse of notation.

Let us consider a subspace S of ${}^*\mathsf{Graphs}_M^{z_3}$ spanned by graphs having at least one internal 1– or 2–valent vertex. Recall that decorations count as valence and there are no 0–valent internal vertices in ${}^*\mathsf{Graphs}_M$.

[7] On $\mathrm{GC}^{\geq 3}_{H^\bullet(M)}$ this filtration is quite trivial.

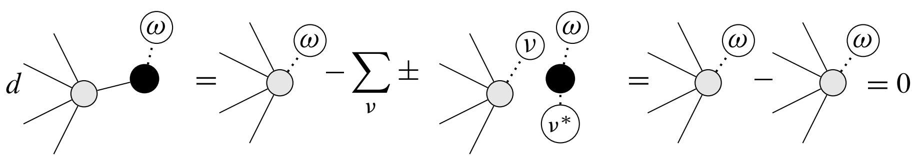
Lemma 48 *The space S described above is a subcomplex of* ${}^*\mathsf{Graphs}_M^{z_3}$.

Proof Recall that the differential has two pieces, a first one that contracts an edge connected to an internal vertex and a second one that either cuts an edge into the diagonal class or deletes a subgraph of internal vertices producing a factor given by the image of such subgraph under Z_M. Due to the Maurer–Cartan element z_3 containing only ≥ 3–valent diagrams, the differential cannot cut out a subgraph containing a bivalent internal vertex. Let us consider a graph with a 2–valent internal vertex that is adjacent to two other vertices. There, the differential acts as follows:



The contributions of contracting both edges appear with opposite signs and therefore cancel. Notice that 1–valent internal vertices are produced on the other summands when the decoration of the internal vertex takes the value 1.

If there is a 2–valent internal vertex that is adjacent to only one other vertex and has one decoration, the action of the differential there is:



It is easy to see that if there is one 1–valent internal vertex the two pieces of the differential cancel each other, thus concluding the proof. □

The following proof is an adaptation of [45, Proposition 3.4].

Theorem 49 *The projection map* ${}^*\mathsf{Graphs}_M^{z_3} \to {}^*\mathsf{graphs}_M := {}^*\mathsf{Graphs}_M^{z_3}/S$ *is a quasi-isomorphism of dg Hopf right* ${}^*\mathsf{Graphs}_D$*–comodules.*

Proof It suffices to show that $H(S) = 0$. If we set up a filtration on S by the total number of decorations, on the zeroth page of the spectral sequence we recover d_0 as the contracting piece and a piece that cuts out a connected component of internal vertices with a factor given by an integral. We claim that the spectral sequence collapses already on the first page.

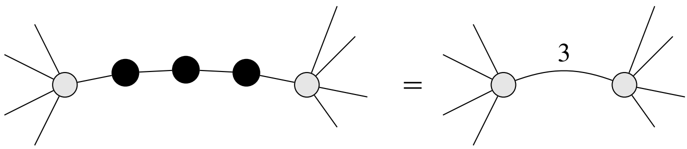


Figure 3: Replacing bivalent internal vertices by a single labeled edge.

Notice that d_0 cannot produce 1–valent internal vertices from 2–valent internal vertices and it follows from the proof of Lemma 48 that a 1–valent internal vertex cannot be destroyed.

It follows that on the zeroth page S decomposes as a sum of complexes $S = S_1 \oplus S_2$, where S_1 is spanned by graphs with at least one 1–valent internal vertex and S_2 is spanned by graphs whose internal vertices are at least 2–valent.

To see that S_1 is acyclic one can look at "antennas" of the graphs, ie maximal connected subgraphs consisting of one 1–valent and some 2–valent internal vertices. By setting a spectral sequence whose differential decreases only the length of antennas, one can construct a contracting homotopy that increases this length; thus showing $H(S_1) = 0$.

As for S_2, the same idea can used by replacing every path on the graph consisting of 2–valent internal vertices by single edges labeled by their length; see Figure 3.

By considering a spectral sequence whose differential on the zeroth page only reduces the numbers on the labels, being careful with the signs one can construct a contracting homotopy which gives $H(S_2) = 0$. □

Overall, we conclude that ${}^*\mathsf{graphs}_M$ is a dgca model for FM_M, by the explicit zigzag

$${}^*\mathsf{graphs}_M \xleftarrow{\sim} {}^*\mathsf{Graphs}_M^{z_3} \xleftarrow{\cong} {}^*\mathsf{Graphs}_M^{z'_M} \xrightarrow{\sim} \Omega_{\mathrm{PA}}(\mathsf{FM}_M).$$

Moreover, the above maps are morphisms of dg Hopf right comodules.

If M is not parallelizable, one can construct the space ${}^*\mathsf{graphs}_M^{\varnothing}$ as the analogous quotient of ${}^*\mathsf{Graphs}_M^{z_3}$. The same proof allows us to conclude that ${}^*\mathsf{graphs}_M^{\varnothing}$ is a dgca model for the collection of topological spaces FM_M by a similar zigzag.

Remark 50 The smaller model ${}^*\mathsf{graphs}_M$ (as well as ${}^*\mathsf{graphs}_M^{\varnothing}$) has the advantage that for $D \geq 3$ it is connected in the sense that each dgca ${}^*\mathsf{graphs}_M(r)$ is concentrated in nonnegative cohomological degrees, and one-dimensional in degree 0. This can be shown by a degree counting argument similar to Lemma 54, using the trivalence condition and the existence of at least one external vertex per connected component. Similarly, one sees that if in addition $H^1(M) = 0$, then ${}^*\mathsf{graphs}_M(r)$ is finite-dimensional in each cohomological degree.

Remark 51 The propagator ϕ_{12} established in Proposition 8 can be chosen so that $\int_2 \phi_{12}\alpha = 0$, where the integration is conducted along the fiber of the forgetful map $p_2\colon \mathsf{FM}_M(2) \to M$, and where α is any of the chosen representative forms for the cohomology; see Convention 7 (also [13]). It would be desirable to show that ϕ_{12} may be chosen so that in addition $\int_3 \phi_{13}\phi_{32} = 0$, where the integration is performed along the fiber of the forgetful map $p_3\colon \mathsf{FM}_M(3) \to \mathsf{FM}_M(2)$. In that case the above discussion could be considerably simplified, since the extra condition immediately renders the integral weights of all graphs with bivalent vertices zero. A propagator with this desired property has been constructed in the smooth setting in [13, Lemma 4]. We expect that the proof carries over to the semialgebraic setting. However, there is a technical difficulty due to our use of PA instead of smooth forms, whose resolution we leave to future work. Roughly speaking, the technical problem is that for a PA form $\beta \in \Omega(M \times N)$ one has to define a good notion of "de Rham differential in the first slot" $d_N\beta$.

7.4 Computing the cohomology and loop orders

Above we construct real dgca models ${}^*\mathsf{Graphs}_M$ and ${}^*\mathsf{graphs}_M$ for configuration spaces of points on a manifold M, which depend on M only through the Maurer–Cartan element $z_M \in \mathrm{GC}_{H^\bullet(M)}$. Note that $\mathrm{GC}_{H^\bullet(M)}$ is naturally filtered by the loop order of graphs. We can decompose the Maurer–Cartan element

$$z_M = z_M^0 + z_M^1 + \cdots$$

accordingly into pieces of various loop orders.

The differential on ${}^*\mathsf{graphs}_M(n)$ can only maintain or decrease the number of loops (genus) of the graphs. It follows that the subspace ${}^*\mathsf{graphs}_M^{\mathrm{for}}(n) \subset {}^*\mathsf{graphs}_M(n)$ spanned by graphs of genus zero, ie forests, is a subcomplex and a dg subalgebra for $n = 1$. Notice that however it is not a subalgebra if $n > 1$. In any case the object ${}^*\mathsf{graphs}_M^{\mathrm{for}}$ depends on M only through the tree-level piece z_M^0 of our Maurer–Cartan element z_M.

Lemma 52 *The inclusion of ${}^*\mathsf{graphs}_M^{\mathrm{for}}$ in ${}^*\mathsf{graphs}_M$ is a quasi-isomorphism* (*of symmetric sequences of complexes*).

Proof The proof follows essentially from the spectral sequence argument given in Lemma 39.

The differential in ${}^*\mathsf{graphs}_M$ cannot decrease the number of connected components of a graph, so by considering a filtration by the number of connected components of the graphs we obtain the respective associated graded complexes $\mathrm{gr}\,{}^*\mathsf{graphs}^{\mathrm{for}}$ and $\mathrm{gr}\,{}^*\mathsf{graphs}_M$. Then we notice that the number $\#\,\mathrm{edges} - \#\mathrm{vertices}$ cannot increase so we take the respective filtration obtaining the associated graded complexes $\mathrm{gr}'\,\mathrm{gr}\,{}^*\mathsf{graphs}^{\mathrm{for}}$ and $\mathrm{gr}'\,\mathrm{gr}\,{}^*\mathsf{graphs}_M$ (notice that this filtration is bounded below since there are no connected components of only internal vertices). After this, the only piece of the differential remaining is the one cutting out a (decorated) tree of internal vertices and evaluating the partition function on it.

At last, filtering by #internal vertices − degree, we obtain in the associated graded complexes $\mathrm{gr}''\,\mathrm{gr}'\,\mathrm{gr}\,{}^*\mathsf{graphs}^{\mathrm{for}}$ and $\mathrm{gr}''\,\mathrm{gr}'\,\mathrm{gr}\,{}^*\mathsf{graphs}_M$ a the piece of the differential that reduces the number of internal vertices exactly by 1, ie the differential contracts one edge connected to one or two internal vertices or cuts out a tree consisting only of a single decorated internal vertex.

We claim that the induced inclusion map is a quasi-isomorphism at this level. As in Lemma 39, by induction on n one can show that the homology of

$$V(n) = \mathrm{gr}''\,\mathrm{gr}'\,\mathrm{gr}\,{}^*\mathsf{graphs}_M(n)$$

satisfies

$$H(V(n)) = H(V(n-1)) \otimes H^\bullet(M) \oplus H(V(n-1))[1-D]^{\oplus n-1},$$

but the same proof gives the same result for the homology of $\mathrm{gr}''\,\mathrm{gr}'\,\mathrm{gr}\,{}^*\mathsf{graphs}_M^{\mathrm{for}}$, so the result follows. □

In particular we see the following:

(1) The dgca ${}^*\mathsf{graphs}_M^{\mathrm{for}}(1)$ is a real model for M, so that the tree-level piece of z_M encodes the real homotopy type of M.

(2) Knowledge of the tree-level piece of z_M suffices to compute the real cohomology of $\mathsf{FM}_M(n)$, as a graded vector space, for all n.

8 The real homotopy type of M and FM_M

The goal of this section is to compare the information contained in the partition function z_M from above to the real homotopy type of M. By the latter, we mean the isomorphism type of a homotopy commutative (C_∞) algebra structure on the cohomology $H(M)$.

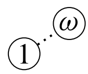
The end result will be that the knowledge of the real homotopy type of M suffices to recover z_M (up to gauge equivalence) in the case that $D \geq 4$ and $H^1(M) = 0$.

Let us first see how the C_∞–algebra structure on $H(M)$ can be obtained from our graphical models. For every closed oriented connected manifold M we fix the following *homotopy data* of chain complexes,

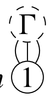
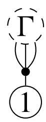
$$\begin{tikzcd} H^\bullet(M) \arrow[r, shift right, "i"'] & {}^*\mathsf{graphs}_M^{\mathrm{for}}(1) \arrow[l, shift right, "p"'] \arrow[loop right, "h"] \end{tikzcd}$$

$$pi = \mathrm{id}, \quad \mathrm{id} - ip = dh + hd,$$

where the map i is defined so that

$$i(\omega) =$$ 

and the map h is defined so that

h  $=$ 

and it vanishes on graphs with a ≤ 1–valent external vertex.

Finally, p is defined so that for every $\Gamma \in {}^*\mathsf{graphs}_M^{\mathrm{for}}$, $p(\Gamma) = \sum_i e_i \int_M e_i^* \wedge f(\Gamma)$, where the $\{e_i\}$ form a basis of $H^\bullet(M)$ and $\{e_i^*\}$ the respective dual basis and

$$f : {}^*\mathsf{graphs}_M^{\mathrm{for}}(1) \to \Omega_{\mathrm{PA}}(M)$$

is the map induced by the one constructed in Section 3.

By the homotopy transfer theorem [33, Section 10.3] such homotopy data defines a C_∞–structure on $H^\bullet(M)$ and such structure retains the real homotopy type of M.

Notice that C_∞–structures on $H^\bullet(M)$ are identified with Maurer–Cartan elements in the Harrison complex

$$\begin{aligned} \mathrm{Harr}(H^\bullet(M), H^\bullet(M)) &= \mathrm{Hom}_{\mathbb{S}}(\mathsf{Lie}^c\{1\}[-1] \circ H^\bullet(M), H^\bullet(M)) \\ &= \prod_{n \in \mathbb{N}} \mathsf{Lie}(n) \otimes_{\mathbb{S}_n} H_\bullet(M)^{\otimes n} \otimes H^\bullet(M)[n]. \end{aligned}$$

Proposition 53 [35, Proposition 1.6.5] *The projection map*

$$\mathrm{Harr}(H^\bullet(M), H^\bullet(M)) \to \mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M))$$

is a quasi-isomorphism of Lie algebras.

Lemma 54 *If M is a connected manifold of dimension at least $D \geq 4$ such that $H^1(M) := H^1(M,\mathbb{R}) = 0$, then all the degree* 0 *graphs in* ${}^*\mathrm{GC}_M^{\geq 3}$ *are trees.*

Proof The proof is a simple combinatorial argument. Let $\Gamma \in {}^*\mathrm{GC}_M^{\geq 3}$ be a nontree graph with E edges and V vertices. We denote the sum of degrees of the decorations of a vertex v_i by $\deg\mathrm{dec}(v_i)$ and the number of incident vertices at v_i by $\mathrm{edges}(v_i)$.

From the relation $\sum_{i=1}^V \mathrm{edges}(v_i) = 2E$, it follows that

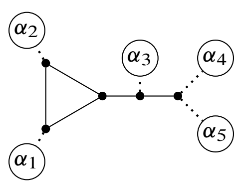
$$\begin{aligned}\deg(\Gamma) &= (D-1)E - DV + \sum_{i=1}^{V} \deg\mathrm{dec}(v_i)\\ &= (D-3)(E-V) + \sum_{i=1}^{V}(\deg\mathrm{dec}(v_i) + \mathrm{edges}(v_i) - 3).\end{aligned}$$

Because of the ≥ 3–valence condition, each term $\deg\mathrm{dec}(v_i) + \mathrm{edges}(v_i) - 3$ must be greater than or equal te zero. In fact, since decorations have degree at least 2 if there is at least one decoration in Γ, the sum $\sum_{i=1}^V(\deg\mathrm{dec}(v_i) + \mathrm{edges}(v_i) - 3)$ is strictly positive.

Now notice that since Γ is a not a tree, we have $E \geq V$ and in case of equality there must be at least one decoration. In any of those cases it follows that $\deg\Gamma > 0$. □

Remark 55 From the proof we also observe the following:

- If $D = 3$ and $H^1(M) = 0$, the only nontree graphs of degree 0 have no decorations and every vertex is exactly trivalent. These graphs are also called simple cubic graphs.
- For $D \geq 4$ but $H^1(M) \neq 0$, there are nontree graphs of degree zero but they take on a very simple form: Besides trees, there are only graphs of genus 1 that are trivalent and decorated only by 1–forms. Such graphs are given by a "fundamental loop" such that every vertex has a decorated trivalent tree attached. Here is an example:



From now on, let us suppose M to be simply connected and of dimension $D \geq 4$.

Proposition 56 *The dgla* $\mathrm{GC}_{\overline{M}}^{\geq 3,\mathrm{tree}}$ *is the quotient of* $\mathrm{GC}_{\overline{M}}^{\geq 3}$ *by the dg Lie ideal spanned by graphs with at least one loop.*

Proof First notice that the Lie bracket of two graphs $\Gamma, \Gamma' \in \mathrm{GC}_{\overline{M}}^{\geq 3}$ will be a sum of graphs with loop order given by the sum of the loop orders of Γ and Γ'. It follows that the subspace spanned by graphs with at least one loop is a Lie ideal.

The splitting part of the differential preserves the loop order and the part of the differential that connects decorations increases the loop order by one and the twisted piece of the differential does not reduce loops. It follows that the differential preserves the ideal. □

Proposition 57 *The dgla* $\mathrm{GC}_{M}^{\mathsf{Lie}}$ *is defined as the quotient of* $\mathrm{GC}_{\overline{M}}^{\geq 3,\mathrm{tree}}$ *by the ideal generated by trees with vertices* ≥ 4*–valent and the IHX* (*or Jacobi*) *relations that originate from the splitting differential of a* 4*–valent vertex.*

The quotient map $\mathrm{GC}_{\overline{M}}^{\geq 3,\mathrm{tree}} \to \mathrm{GC}_{M}^{\mathsf{Lie}}$ *is a quasi-isomorphism.*

Proof It is clear that the differential preserves the ideal.

To see that the quotient map is a quasi-isomorphism, consider first a filtration by $\deg - \#\,\mathrm{edges}$ such that on the associated graded the differential cannot increase the number of vertices by more than one. Then, take a second filtration by the number of decorations and notice that on the associated graded we obtain (the cyclic version of) the quasi-isomorphism $\Lambda^{-D-1}\mathsf{L}_\infty \to \Lambda^{-D-1}\mathsf{Lie}$. □

The dgla $\mathrm{GC}_{M}^{\mathsf{Lie}}$ is a cyclic variant of the Harrison complex of $H^\bullet(M)$. Indeed, let us consider more generally a graded vector space $A = \bar{A} \oplus \mathbb{R}$, with a degree $-D$ pairing. A C_∞–structure on A is given by a Maurer–Cartan element in $\mathrm{Hom}(\mathsf{Lie}^c\{1\}[-1] \circ A, A)$ which, via the pairing, can be identified with the space

$$\mathrm{Hom}\big(A^{\bullet - D} \otimes (\mathsf{Lie}^c\{1\}[-1] \circ A^\bullet), \mathbb{R}\big).$$

There is a map $A \otimes (\mathsf{Lie}^c\{1\}[-1] \circ A)[-D] \to {}^*\mathrm{GC}_{\bar{A}}^{\mathsf{Lie}}$ determined in the following way: A basis of the cooperad Lie^c can be identified with rooted planar trivalent trees modulo the Jacobi (co)relations. Forgetting about the position of the root and considering it as any other leaf, and replacing every leaf with a decoration by A, we obtain an element in ${}^*\mathrm{GC}_{\bar{A}}^{\mathsf{Lie}}$.

Definition 58 Let $A = \bar{A} \oplus \mathbb{R}$ be a graded vector space with a nondegenerate pairing of degree $-D$. A cyclic C_∞–algebra structure on A is a Maurer–Cartan element in $\mathrm{GC}_{\bar{A}}^{\mathsf{Lie}}$.

If such a cyclic C_∞–algebra structure z maps into a C_∞–structure μ via the dual of the map described before Definition 58, we say that z *extends* μ.

Remark 59 Due to the implicit usage of the degree $-D$ pairing, such structure would be more appropriately called a "D–cyclic C_∞–algebra".

Proposition 60 *An orientable closed manifold M determines a cyclic C_∞–algebra structure on its cohomology $H^\bullet(M)$ extending the one arising from the homotopy transfer theorem.*

Proof The C_∞–structure on $H^\bullet(M)$ is given by a map in

$$\mathrm{Hom}\big(\mathsf{Lie}^c\{1\}[-1]\circ H^\bullet(M), H^\bullet(M)\big)$$

which, by the Poincaré duality pairing, is equivalent to an element

$$f \in \mathrm{Hom}\big(H^\bullet(M)\otimes(\mathsf{Lie}^c\{1\}[-1]\circ H^\bullet(M)), \mathbb{R}\big).$$

We claim that there is a factorization of f by

$$\begin{tikzcd} H^\bullet(M)\otimes(\mathsf{Lie}^c\{1\}[-1]\circ H^\bullet(M)) \arrow[r, "f"] \arrow[d, "g"'] & \mathbb{R} \\ {}^*\mathrm{GC}_M^{\mathsf{Lie}} \arrow[ur, dashed, "Z"'] & \end{tikzcd}$$

and the dashed arrow corresponds to a Maurer–Cartan $Z\in\mathrm{GC}_M^{\mathsf{Lie}}$ which is gauge equivalent to the image of $Z_M^3\in\mathrm{GC}_M^{\geq 3,\mathrm{tree}}$.

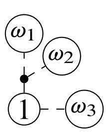
To show that f factors through g it is sufficient to show that for every $\mu\in\mathsf{Lie}^c\{1\}[-1](n)$ and $\omega_0,\dots,\omega_n\in H^\bullet(M)$,

$$f(\omega_0\otimes\mu\otimes\omega_1\otimes\cdots\otimes\omega_n) = f(\omega_n\otimes\mu\otimes\omega_0\otimes\cdots\otimes\omega_{n-1}),$$

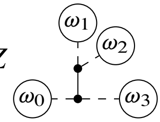
but this follows from the explicit formula the C_∞–action given by the homotopy transfer theorem. This corresponds to computing the partition function on the trivalent graph given by the C_∞ operation μ where the root is replaced by a decoration by the element ω_0, which is clearly cyclically invariant.

As an example, suppose that μ corresponds to $\mu_2\circ_1\mu_2\in\mathsf{Lie}^c(3)$. Then

$$\mu(\omega_1,\omega_2,\omega_3) = p(h(i(\omega_1)i(\omega_2))i(\omega_3))$$

$= p$  $= \sum_i e_i \int_{1,2} \pi_1^*(e_i^*)\pi_1^*(\omega_3)\phi_{1,2}\pi_2^*(\omega_1)\pi_2^*(\omega_2).$

Therefore,

$$f(\omega_0,\mu(\omega_1,\omega_2,\omega_3)) = \int_{1,2} \pi_1^*(\omega_0)\pi_1^*(\omega_3)\phi_{1,2}\pi_2^*(\omega_1)\pi_2^*(\omega_2) = Z$$


□

Remark 61 For simply connected ≥ 4–dimensional M, the cyclic C_∞–structure on $H^\bullet(M)$ determines the spaces $^*\mathsf{graphs}_M^{\varnothing}(n)$, which encode the real homotopy type of $\mathsf{FM}_M(n)$. Moreover, if M is parallelized, the cyclic C_∞–structure determines the Hopf comodule structure of $^*\mathsf{graphs}_M$, which encodes the real homotopy type of FM_M seen as a right FM_D–module.

Finally, one can check that the isomorphism type of the (noncyclic) C_∞–algebra structure on $H(M)$ already determines the cyclic C_∞–algebra structure. In other words, the cyclicity is not to be seen as extra data on, but rather a property of the real homotopy type, reflecting Poincaré duality. More concretely, the following result has been shown in [23, Theorems 5.5 and 5.8]. We also sketch a short proof here for completeness.

Proposition 62 *The real homotopy type of a closed orientable manifold determines its cyclic homotopy type. More precisely, given two cyclic C_∞–algebra structures on $H(M)$ that are C_∞ isomorphic as noncyclic C_∞–structures, they are also isomorphic as cyclic C_∞–structures.*

Proof sketch We are given two cyclic C_∞–structures on $H(M)$ and a C_∞ isomorphism between them. We may assume that the linear part of the C_∞ isomorphism is the identity, otherwise we just pull back one cyclic C_∞–structure along this linear part. Note also that the implicit underlying nondegenerate pairing on $H(M)$ is determined by the product up to an unimportant scale factor, so we may assume it is the same for both our cyclic C_∞–structures.

We denote by μ_1 and μ_2 the two Maurer–Cartan elements in $\mathrm{GC}^{\mathsf{Lie}}_{H(M)}$ encoding our cyclic C_∞–structures. The underlying (noncyclic) C_∞–structure is encoded by the images of μ_1 and μ_2 under the natural inclusion of dg Lie algebras into the reduced Harrison complex

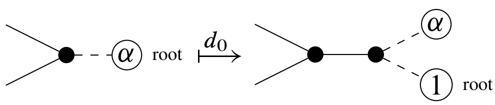
$$\mathrm{root}\colon \mathrm{GC}^{\mathsf{Lie}}_{H(M)} \to \mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M)).$$

Graphically, elements on the left-hand side can be interpreted as linear combinations of nonrooted Lie trees, and elements of the right-hand sides can be seen as rooted

Lie trees as above, and the map root is defined by summing over all possible ways of making one leaf into the root.

The C_∞ morphism between our two C_∞–structures (with linear term being the identity) then yields a gauge equivalence between the MC elements $\mathrm{root}(\mu_1)$ and $\mathrm{root}(\mu_2)$ in $\mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M))$. We desire to check that this implies that μ_1 and μ_2 are already gauge equivalent in $\mathrm{GC}^{\mathsf{Lie}}_{H(M)}$. To this end we can employ the Goldman–Millson theorem [16]. To check the conditions of this theorem we consider a filtration such that $\mathcal{F}^p\,\mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M))$ is spanned by rooted Lie trees with $\geq p$ leaves that are decorated by classes of nonzero degree.

On the associated graded the only piece of the differential that survives replaces the root (say decorated by some $\alpha \in H_k(M)$) by two leaves, with one decorated α and the new root decorated with $1 \in H_0(M)$:



It is an easy exercise to check that the cohomology of the p^{th} graded piece of the Harrison complex is identified for $p \geq 3$ precisely with nonrooted trees all of whose leaves are decorated by elements of $\overline{H}_\bullet(M)$. But this is precisely the image of $\mathrm{GC}^{\mathsf{Lie}}_{H(M)}$ under the map root.

Hence the Goldman–Millson theorem is applicable to the inclusion of dg Lie algebras $\mathrm{root}\colon \mathrm{GC}^{\mathsf{Lie}}_{H(M)} \to \mathcal{F}^2\,\mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M))$. To conclude the desired result we then just need to remark that our gauge equivalence between $\mathrm{root}(\mu_1)$ and $\mathrm{root}(\mu_2)$ in $\mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M))$ may actually be taken in $\mathcal{F}^2\,\mathrm{Harr}(\overline{H^\bullet(M)}, H^\bullet(M))$. To see this in turn one also computes the p^{th} graded piece of the Harrison complex for $p = 2$, and sees that there is no cohomology in the at least quadratic part. But since the underlying C_∞ morphism has trivial linear part, we may always remove the parts in the 2–graded piece by adding an exact terms, to yield the required gauge equivalence in $\mathcal{F}^2$. □

The real homotopy type of a manifold determines its cyclic homotopy type by the previous proposition. This in turn determines the (gauge equivalence class of) the Maurer–Cartan element z_M by Propositions 46 and 57, and Lemma 54 which itself determines the quasi-isomorphism type of the graph complex by the discussion in Section 7.2. We obtain thus the following theorem as a corollary:

Theorem 63 *Let M be an orientable compact manifold without boundary of dimension $D \geq 4$, such that $H^1(M,\mathbb{R}) = 0$. Then the real homotopy type of FM_M depends only on the real homotopy type of M. By this statement we mean that there is a zigzag of quasi-isomorphisms of symmetric sequences of dgcas over $\mathbb{R}$*

$$\Omega_{\mathrm{PA}}(\mathsf{FM}_M) \to \cdot \leftarrow X$$

with X being a sequence of dgcas defined using only knowledge of the quasi-isomorphism class of $\Omega_{\mathrm{PA}}(M)$ as a real dgca.

Remark 64 We generally work with unbounded cochain complexes, and a priori in the zigzag as constructed above there will occur dgcas which have unbounded degrees. However, the concrete X we use is (see above) $X = {}^*\mathsf{Graphs}_M^{\geq 3}$, which is concentrated in nonnegative degrees. Furthermore, X is cofibrant in the category of sequences of (unbounded) dgcas, and by homotopy lifting of the zigzag we may in fact construct a quasi-isomorphism of dgcas $X \to \Omega(\mathsf{FM}_M)$. For the statement above it is hence inessential whether we work over nonnegatively graded cochain complexes or cochain complexes of unbounded degrees.

Moreover, if we suppose M to be parallelized, the action of the Lie algebra GC_M on Graphs_M is compatible with the right Graphs_D–module structure. In this case, the (real homotopy type) of Graphs_M as a right Graphs_D–module is determined by (the gauge equivalence class of) the Maurer–Cartan element z_M. In that case, by the same argument we obtain a stronger version of the previous theorem.

Theorem 65 *Let M be a parallelizable compact manifold without boundary of dimension $D \geq 4$, such that $H^1(M,\mathbb{R}) = 0$. Then the real homotopy type of the operadic right module $\mathsf{FM}_M \circlearrowleft \mathsf{FM}_D$ depends only on the real homotopy type of M, in the sense that there is a zigzag of quasi-isomorphisms of right dg Hopf comodules connecting $\Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ and some X, with X depending only on the quasi-isomorphism type of the dgca $\Omega_{\mathrm{PA}}(M)$.*

We note again that we abuse slightly the notation since $\Omega_{\mathrm{PA}}(\mathsf{FM}_D)$ is not (strictly speaking) a dg Hopf cooperad and $\Omega_{\mathrm{PA}}(\mathsf{FM}_M)$ is not a right comodule; see Remark 15. The cleaner variant of stating the above theorem is to work in a category of homotopy cooperads and homotopy comodules, whose construction we however leave to future work; see [33, Section 3].

9 The framed case in dimension $D = 2$

In Section 3 we considered parallelized manifolds since a trivialization of the tangent bundle is needed to define the right operadic FM_D–module structure. Informally, to define the action one needs to know in which direction to insert, and the parallelization provides us the direction of the insertion.

In this section we wish to focus on the 2–dimensional case where unfortunately the only parallelizable (connected closed) manifold is the torus.

To go around the problem of not having a consistent choice of direction of insertion, instead of working with configuration spaces of points, we consider the framed configuration spaces. In other words, at every point of the configuration there is the additional datum of a direction, ie an element of the Lie group $\mathrm{SO}(2) = S^1$.

In this section Σ shall denote a connected oriented closed surface with a smooth and semialgebraic manifold structure. Most results will be an adaptation of the arguments in the previous sections to the framed case.

9.1 Definitions

In this section we introduce the compactification of the configuration space of framed points on Σ. A more detailed introduction to the subject can be found in [39].

9.1.1 The operad of configurations of framed points The construction of the operad of the framed version of FM_2 is a special case of the notion of the semidirect product of an operad and a group, as described below.

Definition 66 Let $\mathcal{P}$ be a topological operad such that there is an action of a topological group G on every space $\mathcal{P}(n)$ and the operadic compositions are G–equivariant. The semidirect product $\mathcal{P} \rtimes G$ is a topological operad with n–spaces

$$(\mathcal{P} \rtimes G)(n) = G^n \times \mathcal{P}(n),$$

and composition given by

$$(\bar{g}, p) \circ_i (\bar{g}', p') = (g_1, \dots, g_{i-1}, g_i g_1', \dots, g_i g_m', g_{i+1}, \dots, g_n, p \circ_i (g_i \cdot p')),$$

where $\bar{g} = (g_1, \dots, g_n)$ and $\bar{g}' = (g_1', \dots, g_m')$.

The group $\mathrm{SO}(2)$ has a well-defined action on FM_2 given by rotation.

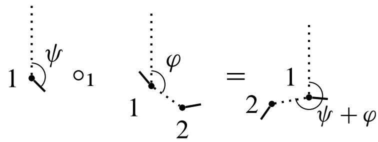


Figure 4: Operadic composition in FFM_2.

Definition 67 The framed Fulton–MacPherson topological operad FFM_2 is the semidirect product $\mathsf{FM}_2 \rtimes \mathrm{SO}(2)$.

When the operadic composition is performed, the configuration inserted rotates according to the frame on the point of insertion, as depicted in Figure 4, where at every point we draw a small line indicating the associated element of $\mathrm{SO}(2)$.

9.1.2 Configurations of framed points on a surface

Definition 68 The Fulton–MacPherson compactification of the configuration spaces of points on the surface Σ, FFM_Σ, is a symmetric sequence in semialgebraic smooth manifolds which is given as the pullback of the diagram

$$\begin{tikzcd} & \mathrm{SO}(\Sigma)^{\times n} \arrow[d, "\pi^n"] \\ \mathsf{FM}_\Sigma(n) \arrow[r] & \Sigma^{\times n} \end{tikzcd}$$

where $\pi : \mathrm{SO}(\Sigma) \to \Sigma$ is the frame bundle over Σ (assuming some Riemannian metric).

As in the nonframed case, the space $\mathsf{FFM}_\Sigma(n)$ is a manifold with corners. The interior of this manifold is the framed configuration space of points and is denoted by $\mathsf{F}\,\mathrm{Conf}_n(\Sigma)$.

Proposition 69 *The insertion of points at the boundary of* FFM_Σ *according to the direction of the frame defines a right* FFM_2 *operadic module structure on* FFM_Σ.

The associativity of the operadic module structure is clear.

9.2 Graphs

In this subsection we work with the operadic module $\mathsf{BVGraphs}_\Sigma \circlearrowleft \mathsf{BVGraphs}_2$ which is the version of $\mathsf{Graphs}_\Sigma \circlearrowleft \mathsf{Graphs}_2$ adapted to the framed case.

Informally, the difference between Graphs_Σ (resp. Graphs_2) and $\mathsf{BVGraphs}_\Sigma$ (resp. $\mathsf{BVGraphs}_2$) is that we now allow tadpoles (edges connecting a vertex to itself) at external vertices but graphs with tadpoles at internal vertices are considered to be 0.

This can be done by considering the subalgebra ${}^*\mathsf{BVGraphs}_\Sigma \subset^* \mathsf{Graphs}_\Sigma$ of graphs with no tadpoles on internal vertices or dually defining $\mathsf{BVGraphs}_\Sigma$ as a quotient of Graphs_Σ. A precise definition of $\mathsf{BVGraphs}_2$ can be found in [11].

The nontwisted analog of ${}^*\mathsf{BVGraphs}(n)$ is ${}^*\mathsf{BVGra}(n)$, the symmetric algebra on symbols $s^{ij} = s^{ji}$ for $1 \le i, j \le n$. One can also consider the nontwisted analog ${}^*\mathsf{BVGra}_\Sigma$, but notice that this is just the same space as ${}^*\mathsf{Gra}_\Sigma$ as tadpoles are not forbidden in ${}^*\mathsf{Gra}_\Sigma$ and there are no internal vertices upon which we can impose any condition.

Let $\phi \in \Omega^1_{\mathrm{triv}}(\mathsf{FFM}_\Sigma(1))$ be a global angular form of the S^1–bundle

$$\pi : \mathsf{FFM}_\Sigma(1) = \mathrm{SO}(\Sigma) \to \Sigma.$$

Such a form satisfies $d\phi = \pi^*(e)$, where $e \in \Omega^2_{\mathrm{triv}}(\Sigma)$ is the Euler class of the circle bundle.

Let $1 \le i \le n$. We denote by $\phi_{ii} \in \Omega^1_{\mathrm{triv}}(\mathsf{FFM}_\Sigma(n))$ the form $\pi_i^*(\phi)$, where

$$\pi_i : \mathsf{FFM}_\Sigma(n) \to \mathsf{FFM}_\Sigma(1)$$

is the map that remembers only the point labeled by 1.

We define a map ${}^*\mathsf{BVGra}_\Sigma(n) \to \Omega_{\mathrm{triv}}(\mathsf{FFM}_\Sigma(n))$ as a morphism of algebras sending s^{ij} to ϕ_{ij}, where if $i \neq j$, ϕ_{ij} is the form constructed in Section 2 and sends $[\omega]^j \in^* \mathsf{BVGra}_\Sigma$ to $p_j^*(\iota([\omega]))$, where $p_j : \mathsf{FFM}(n) \to M$ is the map that remembers only the point labeled by j.

Similarly one defines a map ${}^*\mathsf{BVGra}_2(n) \to \Omega_{\mathrm{triv}}(\mathsf{FFM}_2(n)) = \Omega_{\mathrm{triv}}(\mathsf{FM}_2(n) \times \mathrm{SO}(2)^{\times n})$ as a morphism of algebras sending a tadpole at the vertex i to the volume form of the i^{th} SO(2).

Lemma 70 *This defines a morphism of cooperadic comodules*

$${}^*\mathsf{BVGra}_\Sigma \circlearrowleft {}^*\mathsf{BVGra}_2 \to \Omega_{\mathrm{triv}}(\mathsf{FFM}_\Sigma) \circlearrowleft \Omega_{\mathrm{triv}}(\mathsf{FFM}_2).$$

Proof Regarding the compatibility with the differentials, the only case not covered in Lemma 16 is ϕ_{ii}, but this follows from the fact that the Euler form can be expressed as $\sum_{i,j} g^{ij} e_i \wedge e_j$.

For the compatibility with the cooperadic comodule structure it remains to check it for the elements $s^{ii} \in \mathsf{BVGraphs}_\Sigma(n)$. For simplicity of notation, we consider the element $s^{11} \in \mathsf{BVGraphs}_\Sigma(1)$ which is sent to $\phi_{11} \in \Omega^1_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(1))$ whose coaction gives $\phi_{11} \otimes 1 + 1 \otimes \mathrm{vol}_{S^1} \in \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(1)) \otimes \Omega_{\mathrm{PA}}(\mathsf{FFM}_2(1))$.

On the other hand, the coaction on $s^{11} \in \mathsf{BVGraphs}_\Sigma(1)$ gives us

$$s^{11} \otimes 1 + 1 \otimes s^{11} \in \mathsf{BVGraphs}_\Sigma(1) \otimes \mathsf{BVGraphs}_2(1),$$

from which the compatibility follows. □

Similarly to what was done in Section 4, one can prove the following proposition:

Proposition 71 *There is a morphism of cooperadic modules*

$$^*\mathsf{BVGraphs}_\Sigma \circlearrowleft {}^*\mathsf{BVGraphs}_2 \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma) \circlearrowleft \Omega_{\mathrm{PA}}(\mathsf{FFM}_2)$$

extending the morphism from Lemma 70.

The only difference relatively to the nonframed case is that the map

$$^*\mathsf{BVGraphs}_\Sigma(n) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(n))$$

evaluated at a graph $\Gamma \in \mathsf{BVGraphs}_\Sigma$ with k internal vertices is given by an integral over the fiber of $\mathsf{FFM}_\Sigma(n,k) \to \mathsf{FFM}_\Sigma(n)$, where the space $\mathsf{FFM}_\Sigma(n,k)$ is the (compactification of the) configuration space of n framed points and k unframed points corresponding respectively to the external vertices and the internal vertices of Γ.

A similar procedure is done for the map $^*\mathsf{BVGraphs}_2(n) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_2(n))$.

The goal of this section is to prove the following theorem.

Theorem 72 *The map* $^*\mathsf{BVGraphs}_\Sigma \circlearrowleft {}^*\mathsf{BVGraphs}_2 \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma) \circlearrowleft \Omega_{\mathrm{PA}}(\mathsf{FFM}_2)$ *is a quasi-isomorphism of Hopf cooperadic comodules.*

Proposition 73 *The map* $^*\mathsf{BVGraphs}_2 \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_2)$ *is a quasi-isomorphism.*

Proof On the one hand,

$$\begin{aligned} H^\bullet(\mathsf{FFM}_2(n)) = H^\bullet(\mathsf{FM}_2(n) \times \mathrm{SO}(2)^{\times n}) &= H^\bullet(\mathsf{FM}_2(n)) \otimes H^\bullet(\mathrm{SO}(2))^{\otimes n} \\ &= H^\bullet(\mathsf{FM}_2(n)) \otimes (\mathbb{R} \oplus \mathbb{R}[-1])^{\otimes n} \end{aligned}$$

by the Künneth formula. On the other hand, notice that as dg symmetric sequences $\mathsf{BVGraphs}_2 = \mathsf{Graphs}_2 \circ (\mathbb{R}[-1] \oplus \mathbb{R})$; therefore,

$$\begin{aligned} H(^*\mathsf{BVGraphs}_2(n)) &= H(^*\mathsf{Graphs}_2(n) \otimes (\mathbb{R} \oplus \mathbb{R}[-1])^{\otimes n}) \\ &= H(^*\mathsf{Graphs}_2(n)) \otimes (\mathbb{R} \oplus \mathbb{R}[-1])^{\otimes n}. \end{aligned}$$

Since a tadpole at the vertex labeled by i is sent to the volume form of the i^{th} $\mathrm{SO}(2)$, which is the generator of $H^1(\mathrm{SO}(2))$, we have that at the cohomology level the map

$$H({}^*\mathsf{BVGraphs}_2) = H({}^*\mathsf{Graphs}_2(n)) \otimes (\mathbb{R} \oplus \mathbb{R}[-1])^{\otimes n} \\ \to H^\bullet(\mathsf{FFM}_2(n)) = H^\bullet(\mathsf{FM}_2(n)) \otimes (\mathbb{R} \oplus \mathbb{R}[-1])^{\otimes n}$$

is just the map $f_* \otimes \mathrm{id}$, where $f\colon {}^*\mathsf{Graphs}_2 \to \Omega_{\mathrm{PA}}(\mathsf{FM}_2)$ is the quasi-isomorphism from Proposition 18, whence the result follows. □

9.3 Proof of Theorem 72

Let $n, k \geq 0$ and let us consider an auxiliary differential graded vector space $G(n,k)$ that is the subcomplex of ${}^*\mathsf{BVGraphs}_\Sigma(n+k)$ in which the points labeled $n+1, \dots, n+k$ cannot have tadpoles. This should be seen as the algebraic analog of the space $\mathsf{FFM}_\Sigma(n,k)$, the compactification of the configuration space of n framed points and k unframed points in Σ.

The map ${}^*\mathsf{BVGraphs}_\Sigma(n+k) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(n+k))$ restricts naturally to a map $G(n,k) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(n,k))$. We will show that this map is a quasi-isomorphism; thus proving Theorem 72 which corresponds to the cases with $k=0$. The proof will be done by induction on n. The case $n=0$ was already proven in Theorem 42.

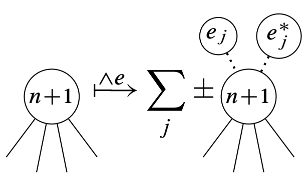
9.3.1 A long exact sequence of graphs Let us prove the following auxiliary result.

Proposition 74 *There is a long exact sequence of graded vector spaces*

$$\cdots \to H^d(G(n+1,k-1)) \xrightarrow{f} H^{d-1}(G(n,k)) \\ \xrightarrow{\wedge e} H^{d+1}(G(n,k)) \xrightarrow{i_*} H^{d+1}(G(n+1,k-1)) \to \cdots,$$

where the map i_* is induced by the inclusion of $G(n,k)$ in $G(n+1,k-1)$.

Proof Let us clarify the undescribed maps. The map f removes a tadpole on the vertex labeled by $n+1$ if there exists one, otherwise it sends a graph to zero. The map $\wedge e$ decorates the vertex $n+1$ with the "Euler form":



It is not clear that these maps are well defined at the cohomology level, but this will become clear by the construction of the sequence.

Let us consider the decomposition of $G(n+1,k-1)$

$$G(n+1,k-1) = \underset{\overset{d_0}{\circlearrowright}}{G(n,k)[-1]} \xrightarrow{d_1} \oplus \underset{\overset{d_0}{\circlearrowright}}{G(n,k)}$$

where the first summand corresponds to graphs in which the vertex labeled by $n+1$ has a tadpole and the second summand corresponds to graphs in which the vertex labeled by $n+1$ does not have a tadpole. The differential splits into two terms, d_0 and d_1, as in the picture. Let us consider a two-level filtration on the number of tadpoles at the vertex $n+1$. On the zeroth page of the spectral sequence the differential is d_0, which acts as the ordinary differential of $G(n,k)$.

The differential on the second page is induced by d_1 and is the map that was denoted by $\wedge e$,

$$\wedge e\colon H^\bullet(G(n,k)[-1]) = H^{\bullet-1}(G(n,k)) \to H^{\bullet+1}(G(n,k)).$$

The spectral sequence converges at the second page since we considered a two-level filtration; therefore

$$H^\bullet(G(n+1,k-1)) = \ker(\wedge e) \oplus \operatorname{coker}(\wedge e).$$

The map f is defined to be the composition

$$H^\bullet(G(n+1,k-1)) \twoheadrightarrow \ker(\wedge e) \hookrightarrow H^{\bullet-1}(G(n,k)).$$

It is then clear that $\operatorname{Im}(f) = \ker(\wedge e)$, which gives us exactness at $H^{d-1}(G(n,k))$.

The map i_* is given by the composition

$$H^\bullet(G(n,k)) \twoheadrightarrow \operatorname{coker}(\wedge e) \hookrightarrow H^{\bullet-1}(G(n+1,k-1)).$$

Therefore its image coincides with the kernel of f, which shows exactness at

$$H^{d+1}(G(n+1,k-1)).$$

Since i_* is the projection to the cokernel of $\wedge e$, its kernel is precisely the image of $\wedge e$, which shows the remaining exactness. □

9.3.2 The Gysin sequence The map $\pi\colon \mathsf{FFM}_\Sigma(n+1,k-1) \to \mathsf{FFM}_\Sigma(n,k)$ that forgets the frame at the point $n+1$ is a circle bundle. We denote by $e \in \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(n,k))$ the Euler form of the circle bundle. The Gysin sequence of this circle bundle is the long exact sequence

$$
\begin{aligned}
(14)\quad H^d(\mathsf{FFM}_\Sigma(n+1,k-1)) &\xrightarrow{\int_\pi} H^{d-1}(\mathsf{FFM}_\Sigma(n,k)) \\
&\xrightarrow{\wedge e} H^{d+1}(\mathsf{FFM}_\Sigma(n,k)) \xrightarrow{\pi^*} H^{d+1}(\mathsf{FFM}_\Sigma(n+1,k-1)) \to \cdots .
\end{aligned}
$$

Using the maps $G(a,b) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(a,b))$, we obtain the morphism of exact sequences

$$
\begin{tikzcd}
H^d(\mathsf{FFM}_\Sigma(n+1,k-1)) \arrow[dr, "\int_\pi"] & & & \\
H^d(G(n+1,k-1)) \arrow[u] \arrow[dr, "f"] & H^{d-1}(\mathsf{FFM}_\Sigma(n,k)) \arrow[dr, "\wedge e"] & & \\
& H^{d-1}(G(n,k)) \arrow[u] \arrow[dr, "\wedge e"] & H^{d+1}(\mathsf{FFM}_\Sigma(n,k)) \arrow[dr, "\pi^*"] & \\
& & H^{d+1}(G(n,k)) \arrow[u] \arrow[dr, "i_*"] & H^{d+1}(\mathsf{FFM}_\Sigma(n+1,k-1)) \\
& & & H^{d+1}(G(n+1,k-1)) \arrow[u]
\end{tikzcd}
$$

Since by induction $G(n,k) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(n,k))$ is a quasi-isomorphism, the five lemma implies that $G(n+1,k-1) \to \Omega_{\mathrm{PA}}(\mathsf{FFM}_\Sigma(n+1,k-1))$ is a quasi-isomorphism as well; thus concluding the proof of Theorem 72.

Appendix A Comparison to the Lambrechts–Stanley model through cyclic C_∞–algebras

In this appendix we show how to obtain from the ${}^*\mathsf{Graphs}_M$ model a proof that the Lambrechts–Stanley algebra is a dgca model for the FM_M (Conjecture 76).

Definition 75 [31] A Poincaré duality algebra of dimension D is a nonnegatively graded connected dgca A together with a linear map

$$\epsilon\colon A^D \to \mathbb{R}$$

such that $\epsilon \circ d = 0$ and such that the bilinear maps

$$A \otimes A \to \mathbb{R}[-D], \quad a \otimes b \mapsto \epsilon(a,b)$$

are nondegenerate.

Note that by the connectivity assumption necessarily $A^D = \mathbb{R}$ and hence ϵ is unique up to scale, if it exists. Note that a Poincaré duality algebra is a particular case of a cyclic C_∞–algebra.

A Poincaré duality model for a manifold M is a Poincaré duality algebra weakly equivalent (as a dgca) to $\Omega(M)$. It is shown in [31] that such a Poincaré duality model always exists for simply connected compact orientable manifolds.

Lambrechts and Stanley furthermore define the following family of dgcas from a Poincaré duality algebra A, generalizing earlier work by Kriz [30] and Totaro [44]. Consider the algebra

$$A^{\otimes n}[\omega_{ij}; 1 \le i \ne j \le n].$$

For $a \in A$ let $p_j^*(a)$ be the element $1 \otimes \cdots \otimes a \otimes \cdots \otimes 1$, with a in the j^{th} slot. Then one imposes on the above algebra the relations

(1) $\omega_{ij} = (-1)^D \omega_{ji}$,

(2) $\omega_{ij}^2 = 0$,

(3) $\omega_{ij}\omega_{ik} + \omega_{jk}\omega_{ji} + \omega_{ki}\omega_{kj} = 0$ for distinct i, j and k,

(4) $(p_i^*(a) - p_j^*(a))\omega_{ij} = 0$.

Let us define for A a Poincaré duality algebra as above the diagonal $\Delta \in A \otimes A$ to be the inverse of the nondegenerate bilinear pairing. Let us further denote by Δ_{ij} the corresponding element in $A^{\otimes n}$, the two "nontrivial" factors of A situated in positions i and j. Then one defines

$$(A^{\otimes n}[\omega_{ij}; 1 \le i \ne j \le n]/\sim, d_A + \nabla),$$

where the differential d_A is that induced by the differential on A and ∇ is defined as

$$\nabla \omega_{ij} = \Delta_{ij}.$$

One readily checks that the ideal generated by the relation is closed under this differential. Furthermore, if the Euler class of A, ie the image Δ under the multiplication, vanishes, then the $F(A,-)$ naturally assemble into a right Pois_D^* cooperadic comodule.

Lambrechts and Stanley [32] show that for A a Poincaré duality model for M, we have that $H(F(A,n)) = H(\mathsf{FM}_M(n))$, and furthermore raise the following conjecture.

Conjecture 76 [32] *If A is a Poincaré duality model for the simply connected compact orientable manifold M then $F(A,n)$ is a dgca model for* $\mathrm{Conf}(M,n)$.

A proof of (a slightly weaker form of) this statement is given in [25], using methods similar to ours. While in this paper we work with cyclic C_∞–structures on $H(M)$, rather than Poincaré duality models to capture the real homotopy type "with Poincaré duality" for M, one can still deduce the conjecture of Lambrechts and Stanley from our methods, at least in the case that the dimension of M is at least 4. (The case $M = S^2$ also follows from the computation in Appendix B, leaving only the case $M = S^3$.) Let us sketch this reduction.

First let V be a finite-dimensional differential nonnegatively graded vector space with the subspace of degree 0 elements $V_0 = \mathbb{R}$ and a nondegenerate symmetric bilinear pairing of degree D

$$V \otimes V \to \mathbb{R}[-D].$$

We denote by $\Delta \in V \otimes V$ the corresponding dual degree D element (the "diagonal") as above. Then we may define a graph complex (and dg Lie algebra) GC_V akin to GC_M above, just replacing each occurrence of $H^*(M)$ by V and with an additional piece of the differential coming from d_V. Concretely, this means that vertices in graphs of GC_V may be decorated by copies of $\overline{V}^*$. Furthermore, suppose a cyclic C_∞–structure is given on V, for the above bilinear form. We may see this structure as a Maurer–Cartan element $Z \in \mathrm{GC}_V$, all of whose coefficients in front of nontree graphs vanish. We may furthermore use it to define a Graph complex ${}^*\mathsf{Graphs}_V$ analogously to ${}^*\mathsf{Graphs}_M$ above, replacing each occurrence of $H(M)$ by V, and using the given Z in place of the partition function.

Next, fix representatives of the cohomology of V by providing a map

$$H(V) \hookrightarrow V. \tag{15}$$

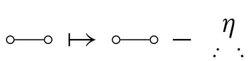
The pairing on V induces a pairing on $H(V)$, independent of the representatives chosen. We denote the corresponding diagonal by $\Delta_H \in H(V) \otimes H(V)$. Via the chosen embedding we may as well consider Δ_H as an element in $V \otimes V$, in which case it becomes cohomologous to Δ. We may hence choose $\eta \in A \otimes A$ (of the same symmetry under exchange of the two A's as Δ) such that

$$\Delta_H = \Delta - d_V \eta. \tag{16}$$

We may then define a natural map of dg cooperadic comodules

$${}^*\mathsf{Gra}_{H(V)} \to {}^*\mathsf{Gra}_V \tag{17}$$

by sending the decorations in $H(V)$ to V using our map (15), and by sending an edge between vertices i and j to the same edge, minus the element η, considered as decoration at vertices i and j. In pictures:



Equation (16) implies that the map (17) is indeed compatible with the differentials.

Following the construction of GC_V, (17) induces an L_∞–morphism of dg Lie algebras

$$\mathrm{GC}_V \to \mathrm{GC}_{H(V)},$$

and we can hence transfer the Maurer–Cartan element $Z \in \mathrm{GC}_V$ inducing the cyclic C_∞–structure on V to a Maurer–Cartan element $Z_H \in \mathrm{GC}_H$. (The MC element Z_H is still supported on trees, and encodes the cyclic C_∞–structure on $H(V)$ induced by homotopy transfer.) Furthermore, we obtain from (17) a map

$${}^*\mathsf{Graphs}_{H(V)} \to {}^*\mathsf{Graphs}_V,$$

that one can check to be a quasi-isomorphism by an easy spectral sequence argument.

In particular, let us take for V a Poincaré duality model for the simply connected manifold M. Then if the dimension D of M is at least 4, the Maurer–Cartan element Z_H is gauge equivalent to the partition function Z_M constructed above. This is because by degree reasons there cannot be loop order ≥ 1 contributions to this partition function, and the tree part of Z_M encodes the real homotopy type of M (in the form of a cyclic C_∞–structure on $H(V) = H(M)$), and hence must be gauge equivalent to Z_H, which also encodes the real homotopy type by construction. Hence we can conclude that ${}^*\mathsf{Graphs}_V$ is quasi-isomorphic to ${}^*\mathsf{Graphs}_M$ and is hence a dgca model for FM_M, with the partition function concentrated on trees with one vertex. Furthermore, in this case we have a direct map

$$(18) \qquad {}^*\mathsf{Graphs}_V \to F(V, -)$$

to the Lambrechts–Stanley algebra, by sending all graphs with internal vertices to zero, and imposing the defining relations. Again, by a spectral sequence argument, the map (18) can be seen to be a quasi-isomorphism. Furthermore, it is evidently compatible with the right Pois_D^* cooperadic comodule structures, in the case the Euler class vanishes. This shows that $F(V, -)$ is quasi-isomorphic to ${}^*\mathsf{Graphs}_M$, ie to a dgca model for FM_n. Hence Conjecture 76 follows, in dimension $D \geq 4$.

Appendix B Example computation: the partition function of the 2–sphere

As an illustration, let us show that the partition function of the two-sphere is essentially trivial. We cover S^2 by two coordinate charts $\mathbb{C}$ via stereographic projection as usual.

The coordinate transformation relating the two charts is then

$$\Phi\colon \mathbb{C}\setminus\{0\}\to\mathbb{C},\quad z\mapsto \frac{1}{z}.$$

We take a basis $1\in H^0(S^2)$, $\omega\in H^2(S^2)$ of the cohomology, with $\int\omega=1$. Take as a representative for ω any compactly supported top form of volume 1, which we also denote by ω. In fact, to abuse the notation further, denote by $\omega\in\Omega^2(\mathbb{C})$ also the coordinate expression in one of our charts. To achieve somewhat nicer formulas later, let us also assume that this ω is supported away from the origin and that

$$\Phi^*\omega=\omega. \tag{19}$$

Let ϕ_0 be the propagator on $\mathbb{C}$, ie

$$\phi_0(z,w)=\frac{1}{2\pi}\Im d\log(z-w).$$

Note that

$$\phi_0\Big(\frac{1}{z},\frac{1}{w}\Big)=\frac{1}{2\pi}\Im d\log\Big(\frac{w-z}{wz}\Big)=\phi_0(z,w)-\phi_0(z,0)-\phi_0(w,0). \tag{20}$$

Then we will take as propagator of the sphere[8]

$$\phi(z,w)=\phi_0(z,w)-\int_u\phi_0(z,u)\omega(u)-\int_u\phi_0(w,u)\omega(u).$$

Let us first verify that this 2–form extends from our coordinate chart to $\mathsf{FM}_2(S^2)$. To this end, apply the coordinate transformation Φ and compute

$$\phi\Big(\frac{1}{z},\frac{1}{w}\Big)=\phi_0\Big(\frac{1}{z},\frac{1}{w}\Big)-\int_u\phi_0\Big(\frac{1}{z},u\Big)\omega(u)-\int_u\phi_0\Big(\frac{1}{w},u\Big)\omega(u).$$

Changing the integration variable from u to $1/u$, using (19) and applying (20) three times, we obtain

$$\begin{aligned}\phi\Big(\frac{1}{z},\frac{1}{w}\Big)&=\phi_0(z,w)-\phi_0(z,0)-\phi_0(w,0)-\int_u(\phi_0(z,u)-\phi_0(z,0)-\phi_0(u,0))\omega(u)\\&\qquad-\int_u(\phi_0(w,u)-\phi_0(w,0)-\phi_0(w,0))\omega(u)\\&=\phi(z,w)-\phi_0(z,0)-\phi_0(w,0)+\phi_0(z,0)\int_u\omega(u)+\phi_0(w,0)\int_u\omega(u)\\&=\phi(z,w).\end{aligned}$$

[8]In Proposition 8 the propagator has been denoted ϕ_{12}. Here we choose to drop the subscript 12 for brevity.

Hence the propagator has the same form in the other coordinate chart, and in particular it has no singularity at the coordinate origin, and hence readily extends to $\mathsf{FM}_2(S^2)$.

Furthermore one checks the following properties:

- Clearly $\phi(z,w) = \phi(w,z)$.
- By Stokes' theorem,
$$d\phi(z,w) = \omega(z) + \omega(w)$$
as required.
- By degree reasons,
$$\int_v \phi(z,v) = 0.$$
Furthermore,
$$\begin{aligned}\int_v \phi(z,v)\omega(v) &= \int_v \phi_0(z,v)\omega(v) - \int_v\int_u \phi_0(z,u)\omega(u)\omega(v) - \int_v\int_u \phi_0(v,u)\omega(u)\omega(v)\\ &= \int_v \phi_0(z,v)\omega(v) - \int_u \phi_0(z,u)\omega(u) - 0\\ &= 0.\end{aligned}$$
Here the third term on the right-hand side vanishes by degree reasons. (One integrates a 5–form over a 4–dimensional space.)
- We have
$$\begin{aligned}&\int_v \phi(z,v)\phi(u,w)\\ &\quad= \int_v \phi_0(z,v)\phi_0(v,w) - \int_v\int_{u_1} \phi_0(z,u_1)\omega(u_1)\phi_0(v,w) - \int_v\int_{u_2} \phi_0(v,w)\phi_0(w,u_2)\omega(u_2)\\ &\qquad\qquad - \int_v\int_{u_1} \phi_0(v,u_1)\omega(u_1)\phi_0(v,w) - \int_v\int_{u_2} \phi_0(v,w)\phi_0(v,u_2)\omega(u_2)\\ &\qquad + \int_v\int_{u_1}\int_{u_2} \phi_0(z,u_1)\omega(u_1)\phi_0(w,u_2)\omega(u_2) + \int_v\int_{u_1}\int_{u_2} \phi_0(v,u_1)\omega(u_1)\phi_0(w,u_2)\omega(u_2)\\ &\qquad + \int_v\int_{u_1}\int_{u_2} \phi_0(z,u_1)\omega(u_1)\phi_0(v,u_2)\omega(u_2) + \int_v\int_{u_1}\int_{u_2} \phi_0(v,u_1)\omega(u_1)\phi_0(v,u_2)\omega(u_2).\end{aligned}$$
The first term on the right-hand side vanishes by a standard vanishing lemma of Kontsevich. The fourth, fifth and last terms vanish by the same reason. The remaining terms vanish by degree reasons: their forms with v–dependence are of degree ≤ 1. Hence we conclude that the whole expression is zero, and graph weights computed using our propagator will be zero for graphs with bivalent vertices.

- Identify the pullback of $\partial\mathsf{FM}_2(S^2)$ to our coordinate chart with $\mathbb{C}\times S^1$, and fix the standard coordinate φ on the S^1 factor. Then restricting ϕ to the boundary $\partial\mathsf{FM}_2(S^2)$, (ie we take the limit $w\to z$ in our coordinate chart) we obtain the form

$$\frac{1}{2\pi}d\varphi+\eta(z),$$

where

$$\eta=-2\int_u \phi_0(z,u)\omega(u)$$

depends only on z, and not on φ, as desired.

B.1 Vanishing of integrals

Proposition 77 *Using the propagator ϕ and the top form ω as above, the partition function becomes*

$$z_{S^2}=\overset{\omega}{\bullet}. \tag{21}$$

In other words, the weights of all graphs with more than one vertex vanish.

Proof By the properties above, all graphs vanish if either some vertex has valence 2 or some vertex has more than one decoration by ω or some vertex has valence one, and there is one incident edge. The only connected graph with a vertex of valence one is the one appearing in (21). All other graphs with potentially nonvanishing weight must hence be of the following kind:

(1) There are ≥ 2 edges incident to any vertex, and at most one decoration ω.

(2) If there are exactly 2 edges incident on some vertex, it must come with a decoration ω.

From an admissible graph Γ, we can build another linear combination of admissible graphs Γ_0 by formally replacing each edge by the linear combination

$$\circ\!\!-\!\!\!-\!\!\circ \;\mapsto\; \circ\!\!-\!\!\!-\!\!\circ \;-\; \overset{\omega}{\circ\!\!-\!\!\bullet}\;\circ \;-\; \circ\;\overset{\omega}{\bullet\!\!-\!\!\circ}\,.$$

Clearly,

$$\int_{\mathsf{FM}_d(|V\Gamma|)}\omega_\Gamma=\int_{\mathsf{FM}_d(|V\Gamma_0|)}\omega^0_{\Gamma_0}$$

where now the weight form $\omega^0_\bullet$ is defined just like $\omega_\bullet$ above, but using the Euclidean propagator ϕ_0 instead of ϕ.

It hence suffices to show that for each admissible graph Γ with more than one vertex,

$$\int_{\mathsf{FM}_d(|V\Gamma|)} \omega_\Gamma^0 = 0.$$

We may assume that the vertices are numbered such that the vertices decorated by ω have indices $1,\dots,k$ for some $k \geq 0$. Then the above integral factorizes as

$$\int_{\mathsf{FM}_d(|V\Gamma|)} \omega_\Gamma^0 = \int_{\mathsf{FM}_d(k)} \omega(x_1)\omega(x_2)\cdots\omega(x_k) \underbrace{\int_{\mathsf{FM}_d(|V\Gamma|-k)} \omega_\Gamma^0}_{=:f(x_1,\dots,x_k)}.$$

Note that here $f(x_1,\dots,x_k)$ is a function associated to a graph with decorations ω. (There can be no form piece in $f(\dots)$, because the remainder of the integrand is already a top form.) Hence by the Kontsevich vanishing lemma [28, Lemma 6.4] $f(x_1,\dots,x_k) \equiv 0$. Hence the desired vanishing result follows. $\square$

Appendix C Pushforward of PA forms

Given an SA bundle $p\colon M \to N$ of rank l, the pushforward map of "integration along the fiber" defined in [24] is a map $p_*\colon \Omega^\bullet_{\mathrm{min}}(M) \to \Omega^{\bullet-l}_{\mathrm{PA}}(N)$. This map is only defined on minimal forms as the natural extension to the full algebra of PA forms is not well defined due to the failure of the relevant semialgebraic chain to be continuous; see the discussion on [24, Section 9].[9]

For our purposes we need to consider pushforwards of the propagator

$$\phi_{12} \in \Omega_{\mathrm{PA}}(\mathsf{FM}_M(2))$$

constructed in Proposition 8. Since we cannot construct the propagator in such a way that $\phi_{12} \in \Omega_{\mathrm{min}}(\mathsf{FM}_M(2))$, in this section we consider a different space of forms, Ω_{triv} such that $\Omega_{\mathrm{PA}} \supset \Omega_{\mathrm{triv}} \supset \Omega_{\mathrm{min}}$, to which the pushforward map can be extended and still satisfies Stokes' theorem.

Recall that for F a compact oriented semialgebraic manifold and M a semialgebraic manifold, the *constant continuous chain* $\widehat{F} \in C^{\mathrm{str}}(M \times F \to M)$ is defined by

$$\widehat{F}(x) = [\![\{x\} \times F]\!].$$

[9]We note that in the original sketch of the construction of PA forms by Kontsevich and Soibelman [29], the pushforward was (claimed to be) defined for all PA forms, for a slightly laxer definition of PA forms compared to [24].

Definition 78 Let M be a semialgebraic manifold. The space $\Omega_{\mathrm{triv}}(M)$ of *trivial forms* is the subvector space of $\Omega_{\mathrm{PA}}(M)$ spanned by forms of the type $\fint_{\widehat{F}} \mu$, where $\mu \in \Omega_{\min}(M \times F)$ and $\widehat{F}$ is a constant continuous chain.

Lemma 79 *The subspace $\Omega_{\mathrm{triv}}(M) \subset \Omega_{\mathrm{PA}}(M)$ is a dg commutative subalgebra.*

Proof $\Omega_{\mathrm{triv}}(M) \subset \Omega_{\mathrm{PA}}(M)$ is closed under the differential by the fiberwise Stokes' theorem [24, Proposition 8.12] and since the fiberwise boundary of a trivial bundle is again a trivial bundle. Furthermore, the subspace $\Omega_{\mathrm{triv}}(M)$ is closed under addition and the commutative product on $\Omega_{\mathrm{PA}}(M)$ because the union and product of trivial bundles is again trivial; see the construction of these operations in [24, Section 5]. □

Let us consider a strongly continuous chain $\Phi \in C_l^{\mathrm{str}}(E \xrightarrow{f} B)$ along a semialgebraic map $f \colon E \to B$. Let $E \times F$ be the trivial bundle over E with fiber F, a compact oriented semialgebraic k manifold.

Proposition 80 *Under the previous conditions, there is a strongly continuous chain*

$$\Phi \ltimes \widehat{F} \in C_{k+l}^{\mathrm{str}}(E \times F \xrightarrow{f \circ \mathrm{pr}_2} B)$$

defined by $(\Phi \ltimes \widehat{F})(b) := \Phi(b) \times F$.

Proof If we consider the family $\{(S_\alpha, F_\alpha, g_\alpha)_{\alpha \in I}\}$ that trivializes the continuous chain Φ, it is easy to see that $\{(S_\alpha, F_\alpha \times F, g_\alpha \times \mathrm{id}_F)_{\alpha \in I}\}$ trivializes $\Phi \ltimes \widehat{F}$ since, by hypothesis, the two squares

$$\begin{tikzcd}
\overline{S}_\alpha \times F_\alpha \times F \arrow[r, "g_\alpha \times \mathrm{id}_F"] \arrow[d] & E \times F \arrow[d, "\mathrm{pr}_2"] \\
\overline{S}_\alpha \times F_\alpha \arrow[r, "g_\alpha"] \arrow[d, "f"] & E \arrow[d] \\
\overline{S}_\alpha \arrow[r] & B
\end{tikzcd}$$

commute. □

Corollary 81 *Let $p \colon Y \to X$ be an oriented SA bundle and $\Phi \in C_l^{\mathrm{str}}(Y \to X)$ be the associated strongly continuous chain. Then there is a well-defined map*

$$p_* \colon \Omega_{\mathrm{triv}}^{\bullet}(Y) \to \Omega_{\mathrm{PA}}^{\bullet - l}(X)$$

extending the one on minimal forms, given by $p_*(\omega) = \fint_{\Phi \ltimes \widehat{F}} \omega$.

Remark 82 Recall that the proof of the fiberwise Stokes' theorem relies essentially on the fact that for $\gamma \in C_k(X)$ and $\Psi \in C_l^{\mathrm{str}}(Y \to X)$,

$$\partial(\gamma \ltimes \Psi) = \partial\gamma \ltimes \Psi + (-1)^{\deg \gamma} \gamma \ltimes \partial\Psi.$$

With the same proof as [24, Proposition 5.17] we see that this formula is still valid if we take Ψ and γ to be Φ and $\widehat{F}$ as above, and therefore Stokes' theorem is also valid for pushforwards of trivial forms.

We prove now the Poincaré lemma for the sheaf of complexes Ω_{triv}.

Proposition 83 *If U is a contractible semialgebraic set, then $H(\Omega_{\mathrm{triv}}(U))$ is one-dimensional and concentrated in degree zero.*

Proof Let $h\colon [0,1] \times U \to U$ be a contraction of U such that $h(1,x) = x$ and $h(0,x) = x_0$ for some fixed $x_0 \in U$. Suppose $\omega \in \Omega_{\mathrm{triv}}(U)$ is a closed form of degree at least 1. By Stokes' formula,

$$d \int_I h^*\omega = \int_I h^* d\omega \pm (\omega - \omega_{x_0}) = \pm\omega,$$

whence it follows that ω is exact. □

We can now conclude more generally that the cohomology of a semialgebraic manifold M agrees with the homology of $\Omega_{\mathrm{triv}}(M)$.

Corollary 84 *Let M be a compact semialgebraic manifold, possibly with corners. The inclusion $\Omega_{\mathrm{triv}}(M) \to \Omega_{\mathrm{PA}}(M)$ is a quasi-isomorphism of commutative algebras.*

Proof Every compact semialgebraic manifold admits a good cover: Indeed, every compact semialgebraic set has a finite semialgebraic triangulation [5, Theorem 9.2.1], and can hence be identified with a finite simplicial complex; see also the discussion in [24, Section 2]. Given a semialgebraic triangulation, one can construct a semialgebraic good cover $\{U_\alpha\}$ by taking the open stars of the vertices of a refinement of the triangulation.[10]

We also choose a subordinate semialgebraic partition of unity $\{\rho_\alpha\}$. For convenience we shall also pick cutoff functions σ_α with support in U_α such that $\sigma_\alpha(x) = 1$ on the

[10] The star of a vertex v is the union of the interiors of faces that contain v.

support of ρ_α. (We may slightly enlarge the U_α to this purpose or alter the partition of unity; see also the proof of [24, Proposition 6.7].)

This allows us to run the standard Čech–de Rham argument with respect to such a good cover to conclude by the Poincaré lemma that the homology of $\Omega_{\mathrm{triv}}(M)$ coincides with the (Čech) cohomology of M; see for instance [8, Example 14.16].

To be concrete, we consider the Čech–de Rham complex

$$C := \Big(\prod \Omega_{\mathrm{triv}}(U_{\alpha_0\dots\alpha_p})[-p], d+\delta\Big),$$

where

$$U_{\alpha_0\dots\alpha_p} = U_{\alpha_0} \cap \dots \cap U_{\alpha_p},$$

d is induced by the differential on the factors $\Omega_{\mathrm{triv}}(U_{\alpha_0\dots\alpha_p})$, and δ is the Čech part of the differential, defined on a cochain $\omega = (\omega_{\alpha_0\dots\alpha_p})$ with $\omega_{\alpha_0\dots\alpha_p} \in U_{\alpha_0\dots\alpha_p}$ by

$$(\delta\omega)_{\alpha_0\dots\alpha_p} = \sum_{i=0}^{p} (-1)^i \omega_{\alpha_0\dots\hat{\alpha}_i\dots\alpha_p}.$$

The Čech–de Rham complex C is a first quadrant double complex, and one compares the two convergent spectral sequences associated to this complex.

The first ("columnwise") spectral sequence has the complex (C, d) as its E^0–page. By the Poincaré lemma (Proposition 83), the E^1–page is then identified with the Čech complex associated to the constant sheaf $\mathbb{R}$. The E^2–page is hence the cohomology $H(M)$, and the spectral sequence abuts at this point by degree reasons.

The other ("rowwise") spectral sequence has first page (C,δ). We claim that the cohomology of this page is identified with $\Omega_{\mathrm{triv}}(M)$. This can in fact be shown identically to [8, Proposition 8.5]. Concretely, one may naturally extend (C,δ) to a complex

$$\widetilde{C} := (\Omega_{\mathrm{triv}}(M) \xrightarrow{\delta} C),$$

with the map

$$\delta \colon \Omega_{\mathrm{triv}}(M) \to \prod_\alpha \Omega_{\mathrm{triv}}(U_\alpha) \subset C$$

given by the natural restriction. One then checks that $(\widetilde{C},\delta)$ is acyclic by providing an explicit homotopy. Concretely, for a p–cocycle $\omega = (\omega_{\alpha_0\dots\alpha_p}) \in \widetilde{C}$, one defines the $(p-1)$–cochain τ such that

$$\tau_{\alpha_0\dots\alpha_{p-1}} = \sum_\alpha \rho_\alpha \omega_{\alpha\alpha_0\dots\alpha_{p-1}}.$$

Note that here we extend $\rho_\alpha \omega_{\alpha\alpha_0\dots\alpha_{p-1}} \in \Omega_{\mathrm{triv}}(U_{\alpha\alpha_0\dots\alpha_{p-1}})$ by zero to an element (abusively also denoted by) $\rho_\alpha \omega_{\alpha\alpha_0\dots\alpha_{p-1}}$ of $\Omega_{\mathrm{triv}}(U_{\alpha_0\dots\alpha_{p-1}})$. To be precise, this extension by zero may be defined as follows. Suppose

$$\omega_{\alpha\alpha_0\dots\alpha_{p-1}} = \int_Y \beta$$

is given by a fiber integral associated to the trivial bundle $Y \times U_{\alpha\alpha_0\dots\alpha_{p-1}} \to U_{\alpha\alpha_0\dots\alpha_{p-1}}$, with $\beta \in \Omega_{\min}(Y \times U_{\alpha\alpha_0\dots\alpha_{p-1}})$. Then we extend $\rho_\alpha\beta$ (by zero) to a minimal form on $U_{\alpha_0\dots\alpha_{p-1}}$, which we (abusively) also denote by $\rho_\alpha\beta$. For example, if $\beta = (f_0, \dots, f_k)$ in the notation of [24, Section 5.2], we may take $\rho_\alpha\beta := (\rho_\alpha f_0, \sigma_\alpha f_1, \dots, \sigma_\alpha f_k)$, with all appearing semialgebraic functions extended by zero, using our cutoff functions σ_α. Then one sets

$$\rho_\alpha \omega_{\alpha\alpha_0\dots\alpha_{p-1}} = \int_Y \rho_\alpha\beta,$$

with the fiber integral now being the one associated to the trivial semialgebraic bundle $Y \times U_{\alpha_0\dots\alpha_{p-1}} \to U_{\alpha_0\dots\alpha_{p-1}}$.

Having defined the cochain τ above one then checks as in the proof [8, Proposition 8.5] that $\delta\tau = \omega$, using that $\delta\omega = 0$. Overall, we have then shown that the second (E^1–)page of the "rowwise" spectral sequence is identified with $(\Omega_{\mathrm{triv}}(M), d)$.

We also note that this step of the proof is closely analogous to that of [24, Lemma 6.7], but slightly simpler since trivial bundles can be extended trivially.

The next page of the "rowwise" spectral sequence is then $H(\Omega_{\mathrm{triv}}(M), d)$, and the spectral sequence converges at this point by degree reasons. Hence

$$H(\Omega_{\mathrm{triv}}(M), d) \cong H^\bullet(M).$$

It is shown in [24] that $H(\Omega_{\mathrm{PA}}(M), d) \cong H(M)$. To see that the inclusion

$$\Omega_{\mathrm{triv}}(M) \subset \Omega_{\mathrm{PA}}(M)$$

induces the isomorphism on cohomology one may consider the PA Čech–de Rham complex C_{PA}, defined by replacing Ω_{triv} by Ω_{PA} in the definition of C above. Using the PA Poincaré lemma [24, Lemma 6.3] it is then clear that the natural inclusion $C \to C_{\mathrm{PA}}$ induces an isomorphism on the E^2–page of the "columnwise" spectral sequences on both sides, and hence is a quasi-isomorphism. □

We note that in fact in the definition of Ω_{triv} we do not need globally trivial bundles, local triviality suffices.

Proposition 85 *Let M be a compact semialgebraic manifold and let $p\colon E \to M$ be an oriented SA bundle; see* [24, Definition 8.1]. *Let $\omega \in \Omega_{\mathrm{triv}}(E)$. Then the corresponding fiber integral $\int_{E\to M} \omega \in \Omega_{\mathrm{PA}}(M)$ is an element of $\Omega_{\mathrm{triv}}(M) \subset \Omega_{\mathrm{PA}}(M)$.*

Proof We may assume that $\omega \in \Omega_{\min}(M)$ by replacing E with a product of E with some trivial bundle if needed. We pick a finite trivializing cover $\{U_i\}$, a semialgebraic partition of unity ρ_i, and cutoff functions σ_i as in the proof of Corollary 84.

We then rewrite

$$\int_{E\to M} \omega = \sum_i \rho_i \int_{E\to M} \omega = \sum_i \int_{E\to M} \rho_i \omega.$$

For the last equality we abused notation and defined $\rho_i := p^*\rho_i$, and we implicitly used [24, Proposition 8.9]. Let the local trivialization of the bundle on U_i be denoted by $h_i : U_i \times F \xrightarrow{\cong} p^{-1}(U_i)$. As in the previous proof we extend the minimal form $\rho_i h_i^*\omega \in \Omega_{\min}(U_i \times F)$ to a minimal form $\rho_i h_i^*\omega \in \Omega_{\min}(M \times F)$, which we abusively denote by the same symbols. We then claim that

$$\int_{E\to M} \rho_i \omega = \int_{M\times F\to M} \rho_i h_i^* \omega. \tag{22}$$

Since the right-hand side is a fiber integral over a trivial bundle the proposition then follows.

To check (22) we need to consider a trivializing stratification $\{S_\alpha\}$ for the strongly continuous chain Φ corresponding to the bundle $E \to M$. The stratification can be taken such that the closure of each stratum is contained in one of the U_j as in the proof of [24, Proposition 8.2]. We can furthermore refine it so that each $\overline{S}_\alpha$ is either contained in U_i or disjoint from the support of ρ_i. (For example, refine the stratification by intersecting the strata with $\{x \mid \sigma_i(x) \geq 0.9\}$ and the closure of its complement.)

Now consider some stratum S_α, and the restriction of (22) to its closure. If $\overline{S}_\alpha$ is disjoint from the support of ρ_i then trivially both sides of (22) vanish on it. Otherwise we may assume that $\overline{S}_\alpha \subset U_i$. But the bundle isomorphism h_i transforms one side of (22) into the other; see [24, Proposition 8.10]. □

References

[1] **M Alexandrov**, **A Schwarz**, **O Zaboronsky**, **M Kontsevich**, *The geometry of the master equation and topological quantum field theory*, Internat. J. Modern Phys. A 12 (1997) 1405–1429 MR Zbl

[2] **S Axelrod**, **I M Singer**, *Chern–Simons perturbation theory, II*, J. Differential Geom. 39 (1994) 173–213 MR Zbl

[3] **D Ayala**, **J Francis**, *Factorization homology of topological manifolds*, J. Topol. 8 (2015) 1045–1084 MR Zbl

[4] **M Bendersky**, **S Gitler**, *The cohomology of certain function spaces*, Trans. Amer. Math. Soc. 326 (1991) 423–440 MR Zbl

[5] **J Bochnak**, **M Coste**, **M-F Roy**, *Real algebraic geometry*, Ergebnisse der Math. (3) 36, Springer (1998) MR Zbl

[6] **F Bonechi**, **A S Cattaneo**, **P Mnev**, *The Poisson sigma model on closed surfaces*, J. High Energy Phys. (2012) art. no. 099 MR Zbl

[7] **R Bott**, **A S Cattaneo**, *Integral invariants of 3–manifolds*, J. Differential Geom. 48 (1998) 91–133 MR Zbl

[8] **R Bott**, **L W Tu**, *Differential forms in algebraic topology*, Graduate Texts in Math. 82, Springer (1982) MR Zbl

[9] **P Boavida de Brito**, **M Weiss**, *Manifold calculus and homotopy sheaves*, Homology Homotopy Appl. 15 (2013) 361–383 MR Zbl

[10] **P Boavida de Brito**, **M Weiss**, *Spaces of smooth embeddings and configuration categories*, J. Topol. 11 (2018) 65–143 MR Zbl

[11] **R Campos**, *BV formality*, Adv. Math. 306 (2017) 807–851 MR Zbl

[12] **A S Cattaneo**, **G Felder**, *Effective Batalin–Vilkovisky theories, equivariant configuration spaces and cyclic chains*, from "Higher structures in geometry and physics" (A S Cattaneo, A Giaquinto, P Xu, editors), Progr. Math. 287, Springer (2011) 111–137 MR Zbl

[13] **A S Cattaneo**, **P Mnëv**, *Remarks on Chern–Simons invariants*, Comm. Math. Phys. 293 (2010) 803–836 MR Zbl

[14] **A S Cattaneo**, **C A Rossi**, *Higher-dimensional BF theories in the Batalin–Vilkovisky formalism: the BV action and generalized Wilson loops*, Comm. Math. Phys. 221 (2001) 591–657 MR Zbl

[15] **F R Cohen**, **L R Taylor**, *Computations of Gelfand–Fuks cohomology, the cohomology of function spaces, and the cohomology of configuration spaces*, from "Geometric applications of homotopy theory, I" (M G Barratt, M E Mahowald, editors), Lecture Notes in Math. 657, Springer (1978) 106–143 MR Zbl

[16] **V A Dolgushev**, **C L Rogers**, *A version of the Goldman–Millson theorem for filtered L_∞–algebras*, J. Algebra 430 (2015) 260–302 MR Zbl

[17] **V Dolgushev**, **T Willwacher**, *Operadic twisting — with an application to Deligne's conjecture*, J. Pure Appl. Algebra 219 (2015) 1349–1428 MR Zbl

[18] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Mathematics 205, Springer (2001) MR Zbl

[19] **Y Felix**, **J-C Thomas**, *Configuration spaces and Massey products*, Int. Math. Res. Not. 2004 (2004) 1685–1702 MR Zbl

[20] **B Fresse**, **T Willwacher**, *The intrinsic formality of E_n–operads*, J. Eur. Math. Soc. 22 (2020) 2047–2133 MR Zbl

[21] **E Getzler**, **J D S Jones**, *Operads, homotopy algebra and iterated integrals for double loop spaces*, preprint (1994) arXiv hep-th/9403055

[22] **T G Goodwillie**, **M Weiss**, *Embeddings from the point of view of immersion theory, II*, Geom. Topol. 3 (1999) 103–118 MR Zbl

[23] **A Hamilton**, **A Lazarev**, *Symplectic C_∞–algebras*, Mosc. Math. J. 8 (2008) 443–475 MR Zbl

[24] **R Hardt**, **P Lambrechts**, **V Turchin**, **I Volić**, *Real homotopy theory of semi-algebraic sets*, Algebr. Geom. Topol. 11 (2011) 2477–2545 MR Zbl

[25] **N Idrissi**, *The Lambrechts–Stanley model of configuration spaces*, Invent. Math. 216 (2019) 1–68 MR Zbl

[26] **M Kontsevich**, *Feynman diagrams and low-dimensional topology*, from "First European Congress of Mathematics, II" (A Joseph, F Mignot, F Murat, B Prum, R Rentschler, editors), Progr. Math. 120, Birkhäuser, Basel (1994) 97–121 MR Zbl

[27] **M Kontsevich**, *Operads and motives in deformation quantization*, Lett. Math. Phys. 48 (1999) 35–72 MR Zbl

[28] **M Kontsevich**, *Deformation quantization of Poisson manifolds*, Lett. Math. Phys. 66 (2003) 157–216 MR Zbl

[29] **M Kontsevich**, **Y Soibelman**, *Deformations of algebras over operads and the Deligne conjecture*, from "Conférence Moshé Flato, I" (G Dito, D Sternheimer, editors), Math. Phys. Stud. 21, Kluwer Acad., Dordrecht (2000) 255–307 MR Zbl

[30] **I Kriz**, *On the rational homotopy type of configuration spaces*, Ann. of Math. 139 (1994) 227–237 MR Zbl

[31] **P Lambrechts**, **D Stanley**, *Poincaré duality and commutative differential graded algebras*, Ann. Sci. Éc. Norm. Supér. 41 (2008) 495–509 MR Zbl

[32] **P Lambrechts**, **D Stanley**, *A remarkable DGmodule model for configuration spaces*, Algebr. Geom. Topol. 8 (2008) 1191–1222 MR Zbl

[33] **P Lambrechts**, **I Volić**, *Formality of the little N–disks operad*, Mem. Amer. Math. Soc. 1079, Amer. Math. Soc., Providence, RI (2014) MR Zbl

[34] **N Levitt**, *Spaces of arcs and configuration spaces of manifolds*, Topology 34 (1995) 217–230 MR Zbl

[35] **J-L Loday**, *Cyclic homology*, 2nd edition, Grundl. Math. Wissen. 301, Springer (1998) MR Zbl

[36] **J-L Loday**, **B Vallette**, *Algebraic operads*, Grundl. Math. Wissen. 346, Springer (2012) MR Zbl

[37] **R Longoni**, **P Salvatore**, *Configuration spaces are not homotopy invariant*, Topology 44 (2005) 375–380 MR Zbl

[38] **N Markarian**, **H L Tanaka**, *Factorization homology in 3–dimensional topology*, from "Mathematical aspects of quantum field theories", Springer (2015) 213–231 MR Zbl

[39] **M Markl**, **S Shnider**, **J Stasheff**, *Operads in algebra, topology and physics*, Mathematical Surveys and Monographs 96, Amer. Math. Soc., Providence, RI (2002) MR Zbl

[40] **M Shiota**, *Nash manifolds*, Lecture Notes in Math. 1269, Springer (1987) MR Zbl

[41] **D P Sinha**, *Manifold-theoretic compactifications of configuration spaces*, Selecta Math. 10 (2004) 391–428 MR Zbl

[42] **D Sullivan**, *Infinitesimal computations in topology*, Inst. Hautes Études Sci. Publ. Math. 47 (1977) 269–331 MR Zbl

[43] **A Tognoli**, *Su una congettura di Nash*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 27 (1973) 167–185 MR Zbl

[44] **B Totaro**, *Configuration spaces of algebraic varieties*, Topology 35 (1996) 1057–1067 MR Zbl

[45] **T Willwacher**, *M Kontsevich's graph complex and the Grothendieck–Teichmüller Lie algebra*, Invent. Math. 200 (2015) 671–760 MR Zbl

University of Montpellier
Montpellier, France

Department of Mathematics, ETH Zurich
Zurich, Switzerland

ricardo.campos@math.univ-toulouse.fr, thomas.willwacher@math.ethz.ch

Received: 3 September 2020 Revised: 20 September 2021

Algebraic & Geometric Topology **23**:5 (2023) 2107–2140
DOI: 10.2140/agt.2023.23.2107
Published: 25 July 2023

The Hurewicz theorem in homotopy type theory

J Daniel Christensen
Luis Scoccola

We prove the Hurewicz theorem in homotopy type theory, ie that for X a pointed, $(n-1)$–connected type, with $n \geq 1$, and A an abelian group, there is a natural isomorphism $\pi_n(X)^{\mathrm{ab}} \otimes A \cong \widetilde{H}_n(X; A)$ relating the abelianization of the homotopy groups with the homology. We also compute the connectivity of a smash product of types and express the lowest nontrivial homotopy group as a tensor product. Along the way, we study magmas, loop spaces, connected covers and prespectra, and we use 1–coherent categories to express naturality and for the Yoneda lemma.

As homotopy type theory has models in all ∞–toposes, our results can be viewed as extending known results about spaces to all other ∞–toposes.

55Q99; 03B38, 18N60, 55N99

1 Introduction

Homotopy type theory is a formal system which has models in all ∞–toposes [2; 3; 11; 15; 19].[1] As such, it provides a convenient way to prove theorems for all ∞–toposes. In addition, homotopy type theory is well suited to being formalized in a proof assistant [1; 8].

Working in homotopy type theory as described in the book [16], we prove the Hurewicz theorem:

Theorem H (Theorem 3.12) *For $n \geq 1$, X a pointed, $(n-1)$–connected type, and A an abelian group, there is a natural isomorphism*

$$\pi_n(X)^{\mathrm{ab}} \otimes A \cong \widetilde{H}_n(X; A),$$

[1]The initiality and semantics of higher inductive types still need to be fully worked out.

where on the left-hand side we take the abelianization (*which only matters when* $n = 1$). *In particular*, *when* A *is the integers*, *this specializes to an isomorphism* $\pi_n(X)^{\mathrm{ab}} \cong \widetilde{H}_n(X)$.

As mentioned above, this holds in any ∞–topos, and so is more general than the well-known Hurewicz theorem in topology. Interpreting the statement in an ∞–topos is somewhat subtle. The groups that appear in the statement are internal group objects whose underlying object is 0–truncated (a "set", internally). The quantification over n means that there is a map $h\colon H \to \mathbb{N}$ in the ∞–topos representing a family of objects over the natural numbers object, and that this map has a section. In particular, since each ordinary natural number gives a global element of $\mathbb{N}$, it follows that the fibre of h over that element must itself have a global element. Continuing in this way, we deduce that for given objects X and A as in the statement, the two internal group objects shown are equivalent as group objects. For more on the interpretation of type theory, see Shulman [18, Section 4.2; 19] for the interpretation in arbitrary ∞–topoi, and Kapulkin and Lumsdaine [12] for a more explicit interpretation in simplicial sets.

Since we prove this theorem for an arbitrary ∞–topos, we must be careful to use arguments that apply in this generality. For example, it is not true in every ∞–topos that a surjective map of sets has a section, so we cannot use the axiom of choice. Similarly, the law of excluded middle and Whitehead's theorem can both fail, so we must not use these results either. Because of this, our proof is different from other known proofs.

Before giving more details, we give some motivation for the interest in this result, for those less familiar with traditional homotopy theory.

Motivation

In topology, homotopy groups are in a certain sense the strongest invariants of a topological space, and so their computation is an important tool when trying to classify spaces up to homotopy. In homotopy type theory, homotopy groups play a fundamental role in that they capture information about iterated identity types. Unfortunately, even in classical topology, the computation of homotopy groups is a notoriously difficult problem. Nevertheless, topologists have come up with a variety of powerful tools for attacking this problem, and one of the most basic tools is the Hurewicz theorem. In most cases, it is much easier to compute homology groups than homotopy groups, and so one can use the isomorphism from right to left (with A taken to be the integers) to compute certain homotopy groups. Moreover, one can apply the theorem even when X

is not $(n-1)$–connected using the following technique. Let $X\langle n-1\rangle$ denote the fibre of the truncation map $X \to \|X\|_{n-1}$ over the image of the basepoint. Then $X\langle n-1\rangle$ is $(n-1)$–connected and $\pi_n(X\langle n-1\rangle) \cong \pi_n(X)$, so $\pi_n(X)^{\mathrm{ab}} \cong \widetilde{H}_n(X\langle n-1\rangle)$. The Serre spectral sequence can often be used to compute the required homology group.

Techniques and main results

We first recall that for $n \geq 1$, the n^{th} homology group $\widetilde{H}_n(X;A)$ of a type X with coefficients in an abelian group A is defined to be the colimit of a certain sequential diagram,

$$\pi_{n+1}(X \wedge K(A,1)) \longrightarrow \pi_{n+2}(X \wedge K(A,2)) \longrightarrow \pi_{n+3}(X \wedge K(A,3)) \longrightarrow \cdots . \tag{1-1}$$

Here $\wedge$ denotes the smash product and $K(A,m)$ is the Eilenberg–Mac Lane space constructed by Licata and Finster in [13], which is an m–truncated, $(m-1)$–connected, pointed type with a canonical isomorphism $\pi_m(K(A,m)) \cong A$.

We now state one of our main results, which is used to prove the Hurewicz theorem, and also has other consequences:

Theorem S (Corollary 2.32 and Theorem 2.38) *If X is a pointed, $(n-1)$–connected type with $n \geq 1$ and Y is a pointed, $(m-1)$–connected type with $m \geq 1$, then $X \wedge Y$ is $(n+m-1)$–connected and $\pi_{n+m}(X \wedge Y)$ is the tensor product of $\pi_n(X)^{\mathrm{ab}}$ and $\pi_m(Y)^{\mathrm{ab}}$ in a natural way.*

Taking Y to be $K(A,m)$ in this result shows that the groups appearing in the sequential diagram (1-1) are tensor products of $\pi_n(X)^{\mathrm{ab}}$ and A. The proof of the Hurewicz theorem follows from showing that the induced maps are isomorphisms, which we do in Lemma 3.11. With this ingredient, we prove the Hurewicz theorem as Theorem 3.12.

In order to define the isomorphism appearing in Theorem S, we must give a bilinear map $\pi_n(X) \to_{\mathsf{Grp}} \pi_m(Y) \to_{\mathsf{Grp}} \pi_{n+m}(X \wedge Y)$. To do so, we define and study a more general natural map

$$\mathsf{smashing}\colon (X \to_\bullet Y \to_\bullet Z) \longrightarrow \big(\pi_n(X) \to_{\mathsf{Grp}} \pi_m(Y) \to_{\mathsf{Grp}} \pi_{n+m}(Z)\big)$$

for any pointed types X, Y and Z and any $n, m \geq 1$. The map we require is obtained by applying $\mathsf{smashing}$ to the natural map $X \to_\bullet Y \to_\bullet X \wedge Y$.

Constructing the map $\mathsf{smashing}$ requires some work. While it lands in group homomorphisms between (0–truncated) groups, in order to construct it, we pass through *magmas*. A magma is a (not necessarily truncated) type M with a binary operation $\cdot\colon M \times M \to M$, with no conditions or coherence laws. As a technical trick which simplifies the formalization, we work with *weak magma morphisms*. A weak magma morphism from a magma M to a magma N is a map $f\colon M \to N$ which *merely* has the property that it respects the operations. This is sufficient for our purposes, because when M and N are groups, it reproduces the notion of group homomorphism. All loop spaces are magmas under path concatenation, and many natural maps involving loop spaces are weak magma morphisms. By working with magmas, we can factor the map $\mathsf{smashing}$ into simpler pieces, and still land in group homomorphisms at the end, without keeping track of higher coherences.

Proving the rest of Theorem S requires a number of results that build on work of Buchholtz, van Doorn and Rijke [4]. For example, Lemma 2.15 and Theorem 2.19 are results of [4], which we use to prove Proposition 2.23: for $n \geq 1$, X a pointed, $(n-1)$–connected type, and Y a pointed, n–truncated type, the map

$$\Omega^n\colon (X \to_\bullet Y) \to (\Omega^n X \to_{\mathsf{Mgm}} \Omega^n Y)$$

is an equivalence. In order to prove this, we prove results about connected covers in Section 2.3.

We go on to define a natural Hurewicz homomorphism $h_n\colon \pi_n(X)^{\mathrm{ab}} \otimes A \to \widetilde{H}_n(X; A)$, without assuming any connectivity hypothesis on X, and show that it is unique up to a sign among such natural transformations that give isomorphisms for $X \equiv S^n$ and $A \equiv \mathbb{Z}$ (Theorem 3.16).

Homology

The theory of homology in homotopy type theory is currently limited by the absence of some important tools and facts that would make it easier to compute. For example, we don't have complete proofs that homology satisfies the Eilenberg–Steenrod axioms, although partial work was done by Graham [9]. The Serre spectral sequence for homology has not been formalized, but high level arguments can be found in [8] and it is expected that techniques similar to those used for cohomology will go through. We are also missing the fact that the homology of a cellular space can be computed cellularly (which is done for cohomology in [5]), the universal coefficient theorem, and the relationship between homology and localization (developed in homotopy type theory in [7; 17]).

Structure of the paper

Section 2 contains our work on smash products and tensor products. After listing our conventions in Section 2.1, we give the basic theory of 1–coherent categories in Section 2.2. We use this theory to express and reason about natural transformations, and we make use of the Yoneda lemma in this setting. In Section 2.3 we study connected covers. Section 2.4 introduces magmas and weak magma morphisms, and proves a variety of results about loop spaces, including Proposition 2.23, mentioned above. We also define the map $\mathsf{smashing}$ in this section. We introduce smash products in Section 2.5 and prove the connectivity part of Theorem S here. Section 2.6 is a short section that defines abelianization and gives a particularly efficient construction of the abelianization of a group as a higher inductive type. In Section 2.7, we define tensor products of abelian groups and prove the second part of Theorem S. Section 2.8 proves results about smash products, truncation and suspension that are needed in Section 3.

Section 3 applies the results of Section 2 to homology, leading up to the Hurewicz theorem and its consequences. In Section 3.1, we define prespectra and their stable homotopy groups, and use this to define homology. The Hurewicz theorem is proved in Section 3.2, and we describe the Hurewicz homomorphism and its uniqueness up to sign in Section 3.3. In Section 3.4, we give some applications of our main results.

Formalization

Formalization of these results is in progress, with help from Ali Caglayan, using the Coq HoTT library [10]. The current status can be seen at [6], where the `README.md` file explains where results from the paper can be found. Currently, we have formalized much of Section 2 but none of Section 3. In Section 2, the only substantial result that is missing is Theorem 2.38. Also missing are Theorem 2.28 and the naturality of many of the maps defined in this section. In our formalization, we take as axioms several results that have been formalized in other proof assistants.

2 Smash products and tensor products

In this section, we give a variety of results about loop spaces, magmas, smash products and tensor products, including the proof of Theorem S. None of the results in this section depend on the definition of homology, but these results are used in the next section to prove the Hurewicz theorem.

2.1 Background and conventions

We follow the conventions and notation used in [16]. We assume we have a univalent universe $\mathcal{U}$ closed under higher inductive types (HITs) and contained in another universe $\mathcal{U}'$. In fact, the higher inductive types we use can all be described using pushouts and truncations. By convention, all types live in the lower universe $\mathcal{U}$, unless explicitly stated. We implicitly use function extensionality for $\mathcal{U}$ throughout.

A *pointed type* is a type X and a choice of $x_0 : X$, and the type of pointed types is denoted by $\mathcal{U}_\bullet :\equiv \sum_{(X:\mathcal{U})} X$. We often keep the choice of basepoint implicit. A *pointed map* between pointed types X and Y is a map $f : X \to Y$ and a path $p : f(x_0) = y_0$. The type of pointed maps is denoted by $X \to_\bullet Y :\equiv \sum_{(f : X \to Y)} f(x_0) = y_0$.

We frequently make use of functions of type $X \to Y \to Z$, and remind the reader that this associates as $X \to (Y \to Z)$, which is the curried form of a function $X \times Y \to Z$.

In the paper, we define the sum $m + n$ of natural numbers by induction on n, so that $m + 1$ is the successor of m. In the HoTT library, the other convention is used, so to translate between the paper and the formalization, one must change $m + n$ to $n + m$ everywhere.

2.2 1–coherent categories

In this section, we briefly discuss the notion of a 1–coherent category, which we use to express that various constructions are natural. The definitions generalize those of [8, Section 4.3.1], which deals with the 1–coherent category of pointed types, except that our hom types are unpointed. A more general notion of wild category has been formalized in the HoTT library [10] by Ali Caglayan, tslil clingman, Floris van Doorn, Morgan Opie, Mike Shulman and Emily Riehl.

Recall that $\mathcal{U}'$ is a universe such that $\mathcal{U} : \mathcal{U}'$.

Definition 2.1 A 1–*coherent category* C consists of a type $C_0 : \mathcal{U}'$, a map

$$\hom_C : C_0 \to C_0 \to \mathcal{U},$$

maps

$$\mathsf{id} : \prod_{a:C_0} \hom_C(a,a),$$

$$- \circ - : \prod_{a,b,c:C_0} \hom_C(b,c) \to \hom_C(a,b) \to \hom_C(a,c),$$

and equalities

$$\mathsf{unitl}\colon \prod_{a,b:C_0} \prod_{f:\hom_C(a,b)} \mathsf{id}_b \circ f = f,$$

$$\mathsf{unitr}\colon \prod_{a,b:C_0} \prod_{f:\hom_C(a,b)} f \circ \mathsf{id}_a = f,$$

$$\mathsf{assoc}\colon \prod_{a,b,c,d:C_0} \prod_{f:\hom_C(a,b)} \prod_{g:\hom_C(b,c)} \prod_{h:\hom_C(c,d)} (h \circ g) \circ f = h \circ (g \circ f),$$

witnessing left and right unitality and associativity, respectively. We do not assume coherence laws or that any of the types are truncated.

If C is a 1–coherent category, the elements of C_0 are called *objects* and, for objects $a, b : C_0$, the elements of $\hom_C(a, b)$ are called *morphisms* from a to b.

The wild 1–categories considered in [10] allow 2–cells to be specified, which are then used in place of the identity types in the above equalities. For simplicity, we use the identity types.

Example 2.2 There is a 1–coherent category U of types, with $\mathsf{U}_0 :\equiv \mathcal{U}$ and

$$\hom_{\mathsf{U}}(X, Y) :\equiv X \to Y$$

for every pair of types $X, Y : \mathcal{U}$. Identity morphisms, composition, unitalities, and associativity all work in the expected way.

Example 2.3 There is a 1–coherent category Grp of groups whose objects are the set-level groups, that is, 0–truncated types equipped with an associative binary operation, a unit and inverses. The morphisms are standard group homomorphisms.

Similarly, there is 1–coherent category Ab of abelian groups.

Example 2.4 Any precategory in the sense of [16, Definition 9.1.1] gives rise to a 1–coherent category, simply by forgetting that its hom types are sets. Moreover, the notions of isomorphism, functor, and natural transformation given in [16, Section 9] are equivalent to the notions we give in this section, in the case of precategories.

Many constructions one can carry out with categories are easy to extend to 1–coherent categories. We mention two that are particularly important for us. Given a 1–coherent category C, we can form the *opposite* 1–coherent category C^{op} by letting the type of objects of C^{op} be C_0, and $\hom_{C^{\mathsf{op}}}(a, b) :\equiv \hom_C(b, a)$ for all $a, b : C_0$. The rest of the structure is straightforward to define.

Given 1–coherent categories C and D, one can form a *product* 1–coherent category, denoted by $C \times D$. The underlying type of $C \times D$ is simply $C_0 \times D_0$, and $\mathrm{hom}_{C\times D}((c,d),(c',d')) :\equiv \mathrm{hom}_C(c,c') \times \mathrm{hom}_D(d,d')$. The rest of the structure is again straightforward to define.

Definition 2.5 Let C be a 1–coherent category, $a, b : C_0$, and $f : \mathrm{hom}_C(a,b)$. An *isomorphism structure* for f is given by morphisms $g, h : \mathrm{hom}_C(b,a)$ together with paths $l : g \circ f = \mathsf{id}_a$ and $r : f \circ h = \mathsf{id}_b$.

In many cases, such as in the 1–coherent category U, being an isomorphism is a mere property of a morphism. The wild 1–categories considered in [10] allow biinvertibility to be replaced by more general notions of isomorphism.

Definition 2.6 A 1*–coherent functor* F between 1–coherent categories C and D, usually denoted by $F\colon C \to D$, consists of a map $F_0\colon C_0 \to D_0$, a map

$$F_1\colon \prod_{a,b:C_0} \mathrm{hom}_C(a,b) \to \mathrm{hom}_D(F_0(a), F_0(b)),$$

and equalities

$$F_{\mathsf{id}} : \prod_{a:C_0} F(\mathsf{id}_a) = \mathsf{id}_{F(a)},$$

$$F_\circ : \prod_{a,b,c:C_0} \prod_{f:\mathrm{hom}_C(a,b)} \prod_{g:\mathrm{hom}_C(b,c)} F_1(g) \circ F_1(f) = F(g \circ f),$$

witnessing the functoriality of F.

Example 2.7 For a 1–coherent category C and an object $a : C_0$, we can define a 1–coherent corepresentable functor $\mathcal{Y}^a\colon C \to \mathsf{U}$. On objects, $\mathcal{Y}^a_0(b) :\equiv \mathrm{hom}_C(a,b)$. The action on morphisms is defined as $\mathcal{Y}^a_1(f) :\equiv \lambda g.\, f \circ g : \mathrm{hom}_C(a,b) \to \mathrm{hom}_C(a,c)$ for $f : \mathrm{hom}_C(b,c)$. The witnesses of functoriality, that is $\mathcal{Y}^a_{\mathsf{id}}$ and $\mathcal{Y}^a_\circ$, are defined using the equalities unitl and assoc of C, respectively.

Definition 2.8 Let C and D be 1–coherent categories and let $F, G\colon C \to D$ be 1–coherent functors. A 1*–coherent natural transformation* α from F to G, usually denoted by $\alpha\colon F \to G$, consists of a map

$$\alpha_0 : \prod_{a:C_0} \mathrm{hom}_D(F(a), G(a)),$$

and equalities

$$\alpha_1 : \prod_{a,b:C_0} \prod_{f:\mathrm{hom}_C(a,b)} \alpha_0(b) \circ F_1(f) = G_1(f) \circ \alpha_0(a).$$

Definition 2.9 Let $\alpha : F \to G$ be a 1–coherent natural transformation between 1–coherent functors $F, G : C \to D$, for C and D 1–coherent categories. An *isomorphism structure* for α is given by an isomorphism structure for each of its components. A *natural isomorphism* is given by a natural transformation together with an isomorphism structure.

The following lemma is straightforward.

Lemma 2.10 *Let C and D be 1–coherent categories, $F, G, H : C \to D$ 1–coherent functors, and $\alpha : F \to G$ and $\beta : G \to H$ 1–coherent natural transformations. Then, by defining $(\beta \circ \alpha)(c) :\equiv \beta(c) \circ \alpha(c)$ and the naturality squares by composing the naturality squares of α and β, one obtains a natural transformation $\beta \circ \alpha : F \to H$. Moreover, if both α and β are natural isomorphisms, so is $\beta \circ \alpha$.* □

The following is a 1–coherent version of the fact that the Yoneda functor is an embedding.

Proposition 2.11 [10] *Let C be a 1–coherent category and let $a, b : C_0$. Assume given a 1–coherent natural isomorphism $\alpha : \mathcal{Y}^b \to \mathcal{Y}^a$. Then $i :\equiv \alpha_0(b)(\mathsf{id}_b) : \hom_C(a, b)$ is part of an isomorphism between a and b, and it satisfies, for every $c : C_0$,*

$$\alpha_0(c) = \lambda g.g \circ i$$

as maps $\hom_C(b, c) \to \hom_C(a, c)$.

The proof is the same as the usual proof, and has been formalized in the HoTT library [10]. Note that we are not claiming that the naturality proofs for α can be recovered using the associativity of composition.

2.3 Connected covers

In order to generalize a result of Buchholtz, van Doorn and Rijke (see Theorem 2.19) to the case where Y has no connectivity assumption, we prove some results about connected covers. In this section, we fix $n \geq -1$.

Definition 2.12 A type X is *n–connected* if $\|X\|_n$ is contractible.

For X pointed, it is equivalent to require that $\pi_i(X)$ be trivial for all $i \leq n$. Every pointed type is (-1)–connected.

Definition 2.13 Let X be a pointed type. The *n–connected cover* $X\langle n\rangle$ of X is defined to be the fibre of the pointed map $X \to_\bullet \|X\|_n$.

Note that $X\langle n\rangle$ is indeed n–connected and that we have a canonical pointed map $i : X\langle n\rangle \to_\bullet X$ which induces an equivalence on the homotopy groups π_k for $k > n$. In fact, this map has a stronger universal property:

Definition 2.14 A pointed map $f : X \to_\bullet Y$ is an *$\langle n\rangle$–equivalence* if for any pointed, n–connected type Z, postcomposition by f gives an equivalence

$$(Z \to_\bullet X) \xrightarrow{\sim} (Z \to_\bullet Y).$$

Lemma 2.15 [4, Lemma 6.2] *Let X be a pointed type. Then the map $i : X\langle n\rangle \to_\bullet X$ is an $\langle n\rangle$–equivalence.*

It follows that the operation sending X to $X\langle n\rangle$ is functorial in a unique way making $i : X\langle n\rangle \to_\bullet X$ natural, and that a map f is an $\langle n\rangle$–equivalence if and only if $f\langle n\rangle$ is an equivalence.

Note that there is a 1–coherent category with objects all pointed types and morphisms given by pointed functions. We denote this 1–coherent category by $\mathsf{U}_\bullet$. There are 1–coherent functors $\Sigma, \Omega : \mathsf{U}_\bullet \to \mathsf{U}_\bullet$ forming a 1–coherent adjunction, in the following sense.

Lemma 2.16 [16, Lemma 6.5.4] *Let X and Y be pointed types. There is an equivalence*

$$(\Sigma X \to_\bullet Y) \simeq (X \to_\bullet \Omega Y),$$

natural in X and Y. Here, we are interpreting $(\Sigma(-) \to_\bullet -)$ and $(- \to_\bullet \Omega(-))$ as 1–coherent functors $\mathsf{U}_\bullet^{\mathsf{op}} \times \mathsf{U}_\bullet \to \mathsf{U}$.

The naturality is not proven in [16], but is proven in the HoTT library [10].

The following two facts will be used in Proposition 2.23.

Proposition 2.17 *Let $f : X \to_\bullet Y$ be a pointed map. If f is an $\langle n+1\rangle$–equivalence, then Ωf is an $\langle n\rangle$–equivalence.*

Proof Let A be an n–connected, pointed type. By naturality of the adjunction between suspension and loops (Lemma 2.16), we have a commutative square

$$\begin{array}{ccc}
(\Sigma A \to_\bullet X) & \xrightarrow{f\circ -} & (\Sigma A \to_\bullet Y) \\
{\scriptstyle\sim}\downarrow & & \downarrow{\scriptstyle\sim} \\
(A \to_\bullet \Omega X) & \xrightarrow{\Omega f\circ -} & (A \to_\bullet \Omega Y)
\end{array}$$

in which the vertical maps are equivalences. Since the suspension of an n–connected type is $(n+1)$–connected, the top map is also an equivalence. Therefore, the bottom map is an equivalence, as required. □

Proposition 2.18 *Let $f : X \to_\bullet Y$ be a pointed map. If f is a $\langle -1\rangle$–equivalence, then f is an equivalence.*

Proof Since S^0 is (-1)–connected, we know that f induces an equivalence

$$(S^0 \to_\bullet X) \to_\bullet (S^0 \to_\bullet Y).$$

Moreover, $(S^0 \to_\bullet Z)$ is equivalent to Z for any pointed type Z, and this equivalence is natural. It follows that f is an equivalence. □

This also follows from the facts that $Z\langle -1\rangle \to_\bullet Z$ is an equivalence for any pointed Z, and that $f\langle -1\rangle$ is an equivalence.

2.4 Loop spaces and magmas

In this section, we study loop spaces and the natural magma structures that they carry and define the map $\mathsf{smashing}$ that plays an important role in this paper. We begin by generalizing the following result of Buchholtz, van Doorn and Rijke.

Theorem 2.19 [4, Theorem 5.1] *Let $n \geq 1$. For X and Y pointed, $(n-1)$–connected, n–truncated types, the map*

$$\Omega^n : (X \to_\bullet Y) \longrightarrow_\bullet (\Omega^n X \to_{\mathsf{Grp}} \Omega^n Y).$$

is an equivalence.

In order to state our generalization, we introduce the notion of magma.

Definition 2.20 A *magma* is given by a type X together with an operation

$$\cdot_X : X \times X \to X.$$

A *map of magmas* is given by a map $f : X \to Y$ between the underlying types that merely respects the operations. More formally, we define

$$X \to_{\mathsf{Mgm}} Y := \sum_{(f : X \to Y)} \Big\| \prod_{(x,x' : X)} f(x \cdot_X x') = f(x) \cdot_Y f(x') \Big\|_{-1}.$$

An *equivalence of magmas* is a map of magmas whose underlying map is an equivalence. We write $X \simeq_{\mathsf{Mgm}} Y$ for the type of magma equivalences from X to Y. Magmas form a 1–coherent category that we denote by Mgm. We will omit the subscript on the operation $\cdot$ when it is clear from context.

The propositional truncation in the definition of magma map is a technical trick to simplify the formalization. With our definition, the type of equalities between magma maps is equivalent to the type of equalities between the underlying maps. All of our results should go through without the truncation, but omitting it leads to path algebra that is not needed in order to get our later results. The maps we are considering should be called "weak magma maps", but since they are the only maps we use, we simply call them "magma maps" in this paper.

Definition 2.21 A *pointed magma* is a magma X with a chosen point $x_0 : X$ and an equality $x_0 \cdot x_0 = x_0$. A *map of pointed magmas* is a pointed map $f : X \to_\bullet Y$ whose underlying map $f : X \to Y$ is a map of magmas. We write $X \to_{\mathsf{Mgm}_\bullet} Y$ for the type of pointed magma maps. An *equivalence of pointed magmas* is a map of pointed magmas whose underlying map is an equivalence. We write $X \simeq_{\mathsf{Mgm}_\bullet} Y$ for the type of pointed magma equivalences. Pointed magmas form a 1–coherent category, which we denote by $\mathsf{Mgm}_\bullet$.

There are no propositional truncations in the above definition, except for the one in the definition of magma map.

Remark 2.22 The loop space ΩX is a pointed magma for any pointed type X, with path concatenation as the operation, reflexivity as the basepoint, and a higher reflexivity as the proof that the basepoint is idempotent. There is a natural map $\Omega\colon (X \to_\bullet Y) \to_\bullet (\Omega X \to_{\mathsf{Mgm}_\bullet} \Omega Y)$, which can be iterated. Any magma map $\Omega X \to_{\mathsf{Mgm}} \Omega Y$ induces a group homomorphism $\pi_1(X) \to_{\mathsf{Grp}} \pi_1(Y)$. Also note that for groups G and H, $(G \to_{\mathsf{Grp}} H) \simeq (G \to_{\mathsf{Mgm}} H)$, where we write $G \to_{\mathsf{Grp}} H$ for the type of group homomorphisms. (We assume that all groups have an underlying type that is a set, which means that the propositional truncation can be removed.)

When X is a pointed magma and G is a group, every magma map $X \to_{\mathsf{Mgm}} G$ can be made pointed in a unique way, so the forgetful map $(X \to_{\mathsf{Mgm}_\bullet} G) \to (X \to_{\mathsf{Mgm}} G)$ is an equivalence.

When A is a pointed type and X is a pointed magma, the type $A \to_\bullet X$ of pointed maps is a pointed magma under the pointwise operation. The requirement that the basepoint $x_0 : X$ be idempotent ensures that for $f, g\colon A \to_\bullet X$, $f \cdot g$ is again pointed: $(f \cdot g)(a_0) \equiv f(a_0) \cdot g(a_0) = x_0 \cdot x_0 = x_0$.

Similarly, when Y is a pointed magma and Z is a pointed type, the type $Y \to_{\mathsf{Mgm}_\bullet} \Omega^2 Z$ of pointed magma maps and the type $Y \to_{\mathsf{Mgm}} \Omega^2 Z$ of all magma maps are pointed magmas under the pointwise operation. This uses that path composition in the double loop space is commutative (by Eckmann–Hilton) and associative. (More precisely, we only use that the operation is merely commutative and merely associative, which will be convenient in Definition 2.26.)

With this background, we can now state our first generalization of Theorem 2.19.

Proposition 2.23 *Let $n \geq 1$, let X be a pointed, $(n-1)$–connected type, and let Y be a pointed, n–truncated type. Then the map*

$$\Omega^n : (X \to_\bullet Y) \xrightarrow{\sim} (\Omega^n X \to_{\mathsf{Mgm}} \Omega^n Y)$$

is an equivalence, natural in X and Y. Similarly,

$$\Omega^n : (X \to_\bullet Y) \xrightarrow{\sim} (\Omega^n X \to_{\mathsf{Mgm}_\bullet} \Omega^n Y)$$

is a natural equivalence.

Proof Since $\Omega^n Y$ is a group, the second equivalence follows from the first, using Remark 2.22, so we focus on the first one. By the functoriality of Ω^n, the diagram

$$\begin{tikzcd}
(X \to_\bullet Y) \arrow[r] & (\Omega^n X \to_{\mathsf{Mgm}} \Omega^n Y) \\
(\|X\|_n \to_\bullet Y) \arrow[r] \arrow[u] & (\Omega^n(\|X\|_n) \to_{\mathsf{Mgm}} \Omega^n Y) \arrow[u] \\
(\|X\|_n \to_\bullet Y\langle n-1\rangle) \arrow[r] \arrow[u] & (\Omega^n(\|X\|_n) \to_{\mathsf{Mgm}} \Omega^n(Y\langle n-1\rangle)) \arrow[u]
\end{tikzcd}$$

commutes, where the vertical maps are induced by the maps $|{-}|_n : X \to_\bullet \|X\|_n$ and $i : Y\langle n-1\rangle \to_\bullet Y$. The vertical maps on the left are equivalences by the universal properties of truncations and of connected covers.

To see that the upper vertical map on the right is an equivalence, let f denote the map $\Omega^n(|{-}|_n) : \Omega^n X \to_\bullet \Omega^n(\|X\|_n)$. This map is 0–connected, since $|{-}|_n$ is n–connected and Ω decreases connectivity. Since $\Omega^n Y$ is a set, it follows that f induces an equivalence $(\Omega^n(\|X\|_n) \to \Omega^n Y) \to (\Omega^n(X) \to \Omega^n Y)$. Given $g : \Omega^n(\|X\|_n) \to \Omega^n Y$, we need to show that g merely preserves the magma structures if and only if $g \circ f$ merely preserves the magma structures. The map f induces an equivalence

$$\Big(\prod_{a,b:\Omega^n(\|X\|_n)} g(a \cdot b) = g(a) \cdot g(b)\Big) \simeq \Big(\prod_{a,b:\Omega^n(X)} g(f(a) \cdot f(b)) = g(f(a)) \cdot g(f(b))\Big)$$

since f is 0–connected and the identity types are sets (in fact, propositions). Note that f, being defined using the functoriality of Ω, preserves the concatenation operation (without any propositional truncation). It follows that the type on the right is equivalent to the type of proofs that $g \circ f$ preserves the magma structure. Therefore, the propositional truncations are also equivalent, so f induces an equivalence on the types of magma maps.

The lower vertical map on the right is an equivalence since $\Omega^n(i)$ is an equivalence of magmas: it is certainly a map of magmas, and the fact that it is an equivalence follows from Propositions 2.17 and 2.18.

The bottom horizontal map is an equivalence by Theorem 2.19, and so the top horizontal map is an equivalence, as required.

The fact that Ω^n is natural in X and Y follows from the functoriality of Ω^n as an operation from pointed maps to magma maps, which is straightforward to check. □

Our next goal is to define the map $\mathsf{smashing}$, using the following lemmas.

Lemma 2.24 *Let $n \geq 1$ and let Y and Z be pointed types. Then there is an equivalence of pointed magmas*

$$\Omega^n(Y \to_\bullet Z) \simeq_{\mathsf{Mgm}_\bullet} (Y \to_\bullet \Omega^n Z),$$

natural in Y and Z. Here we are regarding $\Omega^n(- \to_\bullet -)$ and $- \to_\bullet \Omega^n(-)$ as 1–coherent functors $\mathsf{U}_\bullet \times \mathsf{U}_\bullet \to \mathsf{Mgm}_\bullet$.

On the right-hand side, we are using the pointwise magma structure described in Remark 2.22.

Proof We prove this for $n = 1$, and then iterate, using that the functor Ω sends pointed equivalences to equivalences of pointed magmas.

In order to prove that our equivalence respects the magma structures, it is best to generalize: for $f, g \colon Y \to_\bullet Z$ we define an equivalence

$$\varphi \colon (f = g) \xrightarrow{\sim} \sum_{K : f \sim g} K(y_0) = f_0 \cdot g_0^{-1}.$$

Here K is a homotopy, y_0 is the basepoint of Y, and f_0 and g_0 are the paths witnessing that f and g are pointed. This equivalence is a variant of the standard result that equalities of pointed maps are equivalent to pointed homotopies; the particular choice of the right-hand side means that when f and g are the constant map $Y \to_\bullet Z$ pointed by refl, we obtain a pointed equivalence

$$\Omega(Y \to_\bullet Z) \simeq_\bullet (Y \to_\bullet \Omega Z).$$

Our pointed homotopies can be composed, and we show that φ sends composition of paths to composition of homotopies by first doing induction on the paths to reduce the goal to

$$\varphi(\mathsf{refl}) = \varphi(\mathsf{refl}) \cdot \varphi(\mathsf{refl})$$

and then using path induction to assume that f_0 is refl. We conclude that

$$\Omega(Y \to_\bullet Z) \simeq_{\mathsf{Mgm}_\bullet} (Y \to_\bullet \Omega Z).$$

To prove naturality in Y, consider a pointed map $h\colon Y \to_\bullet Y'$. We must show that the following square commutes:

$$\begin{tikzcd}
\Omega(Y' \to_\bullet Z) \arrow[r, "\Omega(-\circ h)"] \arrow[d, "\varphi"'] & \Omega(Y \to_\bullet Z) \arrow[d, "\varphi"] \\
(Y' \to_\bullet \Omega Z) \arrow[r, "-\circ h"'] & (Y \to_\bullet \Omega Z)
\end{tikzcd}$$

By path induction, we can assume that h is strictly pointed, ie that the given path $h_0 : h(y_0) = y_0'$ is reflexivity. In this case, writing $c\colon Y \to_\bullet Z$ and $c'\colon Y' \to_\bullet Z$ for the constant maps, we have that $c' \circ h$ and c are definitionally equal as pointed maps. Therefore, the corners and vertical maps in the required square are definitionally equal to those in the square

$$\begin{tikzcd}
c' = c' \arrow[r, "\mathsf{ap}_{-\circ h}"] \arrow[d, "\varphi"'] & c' \circ h = c' \circ h \arrow[d, "\varphi"] \\
c' \sim_\bullet c' \arrow[r, "\mathsf{wh}_h"] & c' \circ h \sim_\bullet c' \circ h
\end{tikzcd}$$

where $\sim_\bullet$ denotes the type of pointed homotopies defined above, and wh_h denotes prewhiskering with h. One can check that the horizontal arrows are homotopic to those in the required square, so it remains to show that the new square commutes. To show this, one generalizes from $c' = c'$ to $f = g$, in which case the commutativity follows by path induction.

The proof of naturality in Z is very similar. Since both naturalities have been formalized, we give no further details. □

Lemma 2.25 *Let $n, m \geq 1$ and let Y and Z be pointed types. The action of Ω^m on maps gives a pointed magma map*

$$(Y \to_\bullet \Omega^n Z) \longrightarrow_{\mathsf{Mgm}_\bullet} (\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^m \Omega^n Z).$$

Moreover, the forgetful maps

$$(\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^m \Omega^n Z) \longrightarrow_{\mathsf{Mgm}_\bullet} (\Omega^m Y \to_{\mathsf{Mgm}} \Omega^m \Omega^n Z)$$

and

$$(\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^m \Omega^n Z) \longrightarrow_{\mathsf{Mgm}_\bullet} (\Omega^m Y \to_\bullet \Omega^m \Omega^n Z)$$

are also pointed magma maps. In all cases, we are using the pointwise magma structure described in Remark 2.22. These maps are all natural.

Proof That the forgetful maps are natural pointed magma maps is straightforward, so we focus on the first map. By replacing Z with $\Omega^{n-1} Z$, we can assume that $n = 1$. To prove that Ω^m is a natural pointed magma map, we induct on m. For the inductive step, we define Ω^{m+1} to be the composite

$$(Y \to_\bullet \Omega Z) \xrightarrow{\Omega^m} (\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^m \Omega Z) \longrightarrow (\Omega^m Y \to_\bullet \Omega^m \Omega Z) \xrightarrow{\Omega} (\Omega^{m+1} Y \to_{\mathsf{Mgm}_\bullet} \Omega^{m+1} \Omega Z)$$

so that the claim follows from the inductive hypothesis, the fact that the middle forgetful map is a natural pointed magma map, and the $m = 1$ case.

It remains to prove the $m = 1$ case. It is easy to see that for $f \colon Y \to_\bullet \Omega Z$, Ωf is a pointed magma map. Next we must show that given $f, g \colon Y \to_\bullet \Omega Z$, $\Omega(f \cdot g)$ and $(\Omega f) \cdot (\Omega g)$ are equal as pointed magma maps, where $\cdot$ denotes the pointwise operations. Because we are using weak magma maps, it is equivalent to show that these two maps are equal as pointed maps, or in other words that there is a pointed homotopy $\Omega(f \cdot g) \sim_\bullet (\Omega f) \cdot (\Omega g)$. The underlying homotopy involves some path algebra, and ultimately follows from the fact that horizontal and vertical composition agree in the codomain, which is a double-loop space. The pointedness of the homotopy follows by a simple path induction on the paths $f(y_0) = \mathsf{refl}$ and $g(y_0) = \mathsf{refl}$, after generalizing $f(y_0)$ and $g(y_0)$ to arbitrary loops. The argument in this paragraph has been formalized.

The naturality of Ω follows from the fact that for pointed maps h and k,

$$\Omega(h \circ k) = \Omega(h) \circ \Omega(k)$$

as pointed maps, where again we are taking advantage of the fact that we are using weak magma maps. □

Definition 2.26 For pointed types X, Y, and Z and natural numbers $n, m \geq 1$, we have maps

$$
\begin{aligned}
(X \to_\bullet Y \to_\bullet Z) &\longrightarrow (\Omega^n X \to_{\mathsf{Mgm}} \Omega^n (Y \to_\bullet Z)) \\
&\xrightarrow{\sim} (\Omega^n X \to_{\mathsf{Mgm}} (Y \to_\bullet \Omega^n Z)) \\
&\longrightarrow (\Omega^n X \to_{\mathsf{Mgm}} (\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^m \Omega^n Z)) \\
&\xrightarrow{\sim} (\Omega^n X \to_{\mathsf{Mgm}} (\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^{n+m} Z)) \\
&\longrightarrow (\pi_n(X) \to_{\mathsf{Grp}} \pi_m(Y) \to_{\mathsf{Grp}} \pi_{n+m}(Z)).
\end{aligned}
\tag{2-1}
$$

These maps are natural in X, Y, and Z. The first and third arrows apply Ω^n and Ω^m to morphisms, using Lemma 2.25. The second arrow is an equivalence by Lemma 2.24. To understand the fourth arrow, write $m = k+1$ for some $k : \mathbb{N}$. Then $\Omega^k \Omega^n Z = \Omega^{n+k} Z$ as pointed types. Applying Ω on the outside, we see that $\Omega^m \Omega^n Z = \Omega^{n+m} Z$ as magmas. Since the magma structure on the set of magma maps only uses that the iterated loop space is *merely* commutative and *merely* associative, we can conclude that $(\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^m \Omega^n Z) = (\Omega^m Y \to_{\mathsf{Mgm}_\bullet} \Omega^{n+m} Z)$ as magmas. From this we deduce the required equivalence. The fifth arrow applies 0–truncation on the inside and then on the outside. Let

$$\mathsf{smashing} \colon (X \to_\bullet Y \to_\bullet Z) \longrightarrow (\pi_n(X) \to_{\mathsf{Grp}} \pi_m(Y) \to_{\mathsf{Grp}} \pi_{n+m}(Z))$$

denote the composite.

The map smashing corresponds to the following construction in topology, which uses the smash product from the next section. Given a map $f : X \to_\bullet Y \to_\bullet Z$ and homotopy classes $\alpha : \pi_n(X)$ and $\beta : \pi_m(Y)$, one can smash representatives of the homotopy classes together to get an element $\alpha \wedge \beta : \pi_{n+m}(X \wedge Y)$. The adjoint $X \wedge Y \to_\bullet Z$ of f then induces a map taking this to an element of $\pi_{n+m}(Z)$ which (up to sign) is $\mathsf{smashing}(f, \alpha, \beta)$. This correspondence motivates the name.

Since we'll use it several times, we quote the following result from [4].

Lemma 2.27 [4, Corollary 4.3] *Let $m \geq 0$ and $n \geq -1$. If Y is a pointed, $(m-1)$–connected type and Z is a pointed, $(n+m)$–truncated type, then the type $Y \to_\bullet Z$ is n–truncated.*

The last result in this section plays an important role in our proof, and can be thought of as a generalization of Theorem 2.19 to functions with two arguments.

Theorem 2.28 *Let $n, m \geq 1$. If X is a pointed $(n-1)$–connected type, Y is a pointed $(m-1)$–connected type, and Z is a pointed $(n+m)$–truncated type, then the map* $\mathsf{smashing}$ *is an equivalence.*

Proof The first arrow in (2-1) is an equivalence by Lemma 2.27 and Proposition 2.23. The third arrow is an equivalence by Proposition 2.23. To show that the fifth arrow is an equivalence, one uses the same methods as in the proof of Proposition 2.23, using that $\Omega^{n+m} Z$ is a set. □

2.5 The connectivity of smash products

We recall some basic facts about smash products, and then prove a result about their connectivity.

Definition 2.29 For pointed types X and Y, the *smash product* $X \wedge Y$ is defined to be the higher inductive type with constructors:

- $\mathsf{sm}\colon X \times Y \to X \wedge Y$,
- $\mathsf{auxl} : X \wedge Y$,
- $\mathsf{auxr} : X \wedge Y$,
- $\mathsf{gluel} : \prod_{(y:Y)} \mathsf{sm}(x_0, y) = \mathsf{auxl}$,
- $\mathsf{gluer} : \prod_{(x:X)} \mathsf{sm}(x, y_0) = \mathsf{auxr}$.

The smash product is pointed by $\mathsf{sm}(x_0, y_0)$. It has the expected induction principle.

It is straightforward to see that the smash product is a functor. That is, given pointed maps $f\colon X \to_\bullet X'$ and $g\colon Y \to_\bullet Y'$ between pointed types, there is a pointed map $f \wedge g\colon X \wedge Y \to_\bullet X' \wedge Y'$ defined by induction on the smash product in the evident way, and this operation respects identity maps and composition.

Given pointed types X and Y, the constructors of the smash product $X \wedge Y$ combine to give a map $X \to_\bullet (Y \to_\bullet X \wedge Y)$, which we now describe.

Definition 2.30 Let $X, Y : \mathcal{U}_\bullet$. Currying the constructor sm, we get a map

$$X \to (Y \to X \wedge Y).$$

Using the constructor gluer twice, this map lifts to a map $X \to (Y \to_\bullet X \wedge Y)$. Similarly, using gluel, this last map lifts to a map $\mathsf{sm}_\bullet\colon X \to_\bullet (Y \to_\bullet X \wedge Y)$.

The following adjunction between pointed maps and smash products is fundamental to our work.

Lemma 2.31 [8, Theorem 4.3.28] *Let X, Y, and Z be pointed types. The map*

$$(X \wedge Y \to_\bullet Z) \longrightarrow_\bullet (X \to_\bullet (Y \to_\bullet Z))$$

induced by precomposition with $\mathsf{sm}_\bullet$ is a pointed equivalence, natural in X, Y, and Z. Here, we are interpreting $(- \wedge - \to_\bullet -)$ and $(- \to_\bullet (- \to_\bullet -))$ as 1–coherent functors $\mathsf{U}_\bullet^{\mathsf{op}} \times \mathsf{U}_\bullet^{\mathsf{op}} \times \mathsf{U}_\bullet \to \mathsf{U}_\bullet$.

Note that, by construction, $\mathsf{sm}_\bullet \colon X \to_\bullet Y \to_\bullet X \wedge Y$ is the adjunct of the identity map $X \wedge Y \to_\bullet X \wedge Y$.

In the form stated here, Lemma 2.31 has been formalized [8]. A stronger statement, which roughly involves regarding the category $\mathsf{U}_\bullet$ as being *enriched* over $\mathsf{U}_\bullet$, has not yet been proven, but we do not use this stronger form.

We now give a bound on the connectivity of smash products, proving the first part of Theorem S from the introduction.

Corollary 2.32 *Let $n, m \geq 0$, let X be a pointed, $(n-1)$–connected type, and let Y be a pointed, $(m-1)$–connected type. Then $X \wedge Y$ is $(n+m-1)$–connected.*

Proof It is enough to show that the truncation map $X \wedge Y \to \|X \wedge Y\|_{n+m-1}$ is nullhomotopic. Since the truncation map is pointed, this follows from the following more general fact: for any pointed, $(n+m-1)$–truncated type Z, the type $X \wedge Y \to_\bullet Z$ is contractible. Indeed, by Lemma 2.31, we have $(X \wedge Y \to_\bullet Z) \simeq (X \to_\bullet Y \to_\bullet Z)$. By Lemma 2.27, the type $Y \to_\bullet Z$ is $(n-1)$–truncated. Therefore, using Lemma 2.27 again, we see that the type $X \to_\bullet Y \to_\bullet Z$ is (-1)–truncated, and thus contractible, since any pointed mapping space is inhabited. □

2.6 Abelianization

In this section, we introduce the notion of abelianization, and give an efficient construction of the abelianization of a group.

Definition 2.33 Given a group G, an *abelianization* of G consists of an abelian group A together with a homomorphism $\eta \colon G \to_{\mathsf{Grp}} A$, initial among homomorphisms to abelian groups. In other words, for each abelian group B and homomorphism $h \colon G \to_{\mathsf{Grp}} B$, the type $\sum_{(f \colon A \to B)} h = f \circ \eta$ is contractible.

Since the type of abelianizations of a given group is a mere proposition, we abuse notation and denote any such abelianization by $G \to G^{\mathrm{ab}}$.

Remark 2.34 The existence of abelianizations can be proved in several different ways. One could mimic the classical definition, describing G^{ab} as the quotient of G by the subgroup generated by commutators, but this is awkward to work with constructively.

A second method that clearly works is to define G^{ab} as a higher inductive type with a point constructor $\eta\colon G \to G^{\mathrm{ab}}$, a point constructor giving G^{ab} an identity element, recursive point constructors giving addition and inverses in G^{ab}, recursive path constructors showing that the group laws hold and that the operation is abelian, a path constructor showing that η is a homomorphism, and a recursive path constructor forcing G^{ab} to be a set. While there is no doubt that this will work, it is difficult to use in practice because of the number of constructors and the fact that many of them are recursive.

A much simpler construction is as the higher inductive type with the constructors

- $\eta\colon G \to G^{\mathrm{ab}}$,
- $\mathsf{comm} : \prod_{a,b,c:G} \eta(a\cdot(b\cdot c)) = \eta(a\cdot(c\cdot b))$,
- $\mathsf{isset} : \prod_{x,y:G^{\mathrm{ab}}} \prod_{p,q:x=y} p = q$.

Equivalently, this is the 0–truncation of the coequalizer of the two obvious maps $G\times G\times G \to G$. Using either description, it is straightforward to show that G^{ab} has a unique group structure making η a group homomorphism, that this group structure is abelian, and that η satisfies the universal property. We don't give further details here, since this has been formalized by Ali Caglayan in the HoTT library [10].

Given a group homomorphism $f : G \to_{\mathsf{Grp}} H$, there is a unique group homomorphism $f^{\mathrm{ab}}\colon G^{\mathrm{ab}} \to_{\mathsf{Grp}} H^{\mathrm{ab}}$ making the square

$$\begin{tikzcd} G \arrow[r, "f"] \arrow[d, "\eta"'] & H \arrow[d, "\eta"] \\ G^{\mathrm{ab}} \arrow[r, "f^{\mathrm{ab}}"] & H^{\mathrm{ab}} \end{tikzcd}$$

commute. This makes abelianization into a functor and η into a natural transformation.

2.7 Tensor products

In this section, we define tensor products and use them to complete the proof of Theorem S.

Recall that for a group G and an abelian group H, the set $G \to_{\mathsf{Grp}} H$ is an abelian group. The group operation is given by $(\varphi + \psi)(g) :\equiv \varphi(g) + \psi(g)$, and the inverse by $(-\psi)(g) :\equiv -\psi(g)$, along with the natural proofs that these are homomorphisms.

Definition 2.35 Given abelian groups A and B, a *tensor product* of A and B consists of an abelian group T together with a map $t\colon A \to_{\mathsf{Grp}} B \to_{\mathsf{Grp}} T$ such that for any abelian group C the map

$$t^*\colon (T \to_{\mathsf{Grp}} C) \longrightarrow (A \to_{\mathsf{Grp}} B \to_{\mathsf{Grp}} C)$$

given by composition with t is an equivalence.

One can show in a straightforward way that tensor products exist, although we don't need this, and in fact the existence follows from Theorem 2.38. Moreover, the type of tensor products of a given pair of abelian groups is a mere proposition. We denote any such tensor product by $A \otimes B$. Given $a : A$, and $b : B$, we form the *elementary tensor* $a \otimes b : A \otimes B$ as $a \otimes b :\equiv t(a, b)$.

Example 2.36 Let $A : \mathsf{Ab}$. Then $A \simeq A \otimes \mathbb{Z}$, and the isomorphism is given by mapping $a : A$ to $a \otimes 1$. This follows from the fact that $\mathbb{Z}$ represents the identity; that is, $(\mathbb{Z} \to_{\mathsf{Grp}} C) \simeq_{\mathsf{Grp}} C$ for any $C : \mathsf{Ab}$, where the isomorphism is given by mapping $f : \mathbb{Z} \to_{\mathsf{Grp}} C$ to $f(1)$.

Lemma 2.37 *Let $A, B, C : \mathsf{Ab}$, and $\varphi, \psi : A \otimes B \to_{\mathsf{Grp}} C$. If for every $a : A$ and $b : B$ we have $\varphi(a \otimes b) = \psi(a \otimes b)$, then $\varphi = \psi$.*

Proof By construction, we have $\varphi(a \otimes b) = t^*(\varphi)(a, b)$ and $\psi(a \otimes b) = t^*(\psi)(a, b)$. By assumption and function extensionality, we have $t^*(\varphi) = t^*(\psi)$, and since t^* is an equivalence, we deduce that $\varphi = \psi$. □

A key step towards proving the Hurewicz theorem is constructing a map

$$\pi_n(X)^{\mathrm{ab}} \otimes \pi_m(Y)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_{n+m}(X \wedge Y)$$

natural in the pointed types X and Y, and proving that this map is an equivalence under connectivity assumptions on X and Y. Equivalently, we are looking for a map $\pi_n(X)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_m(Y)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_{n+m}(X \wedge Y)$ that is a tensor product under these assumptions.

In order to do this, observe that, for G and H groups and A an abelian group, we have an equivalence

$$(G^{\mathrm{ab}} \to_{\mathsf{Grp}} H^{\mathrm{ab}} \to_{\mathsf{Grp}} A) \xrightarrow{\sim} (G \to_{\mathsf{Grp}} H \to_{\mathsf{Grp}} A),$$

given by precomposition with the corresponding abelianization maps. Applying the smashing map from Definition 2.26 to the map $\mathsf{sm}_\bullet \colon X \to_\bullet Y \to_\bullet X \wedge Y$ from Definition 2.30 and using the above observation, we get a natural map

$$t_{X,Y} : \pi_n(X)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_m(Y)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_{n+m}(X \wedge Y).$$

Theorem 2.38 *Let $n, m \geq 1$, let X be a pointed, $(n-1)$–connected type, and let Y be a pointed, $(m-1)$–connected type. Then the map $t_{X,Y}$ exhibits $\pi_{n+m}(X \wedge Y)$ as the tensor product of $\pi_n(X)^{\mathrm{ab}}$ and $\pi_m(Y)^{\mathrm{ab}}$.*

This implies in particular that tensor products of abelian groups exist.

Proof Given an abelian group C, we must show that the map

$$t^*_{X,Y} : (\pi_{n+m}(X \wedge Y) \to_{\mathsf{Grp}} C) \longrightarrow (\pi_n(X)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_m(Y)^{\mathrm{ab}} \to_{\mathsf{Grp}} C).$$

is an equivalence. The following diagram will let us show that $t^*_{X,Y}$ is homotopic to a map that is easily proven to be an equivalence. Let $h \colon \pi_{n+m}(X \wedge Y) \to_{\mathsf{Grp}} C$ and consider the diagram

$$\begin{tikzcd}
 & (\pi_{n+m}(X \wedge Y) \to_{\mathsf{Grp}} C) \\
(X \wedge Y \to_\bullet X \wedge Y) \arrow[r, "h'_*"] \arrow[d, "\sim"] & (X \wedge Y \to_\bullet K(C, n+m)) \arrow[u, "\pi_{n+m}", "\sim"'] \arrow[d, "\sim"'] \\
(X \to_\bullet Y \to_\bullet X \wedge Y) \arrow[r, "h'_*"] \arrow[d, "\mathsf{smashing}"'] & (X \to_\bullet Y \to_\bullet K(C, n+m)) \arrow[d, "\mathsf{smashing}", "\sim"'] \\
(\pi_n(X) \to_{\mathsf{Grp}} \pi_m(Y) \to_{\mathsf{Grp}} \pi_{n+m}(X \wedge Y)) \arrow[r, "h_*"] & (\pi_n(X) \to_{\mathsf{Grp}} \pi_m(Y) \to_{\mathsf{Grp}} C) \\
(\pi_n(X)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_m(Y)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_{n+m}(X \wedge Y)) \arrow[r, "h_*"] \arrow[u, "\sim"] & (\pi_n(X)^{\mathrm{ab}} \to_{\mathsf{Grp}} \pi_m(Y)^{\mathrm{ab}} \to_{\mathsf{Grp}} C) \arrow[u, "\sim"']
\end{tikzcd}$$

We explain the diagram. The right-hand vertical arrow at the top is an equivalence by Corollary 2.32 and Proposition 2.23, and also implicitly uses a chosen equivalence $e : \pi_{n+m}(K(C, n+m)) \simeq C$. The unlabeled vertical arrows bordering the first square are the adjunction from Lemma 2.31. The vertical arrows labelled $\mathsf{smashing}$ are from Definition 2.26; the right-hand one uses e and is an equivalence by Theorem 2.28. The unlabeled vertical arrows at the bottom are from the universal property of abelianization. The horizontal maps labelled h_* are postcomposition by h. The horizontal maps labelled h'_* are postcomposition with the map $h' \colon X \wedge Y \to_\bullet K(C, n+m)$ which corresponds to h under the displayed equivalence π_{n+m}. It is straightforward to check that the three squares commute.

The right-hand column is an equivalence which we will show is homotopic to $t^*_{X,Y}$. Consider the identity map $\mathsf{id}_{X\wedge Y}$ at the top of the left-hand side. Its image in the bottom left corner is $t_{X,Y}$, and the image of $t_{X,Y}$ under h_* is equal to the image of h under $t^*_{X,Y}$. By definition of h', the image of $\mathsf{id}_{X\wedge Y}$ in the top-right corner is h. So the right-hand column sends h to $t^*_{X,Y}(h)$. That is, the composite vertical equivalence is homotopic to $t^*_{X,Y}$. □

2.8 Smash products, truncation, and suspension

The goal of this section is to prove a result about the interaction of smash products and truncation, and a result about the interaction of smash products and suspension. Both results make use of the symmetry of the smash product, so we begin with that.

Definition 2.39 Given pointed types X and Y, there is a pointed map

$$\tau \colon X \wedge Y \to_\bullet Y \wedge X$$

defined by induction on the smash product in the following way:

- $\tau(\mathsf{sm}(x, y)) :\equiv \mathsf{sm}(y, x)$,
- $\tau(\mathsf{auxl}) :\equiv \mathsf{auxr}$,
- $\tau(\mathsf{auxr}) :\equiv \mathsf{auxl}$,
- $\mathsf{ap}_\tau(\mathsf{gluel}\, y) := \mathsf{gluer}\, y$,
- $\mathsf{ap}_\tau(\mathsf{gluer}\, x) := \mathsf{gluel}\, x$.

It is pointed by $\mathsf{refl}_{\mathsf{sm}(y_0,x_0)}$.

Lemma 2.40 *For pointed types X and Y, the composite $\tau \circ \tau \colon X \wedge Y \to_\bullet X \wedge Y$ is pointed homotopic to the identity. In particular, the map τ is an equivalence.*

Proof We first show that for every $z : X \wedge Y$, $\tau(\tau(z)) = z$. We prove this using the induction principle for smash products. For the three point-constructors, this holds definitionally. The two 1–dimensional constructors are similar, so we only consider the first one. We must show that for each $y : Y$,

$$\mathsf{transport}^{z \mapsto \tau(\tau(z))=z}(\mathsf{gluel}\ y, \mathsf{refl}_{\mathsf{sm}(x_0,y)}) = \mathsf{refl}_{\mathsf{auxl}}.$$

By a calculation similar to those in [16, Section 2.11], the left-hand side is equal to

$$\mathsf{ap}_\tau(\mathsf{ap}_\tau(\mathsf{gluel}\ y))^{-1} \cdot \mathsf{refl}_{\mathsf{sm}(x_0,y)} \cdot \mathsf{gluel}\ y.$$

By the definition of τ in Definition 2.39, this is equal to

$$(\mathsf{gluel}\ y)^{-1} \cdot \mathsf{refl}_{\mathsf{sm}(x_0,y)} \cdot \mathsf{gluel}\ y,$$

which is equal to $\mathsf{refl}_{\mathsf{auxl}}$, as required.

We must also show that this homotopy is pointed. Up to definitional equality, this amounts to showing that $\mathsf{refl}_{\mathsf{sm}(x_0,y_0)} = \mathsf{refl}_{\mathsf{sm}(x_0,y_0)}$, which is true by reflexivity. □

Next we show that the map τ is natural.

Lemma 2.41 *Given pointed maps $f: X \to_\bullet X'$ and $g: Y \to_\bullet Y'$ between pointed types, the following square of pointed maps commutes*:

$$\begin{array}{ccc} X \wedge Y & \xrightarrow{f \wedge g} & X' \wedge Y' \\ {\scriptstyle\tau}\downarrow & & \downarrow{\scriptstyle\tau} \\ Y \wedge X & \xrightarrow{g \wedge f} & Y' \wedge X' \end{array}$$

Proof By path induction we can reduce to the case that $f(x_0) \equiv x_0'$ and $g(y_0) \equiv y_0'$. Next we use the induction principle for $X \wedge Y$. The square commutes definitionally on the three point constructors of $X \wedge Y$, but requires some straightforward path algebra in the remaining two cases. Since the proof has been formalized, we omit the details. □

Lemma 2.42 *Let $m \geq -1$, let $n \geq 0$, let Y be a pointed type, and let X be a pointed, $(n-1)$–connected type. Then the map*

$$\| |-|_m \wedge \mathsf{id}_X \|_{n+m} : \|Y \wedge X\|_{n+m} \to \big\| \|Y\|_m \wedge X \big\|_{n+m}$$

is an equivalence.

Proof Since the map in the statement is pointed, it is enough to show that for every pointed, $(n+m)$–truncated type T, precomposition with $|-|_m \wedge \mathsf{id}_Y$ induces an equivalence

$$(\|Y\|_m \wedge X \to_\bullet T) \longrightarrow (Y \wedge X \to_\bullet T).$$

By the naturality in the first variable of the adjunction from Lemma 2.31, it is enough to show that precomposition with $|-|_m$ induces an equivalence

$$(\|Y\|_m \to_\bullet X \to_\bullet T) \longrightarrow (Y \to_\bullet X \to_\bullet T),$$

and this follows from the fact that the type $X \to_\bullet T$ is m–truncated (Lemma 2.27). □

Corollary 2.43 *Let $m \geq -1$, let $n \geq 0$, let X be a pointed, $(n-1)$–connected type, and let Y be a pointed type. Then the map*

$$\|\mathsf{id}_X \wedge |-|_m\|_{n+m} \colon \|X \wedge Y\|_{n+m} \to \|X \wedge \|Y\|_m\|_{n+m}$$

is an equivalence.

Proof The square

$$\begin{tikzcd} X \wedge Y \arrow[r, "\mathsf{id}_X \wedge |-|_m"] \arrow[d, "\tau"'] & X \wedge \|Y\|_m \arrow[d, "\tau"] \\ Y \wedge X \arrow[r, "|-|_m \wedge \mathsf{id}_X"] & \|Y\|_m \wedge X \end{tikzcd}$$

commutes by Lemma 2.41. The vertical maps are equivalences by Lemma 2.40. By Lemma 2.42, the bottom map is an equivalence after $(n+m)$–truncation, so the top map must also be an equivalence after truncating. □

We conclude this section with a result letting us commute suspension and smash products.

Lemma 2.44 *Given pointed types X and Y, there is a pointed equivalence*

$$c_\Sigma \colon \Sigma(X \wedge Y) \simeq_\bullet X \wedge \Sigma Y,$$

natural in both X and Y.

Proof By Definition 2.39 and Lemmas 2.40 and 2.41, it is enough construct a natural equivalence $\Sigma(X \wedge Y) \simeq_\bullet \Sigma X \wedge Y$. In order to do this, it suffices to show that, for every pointed type Z, there is an equivalence $(\Sigma(X \wedge Y) \to_\bullet Z) \simeq (\Sigma X \wedge Y \to_\bullet Z)$ natural in X, Y, and Z, by the Yoneda lemma (Proposition 2.11). Given a pointed type Z, we define the equivalence as the composite of natural equivalences,

$$\begin{aligned} (\Sigma(X \wedge Y) \to_\bullet Z) &\simeq (X \wedge Y \to_\bullet \Omega Z) \\ &\simeq (X \to_\bullet Y \to_\bullet \Omega Z) \\ &\simeq (X \to_\bullet \Omega(Y \to_\bullet Z)) \\ &\simeq (\Sigma X \to_\bullet Y \to_\bullet Z) \\ &\simeq (\Sigma X \wedge Y \to_\bullet Z). \end{aligned}$$

The first and fourth equivalences follow from the adjunction between suspension and loops (Lemma 2.16). The second and fifth equivalences use Lemma 2.31. The third equivalence follows from Lemma 2.24. This concludes the proof. □

This result was formalized in the spectral repository [8], but the proof of naturality is not complete.

3 Homology and the Hurewicz theorem

In this section, we begin by defining homology and proving the Hurewicz theorem. Then we define the Hurewicz homomorphism and prove that it is unique up to sign. We conclude by giving some applications about the interaction between homology, connectedness, and truncation.

3.1 Prespectra and homology

In this section, we introduce prespectra as a tool for defining the homology groups of a type.

Definition 3.1 A *prespectrum* (Y, s) is a family of pointed types $Y : \mathbb{N} \to \mathcal{U}_\bullet$ and a family of pointed *structure maps* $s : \prod_{(n:\mathbb{N})} Y_n \to_\bullet \Omega Y_{n+1}$. When the structure maps of Y are clear from the context, we will denote the prespectrum simply by Y.

Definition 3.2 A *map of prespectra* $f : (T, s) \to (T', s')$ consists of a family of pointed maps $f : \prod_{(n:\mathbb{N})} Y_n \to_\bullet Y'_n$, and a family of pointed homotopies

$$\prod_{(n:\mathbb{N})} \Omega s'_n \circ f_n \sim_\bullet \Omega f_{n+1} \circ s_n.$$

Note that a prespectrum can be equivalently defined by giving a family of pointed types $Y : \mathbb{N} \to \mathcal{U}_\bullet$ and a family of pointed maps $\Sigma Y_n \to_\bullet Y_{n+1}$. This is the way that we will specify prespectra.

Example 3.3 Eilenberg–Mac Lane spaces are defined in homotopy type theory in [13]. Given an abelian group A, the *Eilenberg–Mac Lane prespectrum* HA of type A is given by the family $\lambda n.K(A, n)$ of pointed types, where we let $K(A, 0) :\equiv A$, pointed at 0. For $n \geq 1$, the structure map is

$$|-|_{n+1} : \Sigma K(A, n) \to \|\Sigma K(A, n)\|_{n+1} \equiv K(A, n+1).$$

When $n \equiv 0$, we define $\Sigma K(A, 0) \to K(A, 1)$ by induction on suspension, by mapping the north and south poles of $\Sigma K(A, 0)$ to the base point of $K(A, 1)$, and $\mathsf{merid}(a)$ to the loop of $K(A, 1)$ represented by a.

Definition 3.4 Given a pointed type X and a prespectrum (Y, s), we form a prespectrum $X \wedge Y$, called the *smash product* of X and Y, as follows. The type family is given by $(X \wedge Y)_n \equiv X \wedge Y_n$. The structure maps are given by the composite

$$\Sigma(X \wedge Y_n) \xrightarrow{c_\Sigma}_\bullet X \wedge \Sigma Y_n \xrightarrow{\mathsf{id}_X \wedge \bar{s}_n}_\bullet X \wedge Y_{n+1}, \tag{3-1}$$

where $\bar{s}_n : \Sigma Y_n \to_\bullet Y_{n+1}$ corresponds to $s_n : Y_n \to_\bullet \Omega Y_{n+1}$.

Note that, by the naturality of Lemma 2.44 and the functoriality of the smash product on pointed types, the smash product of a pointed type and a prespectrum is functorial.

Definition 3.5 The type of *sequential diagrams of groups* is the type

$$\mathsf{Grp}^{\mathbb{N}} :\equiv \sum_{A : \mathbb{N} \to \mathsf{Grp}} \prod_{n:\mathbb{N}} A_n \to_{\mathsf{Grp}} A_{n+1}.$$

Analogously, we define the type of *sequential diagrams of abelian groups*, which we denote by $\mathsf{Ab}^{\mathbb{N}}$.

The most important example in this paper is given by sequential diagrams of groups that come from prespectra.

Example 3.6 Let (Y, s) be a prespectrum and let $n, k : \mathbb{N}$. The map $s_k : Y_k \to_\bullet \Omega Y_{k+1}$ induces a morphism $\pi_n(s_k) : \pi_n(Y_k) \to_{\mathsf{Grp}} \pi_n(\Omega Y_{k+1}) \simeq \pi_{n+1}(Y_{k+1})$. Iterating this process, we get a sequential diagram of groups $\lambda i.\pi_{n+i}(Y_{k+i}) : \mathbb{N} \to \mathsf{Grp}$. We denote this diagram by $\mathcal{S}^n_k(Y)$. This construction is natural in Y.

Note that, if $n \geq 2$, the diagram $\mathcal{S}^n_k(Y)$ is a sequential diagram of abelian groups.

Definition 3.7 Let $(A, \varphi) : \mathcal{U}^{\mathbb{N}}$ be a sequential diagram of types. We define the *sequential colimit* of (A, φ), denoted by $\mathsf{colim}\, A : \mathcal{U}$, as the higher inductive type generated by the constructors $\iota : \prod_{n:\mathbb{N}} A_n \to \mathsf{colim}\, A$ and

$$\mathsf{glue} : \prod_{n:\mathbb{N}} \prod_{a:A_n} \iota_n(a) = \iota_{n+1}(\varphi_n(a)).$$

Lemma 3.8 *Let $(A, \varphi) : \mathsf{Ab}^{\mathbb{N}}$ be a sequential diagram of abelian groups. Then the sequential colimit $\mathsf{colim}\, A$ of the underlying sets is a set, and it has a canonical abelian group structure such that all of the induced maps $i_n : A_n \to \mathsf{colim}\, A$ are homomorphisms. Moreover, the abelian group $\mathsf{colim}\, A$ has the universal property of the colimit in the category of abelian groups.*

Proof The main difficulty is to show that $\mathsf{colim}\, A$ is 0–truncated. For this, we use [20, Corollary 7.7(1)], which says that a sequential colimit of n–truncated types is n–truncated.

To show that $\mathsf{colim}\, A$ has an abelian group structure we start by using induction to define the operation $+$ on $\mathsf{colim}\, A$. In the case of point constructors, we define $\iota_l(a)+\iota_n(b) :\equiv \iota_m(\varphi_l^m(a)+\varphi_n^m(b))$, where $m \equiv \max(l,n)$ and $\varphi_l^m \colon A_l \to A_m$ and $\varphi_n^m \colon A_n \to A_m$ are defined by iterating φ. The case of a path constructor glue and a point constructor is straightforward, and the case of two path constructors is immediate, since $\mathsf{colim}\, A$ is a set. The fact that, with these operation, $\mathsf{colim}\, A$ is an abelian group is clear.

The map $\iota_n \colon A_n \to \mathsf{colim}\, A$ is a group morphism for every n by construction, and the fact that $\mathsf{colim}\, A$ satisfies the universal property of the colimit follows from the induction principle of $\mathsf{colim}\, A$. □

Definition 3.9 Let Y be a prespectrum, let $n : \mathbb{Z}$, and let $j \equiv \max(0, 2-n)$. We define the n^{th} *stable homotopy group* of Y as

$$\pi_n^s(Y) :\equiv \mathsf{colim}\, \mathcal{S}_j^{n+j}(Y).$$

Note that the stable homotopy groups of a prespectrum are defined for any integer n, and not only for natural numbers. Moreover, by construction, the sequential diagram in the definition of $\pi_n^s(Y)$ is a sequential diagram of abelian groups, so stable homotopy groups are always abelian. As an aside, one can show that any $\mathcal{S}_j^{n+j}(Y)$ with $j \geq \max(0, 2-n)$ will have an isomorphic colimit. Finally, since the construction $\mathcal{S}_k^n(Y)$ is functorial in Y, stable homotopy groups are functorial in the prespectrum.

Definition 3.10 We define the n^{th} *reduced homology* of X with coefficients in Y as

$$\widetilde{H}_n(X;Y) :\equiv \pi_n^s(X \wedge Y).$$

We define the n^{th} (*ordinary*) *reduced homology* of X with coefficients in an abelian group A by

$$\widetilde{H}_n(X;A) :\equiv \widetilde{H}_n(X;HA).$$

Notice that these types carry an abelian group structure, given by the group structure of stable homotopy groups (Definition 3.9).

3.2 The Hurewicz theorem

In this section, we prove our main result, Theorem H. To do so, we first show that when X is sufficiently connected, we can compute $\widetilde{H}_n(X;A)$ without taking a colimit.

Lemma 3.11 *Let $n \geq 1$, let $A : \mathsf{Ab}$, and let X be a pointed, $(n-1)$–connected type. Then the natural homomorphism $\pi_{n+1}(X \wedge K(A,1)) \to \widetilde{H}_n(X;A)$ is an equivalence.*

Proof Recall that

$$\widetilde{H}_n(X;A) \equiv \pi_n^s(X \wedge HA) \equiv \operatorname{colim} \mathcal{S}_j^{n+j}(X \wedge HA),$$

for $j = \max(0, 2-n)$. Since $n \geq 1$, we must consider two cases, $n = 1$ and $n \geq 2$. When $n = 1$, we have $j = 1$, and the sequential diagram that defines $\widetilde{H}_n(X;A)$ starts as

$$\pi_{n+1}(X \wedge K(A,1)) \to \pi_{n+2}(X \wedge K(A,2)) \to \cdots.$$

When $n \geq 2$, we have $j = 0$, and the sequential diagram that defines $\widetilde{H}_n(X;A)$ starts as

$$\pi_n(X \wedge K(A,0)) \to \pi_{n+1}(X \wedge K(A,1)) \longrightarrow \pi_{n+2}(X \wedge K(A,2)) \to \cdots.$$

It suffices to show that in either case the morphism

$$\pi_{n+i}(X \wedge K(A,i)) \to \pi_{n+i+1}(X \wedge K(A,i+1))$$

is an equivalence for $i \geq 1$. To prove this, we use (3-1) to factor the map as

$$\pi_{n+i}(X \wedge K(A,i)) \to \pi_{n+i+1}(X \wedge \Sigma K(A,i)) \to \pi_{n+i+1}(X \wedge K(A,i+1)).$$

Now, the first of these two maps is induced by the Freundenthal map

$$X \wedge K(A,i) \to \Omega\Sigma(X \wedge K(A,i))$$

composed with the equivalence $\Sigma(X \wedge K(A,i)) \simeq (X \wedge \Sigma K(A,i))$. Notice that, by Corollary 2.32, $X \wedge K(A,i)$ is $(n+i-1)$–connected. If $i \geq 1$, we have that $n+i \geq 2$, and thus $(n+i-1)+1 \leq 2(n+i-1)$, so the Freudenthal suspension theorem [16, Theorem 8.6.4] implies that the map $\pi_{n+i}(X \wedge K(A,i)) \to \pi_{n+i+1}(X \wedge \Sigma K(A,i))$ is an equivalence.

The second map is an equivalence by Corollary 2.43, since $K(A,i+1) \equiv \|\Sigma K(A,i)\|_{i+1}$ by definition. $\square$

Theorem 3.12 (Hurewicz theorem) *Given an abelian group A, a natural number $n \geq 1$, and a pointed, $(n-1)$–connected type X, we have an isomorphism*

$$\pi_n(X)^{\mathrm{ab}} \otimes A \simeq_{\mathsf{Grp}} \widetilde{H}_n(X; A),$$

natural in X and A.

By naturality in X, we mean naturality with respect to pointed maps between $(n-1)$–connected types.

Proof By Lemma 3.11, it is enough to show that we have a natural isomorphism $\pi_{n+1}(X \wedge K(A,1)) \simeq_{\mathsf{Grp}} \pi_n(X)^{\mathrm{ab}} \otimes A$, and this follows directly from Theorem 2.38. □

3.3 The Hurewicz homomorphism

In this section we give a construction of the Hurewicz homomorphism and prove that it is unique up to sign.

Let X be a pointed type, A an abelian group, and $n \geq 1$. Applying $\widetilde{H}_n(-; A)$ to the $(n-1)$–connected cover map $X\langle n-1\rangle \to_\bullet X$ we obtain a morphism

$$\widetilde{H}_n(X\langle n-1\rangle; A) \to_{\mathsf{Grp}} \widetilde{H}_n(X; A),$$

natural in X and A. By Theorem 3.12, there is a natural isomorphism

$$\pi_n(X\langle n-1\rangle)^{\mathrm{ab}} \otimes A \simeq_{\mathsf{Grp}} \widetilde{H}_n(X\langle n-1\rangle; A).$$

Since $\pi_n(X\langle n-1\rangle) \to_{\mathsf{Grp}} \pi_n(X)$ is also a natural isomorphism, we can compose with the abelianization of its inverse to obtain a morphism $\pi_n(X)^{\mathrm{ab}} \otimes A \to_{\mathsf{Grp}} \widetilde{H}_n(X; A)$.

Definition 3.13 For every $X : \mathcal{U}_\bullet$, $A : \mathsf{Ab}$, and $n \geq 1$, the morphism

$$h_n : \pi_n(X)^{\mathrm{ab}} \otimes A \to \widetilde{H}_n(X; A)$$

described above is the *n^{th} Hurewicz homomorphism.*

By construction, when X is $(n-1)$–connected, h_n is an isomorphism.

Definition 3.14 Let $n \geq 1$. A morphism of *n–Hurewicz type* is given by a group homomorphism $\pi_n(X)^{\mathrm{ab}} \otimes A \to_{\mathsf{Grp}} \widetilde{H}_n(X; A)$ for each $X : \mathcal{U}_\bullet$ and $A : \mathsf{Ab}$, that is natural in both A and X, and that is an isomorphism when $X \equiv S^n$ and $A \equiv \mathbb{Z}$. Here we are regarding $\pi_n(-)^{\mathrm{ab}} \otimes -$ and $\widetilde{H}_n(-;-)$ as 1–coherent functors $\mathsf{U}_\bullet \times \mathsf{Ab} \to \mathsf{Ab}$.

Example 3.15 For any $n \geq 1$, the n^{th} Hurewicz homomorphism (Definition 3.13) is a morphism of n–Hurewicz type.

Theorem 3.16 *Let $n : \mathbb{N}$ and let F and G be morphisms of n–Hurewicz type. Then either $F(X, A) = G(X, A)$ or $F(X, A) = -G(X, A)$ for every pointed type X and abelian group A. The choice of sign is independent of X and A.*

Proof The morphisms $F(S^n, \mathbb{Z})$ and $G(S^n, \mathbb{Z})$ give us two isomorphisms between $\pi_n(S^n) \otimes \mathbb{Z}$ and $\widetilde{H}_n(S^n; \mathbb{Z})$. We now show that there are exactly two possible isomorphisms between $\pi_n(S^n) \otimes \mathbb{Z}$ and $\widetilde{H}_n(S^n; \mathbb{Z})$, and that these differ by a sign. On the one hand, by [14] (see also [16, Section 8.1]), we know $\pi_n(S^n) \simeq \mathbb{Z}$. On the other hand, we have $\mathbb{Z} \otimes \mathbb{Z} \simeq \mathbb{Z}$ (Example 2.36). So it is enough to show that there are exactly two isomorphisms between $\mathbb{Z}$ and $\mathbb{Z}$, and that they differ by a sign. This is straightforward, using the fact that if two integers n and m satisfy $n \times m = 1$, then $n = m = 1$ or $n = m = -1$, which follows from the fact that $\mathbb{Z}$ has decidable equality.

There are then two cases, $F(S^n, \mathbb{Z}) = G(S^n, \mathbb{Z})$ and $F(S^n, \mathbb{Z}) = -G(S^n, \mathbb{Z})$. We consider only the first case, the second one being analogous. We thus assume that $F(S^n, \mathbb{Z}) = G(S^n, \mathbb{Z})$ and we want to show that for every pointed type X and every abelian group A we have $F(X, A) = G(X, A)$.

By Lemma 2.37, it is enough to check that $F(X, A) = G(X, A)$ when evaluated on elementary tensors. Since the abelianization map is surjective and we are proving a proposition, it is enough to check this on elementary tensors $(\eta\alpha) \otimes \beta$ for $\alpha : \pi_n(X)$ and $\beta : A$. Since we are proving a mere proposition, we can assume that we have a pointed map $\bar{\alpha} : S^n \to_\bullet X$ representing α. Define $\bar{\beta} : \mathbb{Z} \to A$ by sending 1 to β. Consider the following diagram, which commutes by the naturality assumption:

$$\begin{tikzcd}
\pi_n(S^n)^{\mathrm{ab}} \otimes \mathbb{Z} \arrow[r, "F(S^n,\mathbb{Z})"] \arrow[d, "\pi_n(\bar{\alpha})^{\mathrm{ab}} \otimes \bar{\beta}"'] & \widetilde{H}_n(S^n; \mathbb{Z}) \arrow[d, "\widetilde{H}_n(\bar{\alpha}, \bar{\beta})"] \\
\pi_n(X)^{\mathrm{ab}} \otimes A \arrow[r, "F(X,A)"] & \widetilde{H}_n(X; A)
\end{tikzcd}$$

The commutativity of the diagram implies that

$$F(X, A)((\eta\alpha) \otimes \beta) = \widetilde{H}_n(\bar{\alpha}, \bar{\beta})(F(S^n, \mathbb{Z})(\theta \otimes 1)),$$

where $\theta : \pi_n(S^n)$ is represented by the identity map $S^n \to_\bullet S^n$. Similarly, we get that $G(X, A)((\eta\alpha) \otimes \beta) = \widetilde{H}_n(\bar{\alpha}, \bar{\beta})(G(S^n, \mathbb{Z})(\theta \otimes 1))$, and since $F(S^n, \mathbb{Z}) = G(S^n, \mathbb{Z})$, we conclude that $F(X, A)(\alpha \otimes \beta) = G(X, A)(\alpha \otimes \beta)$. $\square$

3.4 Applications

In this section, we give some consequences of the main results in the paper. We start with two immediate applications of the Hurewicz theorem.

Proposition 3.17 *Let $n \geq 1$, let X be a pointed, n–connected type, and let $A : \mathsf{Ab}$. Then $\widetilde{H}_i(X;A) = 0$ for all $i \leq n$. Conversely, if X is a pointed, connected type with abelian fundamental group such that $\widetilde{H}_i(X;\mathbb{Z}) = 0$ for all $i \leq n$, then X is n–connected.* □

Proposition 3.18 *Let $n \geq 1$ and let $A, B : \mathsf{Ab}$. Then $\widetilde{H}_n(K(A,n);B) \simeq A \otimes B$, and, in particular, $\widetilde{H}_n(K(A,n);\mathbb{Z}) \simeq A$.* □

The following result says that truncation does not affect low-dimensional homology.

Proposition 3.19 *Let X be a pointed type and let $m \geq n$ be natural numbers. For every abelian group $A : \mathsf{Ab}$, the truncation map $X \to \|X\|_m$ induces an isomorphism $\widetilde{H}_n(X;A) \xrightarrow{\sim} \widetilde{H}_n(\|X\|_m;A)$.*

Proof The objects in the sequential diagram that defines $\widetilde{H}_n(X;A)$ have the form $\pi_{n+i}(X \wedge K(A,i))$ for $i \geq \min(0, 2-n)$, and the morphism $\widetilde{H}_n(X;A) \to \widetilde{H}_n(\|X\|_m;A)$ is induced by levelwise morphisms $\pi_{n+i}(X \wedge K(A,i)) \to \pi_{n+i}(\|X\|_m \wedge K(A,i))$ given by the functoriality of π_{n+i} and the smash product. We will show that these levelwise morphisms are isomorphisms, which implies that the induced map is an isomorphism.

Consider the commutative square

$$\begin{array}{ccc} \pi_{n+i}(X \wedge K(A,i)) & \longrightarrow & \pi_{n+i}(\|X\|_m \wedge K(A,i)) \\ \downarrow & & \downarrow \\ \pi_{n+i}\big(\|X \wedge K(A,i)\|_{i+m}\big) & \longrightarrow & \pi_{n+i}\big(\big\|\,\|X\|_m \wedge K(A,i)\big\|_{i+m}\big) \end{array}$$

given by functoriality of $(i+m)$–truncation and π_{n+i}. It suffices to show that the bottom map and the vertical maps in the square are isomorphisms. The vertical maps are isomorphisms since $n+i \leq i+m$, and the bottom map is an isomorphism by Lemma 2.42. □

We conclude by showing that ∞–connected maps induce an isomorphism in all homology groups.

Corollary 3.20 *Let $f : X \to_\bullet Y$ be a pointed map between pointed types that induces an isomorphism in π_0 and an isomorphism in π_n for $n \geq 1$ and all choices of basepoint $x_0 : X$. Then f induces an isomorphism in all homology groups for all choices of coefficients.*

Proof Let $A : \mathsf{Ab}$ and let $n \geq 0$. We have a commutative square

$$\begin{tikzcd} \widetilde{H}_n(X; A) \arrow[r] \arrow[d] & \widetilde{H}_n(Y; A) \arrow[d] \\ \widetilde{H}_n(\|X\|_n; A) \arrow[r] & \widetilde{H}_n(\|Y\|_n; A) \end{tikzcd}$$

where the vertical maps are isomorphisms, by Proposition 3.19. The bottom map is induced by $\|f\|_n : \|X\|_n \to \|Y\|_n$, which is an equivalence, by the truncated Whitehead theorem [16, Theorem 8.8.3]. It follows that the top map is an isomorphism, concluding the proof. □

References

[1] **A Bauer**, **J Gross**, **P L Lumsdaine**, **M Shulman**, **M Sozeau**, **B Spitters**, *The HoTT library: A formalization of homotopy type theory in Coq*, from "Proceedings of the 6th ACM SIGPLAN Conference on Certified Programs and Proofs", ACM, New York (2017) 164–172

[2] **M de Boer**, *A proof and formalization of the initiality conjecture of dependent type theory*, licentiate thesis, Stockholm University (2020) `https://su.diva-portal.org/smash/get/diva2:1431287/FULLTEXT01.pdf`

[3] **M de Boer**, **G Brunerie**, *A formalized proof of a version of the initiality conjecture* (2020) `https://github.com/guillaumebrunerie/initiality`

[4] **U Buchholtz**, **F van Doorn**, **E Rijke**, *Higher groups in homotopy type theory*, from "LICS '18—33rd Annual ACM/IEEE Symposium on Logic in Computer Science", ACM, New York (2018) 205–214 MR Zbl

[5] **U Buchholtz**, **K-B Hou**, *Cellular cohomology in homotopy type theory*, from "LICS '18—33rd Annual ACM/IEEE Symposium on Logic in Computer Science", ACM, New York (2018) 521–529 MR Zbl

[6] **J D Christensen**, **A Caglayan**, *Hurewicz formalzation*, fork of [10] (2020) `https://github.com/jdchristensen/HoTT/tree/Hurewicz`

[7] **J D Christensen**, **M Opie**, **E Rijke**, **L Scoccola**, *Localization in homotopy type theory*, High. Struct. 4 (2020) 1–32 MR Zbl

[8] **F van Doorn**, *On the formalization of higher inductive types and synthetic homotopy theory*, PhD thesis, Carnegie Mellon University (2018) arXiv 1808.10690

[9] **R Graham**, *Synthetic homology in homotopy type theory*, preprint (2017) arXiv 1706.01540

[10] *The HoTT library*, a Coq library for homotopy type theory (2012) `https://github.com/HoTT/HoTT`

[11] **K Kapulkin**, **P L Lumsdaine**, *The homotopy theory of type theories*, Adv. Math. 337 (2018) 1–38 MR Zbl

[12] **K Kapulkin**, **P L Lumsdaine**, *The simplicial model of univalent foundations (after Voevodsky)*, J. Eur. Math. Soc. 23 (2021) 2071–2126 MR Zbl

[13] **D R Licata**, **E Finster**, *Eilenberg–Mac Lane spaces in homotopy type theory*, from "Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)", ACM, New York (2014) art. no. 66 MR Zbl

[14] **D R Licata**, **M Shulman**, *Calculating the fundamental group of the circle in homotopy type theory*, from "2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2013)", IEEE Computer Soc., Los Alamitos, CA (2013) 223–232 MR Zbl

[15] **P L Lumsdaine**, **M Shulman**, *Semantics of higher inductive types*, Math. Proc. Cambridge Philos. Soc. 169 (2020) 159–208 MR Zbl

[16] **The Univalent Foundations Program**, *Homotopy type theory: Univalent foundations of mathematics*, Institute for Advanced Study, Princeton, NJ (2013) MR Zbl

[17] **L Scoccola**, *Nilpotent types and fracture squares in homotopy type theory*, Math. Structures Comput. Sci. 30 (2020) 511–544 MR Zbl

[18] **M Shulman**, *Univalence for inverse diagrams and homotopy canonicity*, Math. Structures Comput. Sci. 25 (2015) 1203–1277 MR Zbl

[19] **M Shulman**, *All $(\infty, 1)$–toposes have strict univalent universes*, preprint (2019) arXiv 1904.07004

[20] **K Sojakova**, **F van Doorn**, **E Rijke**, *Sequential colimits in homotopy type theory*, from "Proceedings of the 35th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2020)", ACM, New York (2020) 845–858 MR Zbl

Department of Mathematics, University of Western Ontario
London, ON, Canada

Department of Mathematics, University of Western Ontario
London, ON, Canada

jdc@uwo.ca, lscoccol@uwo.ca

Received: 29 September 2020 Revised: 24 January 2022

msp

Algebraic & Geometric Topology **23**:5 (2023) 2141–2156
DOI: 10.2140/agt.2023.23.2141
Published: 25 July 2023

A concave holomorphic filling of an overtwisted contact 3–sphere

NAOHIKO KASUYA
DANIELE ZUDDAS

We prove that the closed 4–ball admits non-Kähler complex structures with strongly pseudoconcave boundary. Moreover, the induced contact structure on the boundary 3–sphere is overtwisted.

32V40; 32Q55, 57R17

1 Introduction

In [4], Antonio J Di Scala and the authors constructed a family of pairwise inequivalent complex surfaces $E = E(\rho_1, \rho_2)$ together with a holomorphic map $f\colon E \to \mathbb{CP}^1$ admitting compact fibers (the parameters ρ_1 and ρ_2 are such that $1 < \rho_2 < \rho_1^{-1}$). A relevant property of E is that it is diffeomorphic to $\mathbb{R}^4$, giving an extension to real dimension four of a result of Calabi and Eckmann [2].

The compact fibers of f were shown to be smooth elliptic curves and a singular rational curve with one node, and these are the only compact complex curves of E. The existence of embedded compact holomorphic curves implies the nonexistence of a compatible symplectic structure on E. Thus, the complex surface E is non-Kähler.

Further, in [5] we proved that E cannot be realized as a complex domain in any smooth compact complex surface.

In the present paper, we study the structure of E away from a compact subset by providing an exhausting family of embedded strongly pseudoconcave 3–spheres; see Proposition 4.1. From this we derive our main theorem. In order to state our results, we recall the notion of *Calabi–Eckmann type complex manifold* introduced in [4], which was inspired by the results of [2].

Definition 1.1 A complex manifold W is said to be of Calabi–Eckmann type if there exists a compact complex manifold X of positive dimension, and a holomorphic immersion $k\colon X \to W$ which is nullhomotopic as a continuous map.

We also recall the definition of strong pseudoconvexity and pseudoconcavity. Let (W, J) be a complex manifold with complex structure J and complex dimension ≥ 2, and let $M \subset W$ be a smooth real oriented hypersurface. Then, near every point $p \in M$ we can consider a local defining function for M, namely a smooth function $u\colon U \to \mathbb{R}$ defined in a certain open neighborhood U of p in W, such that u has no critical points and $M \cap U = u^{-1}(0)$ is the oriented boundary of the sublevel $u^{-1}(-\infty, 0]$. Moreover, M carries the complex tangencies distribution $\xi = TM \cap J(TM)$, which we assume to be endowed with the canonical complex orientation induced by J.

Definition 1.2 We say that a real oriented hypersurface $M \subset W$ is *strongly pseudoconvex* in W if there exists a strictly plurisubharmonic local defining function for M near every point $p \in M$, namely a defining function u whose complex Hessian Hu is positive definite. The oriented hypersurface M is said to be *strongly pseudoconcave* if it becomes strongly pseudoconvex by reversing its orientation.

In particular, we can consider complex manifolds with strongly pseudoconvex or pseudoconcave boundary. It is a standard fact that when $\dim_{\mathbb{C}} W$ is even, an oriented real hypersurface $M \subset W$ is strongly pseudoconvex (resp. pseudoconcave) if and only if the complex tangencies distribution on M is a positive (resp. negative) contact structure. Since we consider real 3–manifolds embedded in complex surfaces, we mainly refer to strong pseudoconvexity or pseudoconcavity by means of this characterizing property.

Main Theorem *The closed ball B^4 admits a Calabi–Eckmann type complex structure J with strongly pseudoconcave boundary. Moreover, the (negative) contact structure ξ determined on $\partial B^4 = S^3$ by the complex tangencies is overtwisted and homotopic as a plane field to the standard positive tight contact structure on S^3.*

In other words, (B^4, J) is a concave holomorphic filling of the overtwisted contact sphere (S^3, ξ). As far as the authors know, this is the first example of this sort in the literature.

This 4–ball arises as a smooth submanifold of E containing certain compact fibers of the map $f\colon E \to \mathbb{CP}^1$, and so it is evidently of Calabi–Eckmann type.

Our strategy for proving the theorem relies on finding a closed piecewise smooth 3–manifold $M \subset E$ supporting an open book decomposition whose pages are holomorphic annuli and whose monodromy is a left-handed (negative) Dehn twist about the core of the annulus; hence the underlying manifold M is homeomorphic to S^3.

Moreover, we prove that M can be approximated by a 1–parameter family of strongly pseudoconcave smoothly embedded 3–spheres $M_\tau \subset E$, for a suitable parameter $\tau \in (0,1)$. Namely, the complex domain outside the embedded 4–ball with corners $D \subset E$ bounded by M is foliated by strongly pseudoconcave 3–spheres. This implies the existence of a strictly plurisubharmonic function on $E - D$.

As a consequence, the open book decomposition of M is compatible with the contact structure of M_τ given by complex tangencies, which is then overtwisted. For the basics of the three-dimensional contact topology we use throughout the paper, the reader is referred, for example, to the book of Ozbagci and Stipsicz [15, Chapters 4 and 9].

Remark By Eliashberg's classification of overtwisted contact structures on closed oriented 3–manifolds [7], the negative contact structure in the main theorem is uniquely determined up to isotopy.

We point out that in all (odd) dimensions greater than three, a closed co-oriented overtwisted contact manifold (see Borman, Eliashberg and Murphy [1] for the definition) cannot be the strongly pseudoconcave boundary of a complex manifold. Indeed, such a holomorphic filling would give a strongly pseudoconvex CR structure on the contact manifold with reversed orientation. Thus, it can be filled by a Stein space — Rossi's theorem [17] — and therefore it can be filled by a Kähler manifold — Hironaka's theorem [9; 10] — which is impossible for an overtwisted contact manifold. In this sense, our result is particular to dimension three.

Lisca and Matić [13, Theorem 3.2] proved that any Stein filling W of a contact 3–manifold can be realized as a domain in a smooth complex projective surface S. Hence $S - \operatorname{Int} W$ is a concave holomorphic filling of a Stein fillable contact 3–manifold.

On the other hand, Eliashberg in [6] proved that for any closed contact 3–manifold (N, ξ), the 4–manifold $N \times [0,1]$ admits a complex structure such that the height function is strictly plurisubharmonic, providing a holomorphic cobordism of (N, ξ) with itself. However, its proof is not constructive.

Our result gives a rather explicit complex cobordism of an overtwisted contact 3–sphere with itself, by taking $\bigcup_{\tau \in [1/3, 1/2]} M_\tau \cong S^3 \times [0,1]$ as a complex domain in E.

Remark In [11] the authors prove that every closed contact 3–manifold can be filled as the strongly pseudoconcave boundary of a compact complex surface of Calabi–Eckmann type. We point out that this generalization depends on our main theorem.

The paper is organized as follows. In Section 2, we recall the construction of the complex surface E given in [4] and present a holomorphic model of the complement $C = E - \operatorname{Int} D$, which will be helpful for the proof of the main theorem, with D the 4–ball mentioned above. In Section 3, we construct a holomorphic open book decomposition embedded in E. Finally, in Section 4, we prove the main theorem by showing the existence of a strictly plurisubharmonic function near the embedded open book decomposition based on contact topology.

2 The complex surface *E*

In this section, we recall the construction of E, by sketching the original one in [4]. This will be helpful for the proof of our main theorem.

Throughout this paper we make use of the following notation:

$$\Delta(a,b) = \{z \in \mathbb{C} \mid a < |z| < b\},$$
$$\Delta[a,b] = \{z \in \mathbb{C} \mid a \le |z| \le b\},$$
$$\Delta(a) = \{z \in \mathbb{C} \mid |z| < a\},$$

and similarly with mixed brackets. We also denote the closed disk and the circle of radius a in $\mathbb{C}$ by $B^2(a)$ and $S^1(a)$, respectively. When $a = 1$, we drop it from the notation.

According to [4], the construction of $E = E(\rho_1, \rho_2)$ proceeds as follows. Let ρ_1 and ρ_2 be positive numbers such that $1 < \rho_2 < \rho_1^{-1}$, and choose ρ_0 such that $\rho_1\rho_2^{-1} < \rho_0 < \rho_1$.

We want to realize E as the union of two pieces. One of them is the product

$$V = \Delta(1, \rho_2) \times \Delta(\rho_0^{-1}),$$

and the other one is the total space W of a genus-1 holomorphic Lefschetz fibration $h\colon W \to \Delta(\rho_1)$ with only one singular fiber Σ.

In order to define the analytical gluing between V and W, we make use of the following Kodaira model [12]. Consider the elliptic fibration

$$(\mathbb{C}^* \times \Delta(0, \rho_1))/\mathbb{Z} \to \Delta(0, \rho_1),$$

defined by the canonical projection on the quotient space of $\mathbb{C}^* \times \Delta(0,\rho_1)$ with respect to the $\mathbb{Z}$–action given by $n \cdot (w_1, w_2) = (w_1 w_2^n, w_2)$. Then, it canonically extends to a singular elliptic fibration $h\colon W \to \Delta(\rho_1)$, and so we have an identification $W - \Sigma = (\mathbb{C}^* \times \Delta(0,\rho_1))/\mathbb{Z}$. The critical point of h is nondegenerate, namely the complex Hessian is of maximal rank, and so h is a genus-1 holomorphic Lefschetz fibration. In what follows, we shall keep the convention of denoting by (w_1, w_2) the usual complex coordinates of $\mathbb{C}^* \times \Delta(\rho_1) \subset \mathbb{C}^2$ when referring to W (up to the above identification), and by (z_1, z_2) the usual coordinates of $\mathbb{C}^2$ when referring to $V \subset \mathbb{C}^2$.

Now, let us consider the multivalued holomorphic function $\varphi\colon \Delta(0,\rho_1) \to \mathbb{C}^*$ defined by
$$\varphi(w) = \exp\Big(\frac{1}{4\pi i}(\log w)^2 - \frac{1}{2}\log w\Big).$$
We denote by $\Phi\colon U \to W$ the holomorphic map defined by
$$\Phi(z_1, z_2) = [(z_1\varphi(z_2^{-1}), z_2^{-1})],$$
where $U \subset \mathbb{C}^* \times \Delta(\rho_1^{-1}, \rho_0^{-1})$ is a certain open subset that will be specified later. Notice that Φ is single-valued. This depends on the fact that any two branches φ_1 and φ_2 of φ are related by the formula $\varphi_2(w) = w^k\varphi_1(w)$ for some $k \in \mathbb{Z}$, which is compatible with the above $\mathbb{Z}$–action. For the purpose of this section, we take $U = \Delta(1,\rho_2) \times \Delta(\rho_1^{-1}, \rho_0^{-1}) \subset V$.

It follows that Φ is a biholomorphism between $U \subset V$ and $\Phi(U) \subset W$.

We are now ready to holomorphically glue V and W by identifying the open subsets $U \subset V$ and $\Phi(U) \subset W$ by means of Φ. That is, we define the complex surface
$$E = E(\rho_1, \rho_2) = V \cup_\Phi W.$$
We consider V and W as open subsets of E via the quotient map.

By construction, there is a holomorphic map $f\colon E \to \mathbb{CP}^1$ defined by the canonical projection onto the second factor on V and by the elliptic fibration h on W, where $\mathbb{CP}^1$ is regarded as the result of gluing the disks $\Delta(\rho_0^{-1})$ and $\Delta(\rho_1)$ by identifying $\Delta(\rho_1^{-1}, \rho_0^{-1}) \subset \Delta(\rho_0^{-1})$ with $\Delta(\rho_0, \rho_1) \subset \Delta(\rho_1)$ by means of the inversion map $z \mapsto z^{-1}$.

Notice that the resulting complex surface E does not depend on ρ_0, since this parameter determines only the size of the gluing region.

Remark By taking ρ_1' and ρ_2' such that $\rho_2 < \rho_2' < (\rho_1')^{-1} < \rho_1^{-1}$, our construction yields an obvious holomorphic embedding of E in $E' = E(\rho_1', \rho_2')$ as a relatively compact complex domain. The closure $\widehat{E} = \mathrm{Cl}\, E$ in E' has Levi flat piecewise smooth boundary, and $\partial \widehat{E}$ is homeomorphic to S^3. This agrees with the interpretation of the map $f : E \to \mathbb{CP}^1$ given in [4] as the restriction of the Matsumoto–Fukaya fibration $S^4 \to S^2$ [14] to the complement of a neighborhood of the negative critical point in S^4. This also relates to the embedded open book decomposition that we construct in Proposition 3.1.

Let $V' = \Delta(1, s) \times \Delta(\rho_1^{-1}, \rho_0^{-1})$, where the additional parameter s is chosen so that $\rho_0^{-1} < s < \rho_1^{-1}\rho_2$. Let U' be the subset of V' defined by $U' = \{(z_1, z_2) \in V' \mid |z_2| < |z_1|\}$. We put $V'' = V \cup V' \subset \mathbb{C}^2$ and identify a point $(z_1, z_2) \in U'$ with $\psi(z_1, z_2)$, where $\psi : U' \to V'$ is the holomorphic embedding defined by $\psi(z_1, z_2) = (z_1 z_2^{-1}, z_2)$. Let $Y = V''/{\sim}$ be the quotient.

Proposition 2.1 *The manifold $Y = V''/{\sim}$ is biholomorphic to the preimage of the disk $\Delta(\rho_0^{-1}) \subset \mathbb{CP}^1$ by the holomorphic fibration $f : E \to \mathbb{CP}^1$.*

Proof The preimage $f^{-1}(\Delta(\rho_0^{-1}))$ is described as follows. Let $W(\rho_0, \rho_1)$ be the subset of W given, in the Kodaira model above, by

$$W(\rho_0, \rho_1) = (\mathbb{C}^* \times \Delta(\rho_0, \rho_1))/\mathbb{Z} = f^{-1}(\Delta(\rho_0, \rho_1)),$$

being $f = h$ in $W(\rho_0, \rho_1)$. Then, we have $U' \subset W(\rho_0, \rho_1)$, and so

$$f^{-1}(\Delta(\rho_0^{-1})) = V \cup_{U \sim U'} W(\rho_0, \rho_1).$$

Now, we define a map $\Psi : Y \to f^{-1}(\Delta(\rho_0^{-1}))$ by putting $\Psi([(z_1, z_2)]) = (z_1, z_2)$ on $V/{\sim}$ and $\Psi([(z_1, z_2)]) = \Phi(z_1, z_2)$ on $V'/{\sim}$. It is easy to check that Ψ is well defined and is a biholomorphism. □

In order to obtain the complement $C \subset E$ of a 4–ball D containing the singular fiber of f, we remove from Y the subset

$$Z = \{(z_1, z_2) \mid c_1 < |z_1| < c_2\} \subset V,$$

where $s\rho_1 < c_1 < c_2 < \rho_2$. Then, by Proposition 2.1, it is enough to set $C = Y - Z$.

3 The holomorphic open book decomposition

We briefly recall the notion of open book decomposition of a 3–manifold. For a more thorough treatment, the reader is referred to Ozbagci and Stipsicz [15, Chapter 9] and to Rolfsen [16, Chapter 10K].

By an *open book decomposition* of a closed, connected, oriented, manifold M of real dimension three, we mean a smooth map $f: M \to B^2$ such that

(1) the restriction $f|_{\mathrm{Cl}(f^{-1}(\mathrm{Int}\, B^2))}: \mathrm{Cl}(f^{-1}(\mathrm{Int}\, B^2)) \to B^2$ is a (trivial) fiber bundle with fiber a link $L = f^{-1}(0)$, called the *binding* of the open book;

(2) the map $\varphi\colon M - L \to S^1 = \partial B^2$ defined by $\varphi(x) = f(x)/|f(x)|$ is a fiber bundle.

The closure of every fiber $F_\theta = \mathrm{Cl}(\varphi^{-1}(\theta))$, for $\theta \in S^1$, is a compact surface in M, called a *page* of the open book, and $\partial F_\theta = L$. By a little abuse of terminology, we also call the surfaces $f^{-1}(\theta)$, for all $\theta \in S^1 = \partial B^2$, pages of f. The two kinds of pages are ambient isotopic in M to each other.

Given an open book decomposition $f: M \to B^2$, the orientations of M and of B^2 induce an orientation on the pages, and hence on the binding $L = \partial F_\theta$.

For an open book decomposition $f: M \to B^2$, there is an associated *monodromy* ω_f of the bundle φ, which is a diffeomorphism of a page F_* that fixes the boundary pointwise, and it is well defined up to isotopy fixing the boundary.

On the other hand, given an element ω of the mapping class group $\mathrm{Mod}_{g,b}$ of a compact, connected, oriented surface $F_{g,b}$ of genus $g \geq 0$ and with $b \geq 1$ boundary components, there is an open book decomposition $f_\omega\colon M_\omega \to B^2$ with monodromy ω and page $F = F_{g,b}$, and this is uniquely determined up to orientation-preserving diffeomorphisms. The construction goes as follows. Take a representative $\psi\colon F \to F$ of the isotopy class ω and consider the mapping torus $T_\omega = (F \times \mathbb{R})/\mathbb{Z}$, where the $\mathbb{Z}$–action is generated by the diffeomorphism $\tau\colon F \times \mathbb{R} \to F \times \mathbb{R}$ defined by $\tau(x,t) = (\psi(x), t-1)$.

Let M_ω be the result of gluing $\partial F \times B^2$ to T_ω along the boundary, by means of the obvious identifications $\partial(\partial F \times B^2) \cong \partial F \times S^1 \cong \partial F \times (\mathbb{R}/\mathbb{Z}) \cong \partial T_\omega$, where the last identification comes from the fact that ψ is the identity on ∂F. Then, let $f: M_\omega \to B^2$ be the canonical projection $\partial F \times B^2 \to B^2$ on $\partial F \times B^2 \subset M_\omega$, while it is the projection $T_\omega \to \mathbb{R}/\mathbb{Z} \cong \partial B^2$ on $T_\omega \subset M_\omega$.

Consider an oriented surface F and let $\gamma \subset \mathrm{Int}\, F$ be a connected simple closed curve. A *Dehn twist* $\delta_\gamma\colon F \to F$ about the curve γ is a diffeomorphism of F such that away from a tubular neighborhood T of γ in F, δ_γ is the identity, while in $T \cong S^1 \times [0,1]$ the diffeomorphism δ_γ either corresponds to the map $\delta_-\colon S^1 \times [0,1] \to S^1 \times [0,1]$ defined by

$$\delta_-(z,t) = (ze^{-2\pi i t}, t),$$

or to the map $\delta_+ = \delta_-^{-1}$, where $S^1 \times [0,1]$ is endowed with the product orientation and its identification with $T \subset F$ is orientation-preserving. In the former case, δ_γ is called a *left-handed* (*or negative*) *Dehn twist*, while in the latter it is called a *right-handed* (*or positive*) *Dehn twist*. By changing the orientation of F, the two types of Dehn twists are swapped.

The 3–sphere admits an open book decomposition $h_-\colon S^3 \to B^2$ with binding the negative Hopf link H_-, and with page the annulus $S^1 \times [0,1]$. The monodromy is the left-handed Dehn twist about the core circle $\gamma = S^1 \times \{\frac{1}{2}\}$ of the annulus (there is also the positive version $h_+\colon S^3 \to B^2$ of this). This is the well-known realization of the (negative) Hopf link in S^3 as a fibered link, with page the Hopf band [8].

The following proposition will be helpful in the proof of the main theorem. We keep the notation of Section 2.

Proposition 3.1 *There is a piecewise smooth embedded* 3*–sphere* $M \subset E$ *such that the restriction* $f|_M : M \to B^2$ *of the holomorphic map* $f : E \to \mathbb{CP}^1$, *is diffeomorphic to the open book decomposition* h_- *of* S^3 *described above, with* B^2 *a suitable closed disk in* $\Delta(\rho_0^{-1}) \subset \mathbb{CP}^1$. *Every page of* $f|_M$ *is a holomorphic annulus in an elliptic fiber of* f. *Moreover,* M *is not globally smooth, since it has corners along the two linked tori given by* $\partial f|_M^{-1}(\partial B^2)$, *on the complement of which* M *is foliated by holomorphic curves. Thus,* M *is Levi flat in* E.

We endow M with the orientation determined by the open book decomposition, where the pages are oriented by the induced complex structure, and the base disk B^2 takes the orientation from $\mathbb{CP}^1$. By construction, this disk is in the part of $\mathbb{CP}^1$ that corresponds, via the map f, to the Stein open subset $V \subset E$, with the boundary in the gluing region.

Fix two numbers c and ϵ such that $\rho_0 < c < \rho_1$ and

$$0 < \epsilon < \tfrac{1}{2}\min(\rho_1 - \rho_0, \rho_0 - \rho_1\rho_2^{-1}).$$

We put $a = \rho_2 - \epsilon$ and $b = c^{-1} + \epsilon$, and let $A = \Delta[a,b]$. It is then straightforward to check that $(\lambda^k A) \cap A = \varnothing$ for all $\lambda \in \Delta[c,\rho_1]$ and for all $k \in \mathbb{Z} - \{0\}$, with $\lambda^k A = \Delta[\,|\lambda|^k a, |\lambda|^k b\,]$. Moreover, by taking into account the inequalities among the ρ_i's at the beginning of Section 2, we can easily obtain

$$bc < 1 + \frac{c(\rho_1 - \rho_0)}{\rho_0\rho_1} < \rho_2. \tag{1}$$

Proof Consider the set

$$G = f^{-1}(S^1(c)) - \Phi(\Delta(bc,a) \times S^1(c^{-1})) \subset E,$$

with $S^1(c) \subset \Delta(\rho_1) \subset \mathbb{CP}^1$. The map $f_G = f|_G \colon G \to S^1(c) \cong S^1$ is a compact annulus bundle over the circle $S^1(c) \subset \Delta(\rho_1) \subset \mathbb{CP}^1$ of radius c. Here $S^1(c)$ has the clockwise orientation in the disk $\Delta(\rho_1)$, namely it is oriented as the boundary of the disk it bounds in $\Delta(\rho_0^{-1}) \subset \mathbb{CP}^1$. This choice depends on the inversion in the map Φ' below.

This bundle is trivial, and a trivialization is provided by the map $\Phi' \colon A \times S^1 \to G$ defined by

$$\Phi'(w_1, w_2) = \Phi(w_1, c^{-1}w_2) = [(w_1\varphi(cw_2^{-1}), cw_2^{-1})].$$

Notice that Φ' is holomorphic on every fiber.

Now, we construct an open book decomposition of S^3 embedded in E. We begin with an abstract description of this open book, and then we see how it is embedded in E.

Let ψ_1 be the identity map of $S^1(a) \times S^1$, and let

$$\psi_2 \colon S^1(b) \times S^1 \to S^1(b) \times S^1$$

be defined by $\psi_2(w_1, w_2) = (w_1w_2, w_2)$.

We use the diffeomorphism $\psi = \psi_1 \cup \psi_2 \colon \partial(\partial A \times B^2) \to \partial(A \times S^1)$ to construct the oriented 3–manifold

$$M = (\partial A \times B^2) \cup_\psi (A \times S^1)$$

obtained by gluing $\partial A \times B^2$ to $A \times S^1$ along the boundary (these two pieces are oriented in the canonical way).

Let $p \colon M \to B^2$ be defined by $p(w_1, w_2) = w_2$, for (w_1, w_2) in $\partial A \times B^2$ or $A \times S^1$. It is clear that (M, p) is an open book decomposition of M with binding

$$L = \partial A \times \{0\} \subset \partial A \times B^2 \subset M$$

and the annulus A as the page.

Now, we show that the monodromy of p is the diffeomorphism $\delta \colon A \to A$ defined by

$$\delta(z) = ze^{2\pi i \tau(|z|)},$$

where $\tau \colon [a, b] \to [0, 1]$ is an increasing diffeomorphism (for example, the affine one). Thus, δ is the identity on ∂A. Let

$$T(\delta) = \frac{A \times [0, 1]}{(z, 1) \sim (\delta(z), 0)}$$

be the mapping torus of δ.

The open book decomposition with page A and monodromy δ represents a 3–manifold $B(\delta)$ obtained by capping off $T(\delta)$ with $\partial A \times B^2$ glued along the boundary by the identity, up to the obvious identification $\partial B^2 = S^1 \cong [0,1]/(0 \sim 1)$.

Define the map $k\colon T(\delta) \to A \times S^1$ by setting

$$k([(z,t)]) = (ze^{2\pi i \tau(|z|)(t-1)}, e^{2\pi i t}).$$

Then, k is an orientation-preserving fibered diffeomorphism.

The gluing maps ψ_1 and ψ_2 used for building M correspond, by means of k, to the identity of $\partial(T(\delta)) = \partial A \times S^1$. This implies that there is a diffeomorphism $M \cong B(\delta)$, with respect to which the open book p corresponds to that of $B(\delta)$, and so δ is the monodromy of p.

In order to understand δ, we consider the diffeomorphism $q\colon A \to S^1 \times [0,1]$ defined by

$$q(z) = \left(\frac{\bar{z}}{|z|}, \tau(|z|)\right).$$

This is orientation-preserving, as it can be easily shown by writing q in polar coordinates. Moreover, $q^{-1}(w,t) = \tau^{-1}(t)\overline{w}$.

It is now straightforward to prove the identity $\delta_- = q \circ \delta \circ q^{-1}$, where δ_- is the left-handed Dehn twist defined above. Therefore, δ is a left-handed Dehn twist of A about the curve $\gamma \subset A$ of equation $\tau(|z|) = \frac{1}{2}$ (that is, the core of A). It follows that $p\colon M \to B^2$ is equivalent to the open book h_- of S^3, and in particular $M \cong S^3$.

Next, we define an embedding $g\colon M \to E$ by

$$g(z_1,z_2) = \begin{cases} \Phi'(z_1,z_2) & \text{for } (z_1,z_2) \in A \times S^1, \\ j(z_1, c^{-1}z_2) & \text{for } (z_1,z_2) \in S^1(a) \times B^2, \\ j(cz_1, c^{-1}z_2) & \text{for } (z_1,z_2) \in S^1(b) \times B^2, \end{cases}$$

where $j : V \hookrightarrow E$ is the inclusion map.

We show that g is well defined. For $(z_1,z_2) \in S^1(a) \times S^1$,

$$g(z_1,z_2) = j(z_1,c^{-1}z_2) = \Phi(z_1,c^{-1}z_2) = [(z_1\varphi(cz_2^{-1}), cz_2^{-1})] = (\Phi' \circ \psi_1)(z_1,z_2).$$

Finally, we check consistency at $(z_1,z_2) \in S^1(b) \times S^1$. First, $(z_1,z_2) \in S^1(b) \times B^2$ implies $(cz_1, c^{-1}z_2) \in V$ by inequality (1) above, so we can compute $j(cz_1, c^{-1}z_2)$.

We have

$$\begin{aligned} g(z_1, z_2) &= j(cz_1, c^{-1}z_2) \\ &= \Phi(cz_1, c^{-1}z_2) \\ &= [(cz_1\varphi(cz_2^{-1}), cz_2^{-1})] \\ &= [(z_1z_2\varphi(cz_2^{-1}), cz_2^{-1})] \\ &= (\Phi' \circ \psi_2)(z_1, z_2), \end{aligned}$$

where we are using the $\mathbb{Z}$–action considered in Section 2.

By abusing notation, we still denote by $M \subset E$ the image of g. Therefore, M is a piecewise smooth embedded submanifold of E, although it is not globally smooth. Indeed, the two codimension-0 submanifolds of $E \cong \mathbb{R}^4$ bounded by M have corners along $\partial A \times S^1 \subset M$. Away from the corners, M is foliated by holomorphic curves, and hence it is Levi flat. These holomorphic curves are the images of the disks $\{z_1\} \times B^2$ and the images of the annuli $A \times \{z_2\}$ by the embedding g, with $(z_1, z_2) \in \partial A \times S^1$. □

Let $D \subset E$ be the compact submanifold bounded by M, and let C be the noncompact one. Hence, $E = D \cup_M C$.

The argument based on Kirby calculus in [4] proves the following proposition.

Proposition 3.2 *Up to smoothing the corners, D is diffeomorphic to B^4 and C is diffeomorphic to $S^3 \times (0, 1]$.*

The same conclusion follows from the existence of a proper continuous function $u: C \to (0, 1]$, which is smooth, regular (namely, with no critical points) and strictly plurisubharmonic in $\operatorname{Int} C$. In the next section, we show the existence of such a function to prove our main theorem.

4 The proof of the main theorem

In this section we prove the following proposition and then prove our main theorem.

Proposition 4.1 *There exists a smooth 3–sphere $M_1 \subset E$ such that*

(1) *the noncompact submanifold $C_1 \subset E \cong \mathbb{R}^4$ bounded by M_1 admits a proper smooth regular strictly plurisubharmonic function $u: C_1 \to (0, 1]$;*

(2) *the complement $D_1 = E - \operatorname{Int} C_1$ is of Calabi–Eckmann type;*

(3) *M_1 is piecewise smoothly isotopic to M in E.*

Remark Property (3) of the above proposition and Proposition 3.2 imply that M_1 is smoothly standard in E, meaning that there exists a diffeomorphism $E \to \mathbb{R}^4$ mapping M_1 to the standard unit sphere S^3. Thus, $C_1 \cong S^3 \times (0,1]$ and $D_1 = E - \operatorname{Int} C_1 \cong B^4$.

Proposition 4.1 follows from the construction of C in Section 2 and the following well-known facts.

Lemma 4.2 *Let $U \subset \mathbb{C}$ be a nonempty open subset, and let $\psi\colon U \to \mathbb{R}$ be a smooth function. Let $\Omega = \{(z_1, z_2) \in U \times \mathbb{C} \mid |z_2| \le \exp(-\psi(z_1))\} \subset \mathbb{C}^2$. Then the following two conditions are equivalent:*

(1) *$\partial\Omega$ is strongly pseudoconvex (resp. pseudoconcave);*

(2) *ψ (resp. $-\psi$) is a strictly subharmonic function.*

Lemma 4.3 *Let c be a smooth regular curve in $\mathbb{R}^2$. Then the hypersurface*

$$M_c = \{(z_1, z_2) \mid (\log|z_1|, \log|z_2|) \in c\} \subset (\mathbb{C}^*)^2$$

is strongly pseudoconvex if and only if the plane curve c is strictly convex.

Now we construct a strongly pseudoconcave hypersurface M_1 which is a perturbation of the holomorphic open book M. We make use of Proposition 2.1 and of the notation established in Section 2.

Proof of Proposition 4.1 We construct a family $\{M_t\}_{t\in(0,1]}$ of smooth closed hypersurfaces in C as follows. First, for any $t \in (0,1]$ and a sufficiently small positive number δ, we take the two functions $f_t, g_t\colon [0, \rho_1^{-1}] \to (1, \rho_2)$ given by

$$f_t(x) = \log a + t\delta(1 + x^2), \quad g_t(x) = \log(bc) - t\delta(1 + x^2).$$

Recall that (z_1, z_2) are the coordinates on $V = \Delta(1, \rho_2) \times \Delta(\rho_0^{-1}) \subset \mathbb{C}^2$. We then define the hypersurfaces Q_t and R_t in V by $|z_1| = \exp(f_t(|z_2|))$ and $|z_1| = \exp(g_t(|z_2|))$, respectively. By orienting Q_t and R_t as the boundary components of the manifold

$$\{(z_1, z_2) \in V \mid \exp(g_t(|z_2|)) \le |z_1| \le \exp(f_t(|z_2|))\},$$

it turns out that they are both strongly pseudoconcave by Lemma 4.2. Now we retake the coordinates (w_1, w_2) on V' so that $(w_1, w_2) = (z_1, z_2^{-1})$. Then, near Q_t, the coordinate transformation between V and V' is $(w_1, w_2) = (z_1, z_2^{-1})$, and near R_t, it is $(w_1, w_2) = (z_1 z_2, z_2^{-1})$, by taking the embedding $\psi\colon U' \to V'$ into account;

see Section 2. Putting $u_j = \log|z_j|$ and $v_j = \log|w_j|$ for $j = 1, 2$, the coordinate transformation is given by

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \text{near } Q_t, \qquad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \text{near } R_t.$$

Then the defining equations of Q_t and R_t are

$$\begin{aligned} &u_1 = f_t(e^{u_2}) \text{ for } u_2 \le -\log\rho_1 \iff v_1 = \log a + t\delta(1+e^{-2v_2}) && \text{for } v_2 \ge \log\rho_1, \\ &u_1 = g_t(e^{u_2}) \text{ for } u_2 \le -\log\rho_1 \iff v_1 = \log(bc) - v_2 - t\delta(1+e^{-2v_2}) && \text{for } v_2 \ge \log\rho_1, \end{aligned}$$

respectively. Hence, they give plane curves in the $(v_1 v_2)$–plane, say c_{Q_t} and c_{R_t}. Then there exists a smooth family of strictly convex curves c_t satisfying:

(a) each curve c_t is contained in the trapezoid

$$\{(v_1, v_2) \mid v_1 > \log a,\ \log c < v_2 < \log\rho_1,\ v_1 + v_2 < \log(bc)\};$$

(b) c_t, c_{Q_t} and c_{R_t} are smoothly connected to be a regular curve;

(c) the family of curves foliates a subdomain of the trapezoid;

(d) as t goes to 0, the curve c_t piecewise smoothly converges to the polygonal line

$$\{(\log a, v_2) \mid \log c < v_2 < \log\rho_1\} \cup \{(v_1, \log c) \mid \log a \le v_1 \le \log b\} \\ \cup \{(v_1, v_2) \mid v_1 + v_2 = \log(bc),\ \log c < v_2 < \log\rho_1\}.$$

Now we define the hypersurface $S_t \subset V'$ by $S_t = \{(w_1, w_2) \mid (v_1, v_2) \in c_t\}$. Then it is strongly pseudoconvex with one orientation by Lemma 4.3, but with the natural orientation respecting those of Q_t and R_t, it is strongly pseudoconcave. Hence, the three pieces Q_t, R_t and S_t form a smooth closed strongly pseudoconcave hypersurface in Y, which we denote by M_t. Strictly speaking, R_t and the union $H_t = Q_t \cup S_t$ are hypersurfaces in $V'' = V \cup V'$. In the quotient $Y = V''/\sim$, they are glued together to form a smooth closed hypersurface M_t in Y. Since each piece is strongly pseudoconcave, so is M_t. Thus, M_t is a smooth closed strongly pseudoconcave hypersurface in Y. The equations defining Q_t and R_t above and condition (d) of c_t imply that M_t piecewise smoothly converges to M when t goes to 0. In particular, M_1 is a smooth strongly pseudoconcave 3–sphere and satisfies condition (3) of the statement.

Moreover, the smooth 3–sphere M_1 divides the complex manifold E into the two submanifolds, the compact one D_1, which is a closed 4–ball, and the noncompact one C_1. Then condition (2) is automatically fulfilled because D_1 is contractible and contains the singular rational curve of E.

By a similar construction as that of the family $\{M_t\}_{t\in(0,1]}$, we can easily prove that $\operatorname{Int} C_1$ is foliated by a family of strongly pseudoconcave 3–spheres $\{M_t\}_{t\in(1,2)}$. Therefore, the following lemma, which proves the existence of a strictly plurisubharmonic function, concludes the proof. □

Lemma 4.4 *Let $\gamma\colon X \to \mathbb{R}$ be a proper smooth regular function on a complex manifold X such that the complex tangencies define a contact structure on the level sets $\gamma^{-1}(c)$ for all $c \in \gamma(X)$. Then there exists a smooth convex and increasing function $g\colon \gamma(X) \to \mathbb{R}$ such that $g \circ \gamma$ is strictly plurisubharmonic on X.*

Proof See for example [3, Lemma 2.7]. □

Proof of Main Theorem Endow B^4 with the complex structure J induced by an orientation-preserving diffeomorphism $B^4 \cong D_1$, the 4–ball in E bounded by M_1. Then (B^4, J) is of Calabi–Eckmann type and with strongly pseudoconcave boundary (S^3, ξ), where ξ is the induced contact structure.

Since J is homotopic, through almost complex structures, to the standard complex structure of $B^4 \subset \mathbb{C}^2$, the boundary contact structure ξ is homotopic as a plane field to the standard positive tight contact structure of S^3.

We are left to show the compatibility of the contact structure on $M_1 \cong S^3$ with the open book decomposition inherited from M by a suitable diffeomorphism $\varphi\colon M \to M_1$ compatible with the splitting $M = (\partial A \times B^2) \cup_\psi (A \times S^1)$ of the definition of M in Section 3, and the splitting $M_1 = Q_1 \cup R_1 \cup S_1$ above; that is, $\varphi(\partial A \times B^2) = Q_1 \cup R_1$ and $\varphi(A \times S^1) = S_1$. We want to prove that the contact form α is positive on the binding (oriented as the boundary of a page) and that $d\alpha$ is a volume form on the pages (oriented as holomorphic curves of E) of the open book decomposition; see [15, Section 9.2].

Since u is strictly plurisubharmonic on C_1, the 1–form $\alpha = -d^{\mathbb{C}} u$ is a contact form on each level set of u, and the 2–form $d\alpha$ defines a symplectic structure compatible with the complex structure J. The contactness of M_1 is equivalent to the fact that the restriction $(\alpha \wedge d\alpha)|_{TM_1}$ is a volume form. On the other hand, the open book decomposition of M_1 is given by the function

$$\varphi\colon M_1 - L \to S^1, \quad \varphi(z_1, z_2) = \frac{z_2}{|z_2|},$$

where $L \subset M_1$ is the link of equation $z_2 = 0$. The vector $\partial/\partial\theta_1$ is tangent to the binding and the tangent space of the page is spanned by $\partial/\partial\theta_1$ and V, where V is the tangent vector of the curve $\{(e^{v_1}, e^{v_2}) \mid (v_1, v_2) \in c_1\}$. Notice that the binding consists of two components $L_1 = \{(z_1, z_2) \in Q_1 \mid z_2 = 0\}$ and $L_2 = \{(z_1, z_2) \in R_1 \mid z_2 = 0\}$, which are naturally oriented by $-\partial/\partial\theta_1$ and $\partial/\partial\theta_1$, respectively.

Now, we check the compatibility. Since the partial derivative $\partial u/\partial r_1$ is negative near L_1 and positive near L_2,

$$\begin{aligned}\alpha\Big(-\frac{\partial}{\partial\theta_1}\Big)_{r_1=d_1,z_2=0} &= d^{\mathbb{C}}u\Big(\frac{\partial}{\partial\theta_1}\Big)_{r_1=d_1,z_2=0} &&= -r_1\Big(\frac{\partial u}{\partial r_1}\Big)_{r_1=d_1,z_2=0} > 0,\\ \alpha\Big(\frac{\partial}{\partial\theta_1}\Big)_{r_1=d_2,z_2=0} &= -d^{\mathbb{C}}u\Big(\frac{\partial}{\partial\theta_1}\Big)_{r_1=d_2,z_2=0} &&= r_1\Big(\frac{\partial u}{\partial r_1}\Big)_{r_1=d_2,z_2=0} > 0,\end{aligned}$$

which imply the positivity of α along the binding.

In order to see that $d\alpha$ is a volume form on the pages, it is enough to show that the vectors $\partial/\partial\theta_1$, V and R span the tangent space of M_1, where

$$R = J\left(\frac{\nabla u}{\|\nabla u\|}\right)$$

is the Reeb vector field of the contact form $\alpha|_{TM_1}$. Since the r_2 component of the gradient vector is positive except on the binding, so is the θ_2 component of R. Therefore, the three vectors indeed span the tangent space except on the binding. □

Acknowledgements Kasuya is supported by JSPS KAKENHI Grant JP17K14193.

Zuddas acknowledges support of the 2013 ERC Advanced Research Grant 340258 TADMICAMT. He is member of the group GNSAGA of Istituto Nazionale di Alta Matematica "Francesco Severi", Italy.

The authors are indebted to the referee for helpful comments and suggestions.

References

[1] **M S Borman**, **Y Eliashberg**, **E Murphy**, *Existence and classification of overtwisted contact structures in all dimensions*, Acta Math. 215 (2015) 281–361 MR Zbl

[2] **E Calabi**, **B Eckmann**, *A class of compact, complex manifolds which are not algebraic*, Ann. of Math. 58 (1953) 494–500 MR Zbl

[3] **K Cieliebak**, **Y Eliashberg**, *From Stein to Weinstein and back: Symplectic geometry of affine complex manifolds*, American Mathematical Society Colloquium Publications 59, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[4] **A J Di Scala**, **N Kasuya**, **D Zuddas**, *Non-Kähler complex structures on $\mathbb{R}^4$*, Geom. Topol. 21 (2017) 2461–2473 MR Zbl

[5] **A J Di Scala**, **N Kasuya**, **D Zuddas**, *Non-Kähler complex structures on $\mathbb{R}^4$, II*, J. Symplectic Geom. 16 (2018) 631–644 MR Zbl

[6] **Y M Eliashberg**, *Complexification of contact structures on 3–dimensional manifolds*, Uspekhi Mat. Nauk 40 (1985) 161–162 MR Zbl In Russian; translated in Russian Math. Surveys 40 (1985) 123–124

[7] **Y Eliashberg**, *Classification of overtwisted contact structures on 3–manifolds*, Invent. Math. 98 (1989) 623–637 MR Zbl

[8] **J Harer**, *How to construct all fibered knots and links*, Topology 21 (1982) 263–280 MR Zbl

[9] **H Hironaka**, *Resolution of singularities of an algebraic variety over a field of characteristic zero, I*, Ann. of Math. 79 (1964) 109–203 MR Zbl

[10] **H Hironaka**, *Resolution of singularities of an algebraic variety over a field of characteristic zero, II*, Ann. of Math. 79 (1964) 205–326 MR Zbl

[11] **N Kasuya**, **D Zuddas**, *On the strongly pseudoconcave boundary of a compact complex surface*, preprint (2020) arXiv 2012.03018

[12] **K Kodaira**, *On compact analytic surfaces, II*, Ann. of Math. 77 (1963) 563–626 MR Zbl

[13] **P Lisca**, **G Matić**, *Tight contact structures and Seiberg–Witten invariants*, Invent. Math. 129 (1997) 509–525 MR Zbl

[14] **Y Matsumoto**, *On 4–manifolds fibered by tori*, Proc. Japan Acad. Ser. A Math. Sci. 58 (1982) 298–301 MR Zbl

[15] **B Ozbagci**, **A I Stipsicz**, *Surgery on contact 3–manifolds and Stein surfaces*, Bolyai Society Mathematical Studies 13, Springer (2004) MR Zbl

[16] **D Rolfsen**, *Knots and links*, Mathematics Lecture Series 7, Publish or Perish, Houston, TX (1990) MR Zbl

[17] **H Rossi**, *Attaching analytic spaces to an analytic space along a pseudoconcave boundary*, from "Proc. Conf. Complex Analysis" (A Aeppli, E Calabi, H Röhrl, editors), Springer (1965) 242–256 MR Zbl

Department of Mathematics, Hokkaido University
Sapporo, Japan

Dipartimento di Matematica e Geoscienze, University of Trieste
Trieste, Italy

nkasuya@math.sci.hokudai.ac.jp, dzuddas@units.it

https://www.danielezuddas.eu

Received: 4 October 2020 Revised: 14 October 2021

Algebraic & Geometric Topology **23**:5 (2023) 2157–2190
DOI: 10.2140/agt.2023.23.2157
Published: 25 July 2023

Modifications preserving hyperbolicity of link complements

COLIN ADAMS
WILLIAM H MEEKS III
ÁLVARO K RAMOS

Given a link in a 3–manifold such that the complement is hyperbolic, we provide two modifications to the link, called the chain move and the switch move, that preserve hyperbolicity of the complement, with only a relatively small number of manifold-link pair exceptions, which are also classified. These modifications provide a substantial increase in the number of known hyperbolic links in the 3–sphere and other 3–manifolds.

57K10, 57K32

1 Introduction

Thurston proved that every knot in the 3–sphere S^3 is either a torus knot, a satellite knot or a hyperbolic knot; by which we mean that its complement in S^3 admits a complete hyperbolic metric. By the Mostow–Prasad rigidity theorem, the complement of a hyperbolic knot in S^3 has a unique hyperbolic metric, which must have finite volume; hence, a hyperbolic knot in S^3 has associated to it a well-defined set of hyperbolic invariants such as volume, cusp volume, cusp shape, etc. More generally, Thurston proved that a link in a closed, orientable 3–manifold has hyperbolic complement (necessarily of finite volume) if and only if the exterior of the link contains no properly embedded essential disks, spheres, tori or annuli — terms that are described in Definition 2.1.

One would like to be able to identify link complements that satisfy Thurston's criteria, and that therefore possess a hyperbolic metric. In [12], Menasco proved that every non-2–braid prime alternating link in S^3 is hyperbolic. In [2], Adams extended this result to augmented alternating links, where additional nonparallel trivial components wrapping around two adjacent strands in the alternating projection were added to the link. These additional components bound twice-punctured disks, which are totally geodesic in the hyperbolic structure of the complement. By Adams [1], the link complement can be

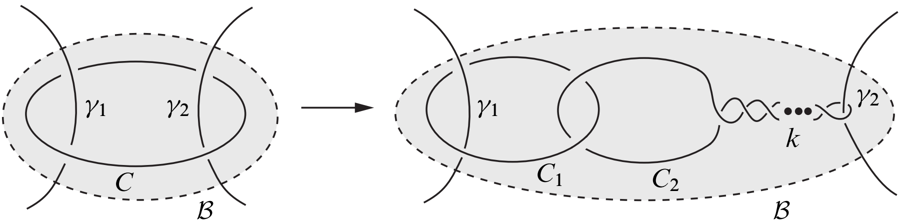


Figure 1: Replacing the left with the right preserves hyperbolicity of the complement.

cut open along such a twice-punctured disk, twisted a half-twist and reglued to obtain another hyperbolic link complement, with identical volume. This operation adds one crossing to the link projection. In many hyperbolic link complements, twice-punctured disks are particularly useful, because they are totally geodesic; see for instance the survey article by Purcell [15] and the references therein.

We consider two moves that one can perform on a link in a 3–manifold with hyperbolic complement. The first move we consider is called the *chain move*. Here, we start with a trivial component bounding a twice-punctured disk in a ball $\mathcal{B}$ as in Figure 1, and we replace the tangle on the left with the tangle on the right in Figure 1, where k is any integer. Assuming that the rest of the manifold outside $\mathcal{B}$ is not the complement of a rational tangle in a 3–ball (see Adams [4, Chapter 2] for this definition), the result is hyperbolic.

There are counterexamples to extending the result to the case where the manifold outside $\mathcal{B}$ is a rational tangle complement in a 3–ball, as demonstrated by the hyperbolic link in the 3–sphere appearing in Figure 2. When the chain move is applied with $k = 3$, the resultant 3–component link is 6^3_3 in Alexander–Rolfsen notation, which is not hyperbolic. However, in Lemma 3.4, we delineate explicitly the only possible exceptions.

The second move is called the *switch move*. Suppose we have a 3–manifold M and a link L in M with hyperbolic complement. Let α be an embedded arc that runs from L to L with interior that is isotopic to an embedded geodesic in the complement, as in Figure 3.

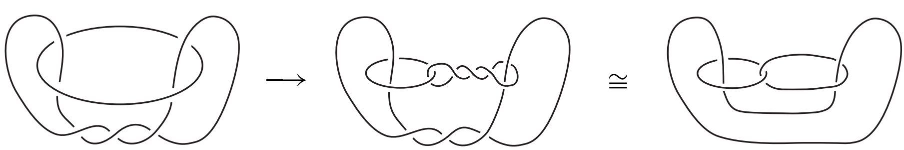


Figure 2: Applying the chain move to this hyperbolic link with $k = 3$ yields the nonhyperbolic link complement 6^3_3.

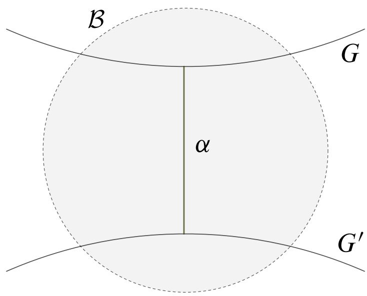


Figure 3: The trace of a geodesic α of $(M \setminus L, h)$ connects one or two components of L to one another, and a neighborhood $\mathcal{B}$ of α intersects L in two arcs.

Such a geodesic always exists since we could take one with minimal length outside fixed cusp boundaries. We consider the possibility that the arc runs from one component of L back to the same component or from one component to a second component. Let $\mathcal{B}$ be a neighborhood of α. Then $\mathcal{B}$ intersects L in two arcs, as in Figure 4, left. The switch move allows us to surger the link and add in a trivial component as in Figure 4, right, while preserving hyperbolicity.

Remark 1.1 The projection depicted in Figure 3 is not well defined, since if the two arcs are skew inside the ball, there are two different projections, depending on point of view. So in fact, for each such geodesic α, there are two switch moves possible. This is equivalent to cutting along the twice-punctured disk D bounded by C and twisting a half-twist in either direction before regluing. Once we prove that the switch move depicted in Figure 4 preserves hyperbolicity, the hyperbolicity of the half-twisted

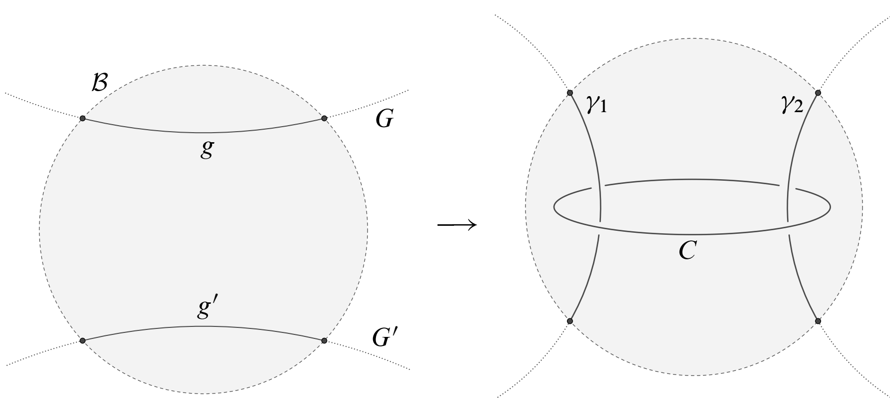


Figure 4: The switch move replaces the arcs g and g' by the tangle $\gamma_1 \cup \gamma_2 \cup C$.

version follows immediately from the previously mentioned results of [1], and the volumes of the resulting manifolds are identical. Further twists give link complements homeomorphic to the original or the half-twisted version.

These moves show that many additional link complements in 3–manifolds are hyperbolic. The authors [6] used the chain move and the switch move, together with the related *switch move gluing operation* described in Section 5, in the proof that for any given surface S of finite topology and negative Euler characteristic and any $H \in [0, 1)$, there exists a proper, totally umbilic embedding of S into some hyperbolic 3–manifold of finite volume with image surface having constant mean curvature H.

Moreover, Adams, Eisenberg, Greenberg, Kapoor, Liang, O'Connor, Pacheco-Tallaj and Wang [5] used the chain move in the proof that a virtual link obtained by taking a reduced classical prime alternating link projection and changing one crossing to be virtual yields a nonclassical virtual link.

We can also use the chain move and the switch move to obtain straightforward proofs of hyperbolicity of well-known classes of links.

Example 1.2 We can show that every chain link of five or more components, no matter how twisted, is hyperbolic. This was first proved by Oertel [14] (or see Neumann and Reid [13] for a proof using explicit hyperbolic structures for manifolds covered by these link complements).

Start with the alternating 4–chain, known to be hyperbolic by Menasco's work in [12]. Then apply the chain move repeatedly. This proves hyperbolicity of any chain link of five or more components with an arbitrary amount of twisting in the chain.

We note that the chain and switch moves apply more broadly than is apparent from Figures 1 and 4. In the case of the chain move, instead of specifying a hyperbolic link complement $M \setminus L$, we can start with a cusped hyperbolic 3–manifold M' containing a two-sided essential embedded thrice-punctured sphere S. Treating two of the boundary curves on the cusps as the meridional punctures of the disk in Figure 4 and the third as the longitudinal boundary of the disk, we can apply the chain move, removing the cusp that contains the longitude by doing a Dehn filling along a curve that crosses the longitude once and adding in the additional two components within a neighborhood of S. In the case that two of the boundaries of S are on the same cusp, they must play the role of the meridional punctures. (Note that if a two-sided thrice-punctured sphere has all three boundaries on the same cusp, no move is possible.)

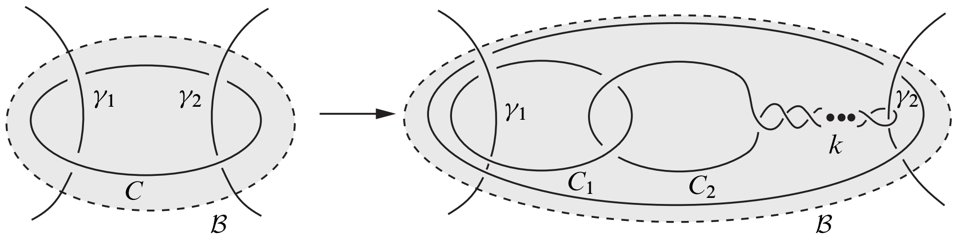


Figure 5: The augmented chain move.

In the case of the switch move, we can again begin with a cusped hyperbolic 3–manifold M'. For two cusps connected by an embedded geodesic, we can choose a nontrivial simple closed curve on each torus corresponding to each cusp. Then by Dehn filling along those curves we obtain a 3–manifold M for which M' is a link complement and the switch move applies.

The same procedure holds for a geodesic from a cusp back to the same cusp, and a specification of a nontrivial simple closed curve on the torus corresponding to the cusp, two copies of which play the role of the meridians around γ_1 and γ_2. Note that when applied to a link complement $M \setminus L$, but with a choice of curve other than meridians, the end result is not a new link complement in the same manifold.

Finally, we point out that there is a variant of the chain move called the *augmented chain move* as in Figure 5 wherein the two new components of the chain move are added in but the previous trivial component is not removed. We prove here that this move also preserves hyperbolicity.

To see this, we consider the link appearing in Figure 6, which is a twisted five-chain.

All five-chains are hyperbolic, as we just proved, so it has a hyperbolic complement. Now we apply the idea of a walnut as in [3]. We can cut the manifold M open along the twice-punctured disk E bounded by C, cut the 5–chain link complement open along the twice punctured disk bounding the bottom component in Figure 6 and then

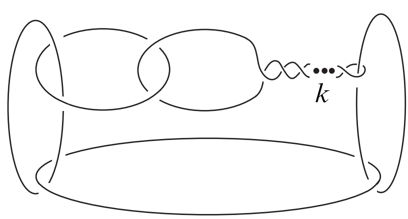


Figure 6: All 5–chains are hyperbolic.

glue copies of the twice-punctured disks to one another to insert the cut-open link complement into M. As in [1], since a twice-punctured disk is totally geodesic with a rigid unique hyperbolic structure, the gluings are isometries and the resulting manifold is hyperbolic with volume the sum of the volumes of the two manifolds.

Next, we explain the organization of the paper. First, we remark that it suffices to demonstrate our results when the ambient manifold is orientable. This property is proved by showing that the oriented cover of a related nonorientable link complement admits a hyperbolic metric and then one applies the Mostow–Prasad rigidity theorem to conclude that the associated order-two covering transformation is an isometry, which in turn implies that the hyperbolic metric on the oriented covering descends. In Section 2, we present some of the background material necessary to the proofs of our main results in the orientable setting. In Section 3, we prove the chain move theorem, stated there as Theorem 3.1. In Section 4, we prove the switch move theorem, Theorem 4.1. In Section 5 we prove the switch move gluing operation, Theorem 5.1, which allows us to glue together two diffeomorphic genus one boundary components from one or two hyperbolic 3–manifolds of finite volume and then operate to generate new hyperbolic 3–manifolds of finite volume.

Acknowledgements Thanks to the referee for very thorough and helpful comments, including a much shorter proof of Lemma 3.4 and a strengthened version of Theorem 5.1. Meeks and Ramos were partially supported by the CNPq Brazil, grant 400966/2014-0.

2 Preliminaries

In this section, we recall some definitions and results that are needed to understand hyperbolic 3–manifolds of finite volume and certain embedded surfaces in such ambient spaces. Our first goal is to understand the statement of Thurston's hyperbolization theorem in our setting. Before stating this result, we first explain some of the definitions and notations we use. Throughout this discussion, P will denote a connected, orientable, compact 3–manifold with nonempty boundary ∂P consisting of tori and $\mathrm{int}(P)$ will denote the interior of P. Moreover, a *surface* Σ *in* P means a properly embedded surface $\Sigma \subset P$, ie Σ is embedded in P with $\partial\Sigma = \Sigma \cap \partial P$.

Definition 2.1 (1) Given a surface Σ in P, a *compression disk for* Σ is a disk $E \subset P$ with $\partial E = E \cap \Sigma$ such that ∂E is homotopically nontrivial in Σ. If Σ does not admit any compression disk, we say Σ is *incompressible*.

(2) Given a surface Σ in P, a *boundary-compression disk for* Σ is a disk $E \subset P$ with $\partial E = E \cap (\Sigma \cup \partial P)$ such that $\partial E = \alpha \cup \beta$, where α and β are arcs intersecting only in their endpoints such that $\alpha = E \cap \Sigma$ and $\beta = E \cap \partial P$ and α does not cut a disk from Σ. If Σ does not admit any boundary-compression disk, we say Σ is *boundary-incompressible*.

(3) A torus T in P is *boundary parallel* if T is isotopic to a boundary component of P.

(4) An annulus A in P is *boundary parallel* if A is isotopic, relative to ∂A, to an annulus $A' \subset \partial P$.

(5) A sphere S in P is *essential* if S does not bound a ball in P.

(6) A disk E in P is *essential* if ∂E is homotopically nontrivial in ∂P.

(7) A torus T is *essential* in P if T is incompressible and not boundary parallel.

(8) An annulus A is *essential* in P if A is incompressible, boundary-incompressible and not boundary parallel.

Using the above definitions, Thurston's hyperbolization theorem implies that a connected, orientable, noncompact 3–manifold N admits a hyperbolic metric of finite volume if and only if N is diffeomorphic to $\operatorname{int}(P)$ as above and there are no essential spheres, disks, tori or annuli properly embedded in P. In this case, we shall say that N is *hyperbolic*. When a link L in a 3–manifold M has hyperbolic complement, we will say either $M \setminus L$ is hyperbolic, or L is hyperbolic.

A useful fact is that if α is an arc with endpoints in a link L in a 3–manifold M such that α corresponds to a geodesic in the hyperbolic link complement $M \setminus L$, then α cannot be homotoped through $M \setminus L$ into L while fixing its endpoints on L. This follows from the fact any such geodesic will lift to geodesics connecting distinct horospheres in the universal cover $\mathbb{H}^3$, whereas an arc that is homotopic into L will lift to arcs, each of which connects one and the same horosphere.

In the case that a manifold M has no essential disks, we say it is *boundary-irreducible*. In the case that a manifold M has no essential spheres, we say it is *irreducible*. Note that if M has only toroidal boundaries and it is not a solid torus, which is the situation we will consider, irreducibility implies boundary-irreducibility. This is because if there exists an essential disk D with boundary in a torus $T \subset \partial M$, then $\partial N(D \cup T)$ is a sphere which must bound a ball to the non-D side, implying M is a solid torus. Here and elsewhere, $N(G)$ denotes a regular neighborhood of a set $G \subset M$.

Given an annulus A properly embedded in an irreducible manifold M with toroidal boundary, we note that if A is boundary-compressible, it is boundary-parallel. This follows because we can surger the annulus along the boundary-compressing disk to obtain a properly embedded disk D, with trivial boundary on ∂M. Then ∂D bounds a disk D' in ∂M, and $D \cup D'$ is a sphere bounding a ball. This allows us to isotope A relative ∂A into ∂M.

Finally, we remark that if S is a two-sided, incompressible surface properly embedded in an irreducible manifold M with toroidal boundary, then either S is boundary-incompressible or S is a boundary parallel annulus; see for instance [11, Lemma 1.10].

3 The chain move theorem

Let L be a hyperbolic link in a 3–manifold M and let $\mathcal{B} \subset M$ be a ball in M that intersects L as in Figure 1, left. In this section we prove the chain move theorem, as stated by Theorem 3.1 below. The proof breaks up into two cases depending on whether or not the pair $(M \setminus \mathcal{B}, L \setminus \mathcal{B})$ is a rational tangle in a 3–ball; see [4, Chapter 2] for this definition and for the representation of a rational tangle by a sequence of integers.

Theorem 3.1 (chain move theorem) *Let L be a link in a 3–manifold M such that the link complement $M \setminus L$ admits a complete hyperbolic metric of finite volume. Suppose that there is a sphere $\mathcal{S}$ in M bounding a ball $\mathcal{B}$ that intersects L as in Figure 1, left. Let L' be the resulting link obtained by replacing $L \cap \mathcal{B}$ by the components as appear in Figure 1, right. Then if $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is not any of the rational tangles $-k$, $-(k+1)$, or $-2-k$ in a 3–ball, then $M \setminus L'$ admits a complete hyperbolic metric of finite volume.*

In Figure 7, top, we see the new link components that are inserted into the ball $\mathcal{B}$. In Figures 7, bottom, we see, for any fixed integer k, the three cases of rational tangles in the exterior 3–ball that do not yield a hyperbolic link complement.

Remark 3.2 The crossings around the single trivial component need not be nonalternating for Theorem 3.1 to apply. If the crossings alternate (as shown in Figure 8, left), we could add a crossing to γ_2 and work in a subball as in Figure 8, right, so that the crossings are those shown in Figure 1, left.

Remark 3.3 Repeated application of the chain move theorem allows us to create a hyperbolic link complement with an arbitrarily long chain of trivial components and

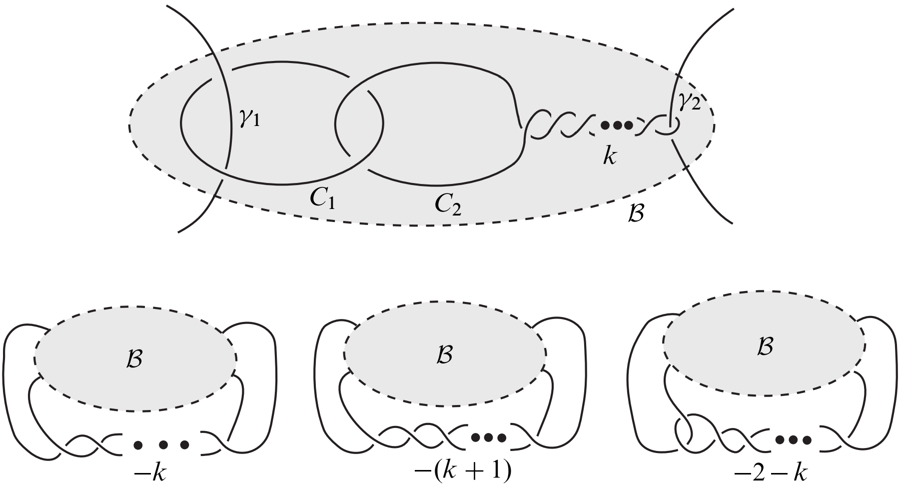


Figure 7: The link components we are inserting in $\mathcal{B}$ and the three rational tangles in an exterior ball that do not generate a hyperbolic link complement.

with any amount of twist. Moreover, if the original exterior tangle is assumed not to be rational, the subsequent exterior tangles to which we apply the move cannot be rational either, so all resulting link complements are hyperbolic. In fact, even if the initial exterior tangle is rational, if our first application of the move results in a hyperbolic link complement, all repeated applications will also be hyperbolic.

We set the stage for the proof of Theorem 3.1 with the following lemma.

Lemma 3.4 *Let L be a link in the* 3*–sphere such that the tangle $R = L \setminus \mathcal{B}$ is a rational tangle and the tangle $L \cap \mathcal{B}$ is the tangle T_k appearing in Figure 1, right, for some*

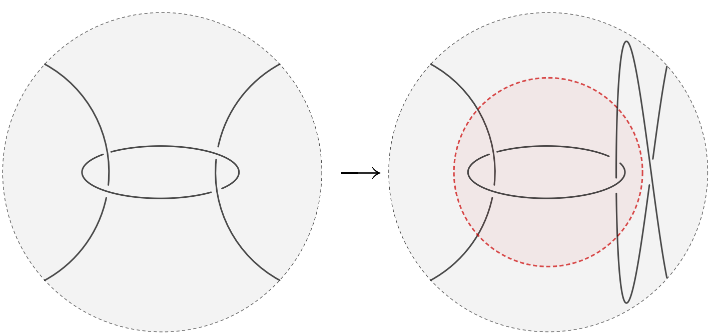


Figure 8: Using an isotopy within $\mathcal{B}$ to obtain a subball where the chain move theorem applies.

integer k. If R is none of the rational tangles ∞, $-k$, $-(k+1)$ or $-2-k$, the link complement is hyperbolic.

Proof We represent rational tangles by a fraction $\frac{p}{q} \in \mathbb{Q} \cup \{\infty\}$. We also use the notation $K(p_1/q_1, \dots, p_n/q_n)$ to denote the Montesinos link created by the tangles $p_1/q_1, \dots, p_n/q_n$. For more details, see [14] or [16].

Note that if L is as stated in Lemma 3.4, then it is a Montesinos link of either three or four components. Furthermore, after untwisting the k half-twists into R, the rational tangles in $\mathcal{B}$ are $-\frac{1}{2}$, $\frac{1}{2}$ and $\frac{1}{2}$ and R is also a rational tangle. Thus, there exists $\frac{p}{q} \in \mathbb{Q} \cup \{\infty\}$ such that L is equivalent to $K\left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{p}{q}\right)$. If $\frac{p}{q} = \infty$, then L is not prime (and not hyperbolic).

Next, we use the classification of all the nonhyperbolic Montesinos links given by work of Bonahon and Siebenmann [8] (or see [9] for a different proof) and Oertel [14, Corollary 5] to analyze the possibilities for $\frac{p}{q} \in \mathbb{Q}$ for which L is not hyperbolic. In [8], the Montesinos "torus links" are determined, all of which are nonhyperbolic. These include torus links in the usual sense but additionally allowing for the inclusion of the core curves of the solid tori to either side of the defining torus. In [14], the nonhyperbolic Montesinos links that are not "torus links" are determined. See [16] (Theorem 4.1 and the following paragraph) for a complete list of the nonhyperbolic Montesinos links.

If L has three components, then γ_1 and γ_2 are in the same component $C_3 \subset L$. But the only nonhyperbolic Montesinos links of three components are $L = K\left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{m}\right)$, for $m \in 2\mathbb{N}$, $L = K\left(-\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$, or their mirror images. Since $\operatorname{lk}(C_1, C_2) = \pm 1$, $\operatorname{lk}(C_2, C_3) = \pm 1$ and $\operatorname{lk}(C_3, C_1) = \pm 1$, the only possibility is $L = K\left(-\frac{1}{2}, \frac{1}{2}, \pm\frac{1}{2}\right)$. In this situation, note that $K\left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{p}{q}\right)$ is equivalent to $K\left(-\frac{1}{2}, \frac{1}{2}, \pm\frac{1}{2}\right)$ if and only if $\frac{p}{q} = 0$ or $\frac{p}{q} = -1$.

In the case when L is a nonhyperbolic 4–component Montesinos link, the only possibility is $L = K\left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$. But suppose it has another description as a 4–tangle Montesinos link. Let $L' = K\left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{r}{s}\right)$, and suppose $L = L'$. Then $s = 2$ by consideration of the Seifert invariants of the double branched cover. Then r is odd, and L' is a chain link. But L is nonhyperbolic, while L' is hyperbolic by [13] unless $r = -1$ (corresponding to chain link $C(4, -2)$ in their notation).

Hence, there are only four possibilities for $\frac{p}{q}$ which make L nonhyperbolic, namely ∞, 0, -1 and $-\frac{1}{2}$. After compensating for the k twists being moved into R, these correspond exactly to the four tangles in the statement of the lemma. $\square$

Proof of Theorem 3.1 Let $X = M \setminus L$ and, for $i = 1, 2$, let Γ_i be the connected component of L containing the arc γ_i (note that possibly $\Gamma_1 = \Gamma_2$). First, we assume that M is orientable. We let L' be the link formed by the replacement stated in Theorem 3.1. As stated in the introduction of this section, we will assume that $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is not a rational tangle in a 3–ball as this special case has been dealt with by Lemma 3.4.

Note that we do not include the rational tangle ∞ as a tangle to exclude in the statement of Theorem 3.1 since, in the case of this tangle, the original link L is not prime and hence X is not hyperbolic. We prove Theorem 3.1 when M is orientable by showing that the resulting link complement $Y = M \setminus L'$ does not admit essential disks, spheres, tori or annuli. In order to do so, we first prove the following.

Claim 3.5 *The four-punctured sphere $\mathcal{Q} = \mathcal{S} \setminus L$ is incompressible and boundary-incompressible in X and also in Y.*

Proof We prove that if $\mathcal{Q}$ is compressible in X or in Y, then $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is a rational tangle in a 3–ball. We first prove this property in X. Let γ be a nontrivial simple closed curve in $\mathcal{Q}$ and assume that there is a compact disk $E \subset X$ with $\partial E = \gamma = E \cap \mathcal{Q}$. Then each of the two disks E_1 and E_2 bounded by γ in $\mathcal{S}$ must contain exactly two of the punctures of $\mathcal{Q}$, otherwise we could attach a one-punctured disk in $\mathcal{Q}$ to E to find an essential disk in X, contradicting its hyperbolicity.

If E were contained in $\mathcal{B}$, then $E \cup E_1$ and $E \cup E_2$ are two spheres in $\mathcal{B}$, each punctured twice by L. Since both punctures in each sphere cannot come from distinct arcs in $L \cap \mathcal{B}$, E separates $\mathcal{B}$ into two balls B_1 and B_2, where $\gamma_1 \subset B_1$ and $\gamma_2 \subset B_2$, and it then follows that C cannot link γ_1 and γ_2 simultaneously, a contradiction.

Next assume that $E \cap \operatorname{int}(\mathcal{B}) = \varnothing$. Let $A_i = E \cup E_i \setminus L$ for $i = 1, 2$. Then each A_i is an annulus in X. Since each A_i is incompressible and X is hyperbolic, A_i is boundary parallel. Therefore, the closure of A_i in M bounds a closed ball $B_i \subset M \setminus \operatorname{int}(\mathcal{B})$ with $\partial B_i = E_i \cup E$ and such that $B_i \cap L$ is an unknotted arc in B_i. Hence, we can isotope $L \cap B_i$ through B_i to the surface $\mathcal{S}$. Then, after the isotopy, $\partial N(\mathcal{B})$ is a sphere in X. Since X is hyperbolic, $\partial N(\mathcal{B})$ bounds a ball which is disjoint from $\mathcal{B}$, and this is a contradiction unless $M = \mathbb{S}^3$.

If $M = \mathbb{S}^3$, then the fact $L \cap B_i$ can be isotoped through B_i to the surface $\mathcal{S}$ implies $L \setminus (\mathcal{B} \cap L)$ can be isotoped to be two disjoint embedded arcs on $\mathcal{S}$. Hence,

$$(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$$

is a rational tangle determined by γ, and up to isotopy, E is the only compression disk for $\mathcal{Q}$ in X.

Note that the above argument implies that $\mathcal{Q}$ is also incompressible in Y, as we next explain. If $E \subset Y$ was a compressing disk with $\partial E = E \cap \mathcal{Q}$, then E necessarily is contained in $\mathcal{B}$, otherwise $E \subset Y \setminus \mathcal{B}$ and $Y \setminus \mathcal{B} = X \setminus \mathcal{B}$ would give a compression disk for $\mathcal{Q}$ in X. Once again, $E \subset \mathcal{B}$ gives that E separates $\mathcal{B}$ into two balls, B_1 and B_2, such that $\gamma_1 \subset B_1$ and $\gamma_2 \subset B_2$. Then, since C_1 links γ_1, $C_1 \subset B_1$. And since C_2 links γ_2, $C_2 \subset B_2$. But then E separates C_1 from C_2 in $\mathcal{B}$, a contradiction to the fact they are linked in $\mathcal{B}$.

To prove boundary-incompressibility of $\mathcal{Q}$ in either X or Y, suppose E is a boundary-compression disk such that $\partial E = \alpha \cup \beta$ with $\alpha = E \cap \mathcal{Q}$. If α connects two distinct punctures of $\mathcal{Q}$ and $N(E)$ is a small neighborhood of E in M, then $\partial N(E) \setminus (\partial N(E) \cap \mathcal{B})$ is a compression disk for $\mathcal{Q}$, a contradiction.

If both endpoints of α are at the same puncture, then, since the interior of β is disjoint from $\mathcal{Q}$, β together with an arc in $\mathcal{Q}$ bound a disk $\widetilde{E}$ in $\partial N(L)$. Thus, $E \cup \widetilde{E}$ is a compression disk for $\mathcal{Q}$, a contradiction. $\square$

Claim 3.6 *Y does not admit any essential spheres or essential disks.*

Proof We argue by contradiction and first suppose that there is an essential sphere S in Y. If S intersects $\mathcal{Q}$, then, by incompressibility of $\mathcal{Q}$, we can exchange disks on S for disks on $\mathcal{Q}$ in order to obtain an essential sphere S' in Y that does not intersect $\mathcal{Q}$. If $S' \subset Y \setminus \mathcal{B}$, then $S' \subset X$, which implies S' is the boundary of a ball $B \subset X$. In this case, B must be disjoint from $\mathcal{B}$, since $C \subset \mathcal{B}$; hence, $B \subset Y$ which contradicts that S' is essential in Y. Thus, we may assume that S' is contained in $\mathcal{B}$, and so it bounds a subball B of $\mathcal{B}$. If B intersects $C_1 \cup C_2$, then, by the linking properties of these circles, $C_1 \cup C_2$ must be contained in B. As γ_1 links C_1 in $\mathcal{B}$, we arrive at a contradiction because the endpoints of γ_1 lie outside of B. This contradiction implies that $L' \cap \mathcal{B}$ is disjoint from B, which in turn implies that $B \subset Y$, contradicting that S' is essential in Y.

Suppose now that there is an essential disk D with boundary in $\partial N(L')$. Then there is a component J of L' such that $\partial D \subset \partial N(J)$, and we let $S = \partial N(D \cup N(J))$. It then follows that S is an essential sphere, as it splits J from the other components of L', contradicting the nonexistence of such spheres. $\square$

For the next arguments in the proof, let $D_1, D_2 \subset Y \cap \mathcal{B}$ denote two twice-punctured disks bounded respectively by $C_1, C_2 \subset L'$ and let $\overline{D}_i$ denote the closure of D_i in M; thus each $\overline{D}_i$ is a disk in $\mathcal{B}$. We prove the following.

Claim 3.7 *The twice punctured disks D_1 and D_2 are incompressible and boundary-incompressible in Y.*

Proof Suppose there were a disk $E \subset Y$, $\operatorname{int}(E) \cap D_i = \varnothing$ with nontrivial boundary in D_i. Since $\mathcal{Q}$ is incompressible and we may assume general position, any component in $E \cap \mathcal{Q}$ is a simple closed curve that is trivial both in E and in $\mathcal{Q}$. Choose an innermost curve $\alpha \subset E \cap \mathcal{Q}$ in the sense that the interior of the disk $E' \subset E$ bounded by α does not intersect $\mathcal{Q}$ and let E'' be the disk bounded by α in $\mathcal{Q}$. Then $E' \cup E''$ is a sphere that is either in the hyperbolic manifold X or in $Y \cap \mathcal{B}$. In either case, $E' \cup E''$ bounds a ball in Y that can be used to isotope E' to E'' and further to remove α from the intersection $E \cap \mathcal{Q}$. After repeating this disk replacement argument a finite number of times, we may assume that $E \subset \mathcal{B}$.

Let E' be the disk in $\overline{D}_i$ bounded by ∂E. Then $E \cup E'$ is a sphere in $\mathcal{B}$ which is punctured only once by at least one of the components in $L' \cap \mathcal{B}$, a contradiction that shows that D_1 and D_2 are incompressible in Y.

To finish the proof of Claim 3.7, we note that D_i is 2–sided and incompressible, and Y is irreducible by Claim 3.6. Thus, as explained in the end of Section 2, D_i is boundary-incompressible. □

Claim 3.8 *Y does not admit essential annuli.*

Proof Arguing by contradiction, assume there exists an essential annulus A in $M \setminus N(L')$. Let α_1 and α_2 denote the two boundary components for A. Then there are components J_1 and J_2 of L' such that $\alpha_1 \subset \partial N(J_1)$ and $\alpha_2 \subset \partial N(J_2)$. After an isotopy of A we will assume without loss of generality that both α_1 and α_2 are taut in the respective tori $\partial N(J_1)$ and $\partial N(J_2)$, in the sense that, in the product structure generated by respective meridional curves in $\partial N(J_i)$, each α_i is transverse to all meridians and also to all longitudes, unless α_i is one of them.

We next rule out the various possibilities for A, starting with the assumption that A does not intersect $D_1 \cup D_2$.

In this case, we may use the fact that $\partial N(D_1 \cup D_2) \setminus N(\gamma_1 \cup \gamma_2)$ is isotopic to $\mathcal{Q}$ to isotope A in $M \setminus N(L')$ to lie outside of $\mathcal{B}$. Thus, A is an annulus in X, and the fact that X is hyperbolic implies that A is either compressible or boundary parallel in X. If A is compressible in X, then we may use the fact that $\mathcal{Q}$ is incompressible in Y and a disk replacement argument to show that A is compressible in Y, a contradiction.

Next, we treat the case when A is boundary parallel in $M \setminus N(L)$. In this case, A defines a product region $W \subset M \setminus N(L)$ through which A is parallel to a subannulus in $\partial N(L)$. Since C lies outside of W, separation properties imply that $\mathcal{B}$ is disjoint from W; hence, $W \subset M \setminus N(L')$ from where it follows that A is boundary parallel in Y, a contradiction.

Now suppose that A intersects $D_1 \cup D_2$ and assume that A has the fewest number of intersection components in $A \cap (D_1 \cup D_2)$ for an essential annulus in Y. Note that for $i = 1, 2$, the intersection curves which may appear in $A \cap D_i$ are either simple closed curves or arcs with endpoints in ∂A.

We next eliminate the possibility that $A \cap D_i$ contains a simple closed curve. Since D_i is incompressible, by minimality of intersection curves, any simple closed curve in the intersection $A \cap D_i$ is nontrivial in A. Note that if $A \cap D_i$ contains a simple closed curve that circles one puncture, we may take an innermost such curve and use the once-punctured disk on D_i that it bounds to surger A to obtain two annuli, each with fewer intersection curves and at least one of them must be essential. So we may assume that all simple closed curves in $A \cap D_i$ circle both punctures of D_i. But then, the outermost of such intersection curves bounds an annulus that again allows us to surger A to obtain an essential annulus with fewer intersection curves. Hence, all curves in $A \cap D_i$ are arcs with endpoints in ∂A.

Next, we show that there are no arcs in $A \cap D_i$ that have endpoints on the same boundary component of A. Assume that α is such an arc and let E_1 be the disk defined by α in A. We assume that α is innermost in the sense that the interior of E_1 is disjoint from D_i. Since Claim 3.7 implies that D_i is boundary-incompressible, it follows that α must cut a disk E_2 from D_i. Then $E = E_1 \cup E_2$ is a disk with boundary $\partial E \subset \partial N(J)$. Since Y does not admit essential disks, it follows that ∂E is trivial in $\partial N(J)$, and we may use the fact that all spheres in Y bound balls to isotope A so that E_1 moves past E_2, thus eliminating the intersection curve α and contradicting minimality of the number of intersection components.

In particular, if A intersects D_i, both α_1 and α_2 must intersect D_i, and none of the intersection arcs on $A \cap D_i$ can cut a disk off D_i, as if they did, A would be boundary-compressible and hence boundary-parallel since Y is irreducible. Note that because there is at least one arc of intersection of A with a D_i, and such arc goes from α_1 to α_2, we have $\partial A \subset (\partial N(C_1) \cup \partial N(C_2) \cup \partial N(\Gamma_1) \cup \partial N(\Gamma_2))$. Moreover, since both α_1 and α_2 intersect $D_1 \cup D_2$ and we assume minimality of intersection components in $\partial A \cap (D_1 \cup D_2)$, no component of ∂A can be a meridian in $\partial N(\Gamma_1)$ or in $\partial N(\Gamma_2)$; hence any closed curve in $A \cap \mathcal{Q}$ must be trivial in A, and, consequently, trivial in $\mathcal{Q}$.

We next consider the case that $\partial A \subset \partial N(C_1) \cup \partial N(C_2)$. Then by incompressibility of $\mathcal{Q}$, we can isotope A to lie inside $\mathcal{B}$. Moreover, $C_1 \cup C_2$ is a Hopf link with complement in the 3–sphere that is a thickened torus $T \times [0,1]$. Thus, $\mathcal{B} \setminus (N(C_1) \cup N(C_2))$ is the complement of a ball B in $T \times [0,1]$, where we identify $\partial N(C_1)$ with $T \times \{0\}$ and $\partial N(C_2)$ with $T \times \{1\}$.

Assume that both boundary components of A are on $\partial N(C_1)$. Then A is an annulus in $(T \times [0,1]) \setminus B$ with both boundaries on $T \times \{0\}$. In particular, in $T \times [0,1]$, A is boundary-parallel through a solid torus V that A cuts from $T \times [0,1]$. Since ∂V is a closed surface in the interior of the three-ball $\mathcal{B}$, it defines a unique compact region disjoint from $\partial \mathcal{B} = \partial B$, from where it follows that B must be disjoint from V. But then both the arcs γ_1 and γ_2, which have endpoints on $\partial \mathcal{B}$, must also be disjoint from V, meaning that $V \subset Y$, and then A is boundary-parallel in Y, a contradiction. By symmetry, the same argument also proves that A cannot have both boundary components on $\partial N(C_2)$.

Next, suppose that one boundary of A is on $\partial N(C_1)$ and the other is on $\partial N(C_2)$. Then again, A can be seen as an annulus in $(T \times [0,1]) \setminus B$, but now its boundary is a pair of nontrivial parallel curves on $T \times \{0\}$ and $T \times \{1\}$. These curves are respectively realized as a (p,q)–curve[1] on $\partial N(C_1)$ and a (q,p)–curve on $\partial N(C_2)$. But there exist arcs $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ on $\mathcal{Q}$ such that the closed curve $\gamma_1 \cup \widetilde{\gamma}_1$ wraps meridionally around C_1 and the closed curve $\gamma_2 \cup \widetilde{\gamma}_2$ wraps meridionally around C_2, where in $T \times [0,1]$, a meridian of $\partial N(C_2)$ corresponds to a longitude of $N(C_1)$. Hence, when we add γ_1 and γ_2 to $T \times [0,1] \setminus B$, one wrapping meridionally around $T \times [0,1]$ and the other wrapping longitudinally, at least one will puncture A, a contradiction.

Thus, at least one boundary component of A, say α_1, must be on $\partial N(\Gamma_i)$, for some $i \in \{1,2\}$. As already explained, α_1 is not a meridian on $\partial N(\Gamma_i)$.

Next, assume that α_2 is on $\partial N(C_1)$ or $\partial N(C_2)$. Since $\mathcal{Q}$ is incompressible and Y is irreducible, after performing a disk replacement argument, we may assume that $A \cap \mathcal{Q}$ is

[1] For given relatively prime integers p and q, a (p,q)–curve is a torus knot that winds p times around the meridian of the torus and q times around its longitude.

a family of pairwise disjoint arcs, each with both endpoints in α_1. Let a be one of such arcs and assume that a cuts an innermost disk D from A, in the sense that $D \cap \mathcal{Q} = a$. If $D \subset \mathcal{B}$, then, if we let $b = \partial D \setminus a$, it follows that $b \subset (\partial N(\gamma_i)) \cap \mathcal{B}$ and our assumptions on α_1 being taut imply that b joins two distinct punctures of $\mathcal{Q}$. But then it follows that ∂D links C_i on $\mathcal{B}$, and D must be punctured by C_i, a contradiction. Hence, it follows that D is to the outside of $\mathcal{B}$. Once again, our assumptions on α_1 imply that a joins two distinct punctures of $\mathcal{Q}$, from where it follows that D is a boundary-compression disk for $\mathcal{Q}$, which contradicts Claim 3.5.

It remains to rule out the case where $\alpha_1 \cup \alpha_2 \subset \partial N(\Gamma_1) \cup \partial N(\Gamma_2)$. Let a be an arc of intersection $A \cap (D_1 \cup D_2)$. Then our previous arguments give that a joins α_1 and α_2 and that a cannot cut a disk off D_i. In particular, a must necessarily intersect the disk D_j for $j \neq i$ and that creates another arc $b \subset A \cap D_j$ which meets a transversely at a point p and joins α_1 and α_2. In particular, $\Gamma_1 = \Gamma_2$. The point p separates both arcs a and b, and that defines a unique disk $D \subset A$ with boundary given by one arc in a, one arc in b and one arc c in α_1. Note that $D \cap D_i \subset a \cup b$, since any arc in $A \cap D_i$ must join α_1 and α_2. Let E be a connected component of $D \setminus \mathcal{B}$ that contains a subarc of c in its boundary. Such component exists because the endpoints of a and b on D are on distinct disks, D_1 and D_2, and hence c cannot be contained in $\mathcal{B}$. Once again, the fact that α_1 is taut gives that $\partial E \cap \mathcal{Q}$ is an arc joining two distinct punctures of $\mathcal{Q}$. But then, E is a boundary-compression disk for $\mathcal{Q}$, a contradiction.

The cases treated above rule out the possibility that Y admits an essential annulus, thereby proving Claim 3.8. □

Claim 3.9 *Y does not admit essential tori.*

Proof We argue by contradiction and suppose that $T \subset Y$ is a torus which is incompressible and not boundary-parallel in Y. First, suppose that T does not intersect $D_1 \cup D_2$. Then we can isotope T in Y to assume that $T \cap \mathcal{B} = \varnothing$, and then $T \subset X$. Since X is hyperbolic, either T admits a compression disk in X or T is boundary parallel in X.

First assume that $E \subset X$ is a compression disk for $T \subset X \setminus \mathcal{B}$. Since $\mathcal{Q}$ is incompressible in X, after disk replacements, we may assume that E is disjoint from $\mathcal{Q}$. In particular, $E \subset X \setminus \mathcal{B} \subset Y$, which is a contradiction.

Next, suppose that T is parallel to the boundary of a neighborhood of one of the components J of L, and let $W \subset X$ be the related proper product region with boundary T.

We claim that $\mathcal{Q}$ must be disjoint from W. Otherwise, $\mathcal{Q} \subset W$ which would imply that $\mathcal{B} \setminus (\gamma_1 \cup \gamma_2 \cup C) \subset W$; this is a contradiction because W has only one end corresponding to a single component of L. Since $\mathcal{Q}$ separates X and is disjoint from W, we have $W \subset Y$, which means T is boundary parallel in Y. This proves that any essential torus in Y must intersect $D_1 \cup D_2$.

Let $T \subset Y$ be an essential torus that intersects D_i, for some $i \in \{1, 2\}$. Next, we prove that Y must contain an essential annulus, which contradicts Claim 3.8. After possibly replacing disks in T by disks in the incompressible surface D_i, we may assume that any component in $T \cap D_i$ is homotopically nontrivial in D_i; let $\gamma \subset T \cap D_i$ be one such component. First assume that γ encircles a single puncture in D_i and choose it to be an innermost such curve in $T \cap D_i$. Using the once-punctured disk bounded by γ in D_i to surger T, we obtain an essential annulus in Y, as claimed. Next, assume that γ encircles both punctures of D_i and that it is an outermost such curve on $T \cap D_i$. In this case, we may use the outer annulus on D_i to surger T in order to obtain an essential annulus in Y, thereby proving Claim 3.9. □

Having proved that there are no essential disks, spheres, tori or annuli in Y, it follows that Y satisfies Thurston's conditions for hyperbolicity, proving Theorem 3.1 when M is orientable.

The case when M is nonorientable can be proved using the orientable case as we next explain. Suppose that M is nonorientable and that L, L' and $\mathcal{B}$ are as stated. Let $\Pi\colon \widehat{M} \to M$ be the oriented two-sheeted covering of M and let $\widehat{L} = \Pi^{-1}(L)$ and $\mathcal{B}_1$ and $\mathcal{B}_2$ be the two connected components of $\Pi^{-1}(\mathcal{B})$. Then, $\widehat{L}$ is a hyperbolic link in $\widehat{M}$ and $\widehat{L} \setminus \mathcal{B}_1$ is not a rational tangle in a 3–ball, since $\widehat{L} \cap \mathcal{B}_2$ is diffeomorphic to $L \cap \mathcal{B}$. Then, we may use the chain move to modify $\widehat{L}$ in $\mathcal{B}_1$, replacing $\widehat{L} \cap \mathcal{B}_1$ by a tangle diffeomorphic to $L' \cap \mathcal{B}$, which creates a hyperbolic link $\widehat{L}'$ in $\widehat{M}$. Then, since $\widehat{L}' \cap \mathcal{B}_2 = \widehat{L} \cap \mathcal{B}_2$ and $\widehat{M} \setminus \widehat{L}'$ is hyperbolic, we can use the chain move in $\mathcal{B}_2$ to replace $\widehat{L}' \cap \mathcal{B}_2$ by a tangle diffeomorphic to $L' \cap \mathcal{B}$ and create another hyperbolic link $\widehat{L}''$ in $\widehat{M}$. Since we may do this second replacement in an equivariant manner with respect to the nontrivial covering transformation σ defined by Π, the restriction of Π to the hyperbolic manifold $\widehat{M} \setminus \widehat{L}''$ is the two-sheeted covering space of $M \setminus L'$. Since σ is an order-two diffeomorphism of $\widehat{M} \setminus \widehat{L}''$, the Mostow–Prasad rigidity theorem implies that we may consider σ to be an isometry of the hyperbolic metric of $\widehat{M} \setminus \widehat{L}''$. Hence, the hyperbolic metric of $\widehat{M} \setminus \widehat{L}''$ descends to $M \setminus L'$ via Π, which finishes the proof of Theorem 3.1. □

4 The switch move theorem

Theorem 4.1 (switch move theorem) *Let L be a link in a 3–manifold M such that $M \setminus L$ admits a complete hyperbolic metric of finite volume. Let $\alpha \subset M$ be a compact arc which intersects L transversely in its two distinct endpoints, and such that $\mathrm{int}(\alpha)$ is a complete, properly embedded geodesic of $M \setminus L$. Let $\mathcal{B}$ be a closed ball in M containing α in its interior and such that $\mathcal{B} \cap L$ is composed of two arcs in L, as in Figure 3. Let L' be the resulting link in M obtained by replacing $L \cap \mathcal{B}$ by the components as appearing in Figure 4, right. Then $M \setminus L'$ admits a complete hyperbolic metric of finite volume.*

Proof We begin the proof by setting the notation. Let G and G' be the connected components of L containing the arcs g and g' respectively, as in Figure 4, left. Note that it can be the case $G = G'$. Let L' be the link formed by replacing $g \cup g'$ in $\mathcal{B}$ by $\gamma_1 \cup \gamma_2 \cup C$. For $i = 1, 2$, let Γ_i be the component of L' containing γ_i. Note that possibly $\Gamma_1 = \Gamma_2$ and let $\Gamma = \Gamma_1 \cup \Gamma_2$.

We split the proof into two cases, depending on whether or not $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is a rational tangle in a 3–ball.

Claim 4.2 *If $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is a rational tangle in a 3–ball, then $M \setminus L'$ is hyperbolic.*

Proof A rational tangle in a 3–ball always has a projection that is alternating; see for instance [10]. Then L is a rational, alternating link in S^3 that is prime, since $M \setminus L$ is hyperbolic. By [12, Corollary 2], a rational, alternating link in S^3 that is prime is hyperbolic if and only if it is nontrivial and not a 2–braid. After forming L', we consider the link L'' obtained from L' by doing a half-twist on the twice-punctured disk bounded by C to add a crossing so that $L'' \setminus C$ has an alternating projection, as in Figure 9. Then L'' is in an augmented alternating link projection obtained from a prime, nonsplit reduced alternating projection. If $L'' \setminus C$ is neither trivial nor a 2–braid, L'' is hyperbolic by [2]. However, if $L'' \setminus C$ is trivial, then L is a 2–braid and hence it does not satisfy the hypothesis that $M \setminus L$ is hyperbolic. And if $L'' \setminus C$ is a 2–braid, then L is a trivial knot, again not satisfying the same hypothesis. So L'' is a hyperbolic link in S^3. But by [1, Theorem 4.1], L'' is hyperbolic if and only if L' is hyperbolic. □

Remark 4.3 If $M \setminus L$ is hyperbolic and $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is a rational tangle in a 3–ball, then L is either a rational link or a rational knot in S^3 which is hyperbolic.

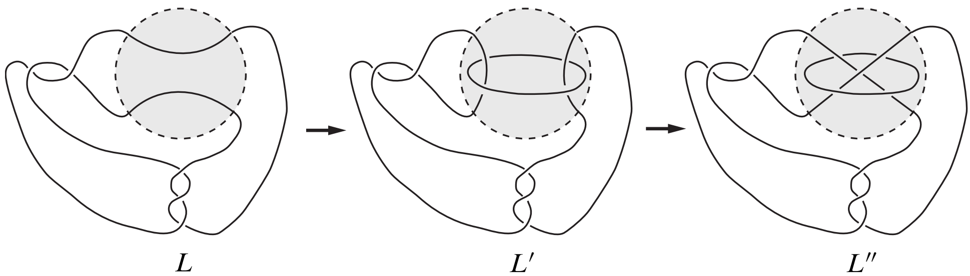


Figure 9: Creating L, L' and L''.

In this case, there is always an arc α as depicted in Figure 3 which is isotopic to a geodesic and hence the switch move can be applied. This follows because α can be chosen to be part of the fixed point set of an involution of the complement, which is realized by an isometry, and fixed-point sets of isometries must be geodesics (see [7] for the details).

From now on, we assume that $(M \setminus \mathcal{B}, L \setminus (\mathcal{B} \cap L))$ is not a rational tangle in a 3–ball. As in the proof of the chain move theorem (Theorem 3.1), we first assume that M is orientable. We also let $X = M \setminus L$ and $Y = M \setminus L'$ and we will prove that Y is hyperbolic by showing that there are no essential disks, spheres, tori or annuli in Y. Once again, we let $\mathcal{S} = \partial\mathcal{B}$, $\mathcal{Q} = \mathcal{S} \setminus L = \mathcal{S} \setminus L'$ and notice that the same arguments used to prove Claim 3.5 can be used to prove that $\mathcal{Q}$ is incompressible and boundary-incompressible in Y; the details are left to the reader.

Claim 4.4 *Y does not admit essential spheres or essential disks.*

Proof We first show that there are no essential spheres in Y. Suppose that $S \subset Y$ is a sphere and first assume that $S \cap \mathcal{B} = \varnothing$. Then $S \subset X$, and, since there are no essential spheres in X, it follows that S bounds a ball $B \subset X$. Since L intersects $\mathcal{B}$, this gives that $B \cap \mathcal{B} = \varnothing$, hence $B \subset Y$, proving that S is not essential in Y.

Next, we treat the case where S intersects $\mathcal{B}$. We can take S to have the least number of intersection curves in $S \cap \mathcal{Q}$ over all essential spheres. If S were contained in $\mathcal{B}$, it bounds a ball in $\mathcal{B}$ which is also a ball in $Y = M \setminus L'$, since $S \cap L' = \varnothing$. Next, we assume that $S \cap \mathcal{Q} \neq \varnothing$. Then there exists a disk $E \subset S$ with $\partial E = E \cap \mathcal{Q}$. After a standard disk replacement argument using that $\mathcal{Q}$ is incompressible and that there are no essential spheres that do not intersect $\mathcal{Q}$, we isotope S to lower the number of components in $S \cap \mathcal{Q}$, which proves that there are no essential spheres in Y.

To prove that there are no essential disks in Y, we argue by contradiction and assume that E is such a disk with boundary on a regular neighborhood of a component J of L'. Then $S = \partial N(E \cup N(J))$ is an essential sphere in Y, as it splits J from the other components of L', a contradiction. □

Let D be the interior of a twice-punctured disk in $\mathcal{B} \setminus L'$ bounded by C and let $\overline{D}$ be its closure in M.

Claim 4.5 *D is incompressible and boundary-incompressible in Y.*

Proof Using the facts that $\mathcal{Q}$ is incompressible in Y, X is hyperbolic and $Y \setminus \mathcal{B} = X \setminus \mathcal{B}$, we may use a disk replacement argument to assume that any compression disk for D is contained in $\mathcal{B} \setminus L'$. Arguing by contradiction, assume that $E \subset \mathcal{B} \setminus L'$ is a disk with $\partial E = E \cap D$, and that ∂E is nontrivial in D. Let $E_1 \subset \overline{D}$ be the subdisk bounded by ∂E in $\overline{D}$. Let $S = E_1 \cup E$. Then, S is a two-sphere in the ball $\mathcal{B}$ which is punctured only once by at least one of the arcs γ_1 or γ_2, which is impossible.

In order to prove that D is boundary-incompressible, we proceed as in the proof of Claim 3.7 and just observe that D is two-sided, incompressible, properly embedded in the irreducible manifold Y. □

Using that both D and $\mathcal{Q}$ are incompressible and boundary incompressible, we next proceed with the proof of Theorem 4.1.

Claim 4.6 *There are no essential annuli in $M \setminus N(L')$.*

Proof Suppose that A is an essential annulus in $M \setminus N(L')$. Our next arguments rule out the several distinct possibilities for A, which are separated into cases.

Case 1 *Assume that $A \cap \mathcal{B} = \varnothing$.*

In this case, $A \subset M \setminus N(L)$ and it must either compress or be boundary-parallel in $M \setminus N(L)$. First, let us assume that $E \subset M \setminus N(L)$ is a compression disk to A with boundary β. Then β separates A into two subannuli A_1 and A_2, and $E \cup A_1$ and $E \cup A_2$ give rise to two essential disks in X, which contradicts its hyperbolicity.

Hence, we may assume that A is boundary-parallel in $M \setminus N(L)$. Then, there is a component J of L and an annulus $A' \subset \partial N(J)$ such that $\partial A' = \partial A$ and $A \cup A'$ bounds a solid torus W in $M \setminus N(L)$, through which A is parallel to A'. If $\mathcal{B} \cap W = \varnothing$, then A is boundary parallel in Y, a contradiction. Hence, we may assume that $\mathcal{B} \cap W \neq \varnothing$. Since $A \cap \mathcal{B} = \varnothing$ and $L \cap W = \varnothing$, then A' must intersect $\mathcal{B}$ and J must be either G or G', which could be the same component. Suppose first that G and G' are distinct. Then if λ is an arc in $\mathcal{B} \setminus N(L)$ with an endpoint in $\partial N(G)$ and the other in $\partial N(G')$, at least one endpoint of λ is not in W. Since $\mathrm{int}(\lambda)$ cannot intersect A, it follows that $G' \subset W$, a contradiction.

Suppose now G and G' are the same component J. Since $A \cap \mathcal{B} = \varnothing$, ∂A is a pair of meridians on $\partial N(J)$. Then, there is a ball B' in $N(J)$ bounded by A' and two meridional disks in $N(J) \setminus \mathcal{B}$ bounded by ∂A. Then $W' = W \cup B'$ is a ball in M, and $J \cap W'$ is an unknotted properly embedded arc within it. Since $\mathcal{B} \cap W \neq \varnothing$, we have $\mathcal{B} \cap W' \neq \varnothing$. But then, the fact that $\partial W' \cap \mathcal{B} = \varnothing$ implies that $\mathcal{B} \subset W'$. Hence α can be homotoped into $\partial N(J)$, contradicting the fact it is a geodesic with endpoints on L.

Case 2 *Assume that $A \subset \mathcal{B}$.*

Let α_1 and α_2 denote the two components of ∂A. First, we assume that $\alpha_1 \subset \partial N(\Gamma)$ and $\alpha_2 \subset \partial N(C)$. Since $A \subset \mathcal{B} \setminus N(L')$, α_1 is either a meridian of $\partial N(\gamma_1)$ or a meridian of $\partial N(\gamma_2)$, and the symmetry between γ_1 and γ_2 allows us to assume $\alpha_1 \subset \partial N(\gamma_1)$. Take a meridional disk E_1 in $N(\gamma_1) \cap \mathcal{B}$ with $\partial E_1 = \alpha_1$. Then $E = A \cup E_1$ is a disk in $\mathcal{B} \setminus N(C)$ with $\partial E = \alpha_2 \subset \partial N(C)$. Hence, α_2 is a longitude of $\partial N(C)$. In particular, α_2 links γ_2 in $\mathcal{B}$, and hence γ_2 must puncture E, which is a contradiction. This contradiction shows that if $A \subset \mathcal{B} \setminus N(L')$ is an essential annulus, then α_1 and α_2 are either both parallel curves on $\partial N(C)$ or both meridians on $\partial N(\Gamma)$.

Assume that A is an essential annulus in $M \setminus N(L')$ such that $A \subset \mathcal{B}$ and α_1 and α_2 are meridians on $\partial N(\Gamma)$. Let E_1 and E_2 be two meridional disks in $N(\Gamma)$ with respective boundaries α_1 and α_2. Then $A \cup E_1 \cup E_2$ is a sphere in $\mathcal{B}$ that bounds a ball $B \subset \mathcal{B}$, which is either punctured once by each γ_1 and γ_2, which is not possible, or twice by one of them, say γ_1. Since A is not boundary parallel, $C \subset B$. However, since C links both γ_1 and γ_2, γ_2 must be contained in B, which is a contradiction.

Still assuming that $A \subset \mathcal{B}$, it remains to obtain a contradiction when both α_1 and α_2 are (p,q)–curves on $\partial N(C)$. In this case, $\mathcal{B} \setminus (N(\gamma_1) \cup N(C))$ is diffeomorphic to $T \times [0,1]$, where $T = \mathbb{S}^1 \times \mathbb{S}^1$ is a torus, and we identify $\partial N(C)$ with $T \times \{0\}$. Since any annulus in $T \times [0,1]$ with boundary in $T \times \{0\}$ is parallel to an annulus in $T \times \{0\}$, it follows that A is parallel to an annulus $A' \subset \partial N(C)$ with $\partial A' = \alpha_1 \cup \alpha_2$, in the sense that there is a solid torus region $W \subset T \times [0,1]$ with $\partial W = A \cup A'$. Since $N(\gamma_2) \subset T \times [0,1]$ and does not intersect ∂W, the fact that the endpoints of γ_2 lie in $T \times \{1\}$ implies that $N(\gamma_2)$ is disjoint from W. Therefore, A is boundary parallel in $\mathcal{B} \setminus N(L')$, contradicting the assumption that A was essential.

Having proved Claim 4.6 in Cases 1 and 2, from now on, we assume that A intersects $\mathcal{Q}$. We also assume that A minimizes the number of intersection curves of an essential annulus of $M \setminus N(L')$ with $\mathcal{Q}$. In particular, since $\mathcal{Q}$ is incompressible, the connected components of $A \setminus \mathcal{Q}$ are either annuli or disks whose boundary intersect ∂A.

Case 3 *Assume that there is an intersection arc in $A \cap \mathcal{Q}$ that cuts a disk from A.*

Let a be an intersection arc in $A \cap \mathcal{Q}$ that cuts a disk E from A. Then both endpoints of a are on the same boundary component of A and $E \cap \mathcal{Q} \subset \partial E$. Because $\mathcal{Q}$ is boundary-incompressible, it must be the case that a cuts a disk E_1 from $\mathcal{Q}$. Then $E_2 = E \cup E_1$ is a disk properly embedded in $M \setminus N(L')$. Since there are no essential disks in $M \setminus N(L')$, then ∂E_2 bounds a disk E_3 in $\partial N(L')$. Then $E_2 \cup E_3$ is a sphere that bounds a ball in $M \setminus N(L')$, through which E can be isotoped to E_1, and just beyond to eliminate a from $A \cap \mathcal{Q}$, contradicting that we assumed a minimal number of intersection components.

Thus, we now know that there are only two possibilities for the intersection curves in $A \cap \mathcal{Q}$. Either they are all parallel nontrivial closed curves on A or they are all arcs with endpoints on distinct boundary components of A.

Case 4 *Assume that $\partial A \cap \mathcal{Q} = \varnothing$, with $A \cap \mathcal{Q} \neq \varnothing$.*

In this case, there are no arcs in $A \cap \mathcal{Q}$. Since A and $\mathcal{Q}$ are incompressible in $M \setminus L'$, the minimality condition on the number of curves in $A \cap \mathcal{Q}$ implies that any curve in $A \cap \mathcal{Q}$ is nontrivial on both A and on $\mathcal{Q}$.

Next, we prove that any curve in $A \cap \mathcal{Q}$ must encircle two of the punctures of $\mathcal{Q}$. Arguing by contradiction, assume that a is a simple closed curve in $A \cap \mathcal{Q}$ and assume that a bounds a once-punctured disk E in $\mathcal{Q}$. Without loss of generality, we may assume that E is innermost in the sense that $E \cap A = a$. Using E to surger A, we obtain two annuli in $M \setminus L'$, where at least one is still essential, and, after a small isotopy, with a lesser number of intersection components with $\mathcal{Q}$, which is a contradiction. Thus, any curve in $A \cap \mathcal{Q}$ encircles two of the punctures of $\mathcal{Q}$ and all intersection curves must be parallel on $\mathcal{Q}$, separating one pair of punctures from the other pair.

Still under the assumption that $\partial A \cap \mathcal{Q} = \varnothing$ and $A \cap \mathcal{Q} \neq \varnothing$, we next rule out the case where at least one boundary component of A, say α_1, lies in $\partial N(C)$. In this case, let A_1 be the connected component of $A \cap \mathcal{B}$ containing α_1 and let $a = \partial A_1 \setminus \alpha_1$ denote the other boundary component of the annulus A_1. Let E be one of the two disks defined by a in $\mathcal{S}$. Then $A_1 \cup E$ is a disk in $\mathcal{B} \setminus N(C)$ which has nontrivial boundary in $\partial N(C)$; hence, α_1 is a longitude. After an isotopy on A_1, we may assume that $\alpha_1 \cap D = \varnothing$, and thus $\partial A_1 \cap D = \varnothing$. Since D is incompressible, we may isotope A_1 in $\mathcal{B} \setminus L'$ to assume that $A_1 \cap D$ does not contain any trivial curves. Moreover, if $\beta \subset A_1 \cap D$ is a nontrivial simple closed curve both in D and in A_1, then β cannot encircle one puncture in D, since this would generate a sphere in $\mathcal{B}$ punctured three times by L', a contradiction. Hence, any curve in $A_1 \cap D$ encircles both punctures of D; this gives rise to solid tori regions in $\mathcal{B} \setminus N(L')$ that can be used to further isotope A_1 in $\mathcal{B} \setminus L'$ to assume that $A_1 \cap D = \varnothing$. In particular, after capping α_1 with a longitudinal disk in $\mathcal{B} \setminus (N(C) \cup A_1)$, it follows that a is the boundary of a disk in $\mathcal{B} \setminus L$.

Since any other curve in $A \cap \mathcal{Q}$ must be parallel to a, $A \cap \mathcal{Q}$ is a family $\{a_1, a_2, \ldots, a_n\}$ of pairwise disjoint simple closed curves, all parallel to each other both in $\mathcal{Q}$ and in A. In particular, for each $i \in \{1, 2, \ldots, n\}$, a_i generates $\pi_1(A)$ and bounds a disk $E_i \subset \mathcal{B} \setminus L$, punctured once by the arc α. Note that $n \geq 2$, since otherwise $\alpha_2 \subset \partial N(J)$, where J is a component of L and then capping A with a disk in $\mathcal{B} \setminus L$ bounded by α_1 would yield an essential disk in X. This implies that there exists a subannulus $A_2 \subset A \setminus \mathcal{B}$ with boundary $\partial A_2 \subset \mathcal{Q}$. Let us assume that $\partial A_2 = a_1 \cup a_2$. Then (after possibly isotoping the disks E_1 and E_2 in $\mathcal{B} \setminus L$ so they become disjoint) $S = A_2 \cup E_1 \cup E_2$ is a sphere in X, which bounds a ball $B \subset X$. Let $V = B \setminus \mathcal{B}$, then V is a solid torus in $X \setminus \mathcal{B} = Y \setminus \mathcal{B}$ and we may use V to isotope A in Y to reduce the number of intersection components in $A \cap \mathcal{Q}$, a contradiction.

At this point in the proof of Case 4 of Claim 4.6, it remains to rule out the case where no boundary component of A is on $\partial N(C)$. Then $\partial A \cap \mathcal{B} = \varnothing$, since otherwise a boundary component of A would be a meridian in $\partial N(\Gamma)$ and we could isotope A to reduce the number of intersection components in $A \cap \mathcal{Q}$. Next, we show that, after an isotopy, $A \cap D = \varnothing$. Indeed, since D and A are both incompressible, after a disk replacement argument we may assume that any curve in $A \cap D$ is a simple closed curve that generates $\pi_1(A)$ and either encircles one or two of the punctures of D. If there is a curve $a \subset A \cap D$, we may assume that either a encircles one puncture of D and is innermost or that a encircles the two punctures of D and is outermost. In either case, we can surger A to obtain two annuli in Y, where at least one is still essential in Y and

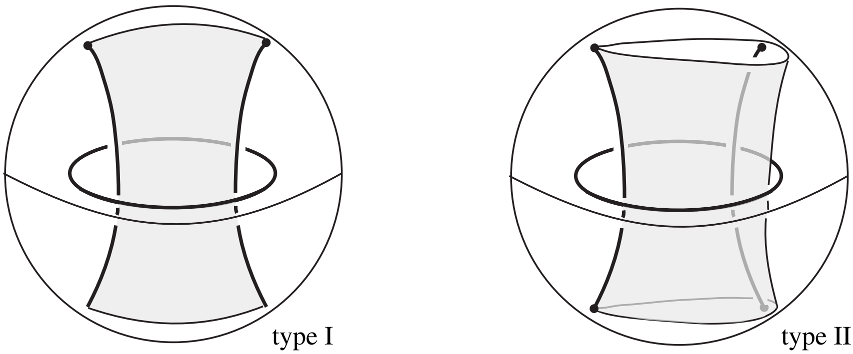


Figure 10: Possibilities for E, a connected component of $A \cap \mathcal{B}$ when all intersections of $A \cap \mathcal{Q}$ are arcs.

with less intersection components with $\mathcal{Q}$, a contradiction that proves that $A \cap D = \varnothing$. Next, using $\mathcal{Q} \times [0,1]$ as a coordinate system for $\mathcal{B} \setminus N(D \cup L')$, we can isotope A in Y to make A disjoint from $\mathcal{B}$. Since we already showed that there are no essential annuli in Y disjoint from $\mathcal{B}$, this is a contradiction.

Case 5 *Assume that each intersection curve in $A \cap \mathcal{Q}$ is an arc with endpoints on distinct boundary components of A.*

The arcs in $A \cap \mathcal{Q}$ cut A into a collection of disks. Because $\mathcal{S}$ separates M, there must be an even number of such arcs and hence such disks, and the disks must alternate between lying inside and outside $\mathcal{B}$.

There are no such arcs that cut a disk from $\mathcal{Q}$. Indeed, if there were such a disk, by choosing an innermost one, we could surger A along this disk to obtain a disk Δ with boundary in $\partial N(L')$. Since there are no essential disks in $M \setminus N(L')$, $\partial\Delta$ must bound a disk Δ' on $\partial N(L')$. Then $\Delta \cup \Delta'$ is a sphere bounding a ball in $M \setminus N(L')$. Thus we can isotope Δ to Δ' through the ball, and hence isotope A to an annulus in $\partial N(L')$, contradicting the fact that A is not boundary parallel in $M \setminus N(L')$.

Let E be a connected component of $A \cap \mathcal{B}$, which necessarily is a disk. Next, we show that there are two possibilities for E up to isotopy and switching the roles of γ_1 and γ_2. These two possibilities are depicted in Figure 10.

Let $\partial E = \beta_1 \cup \mu_1 \cup \beta_2 \cup \mu_2$ where β_1 and β_2 lie in $\mathcal{Q}$ and μ_1 and μ_2 lie in $\partial N(L')$. Note that each of μ_1 and μ_2 must begin and end at distinct components of $\partial N(\Gamma) \cap \mathcal{Q}$, since otherwise we could lower the number of intersection curves of A with $\mathcal{Q}$.

For the arguments that follow, we set coordinates and consider

$$\mathcal{B} = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq 1\},$$

D a horizontal disk in $\{z = 0\}$ and the two arcs γ_1 and γ_2 parallel to the z–axis. Let A' be the annular connected component of $(\mathcal{B} \setminus N(C)) \cap \{z = 0\}$. Then A' is annulus with one boundary component in $\mathcal{Q}$ and the other boundary component a longitude on $\partial N(C)$.

We assume that we have isotoped E in $\mathcal{B} \setminus N(L')$ to minimize the number of intersection curves in $E \cap A'$, and next we prove that $E \cap A' = \varnothing$. First, we claim that $E \cap A'$ does not contain any arc. Indeed, if there were an arc $\phi \subset E \cap A'$, since $\partial E \cap \partial N(C) = \varnothing$, ϕ would cut a disk H_1 from A' and a disk H_2 from E. Let $H_3 = H_1 \cup H_2$. If ϕ has both endpoints in the same β_i, then $\partial H_3 \subset \mathcal{Q}$, which, by incompressibility of $\mathcal{Q}$, implies that ∂H_3 is a trivial curve bounding a disk $H_4 \subset \mathcal{Q}$. Then $H_3 \cup H_4$ is a sphere bounding a ball, through which we can isotope H_1 through H_2, and lower the number of intersection curves in $E \cap A'$, a contradiction.

If ϕ has one endpoint in β_1 and the other in β_2, then ∂H_3 consists of one arc in $\mathcal{Q}$ and one taut arc on $\partial N(\gamma_i)$. Then we can use H_3 to isotope γ_i to $\mathcal{Q}$, a contradiction to the fact that C links γ_i in $\mathcal{B}$. So $E \cap A'$ can only contain simple closed curves.

If $\phi \subset E \cap A'$ is a simple closed curve, then there is a disk $E' \subset E$ with $\partial E' = \phi$. In particular, ϕ is nontrivial in A', since otherwise we could use a disk replacement argument to isotope E removing ϕ from $E \cap A'$. Since ϕ is isotopic to C through A', we could obtain a disk in Y with nontrivial boundary in $\partial N(C)$, a contradiction. Thus, $A' \cap E = \varnothing$.

If μ_1 and μ_2 lie on $\partial N(\gamma_1)$ and $\partial N(\gamma_2)$ respectively, then by an isotopy on

$$\mathcal{Q} \cup \partial N(\gamma_1 \cup \gamma_2),$$

we can assume that μ_1 and μ_2 are vertical arcs that do not wind around $\partial N(\gamma_1)$ or $\partial N(\gamma_2)$. Then because β_1 and β_2 cannot cross the equator $\partial A' \cap \mathcal{Q}$, after possibly reindexing, β_1 connects the top two punctures of $\mathcal{Q}$ and β_2 connects the bottom two punctures. Since ∂E must be trivial as an element of the fundamental group of the handlebody $\mathcal{B} \setminus N(\gamma_1 \cup \gamma_2)$, there can be no twisting around the punctures, and E must appear as in Figure 10, left.

If μ_1 and μ_2 both lie on $\partial N(\gamma_1)$, then β_1 and β_2 are loops on $\mathcal{Q}$ based at a puncture and restricted to the upper and lower hemisphere. Since no arcs in $A \cap \mathcal{Q}$ can cut disks off $\mathcal{Q}$, each β_1 and β_2 circle a puncture in $\mathcal{Q}$. Hence, E must appear as in Figure 10, right. A similar case occurs when μ_1 and μ_2 both lie on $\partial N(\gamma_2)$.

This argument allows us to introduce the following language. If $\beta \subset A \cap \mathcal{Q}$ is any arc, then there is a unique disk $E \subset A \cap \mathcal{B}$ with $\beta \subset \partial E$. If E is a type I disk (where type I and type II are defined as in Figure 10), we shall say that β is a type I arc. Otherwise, we will say that β is a type II arc.

Next, we show that all intersection arcs in $A \cap \mathcal{Q}$ are of the same type. If $\Gamma_1 \neq \Gamma_2$, then, if there exists a type I disk, the two boundaries of A are on different components and only type I disks can occur. If there is not a type I disk, then all disks are type II. On the other hand, if $\Gamma_1 = \Gamma_2$, both components of ∂A intersect $\partial N(\gamma_1)$ and $\partial N(\gamma_2)$ the same equal number of times. In this case, the existence of a type II disk E_1 with boundary intersecting $\partial N(\gamma_1)$ in two arcs, implies that there exists a type II disk E_2 intersecting $\partial N(\gamma_2)$. But E_1 and E_2 would then intersect, a contradiction.

Assume that all arcs in $A \cap \mathcal{Q}$ are of type I and let E be a connected component of $A \setminus \mathcal{B}$. Then, when we switch from L' to L, E can be extended to a disk properly embedded in $M \setminus N(L)$. Thus, there is a trivial component in L, a contradiction to its hyperbolicity.

Our next argument eliminates the last case when all intersections of $A \cap \mathcal{Q}$ are of type II, and $\Gamma_1 \neq \Gamma_2$, since we cannot mix the two types of type II intersections. Until the end of the proof we will assume that $\partial A \subset \partial N(\Gamma_1)$. Let $E \subset A$ be a connected component of $A \setminus \mathcal{B}$, and we label $\partial E = \beta_1 \cup \mu_1 \cup \beta_2 \cup \mu_2$, where β_1 and β_2 lie in $\mathcal{Q}$ and μ_1 and μ_2 lie in $\partial N(\Gamma_1)$. Then μ_1 and μ_2 define two disks, Δ and $\widetilde{\Delta}$, in the annulus $\partial N(\Gamma_1) \setminus \mathcal{B}$. We assume that the disk Δ is the one that makes $\widetilde{A} = E \cup \Delta$ an annulus in $Y \setminus \mathcal{B}$ with both boundary components in $\mathcal{Q}$ parallel to the punctures that come from Γ_2.

After capping $\widetilde{A}$ with the two once-punctured disks bounded by $\partial \widetilde{A}$ in $\mathcal{Q}$, we create an incompressible annulus $\widehat{A}$ in $Y \setminus \mathcal{B}$ which also lives and is incompressible in $X \setminus \mathcal{B}$. Since X is hyperbolic, it follows that $\widehat{A}$ must be boundary-parallel to Γ_2. But this implies that μ_1 is parallel in $X \setminus \mathcal{B}$ to the arc $j_2 = \Gamma_2 \setminus \mathcal{B}$, and there exists a disk $E' \subset X \setminus \mathcal{B}$ with $\partial E' = \mu_1 \cup \nu_1 \cup j_2 \cup \nu_2$, where $\nu_1 \cup \nu_2 = E' \cap \mathcal{Q}$ are two arcs joining the respective two upper punctures and the two lower punctures of $\mathcal{Q}$ which avoid the equator of $\mathcal{Q}$. It then follows that $\nu_1 \cup g$ and $\nu_2 \cup g'$ bound two respective disks in $\mathcal{B} \setminus L$, and the union of those disks with E' is an essential disk in $M \setminus L$, contradicting hyperbolicity of X and finishing the proof of Claim 4.6. □

Claim 4.7 *Y does not admit essential tori.*

Proof We argue by contradiction and suppose that T is a torus which is incompressible and not boundary-parallel in Y. First, suppose that $T \cap D = \varnothing$. Then, after an isotopy in Y, we may assume that $T \cap \mathcal{B} = \varnothing$. Hence, $T \subset X$ and, since X is hyperbolic, T is either compressible or boundary parallel in X. If T is boundary parallel, since both G and G' intersect $\mathcal{B}$, T must be parallel to a component J of L which lives in L', contradicting that T is essential in Y.

Next, we treat the case where T is compressible in X; let $E \subset X$ be a compression disk for T and assume that E has the least number of intersection curves with $\mathcal{Q}$ among compression disks for T. Since T is incompressible in Y, E intersects $\mathcal{B} \cap L'$ and the arc α, which is a complete geodesic in the hyperbolic metric of X. Let $\overline{N}(E) \subset X$ be a closed neighborhood of E with coordinates $E \times [0,1]$ and such that $(\partial E \times [0,1]) \subset T$. Since α is transverse to E, we may choose such a coordinate system on $\overline{N}(E)$ in such a way that, for each $t \in [0,1]$, each component of $\alpha \cap \overline{N}(E)$ intersects $E \times \{t\}$ transversely in a single point.

Let $S = (T \setminus (\partial E \times [0,1])) \cup (E \times \{0\}) \cup (E \times \{1\})$. Then S is a sphere in X and $T \setminus S = \partial E \times (0,1)$. Since X is hyperbolic, S separates and must bound a closed ball $B \subset X$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the arcs in $\alpha \cap \overline{N}(E)$. We claim that each α_i is contained in B. This follows because the endpoints of α are in L, $L \cap B = \varnothing$ and $T \cap \mathcal{B} = \varnothing$. In particular, $\overline{N}(E) \subset B$.

Let $W = B \setminus \overline{N}(E)$. Then $\partial W = T$ and W is a knot exterior in B bounded by T (in fact, we think of W as obtained from B by removing a potentially knotted hole; see Figure 11). Since $T \cap L' = \varnothing$ and L' intersects $\overline{N}(E)$, we have $W \subset Y$. Our next argument is to show that $\overline{N}(E)$ is unknotted in B; thus W is a solid torus bounded by T, which contradicts the essentiality of T in Y.

Let $\Pi\colon \mathbb{H}^3 \to X$ be the Riemannian universal covering map of X. By appropriately choosing a neighborhood $N(L)$, it follows that $\Pi^{-1}(\partial N(L))$ is a collection of horospheres in $\mathbb{H}^3$. Moreover, $\Pi^{-1}(\alpha)$ is a collection of geodesics connecting these horospheres. On the other hand, B lifts to a collection of balls, one of which is a ball $\widetilde{B}$, containing a lift of W, denoted by $\widetilde{W}$.

In order for T to be incompressible in Y, a lift of α, which we denote by $\widetilde{\alpha}$, must pass through the hole $\widetilde{B} \setminus \widetilde{W}$ in $\widetilde{B}$. Since $\widetilde{\alpha}$ is a geodesic in $\mathbb{H}^3$, it follows that $\widetilde{\alpha}$ is unknotted, which implies that $\widetilde{W}$ is a solid torus. Since W is homeomorphic to $\widetilde{W}$, this gives a contradiction, as previously explained.

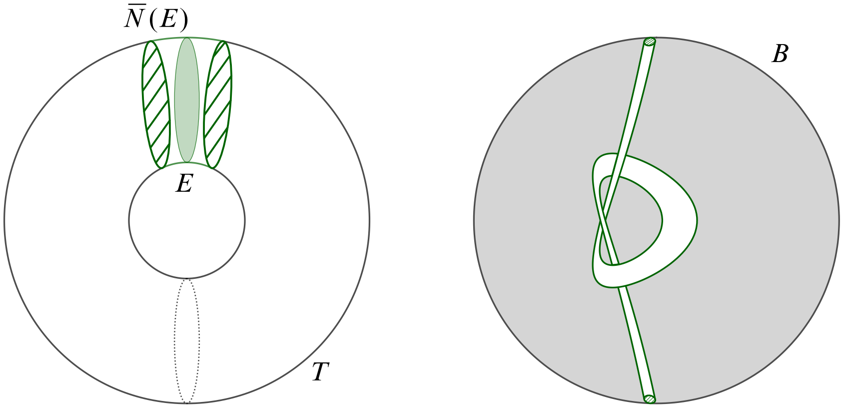


Figure 11: Left: the compressing disk E for the torus T and the neighborhood $\overline{N}(E)$. Right: the sphere S bounding the ball B and the (possibly knotted) region $\overline{N}(E) \subset B$, which defines the highlighted knot exterior $W = B \setminus \overline{N}(E)$.

It remains to prove that there are no essential tori in Y which intersect D. Arguing by contradiction, assume that T is such a torus. Since D is incompressible in Y, a disk replacement argument allows us to further assume that any curve in $T \cap D$ is nontrivial both in T and in D. Let β be a curve in $T \cap D$. If β encircles one puncture of D, take an innermost curve in such intersection and use the one-punctured disk it bounds in D to surger T and obtain an essential annulus in Y with boundary in $\partial N(\Gamma_i)$. If β encircles both punctures of D, take an outermost curve on $T \cap D$ and use the outer annulus on D to surger T and obtain an essential annulus with boundary in $\partial N(C)$. Since Claim 4.6 gives that Y does not admit essential annuli, this proves Claim 4.7. □

Thus, having proved that Y satisfies Thurston's hyperbolicity conditions, Theorem 4.1 follows when M is orientable.

Next, we assume that M is nonorientable and that L, L', α and $\mathcal{B}$ are as before. Let $\Pi\colon \widehat{M} \to M$ be the two-sheeted oriented covering map of M. Then $\widehat{L} = \Pi^{-1}(L)$ is a hyperbolic link in $\widehat{M}$. We also let $\widehat{L}'' = \Pi^{-1}(L')$, $\mathcal{B}_1$ and $\mathcal{B}_2$ be the connected components of $\Pi^{-1}(\mathcal{B})$ and α_1 and α_2 be the connected components of $\Pi^{-1}(\alpha)$. We claim that $\widehat{Y} = \widehat{M} \setminus \widehat{L}''$ is also hyperbolic. Note that, as explained in the proof of the nonorientable setting for the chain move, the fact that $\widehat{Y}$ is hyperbolic implies that $Y = M \setminus L'$ is hyperbolic.

Since α is a complete geodesic in the hyperbolic metric of $M \setminus L$, both α_1 and α_2 are complete geodesics in $\widehat{M} \setminus \widehat{L}$. In particular, since $\widehat{M}$ is orientable, the switch move allows us to replace $\widehat{L} \cap \mathcal{B}_1$ by a tangle diffeomorphic to $L' \cap \mathcal{B}$ to create a new hyperbolic link $\widehat{L}'$ in $\widehat{M}$. Note that $\widehat{L}''$ may be obtained from $\widehat{L}'$ by replacing the tangle $\widehat{L}' \cap \mathcal{B}_2 = \widehat{L} \cap \mathcal{B}_2$ by a tangle diffeomorphic to $L \cap \mathcal{B}$.

Since it might be the case that α_2 is not isotopic to a geodesic in the hyperbolic metric of $\widehat{X} = \widehat{M} \setminus \widehat{L}'$, one cannot directly apply the switch move a second time. However, most of the arguments in its proof can be repeated without change for this setting. We next guide the reader over the steps in the proof that need some adaptation.

First, the arguments in the proof of the orientable case for the switch move can be used to prove that the four-punctured sphere $\mathcal{Q}_1 = \partial\mathcal{B}_1 \setminus \widehat{L}''$ is incompressible and boundary-incompressible in $\widehat{X}$ and in $\widehat{Y} = \widehat{M} \setminus \widehat{L}''$, that $\mathcal{Q}_2 = \partial\mathcal{B}_2 \setminus \widehat{L}''$ is incompressible and boundary-incompressible in $\widehat{Y}$ and that $\widehat{Y}$ does not admit any essential disks and essential spheres.

To prove that $\widehat{Y}$ does not admit any essential annuli, the arguments in Claim 4.6 apply to show that if A is an essential annulus in $\widehat{Y}$, then both boundary components of A are meridians in a component $\widehat{G}'$ of $\widehat{L}'$ that intersects $\mathcal{B}_2$ and that we may isotope A in $\widehat{Y}$ to assume that $A \cap \mathcal{B}_2 = \varnothing$. Since $\widehat{X} \setminus \mathcal{B}_2 = \widehat{Y} \setminus \mathcal{B}_2$, A is an incompressible annulus in $\widehat{X}$, and A must be boundary-parallel in $\widehat{M} \setminus N(\widehat{L}')$. In particular, after an isotopy in $\widehat{Y}$ that does not change the property $A \cap \mathcal{B}_2 = \varnothing$, we may assume that $\partial A \cap \mathcal{B}_1 = \varnothing$ and that if A intersects $\mathcal{B}_1$, then each connected component of $A \cap \mathcal{B}_1$ is an annulus parallel to one of the two arcs in the tangle $\widehat{L}' \cap \mathcal{B}_1$.

If $A \cap \mathcal{B}_1 = \varnothing$, A is an incompressible annulus in the hyperbolic manifold $\widehat{M} \setminus \widehat{L}$, and the same arguments in the proof of Claim 4.6 apply to show that the neighborhood through which A is parallel to an annulus A' in $\partial N(\widehat{L})$ can be capped off by meridional disks to define a ball W' in $\widehat{M}$ that contains both $\mathcal{B}_1$ and $\mathcal{B}_2$ and may be used to homotope the arcs α_1 and α_2 to $\partial N(\widehat{L})$, a contradiction with the fact that both α_1 and α_2 are geodesics in the hyperbolic metric of $\widehat{M} \setminus \widehat{L}$.

Hence, there must exist A_0 a connected component of $A \cap \mathcal{B}_1$. We assume that A_0 is innermost in the sense that no other component of $A \cap \mathcal{B}_1$ lies in the ball region defined by A_0 in $\mathcal{B}_1$. Then each boundary component of A_0 is a curve in $\mathcal{Q}_1$ that encircles one puncture, defining a once-punctured disk in $\mathcal{Q}_1$. Using these two once-punctured disks to surger A gives three incompressible annuli in $\widehat{M} \setminus N(\widehat{L}')$, all disjoint from $\mathcal{B}_2$. One of them lies in $\mathcal{B}_1$ and at least one of the other two must be essential in $\widehat{Y}$. By induction on the number of components in $A \cap \mathcal{B}_1$, this argument yields an essential annulus $\widehat{A}$ in $\widehat{Y}$, with both boundary components being meridians, and that is disjoint both from $\mathcal{B}_1$ and from $\mathcal{B}_2$. As already shown, this is a contradiction that proves that $\widehat{Y}$ does not admit any essential annuli.

The proof that $\widehat{Y}$ does not admit any essential tori uses the arguments in Claim 4.7. Among all possible essential tori, the only case that still needs an adaptation is when V is an essential torus in $\widehat{Y}$ that can be isotoped to be disjoint from $\mathcal{B}_2$. Let D_1 be a twice punctured disk in $\mathcal{B}_1 \setminus \widehat{L}'$ bounded by the trivial component $\widehat{C}_1$ of $\widehat{L}' \cap \mathcal{B}_1$. Then D_1 is incompressible and we may isotope V in $\widehat{Y}$ to assume that there are no trivial curves in $V \cap D_1$. Hence, $V \cap D_1 = \varnothing$, since the existence of a nontrivial curve in $V \cap D_1$ allows us to surger V to produce an essential annulus in $\widehat{Y}$, which we already proved that cannot exist. In particular, V can also be isotoped in $\widehat{Y}$ to be disjoint from $\mathcal{B}_1$, and then V is a torus in the hyperbolic manifold $\widehat{M} \setminus \widehat{L}$. Since V cannot be boundary parallel in $\widehat{M} \setminus \widehat{L}$, there exists a compressing disk E for V in $\widehat{M} \setminus \widehat{L}$, and the fact that V is incompressible in $\widehat{Y}$ implies that E must necessarily intersect the arcs α_1 and α_2, which are geodesics in the hyperbolic metric of $\widehat{M} \setminus \widehat{L}$. Now, the same arguments in the proof of Claim 4.7 apply to show that V bounds a unknotted solid region W in $\widehat{Y}$, contradicting the fact that V is essential in $\widehat{Y}$. This argument finishes the proof that $\widehat{Y}$ satisfies Thurston's hyperbolicity conditions and, as already explained, proves the switch move theorem for the nonorientable case. $\square$

5 The switch move gluing operation

We describe in Theorem 5.1 below a method to obtain new hyperbolic 3–manifolds of finite volume from previously given ones; this method uses a variant of the switch move (Theorem 4.1). Before stating this result, we set the notation.

Let P be a compact 3–manifold with nonempty genus one boundaries and let L be a link in P. We allow for P to consist of one connected manifold or two connected manifolds. Let $M = P \setminus L$ and assume that $\operatorname{int}(M)$ admits a complete hyperbolic metric of finite volume. Let T_1 and T_2 be two distinct, diffeomorphic components of ∂P and, for $i \in \{1, 2\}$, let α_i be a complete geodesic in the hyperbolic metric of $\operatorname{int}(M)$, with one endpoint in T_i and the other in a component J_i of L. Note that J_1 can equal J_2. Let $\Phi\colon T_1 \to T_2$ be a gluing diffeomorphism that maps the endpoint of α_1 in T_1 to the endpoint of α_2 in T_2. Let $P' = P/\Phi$ be the manifold obtained from P by identifying T_1 and T_2 to a genus one surface T using Φ. Note that P' is compact (possibly with empty boundary, if $\partial P = T_1 \cup T_2$), connected and that $L \subset P'$. Let $X = P' \setminus L$.

Theorem 5.1 (switch move gluing operation) *With the above notation, let α be the concatenation of α_1 and α_2^{-1} in P'. Let $\mathcal{B}$ be a ball neighborhood of α in P' that intersects L in two arcs $g \subset J_1$ and $g' \subset J_2$ and intersects T in a disk Δ. Let L' be the resulting link obtained in P' by replacing $g \cup g'$ by the tangle $\gamma_1 \cup \gamma_2 \cup C$ as in Figure 4, right. Then the manifold $Y = P' \setminus L'$ is hyperbolic.*

After choosing ϕ as above, as in the case of the switch move, the operation described above may yield two distinct hyperbolic 3–manifolds depending on the projection of the strands g and g'; see Remark 1.1.

Proof of Theorem 5.1 We first prove the theorem in the case when P is orientable. In this circumstance, the setting in Theorem 5.1 is the same as in the switch move theorem (Theorem 4.1), with the exception that X is no longer hyperbolic. However, X is close to being hyperbolic in the following sense:

Claim 5.2 *X does not admit any essential spheres, essential disks and essential annuli. Moreover, any essential torus in X is isotopic to T.*

Proof Suppose there were an essential disk E in X. Since $\operatorname{int}(M)$ is hyperbolic, it follows that E must intersect T. But because T is incompressible and ∂E is disjoint from T, we may replace subdisks in E by disks in T to obtain an essential disk in M, a contradiction. The same argument shows that an essential sphere in X would generate an essential sphere in M, also a contradiction.

Because $\operatorname{int}(M)$ is hyperbolic, an essential torus in X that does not intersect T must be parallel to T, and hence isotopic to T. To prove that T is the only possible essential torus up to isotopy, we argue by contradiction. Suppose T' is an essential torus in X that is not isotopic to T and has the fewest number of intersection curves with T. Then any curve in $T' \cap T$ is nontrivial both in T and in T'. It follows that there is a component of $T' \setminus T$ that is an essential annulus in $\operatorname{int}(M)$, a contradiction. Analogously, we may show that X does not admit any essential annuli, and this proves Claim 5.2. □

Having proved Claim 5.2, we observe that $L \setminus \mathcal{B}$ is not a rational tangle in a 3–ball as X contains an essential torus that intersects the 3–ball in an essential punctured torus, which cannot exist in a rational tangle complement. Then we note that the arguments in the proof of the switch move theorem apply directly to show that Y does not admit any essential spheres or essential disks and that the four-punctured sphere $\mathcal{Q} = \partial\mathcal{B} \setminus L = \partial\mathcal{B} \setminus L'$ and the twice-punctured disk D bounded by C on T are incompressible and boundary-incompressible in Y.

To prove that Y does not admit any essential annuli, the proof of Claim 4.6 applies directly. Hence, to prove Theorem 5.1 when M is orientable, it remains to show that Y does not admit any essential tori.

We argue by contradiction and assume that V is an essential torus in $M \setminus L'$ that has the least number of intersection components with T among all essential tori in Y. Then, after assuming general position, $T \cap V$ is a finite collection of pairwise disjoint simple closed curves. Let γ be one of such intersection components. If $\gamma \subset D$, then it does not bound a disk in $T \setminus L'$ and either encircles one or two of the punctures of D. Then we can choose a component γ' in $V \cap D$ (if γ encircles one puncture, we choose γ' as an innermost curve, otherwise we choose γ' as an outermost curve) and surger V (along a punctured disk in the first case and an annulus in the second case) to obtain an essential annulus in Y, a contradiction. It then follows that $V \cap D = \varnothing$, and then V can be isotoped through Y to be disjoint from $\mathcal{B}$, without increasing the number of intersection components in $V \cap T$. Hence, V is a torus that is contained in X.

In X, V is not isotopic to T, since $V \subset M \setminus L'$ and any torus isotopic to T is punctured by L'. We claim that $V \cap T = \varnothing$. Argue by contradiction and assume that there exists a curve γ in $V \cap T$. Then γ does not intersect D and there are two possibilities: either γ is a nontrivial curve in T or γ, together with C, bounds an annulus in $T \setminus D$. In the latter case, we may use this annulus to surger V and obtain an essential annulus in Y. Since Y does not admit essential annuli, γ is nontrivial in T. Then there is a component of $V \setminus T$ that is an essential annulus in X, which cannot occur, proving that $V \cap T = \varnothing$.

Thus, V is a torus in the hyperbolic manifold int(M). Note that V cannot be boundary parallel in M, since this either contradicts its essentiality in Y or the fact that it is not isotopic to T. Hence, it must be the case that V is compressible in M. Let $E \subset M$ be a compression disk for V. Then, since V is incompressible in Y, the geodesic α_1 must intersect E. Now, the same arguments used in the proof of Claim 4.7 for the case when T was an essential torus in Y, disjoint from $\mathcal{B}$ and compressible in X apply to obtain that V is compressible in Y, a contradiction that proves Theorem 5.1 when M is orientable.

Next, we sketch the arguments that prove Theorem 5.1 when M is nonorientable, using the notation already introduced. Let $\widehat{X}$ be the oriented double cover of X and let $\Pi\colon \widehat{X} \to X$ be the covering map. Let $\widehat{T} = \Pi^{-1}(T)$ in $\widehat{X}$. Then $\widehat{T}$ consists of one or two tori.

In the case of two tori V_1 and V_2, $\Pi^{-1}(\alpha)$ consists of two arcs μ_1 and μ_2 each of which intersects one of the two tori. Since $\widehat{X}$ is orientable, we may apply Theorem 5.1 twice first for V_1 and μ_1 and then, in an equivariant manner to the first switch move with respect to the covering translation, for V_2 and μ_2, to obtain a hyperbolic link complement. Then, by Mostow–Prasad rigidity, the covering translation can be realized as an isometry, proving that the switch move gluing operation on the original nonorientable manifold M yields a hyperbolic manifold.

In the case that $\widehat{T}$ consists of one torus, both copies μ_1 and μ_2 intersect $\widehat{T}$. In this situation, we may apply Theorem 5.1 on μ_1, obtaining a hyperbolic link complement where Theorem 4.1 can be performed, in an equivariant manner with respect to the covering translation, on a neighborhood of μ_2 and again the result is hyperbolic. Realizing the covering translation as an isometry allows us to prove that the switch move gluing operation on the nonorientable manifold M yields a hyperbolic manifold Y. □

References

[1] **C C Adams**, *Thrice-punctured spheres in hyperbolic 3–manifolds*, Trans. Amer. Math. Soc. 287 (1985) 645–656 MR Zbl

[2] **C C Adams**, *Augmented alternating link complements are hyperbolic*, from "Low-dimensional topology and Kleinian groups" (D B A Epstein, editor), London Math. Soc. Lecture Note Ser. 112, Cambridge Univ. Press (1986) 115–130 MR Zbl

[3] **C C Adams**, *Isometric cusps in hyperbolic 3–manifolds*, Michigan Math. J. 46 (1999) 515–531 MR Zbl

[4] **C C Adams**, *The knot book*: *an elementary introduction to the mathematical theory of knots*, Amer. Math. Soc., Providence, RI (2004) MR Zbl

[5] **C Adams**, **O Eisenberg**, **J Greenberg**, **K Kapoor**, **Z Liang**, **K O'Connor**, **N Pacheco-Tallaj**, **Y Wang**, *tg-hyperbolicity of virtual links*, J. Knot Theory Ramifications 28 (2019) 1950080, 26 MR Zbl

[6] **C Adams**, **W H Meeks III**, **A K Ramos**, *Totally umbilic surfaces in hyperbolic 3–manifolds of finite volume*, preprint (2020) arXiv 2007.03166

[7] **C C Adams**, **A W Reid**, *Unknotting tunnels in two-bridge knot and link complements*, Comment. Math. Helv. 71 (1996) 617–627 MR Zbl

[8] **F Bonahon**, **L C Siebenmann**, *New geometric splittings of classical knots and the classification and symmetries of arborescent knots*, unpublished manuscript (2016) Available at `https://dornsife.usc.edu/assets/sites/1191/docs/Preprints/BonSieb.pdf`

[9] **D Futer**, **F Guéritaud**, *Angled decompositions of arborescent link complements*, Proc. Lond. Math. Soc. 98 (2009) 325–364 MR Zbl

[10] **J R Goldman**, **L H Kauffman**, *Rational tangles*, Adv. in Appl. Math. 18 (1997) 300–332 MR Zbl

[11] **A Hatcher**, *Notes on basic* 3*–manifold topology*, unpublished manuscript (2007) Available at `https://pi.math.cornell.edu/~hatcher/3M/3Mdownloads.html`

[12] **W Menasco**, *Closed incompressible surfaces in alternating knot and link complements*, Topology 23 (1984) 37–44 MR Zbl

[13] **W D Neumann**, **A W Reid**, *Arithmetic of hyperbolic manifolds*, from "Topology '90" (B Apanasov, W D Neumann, A W Reid, L Siebenmann, editors), Ohio State Univ. Math. Res. Inst. Publ. 1, de Gruyter, Berlin (1992) 273–310 MR Zbl

[14] **U Oertel**, *Closed incompressible surfaces in complements of star links*, Pacific J. Math. 111 (1984) 209–230 MR Zbl

[15] **J S Purcell**, *An introduction to fully augmented links*, from "Interactions between hyperbolic geometry, quantum topology and number theory" (A Champanerkar, O Dasbach, E Kalfagianni, I Kofman, W Neumann, N Stoltzfus, editors), Contemp. Math. 541, Amer. Math. Soc., Providence, RI (2011) 205–220 MR Zbl

[16] **K Reif Volz**, *Hyperbolicity of arborescent tangles and arborescent links*, Topology Appl. 156 (2009) 963–978 MR Zbl

Department of Mathematics, Williams College
Williamstown, MA, United States

Department of Mathematics and Statistics, University of Massachusetts
Amherst, MA, United States

Departamento de Matemática Pura e Aplicada, Universidade Federal do Rio Grande do Sul
Porto Alegre, Brazil

`cadams@williams.edu`, `profmeeks@gmail.com`, `alvaro.ramos@ufrgs.br`

Received: 30 December 2020 Revised: 3 February 2022

Algebraic & Geometric Topology **23**:5 (2023) 2191–2212
DOI: 10.2140/agt.2023.23.2191
Published: 25 July 2023

Golod and tight 3–manifolds

KOUYEMON IRIYE
DAISUKE KISHIMOTO

The notions Golodness and tightness for simplicial complexes come from algebra and geometry, respectively. We prove these two notions are equivalent for 3–manifold triangulations, through a topological characterization of a polyhedral product for a tight-neighborly manifold triangulation of dimension ≥ 3.

57Q15; 13F55, 55U10

1 Introduction

Let $\mathbb{F}$ be a field, and let $S = \mathbb{F}[x_1, \dots, x_m]$, where we assume each x_i is of degree 2. Serre [26] proved that for $R = S/I$ where I is a homogeneous ideal of S, there is a coefficientwise inequality

$$P(\mathrm{Tor}^R(\mathbb{F}, \mathbb{F}); t) \leq \frac{(1+t^2)^m}{1 - t(P(\mathrm{Tor}^S(R, \mathbb{F}); t) - 1)},$$

where $P(V; t)$ denotes the Poincaré series of a graded vector space V. In the extreme case that the equality holds, R is called *Golod*. It was Golod who proved that R is Golod if and only if all products and (higher) Massey products in the Koszul homology of R vanish, where the Koszul homology of R is isomorphic to $\mathrm{Tor}^S(R, \mathbb{F})$ as a vector space.

Let K be a simplicial complex with vertex set $[m] = \{1, 2, \dots, m\}$. Let $\mathbb{F}[K]$ denote the Stanley–Reisner ring of K over $\mathbb{F}$, where we assume generators of $\mathbb{F}[K]$ are of degree 2. Then $\mathbb{F}[K]$ expresses combinatorial properties of K, and conversely, it is of particular interest to translate a given algebraic property of the Stanley–Reisner ring $\mathbb{F}[K]$ into a combinatorial property of K. We say that K is $\mathbb{F}$–*Golod* if $\mathbb{F}[K]$ is Golod. We aim to characterize Golod complexes combinatorially.

Recently, a new approach to a combinatorial characterization of Golod complexes has been taken. We can construct a space Z_K, called the *moment–angle complex* for K, in accordance with the combinatorial information of K. Then combinatorial properties are encoded in the topology of Z_K, and in particular, Golodness can be read from a homotopical property of Z_K as follows. Baskakov, Buchstaber and Panov [6] proved that the cohomology of Z_K with coefficients in $\mathbb{F}$ is isomorphic to the Koszul homology of $\mathbb{F}[K]$, where the isomorphism respects products and (higher) Massey products. Then it follows that K is Golod over any field whenever Z_K is a suspension, and so Golod complexes have been studied also in connection with desuspension of Z_K and a more general *polyhedral product*; see Grbić, Panov, Theriault and Wu [10; 11], Grujić and Welker [12], and the authors [14; 15; 16; 17; 18; 19]. See the survey by Bahri, Bendersky and Cohen [4] for more information about moment–angle complexes and polyhedral products. Here we remark that there is a Golod complex K such that Z_K is not a suspension as shown by Yano and the first author [20].

In [15; 17; 19], the authors characterized Golod complexes of dimension one and two in terms of both combinatorial properties of K and desuspension of Z_K. Here we recall the characterization of Golodness of a closed connected surface triangulation, proved in [15]. The original statement in [15] is given in terms of polyhedral products, but here we state in terms of moment–angle complexes, which is easier, as in [17, Theorem 1.3]. Recall that a simplicial complex is called *neighborly* if every pair of vertices forms an edge.

Theorem 1.1 [15, Theorem 1.1] *Let S be a triangulation of a closed connected $\mathbb{F}$–orientable surface. Then the following statements are equivalent*:

(1) *S is $\mathbb{F}$–Golod.*
(2) *S is neighborly.*
(3) *Z_S is a suspension.*

We introduce another notion of simplicial complexes coming from geometry. S-S Chern and R K Lashof proved that the total absolute curvature of an immersion $f: M \to \mathbb{R}^n$ of a compact manifold M is bounded below by the Morse number of some Morse function on M. On the other hand, the Morse number is bounded below by the Betti number. Tightness of an immersion f is defined by the equality between the total absolute curvature of an immersion f and the Betti number of M, which is the case that the total absolute curvature is minimal. See Kühnel and Lutz [22] and Kuiper [23]. It is known that an immersion f is tight if and only if for almost every closed half-space H, the inclusion $f(M) \cap H \to f(M)$ is injective in homology.

Tightness of a simplicial complex is defined as a combinatorial analog of tightness of an immersion. See [22] for details. Let K be a simplicial complex with vertex set $[m]$.

For $\emptyset \neq I \subset [m]$, the full subcomplex of K over I is defined by

$$K_I = \{\sigma \in K \mid \sigma \subset I\}.$$

Definition 1.2 Let K be a connected simplicial complex with vertex set $[m]$. We say that K is $\mathbb{F}$–*tight* if the natural map $H_*(K_I;\mathbb{F}) \to H_*(K;\mathbb{F})$ is injective for each $\emptyset \neq I \subset [m]$.

Golodness and tightness have origins in different fields of mathematics, algebra and geometry, respectively. The aim of this paper is to prove the seemingly irrelevant these two notions are equivalent for 3–manifold triangulations through the topology of Z_K or more general polyhedral products (see Section 5). Now we state the main theorem.

Theorem 1.3 *Let M be a triangulation of a closed connected $\mathbb{F}$–orientable 3–manifold. Then the following statements are equivalent*:

(1) *M is $\mathbb{F}$–Golod.*

(2) *M is $\mathbb{F}$–tight.*

(3) *Z_M is a suspension.*

Recall that a d–manifold triangulation is called *stacked* if it is the boundary of a $(d+1)$–manifold triangulation whose interior simplices are of dimension $\geq d$. Stacked manifold triangulations have been studied in several directions, and we will use its connection to tightness (Section 2). See Bagchi, Datta, Murai and Spreer [3; 9] and [22] for more on stacked manifold triangulations. Bagchi, Datta and Spreer [3] (cf Theorem 2.3) proved that a closed connected $\mathbb{F}$–orientable 3–manifold triangulation is $\mathbb{F}$–tight if and only if it is neighborly and stacked. Then we get the following corollary of Theorem 1.3, which enables us to compare with Theorem 1.1, the 2–dimensional case.

Corollary 1.4 *Let M be a triangulation of a closed connected $\mathbb{F}$–orientable 3–manifold. Then the following statements are equivalent*:

(1) *M is $\mathbb{F}$–Golod.*

(2) *M is neighborly and stacked.*

(3) *Z_M is a suspension.*

We will investigate a relation between Golodness and tightness of d–manifold triangulations for $d \geq 3$, not only for $d = 3$, through tight-neighborliness. We will prove the following theorem, where Theorem 1.3 is its special case $d = 3$.

Theorem 1.5 *Let M be a triangulation of a closed connected $\mathbb{F}$–orientable d–manifold for $d \geq 3$, and consider the following conditions:*

(1) *M is $\mathbb{F}$–Golod.*

(2) *M is $\mathbb{F}$–tight.*

(3) *M is tight-neighborly.*

(4) *the fat-wedge filtration of $\mathbb{R}Z_M$ is trivial.*

Then there are implications

$$(1) \implies (2) \impliedby (3) \implies (4) \implies (1).$$

Moreover, for $d = 3$, the implication $(2) \implies (3)$ also holds, so all conditions are equivalent.

Remarks on Theorem 1.5 are in order. Tight-neighborly triangulations of d–manifolds for $d \geq 3$ will be defined in Section 2. To clarify a connection to Theorem 1.3 and Corollary 1.4, we need to mention that a triangulated manifold of dimension ≥ 3 is tight-neighborly if and only if it is neighborly and stacked as noted soon before Theorem 2.3 below. The space $\mathbb{R}Z_K$ is the real moment–angle complex, and properties of its fat-wedge filtration will be given in Section 5. In particular, we will see that if the fat-wedge filtration of $\mathbb{R}Z_K$ is trivial, then Z_K is a suspension. So Theorem 1.3 is the special case of Theorem 1.5 for $d = 3$ as mentioned above. Datta and Murai [9] proved that if M is tight-neighborly and $d \geq 4$, then it is $\mathbb{F}$–tight and $\beta_i(M;\mathbb{F}) = 0$ for $2 \leq i \leq d-2$, where $\beta_i(M;\mathbb{F}) = \dim H_i(M;\mathbb{F})$ denotes the i^{th} Betti number. So if $\beta_i(M;\mathbb{F}) = 0$ for $2 \leq i \leq d-2$ and $d \geq 4$, then all conditions in Theorem 1.5 are equivalent, where the triviality of the Betti numbers is necessary because as in [2, Example 3.15], there is an $\mathbb{F}$–tight 9–vertex triangulation of $\mathbb{C}P^2$ for any field $\mathbb{F}$, which is not tight-neighborly.

The paper is organized as follows. Section 2 collects properties of tight and tight-neighborly manifold triangulations that will be needed in later sections. Section 3 introduces a weak version of Golodness and proves that weak Golodness implies tightness of orientable manifold triangulations. Section 4 investigates a simplicial complex $F(M)$ constructed from a tight-neighborly d–manifold triangulation M for $d \geq 3$, and Section 5 recalls the fat-wedge filtration technique for polyhedral products, which is the main ingredient in desuspending Z_K. Section 6 applies the results in Sections 4 and 5 to prove Theorem 1.5. Finally, Section 7 poses a problem on Golodness and tightness of d–manifold triangulations for $d \geq 4$, and shows related results.

Acknowledgements Iriye was supported by JSPS KAKENHI grant JP19K03473, and Kishimoto was supported by JSPS KAKENHI grant JP17K05248.

2 Tightness

This section collects facts about tight and tight-neighborly manifold triangulations that we will use. As mentioned in Section 1, tightness of a simplicial complex is a discrete analog of a tight space studied in differential geometry with connection to minimality of the total absolute curvature, and tight complexes have been studied mainly for manifold triangulations. First, we show:

Lemma 2.1 *Every $\mathbb{F}$–tight complex is neighborly.*

Proof Let K be an $\mathbb{F}$–tight complex. Then for two vertices v and w of K, the natural map $H_0(K_{\{v,w\}};\mathbb{F}) \to H_0(K;\mathbb{F})$ is injective. Since K is connected, $H_0(K;\mathbb{F}) \cong \mathbb{F}$, and so $H_0(K_{\{v,w\}};\mathbb{F}) \cong \mathbb{F}$. Then v and w must be joined by an edge. □

Next, we explain a conjecture on tight manifold triangulations. Let K be a simplicial complex. Let $|K|$ denote its geometric realization of K, and let

$$f(K) = (f_0(K), f_1(K), \dots, f_{\dim K}(K))$$

denote the f–vector of K. We say that K is *strongly minimal* if for any simplicial complex L with $|K| \cong |L|$, it holds that

$$f_i(K) \le f_i(L)$$

for each $i \ge 0$. Kühnel and Lutz [22] conjectured that every $\mathbb{F}$–tight triangulation of a closed connected manifold is strongly minimal. Clearly, the only $\mathbb{F}$–tight closed connected 1–manifold triangulation is the boundary of a 2–simplex, so the conjecture is true in dimension 1. Moreover, the 2–dimensional case was verified, as mentioned in [22], and the 3–dimensional case was verified by Bagchi, Datta and Spreer [3]. But the case of dimensions ≥ 4 is still open.

As for minimality of manifold triangulations, we have another notion introduced by Lutz, Sulanke and Swartz [24].

Definition 2.2 A closed connected d–manifold triangulation M with vertex set $[m]$ for $d \ge 3$ is *tight-neighborly* if

$$\binom{m-d-1}{2} = \binom{d+2}{2}\beta_1(M;\mathbb{F}).$$

Tight-neighborly manifold triangulations are known to be vertex minimal. By definition, tight-neighborliness seems to depend on the ground field $\mathbb{F}$, but it is actually independent of the ground field $\mathbb{F}$ as tight-neighborly manifold triangulations are neighborly and stacked. Tightness and tight-neighborliness have the following relation. Let $S^1 \tilde{\times} S^{d-1}$ denote the nontrivial S^{d-1}–bundle over S^1.

Theorem 2.3 *Let M be a closed connected $\mathbb{F}$–orientable d–manifold triangulation for $d \geq 3$, and consider the following conditions*:

1. *M is $\mathbb{F}$–tight.*
2. *M is tight-neighborly.*
3. *M is neighborly and stacked.*
4. *M has the topological type of either*

$$S^d, \quad (S^1 \tilde{\times} S^{d-1})^{\#k}, \quad (S^1 \times S^{d-1})^{\#k}.$$

Then there are implications

$$(1) \impliedby (2) \iff (3) \implies (4).$$

Moreover, the implication $(1) \implies (2)$ *also holds for* $d = 3$.

Proof The implications are shown in [9] for $d \geq 4$ and [3] for $d = 3$. □

Remark The integer k in Theorem 2.3 for $d = 3$ is known to satisfy $80k + 1$ is a perfect square. For $k = 1, 30, 99, 208, 357, 546$, tight-neighborly triangulations of $(S^1 \tilde{\times} S^2)^{\#k}$ are constructed in [8], but no tight-neighborly triangulation of $(S^1 \times S^2)^{\#k}$ is known.

3 Weak Golodness

This section introduces weak Golodness and studies it for manifold triangulations. Let K be a simplicial complex with vertex set $[m]$, and let $\mathcal{H}_*(\mathbb{F}[K])$ denote the Koszul homology of the Stanley–Reisner ring $\mathbb{F}[K]$. As mentioned in Section 1, K is $\mathbb{F}$–Golod if and only if all products and (higher) Massey products in $\mathcal{H}_*(\mathbb{F}[K])$ vanish. Now we define weak Golodness.

Definition 3.1 A simplicial complex K is *weakly* $\mathbb{F}$*–Golod* if all products in $\mathcal{H}_*(\mathbb{F}[K])$ vanish.

Clearly, K is weakly $\mathbb{F}$–Golod whenever it is $\mathbb{F}$–Golod. Berglund and Jöllenbeck [7] stated that Golodness and weak Golodness of every simplicial complex are equivalent, but this was disproved by Katthän [21]. Thus defining weak Golodness makes sense.

We recall a combinatorial description of the multiplication in $\mathcal{H}_*(\mathbb{F}[K])$. For disjoint nonempty subsets $I, J \subset [m]$, there is an inclusion

$$\iota_{I,J}\colon K_{I\sqcup J} \to K_I * K_J, \quad \sigma \mapsto (\sigma \cap I, \sigma \cap J).$$

Baskakov, Buchstaber and Panov proved:

Lemma 3.2 [6, Theorem 1] *There is an isomorphism of vector spaces*

$$\mathcal{H}_i(\mathbb{F}[K]) \cong \bigoplus_{\varnothing \neq I \subset [m]} \widetilde{H}^{i-|I|-1}(K_I;\mathbb{F})$$

for $i > 0$ such that for nonempty subsets $I, J \subset [m]$ the multiplication

$$\widetilde{H}^{i-|I|-1}(K_I;\mathbb{F}) \otimes \widetilde{H}^{j-|J|-1}(K_J;\mathbb{F}) \to \widetilde{H}^{i+j-|I\cup J|-1}(K_{I\cup J};\mathbb{F})$$

is trivial for $I \cap J \neq \varnothing$ and the induced map of $\iota_{I,J}$ for $I \cap J = \varnothing$.

Let M be a triangulation of a closed connected $\mathbb{F}$–oriented d–manifold with vertex set $[m]$. We consider a relation between the inclusion $\iota_{I,J}$ and Poincaré duality. For any subset $I \subset [m]$, Poincaré duality [13, Proposition 3.46] holds such that the map

$$H^i(|M_I|;\mathbb{F}) \to H_{d-i}(|M|, |M| - |M_I|;\mathbb{F}), \quad \alpha \mapsto \alpha \frown [M]$$

is an isomorphism, where $[M]$ denotes the fundamental class of M. By Lemma 70.1 of [25], $|M| - |M_I| \simeq |M_J|$ for $J = [m] - I$. Then there is an isomorphism

$$D_{I,J}\colon H^i(M_I;\mathbb{F}) \xrightarrow{\cong} H_{d-i}(M, M_J;\mathbb{F}).$$

Let $\partial\colon H_*(M, M_J;\mathbb{F}) \to H_{*-1}(M_J;\mathbb{F})$ denote the boundary map of the long exact sequence

$$\cdots \to H_*(M_J;\mathbb{F}) \to H_*(M;\mathbb{F}) \to H_*(M, M_J;\mathbb{F}) \xrightarrow{\partial} H_{*-1}(M_J;\mathbb{F}) \to \cdots.$$

Lemma 3.3 *Let M be a triangulation of a closed connected $\mathbb{F}$–oriented d–manifold with vertex set $[m]$. For any partition $[m] = I \sqcup J$ and $\alpha \in H^i(M_I;\mathbb{F})$,*

$$(\partial \circ D_{I,J})(\alpha) = (-1)^{i+1}(\alpha \otimes 1)((\iota_{I,J})_*([M])) \in H_{d-i-1}(M_J;\mathbb{F}),$$

where we regard $(\iota_{I,J})_*([M])$ as an element of

$$\bigoplus_{i+j=d-1} H_i(M_I;\mathbb{F}) \otimes H_j(M_J;\mathbb{F}) \cong H_d(M_I * M_J;\mathbb{F}).$$

Proof Let $\varphi \in C^i(M_I;\mathbb{F})$ be a representative of α. We define $\bar{\varphi} \in C^i(M;\mathbb{F})$ by

$$\bar{\varphi}(\sigma) = \begin{cases} \varphi(\sigma) & \text{if } \sigma \in M_I, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\alpha \frown [M]$ is represented by $\bar{\varphi} \frown \mu$ where μ represents $[M]$. Let $[v_0,\dots,v_i]$ denote an oriented i–simplex with vertices $v_0,\dots,v_i$. We may set

$$\mu = \sum_k a_k[v_0^k, v_1^k, \dots, v_d^k] \in C_d(M;\mathbb{F})$$

for $a_k \in \mathbb{F}$, where $v_0^k,\dots,v_{n_k}^k \in I$ and $v_{n_k+1}^k,\dots,v_d^k \in J$ for some n_k. Then $(\partial \circ D_{I,J})(\alpha)$ is represented by

$$\partial(\bar{\varphi} \frown \mu) = (\bar{\varphi} \circ \partial) \frown \mu = [(\bar{\varphi} \circ \partial) \frown \mu] = \sum_k a_k \bar{\varphi}(\partial[v_0^k,\dots,v_{i+1}^k])[v_{i+1}^k,\dots,v_d^k].$$

Since $(\bar{\varphi} \circ \partial)|_{C_{i+1}(M_I;\mathbb{F})} = \varphi \circ \partial = 0$, we have $\bar{\varphi}(\partial[v_0^k,\dots,v_{i+1}^k]) \neq 0$ only when $n_k = i$. Then $(\partial \circ D_{I,J})(\alpha)$ is represented by

$$\sum_{n_k=i} a_k \bar{\varphi}(\partial[v_0^k,\dots,v_{i+1}^k])[v_{i+1}^k,\dots,v_d^k] = (-1)^{i+1} \sum_{n_k=i} a_k \varphi([v_0^k,\dots,v_i^k,\widehat{v_{i+1}^k}])[v_{i+1}^k,\dots,v_d^k].$$

On the other hand, since the $C_i(M_I;\mathbb{F}) \otimes C_{d-i-1}(M_J;\mathbb{F})$ part of μ is given by $\sum_{n_k=i} a_k[v_0^k,\dots,v_d^k]$, $(\iota_{I,J})_*([M])$ is represented by

$$\sum_{n_k=i} a_k[v_0^k,\dots,v_i^k] \otimes [v_{i+1}^k,\dots,v_d^k]. \qquad \square$$

Now we are ready to prove:

Theorem 3.4 *If a triangulation of a closed connected $\mathbb{F}$–orientable d–manifold is weakly $\mathbb{F}$–Golod, then it is $\mathbb{F}$–tight.*

Proof Let M be a triangulation of a closed connected $\mathbb{F}$–oriented d–manifold with vertex set $[m]$. Let $[m] = I \sqcup J$ be a partition. Suppose that the map $\iota_{I,J}$ is trivial in cohomology with coefficients in $\mathbb{F}$. Then by the universal coefficient theorem, $\iota_{I,J}$ is trivial in homology with coefficients in $\mathbb{F}$ too. Thus, by Lemma 3.3, the boundary map

$$\partial\colon H_*(M, M_J;\mathbb{F}) \to H_{*-1}(M_J;\mathbb{F})$$

is trivial, and so the natural map $H_*(M_J;\mathbb{F}) \to H_*(M;\mathbb{F})$ is injective, completing the proof. $\square$

4 The complex $F(M)$

Throughout this section, let M be a closed connected tight-neighborly d–manifold triangulation for $d \geq 3$ with vertex set $[m]$. Let K be a simplicial complex with vertex set $[m]$. A subset $I \subset [m]$ is a *minimal nonface* of K if every proper subset of I is a simplex of K and I itself is not a simplex of K. Define a simplicial complex $F(M)$ by filling all minimal nonfaces of cardinality $d+1$ into M. This section investigates the complex $F(M)$.

We set notation. The link of a vertex v in a simplicial complex K is defined by

$$\mathrm{lk}_K(v) = \{\sigma \in K \mid v \notin \sigma \text{ and } \sigma \sqcup v \in K\}.$$

For a finite set S, let $\Delta(S)$ denote the simplex with vertex set S. Then $I \subset [m]$ is a minimal nonface of K if and only if $K_I = \partial\Delta(I)$. Let K_1 and K_2 be simplicial complexes of dimension d such that $K_1 \cap K_2$ is a single d–simplex σ. Then we write

$$K_1 \# K_2 = K_1 \cup K_2 - \sigma \quad \text{and} \quad K_1 \circ K_2 = K_1 \cup K_2.$$

The following lemma may be known, but we produce a proof for completeness of the paper; cf [1; 3; 9].

Lemma 4.1 *For each $v \in [m]$, there exist $V(v,1), \dots, V(v,n_v) \subset [m]$ such that $|V(v,k)| = d+1$ for $1 \leq k \leq n_v$ and*

$$\mathrm{lk}_M(v) = \partial\Delta(V(v,1)) \# \cdots \# \partial\Delta(V(v,n_v)).$$

Proof The case $d=3$ is proved in [3, Proof of Theorem 1.2]. For $d \geq 4$, tight-neighborliness implies local stackedness, that is, every vertex link is a stacked sphere, as in [9]. Moreover, stacked spheres are characterized by Bagchi and Datta [1] such that every stacked $(d-1)$–sphere is of the form $\partial\Delta^d \# \cdots \# \partial\Delta^d$. Then we obtain the result for $d \geq 4$. □

Generalizing neighborliness, we say that a simplicial complex is *k–neighborly* if every $k+1$ vertices form a simplex. So 1–neighborliness is precisely neighborliness.

Lemma 4.2 *For each $v \in [m]$ and $1 \leq k \leq n_v$, $M_{V(v,k)\sqcup v}$ is $(d-1)$–neighborly.*

Proof By Lemma 4.1, $\mathrm{lk}_M(v)_{V(v,k)}$ is $\partial\Delta^d$ with some $(d-1)$–simplices removed, implying it is $(d-2)$–neighborly. So if I is a subset of $V(v,k)$ with $|I| = d-1$, then $I \sqcup v$ is a simplex of M. It remains to show $M_{V(v,k)}$ is $(d-1)$–neighborly. Let J be any subset of $V(v,k)$ with $|J|=d$. Then $\partial\Delta(J)$ is a subcomplex of M. If $M_J=\partial\Delta(J)$, then $M_{J\sqcup v}=\partial\Delta(J)*v$, which is contractible. So the inclusion $M_J\to M_{J\sqcup v}$ is not injective in homology with coefficients in $\mathbb{F}$. By Theorem 2.3, M is $\mathbb{F}$–tight, so we get a contradiction. Thus J must be a simplex of M, completing the proof. □

We prove local properties of the complex $F(M)$.

Proposition 4.3 (1) *For each $v\in[m]$,*

$$\mathrm{lk}_{F(M)}(v)=\partial\Delta(V(v,1))\circ\cdots\circ\partial\Delta(V(v,n_v)).$$

(2) *For each $v\in[m]$ and $1\le k\le n_v$, $V(v,k)\sqcup v$ is a minimal nonface of $F(M)$.*

Proof (1) Let σ be the $(d-1)$–simplex

$$(\partial\Delta(V(v,1))\#\cdots\#\partial\Delta(V(v,k)))\cap\partial\Delta(V(v,k+1)).$$

Then by Lemma 4.2, $\partial\Delta(\sigma\sqcup v)$ is a subcomplex of M, implying $\sigma\sqcup v$ is a simplex of $F(M)$. Then by induction, we get $\partial\Delta(V(v,1))\circ\cdots\circ\partial\Delta(V(v,n_v))\subset\mathrm{lk}_{F(M)}(v)$. The reverse inclusion is obvious by the construction of $F(M)$, completing the proof.

(2) By Lemma 4.2, $V(v,k)$ is a simplex of $F(M)$, so every proper subset I of $V(v,k)\sqcup v$ is a simplex of $F(M)$. By (1), $V(v,k)\sqcup v$ is not a simplex of $F(M)$. □

We compute the homology of $F(M)$. Let

$$S(M)=\{V(v,k)\sqcup v\mid v\in[m]\text{ and }1\le k\le n_v\}.$$

Then $S(M)$ is the set of all subsets $I\subset[m]$ such that $|I|=d+2$ and $\mathrm{lk}_{M_I}(v)$ is $(d-2)$–neighborly for some $v\in I$.

Lemma 4.4

$$F(M)=\bigcup_{I\in S(M)}\partial\Delta(I).$$

Proof Let $K=\bigcup_{I\in S(M)}\partial\Delta(I)$. By Proposition 4.3, $K\subset F(M)$. For any k–simplex σ of $F(M)$ with $0\le k\le d-1$ and $v\in\sigma$, $\sigma-v$ is a simplex of $\mathrm{lk}_M(v)$ because σ is a simplex of M too. Then $\sigma-v\subset V(v,l)$ for some $1\le l\le n_v$, implying σ is a simplex of K. Thus the $(d-1)$–skeleton of $F(M)$ is included in K. Take any d–simplex σ of $F(M)$. Then σ is either a simplex or a minimal nonface of M. In both cases, $\partial\Delta(\sigma-v)$ is a subcomplex of $\mathrm{lk}_M(v)$ for $v\in\sigma$. Then $\sigma-v\subset V(v,l)$ for some $1\le l\le n_v$, implying σ is a simplex of K. Thus $F(M)\subset K$. □

By Lemma 4.4, there is an inclusion $g_I\colon\partial\Delta(I)\to F(M)$ for each $I\in S(M)$. Let $u_I\in H_d(F(M);\mathbb{Z})$ be the Hurewicz image of g_I.

Proposition 4.5 *The integral homology of $F(M)$, except for dimension* 1, *is given by*

$$\widetilde{H}_*(F(M);\mathbb{Z}) = \begin{cases} \mathbb{Z}\langle u_I \mid I \in S(M)\rangle & \text{if } * = d, \\ 0 & \text{if } * \neq 1, d. \end{cases}$$

Proof Since $F(M)$ is obtained from M by attaching d–simplices, we only need to calculate H_{d-1} and H_d by Theorem 2.3. By Proposition 4.3, each component of $\mathrm{lk}_{M_I}(v)$ is $(d-2)$–connected, where $\mathrm{lk}_{M_I}(v) = \mathrm{lk}_M(v)_{I-v}$. Then there is an exact sequence

$$(1)\quad 0 \to \widetilde{H}_d(F(M)_{I-v};\mathbb{Z}) \to H_d(F(M)_I;\mathbb{Z}) \xrightarrow{\partial} H_{d-1}(\mathrm{lk}_{F(M)_I}(v);\mathbb{Z}) \\ \to H_{d-1}(F(M)_{I-v};\mathbb{Z}) \to H_{d-1}(F(M)_I;\mathbb{Z}) \to 0.$$

By Proposition 4.3, there is an inclusion $\partial\Delta(V(v,k)) \to \mathrm{lk}_{F(M)_I}(v)$ for $V(v,k)\sqcup v \subset I$, and we write the Hurewicz image of this inclusion by $\bar{u}_{V(v,k)}$. Then we have

$$H_{d-1}(\mathrm{lk}_{F(M)_I}(v);\mathbb{Z}) = \mathbb{Z}\langle \bar{u}_{V(v,k)} \mid V(v,k)\sqcup v \subset I\rangle$$

such that $\partial(u_{V(v,k)\sqcup v}) = \bar{u}_{V(v,k)}$. Hence the map ∂ in (1) is surjective, so we get an isomorphism

$$H_{d-1}(F(M)_{I-v};\mathbb{Z}) \cong H_{d-1}(F(M)_I;\mathbb{Z}).$$

Thus we obtain $H_{d-1}(F(M)_I;\mathbb{Z}) = 0$ for any $I \subset [m]$ by induction on $|I|$, where $H_{d-1}(F(M)_I;\mathbb{Z}) = 0$ for $|I| = 1$. We also get a split exact sequence

$$0 \to H_d(F(M)_{I-v};\mathbb{Z}) \to H_d(F(M)_I;\mathbb{Z}) \xrightarrow{\partial} H_{d-1}(\mathrm{lk}_{F(M)_I}(v);\mathbb{Z}) \to 0.$$

Then by induction on $|I|$, we also obtain

$$H_d(F(M)_I;\mathbb{Z}) = \mathbb{Z}\langle u_{V(v,k)} \mid V(v,k)\sqcup v \subset I\rangle. \qquad \square$$

By Theorem 2.3, $\pi_1(|M|)$ is a free group. Since $|F(M)|$ is obtained by attaching d–cells to $|M|$, the inclusion $|M| \to |F(M)|$ is an isomorphism in π_1, so $\pi_1(|F(M)|)$ is a free group too. Then there is a map $f\colon B \to |F(M)|$ which is an isomorphism in π_1, where B is a wedge of circles. Let $\widehat{F}(M)$ be the cofiber of f. Since there is an exact sequence

$$\cdots \to H_*(B;\mathbb{Z}) \xrightarrow{f_*} H_*(F(M);\mathbb{Z}) \to \widetilde{H}_*(\widehat{F}(M);\mathbb{Z}) \to \cdots,$$

the natural map $H_*(F(M);\mathbb{Z}) \to H_*(\widehat{F}(M);\mathbb{Z})$ is an isomorphism for $* \neq 1$. Let $\hat{g}_I$ be the composite $|\partial\Delta(I)| \xrightarrow{g_I} |F(M)| \to \widehat{F}(M)$ for $I \in S(M)$, and let $\hat{u}_I$ be the Hurewicz image of $\hat{g}_I$. By Proposition 4.5, we get:

Corollary 4.6 *The reduced homology of $\widehat{F}(M)$ is given by*

$$\widetilde{H}_*(\widehat{F}(M);\mathbb{Z}) = \begin{cases} \mathbb{Z}\langle \hat{u}_I \mid I \in S(M)\rangle & \text{if } * = d, \\ 0 & \text{if } * \neq d. \end{cases}$$

Since $\widehat{F}(M)$ is path-connected, there is a map

$$g\colon \bigvee_{I\in S(M)} |\partial\Delta(I)| \to \widehat{F}(M)$$

such that $g|_{|\partial\Delta(I)|} \simeq \hat{g}_I$ for each $I \in S(M)$. Then by Corollary 4.6 and the Whitehead theorem, we obtain the following.

Corollary 4.7 *The map $g\colon \bigvee_{I\in S(M)} |\partial\Delta(I)| \to \widehat{F}(M)$ is a homotopy equivalence.*

5 Polyhedral product

Throughout this section, let K be a simplicial complex with vertex set $[m]$. Let $(\underline{X},\underline{A}) = \{(X_i, A_i)\}_{i=1}^m$ be a collection of pairs of pointed spaces indexed by vertices of K. For $I \subset [m]$, let

$$(\underline{X},\underline{A})^I = Y_1 \times \cdots \times Y_m$$

where $Y_i = X_i$ for $i \in I$ and $Y_i = A_i$ for $i \notin I$. The *polyhedral product* of $(\underline{X},\underline{A})$ over K is defined by

$$Z_K(\underline{X},\underline{A}) = \bigcup_{\sigma\in K} (\underline{X},\underline{A})^\sigma.$$

For $\varnothing \neq I \subset [m]$, let $(\underline{X}_I,\underline{A}_I) = \{(X_i, A_i)\}_{i\in I}$. Then we can define $Z_{K_I}(\underline{X}_I,\underline{A}_I)$. The following lemma is immediate from the definition of a polyhedral product.

Lemma 5.1 *For each $\varnothing \neq I \subset [m]$, $Z_{K_I}(\underline{X}_I,\underline{A}_I)$ is a retract of $Z_K(\underline{X},\underline{A})$.*

For a collection of pointed spaces $\underline{X} = \{X_i\}_{i=1}^m$, let $(C\underline{X},\underline{X}) = \{(CX_i, X_i)\}_{i=1}^m$. For $0 \le i \le m$, we define a subspace of $Z_K(C\underline{X},\underline{X})$ by

$$Z_K^i(C\underline{X},\underline{X}) = \{(x_1,\dots,x_m) \in Z_K(C\underline{X},\underline{X}) \mid \text{at least } m-i \text{ of } x_1,\dots,x_m \text{ are basepoints}\}.$$

Using the basepoint of each X_i, we regard $Z_{K_I}(C\underline{X}_I,\underline{X}_I)$ as a subspace of $Z_K(C\underline{X},\underline{X})$ so that we can alternatively write

$$Z_K^i(C\underline{X},\underline{X}) = \bigcup_{I\subset[m],\,|I|=i} Z_{K_I}(C\underline{X}_I,\underline{X}_I). \tag{2}$$

There is a filtration

$$* = Z_K^0(C\underline{X},\underline{X}) \subset Z_K^1(C\underline{X},\underline{X}) \subset \cdots \subset Z_K^m(C\underline{X},\underline{X}) = Z_K(C\underline{X},\underline{X}),$$

which we call the *fat-wedge filtration* of $Z_K(C\underline{X},\underline{X})$. By [17, Theorem 4.1],

$$Z_K^i(C\underline{X},\underline{X})/Z_K^{i-1}(C\underline{X},\underline{X}) = \bigvee_{I\subset[m],\,|I|=i} |\Sigma K_I| \wedge \widehat{X}^I,$$

where $\widehat{X}^I = \bigwedge_{i\in I} X_i$. Moreover, it is shown in [17, Corollary 4.2] that the fat-wedge filtration of $Z_K(C\underline{X},\underline{X})$ splits after a suspension, and the decomposition of Bahri, Bendersky, Cohen and Gitler [5, Theorem 2.2.1] is reproduced as:

Theorem 5.2 (BBCG decomposition) *There is a homotopy equivalence*

$$\Sigma Z_K(C\underline{X},\underline{X}) \simeq \Sigma \bigvee_{\varnothing\neq I\subset[m]} |\Sigma K_I| \wedge \widehat{X}^I.$$

In particular, if the BBCG decomposition desuspends, then $Z_K(C\underline{X},\underline{X})$ itself desuspends. Moreover, if each X_i is a connected CW complex, then the BBCG decomposition desuspends whenever $Z_K(C\underline{X},\underline{X})$ desuspends [17]. Then we aim to desuspend the BBCG decomposition. Desuspension of the BBCG decomposition was studied for specific Golod complexes such as shifted complexes [11; 12; 14] by ad hoc methods, and desuspension for much broader classes of simplicial complexes, including the previous specific simplicial complexes, was proved by using the fat-wedge filtration technique [17].

The moment–angle complex Z_K introduced in Section 1 is the polyhedral product $Z_K(D^2,S^1)$. The *real moment–angle complex* $\mathbb{R}Z_K$ is defined to be the polyhedral product $Z_K(D^1,S^0)$, and we denote its fat-wedge filtration by

$$* = \mathbb{R}Z_K^0 \subset \mathbb{R}Z_K^1 \subset \cdots \subset \mathbb{R}Z_K^m = \mathbb{R}Z_K$$

where we choose the basepoint of $S^0=\{-1,+1\}$ to be -1. The fat-wedge filtration of $\mathbb{R}Z_K$ is proved to be a cone decomposition [17, Theorem 3.1]. For $\varnothing\neq I\subset[m]$, let $j_{K_I}\colon \mathbb{R}Z_{K_I}^{|I|-1} \to \mathbb{R}Z_K^{|I|-1}$ denote the inclusion.

Theorem 5.3 [17, Theorem 3.1] *For each $\varnothing\neq I\subset[m]$, there is a map*

$$\varphi_{K_I}\colon |K_I| \to \mathbb{R}Z_{K_I}^{|I|-1}$$

such that

$$\mathbb{R}Z_K^i = \mathbb{R}Z_K^{i-1} \bigcup_{I\subset[m],\,|I|=i} C|K_I|,$$

where the attaching maps are $j_{K_I}\circ\varphi_{K_I}$.

We say that the fat-wedge filtration of $\mathbb{R}Z_K$ is trivial if φ_{K_I} is nullhomotopic for each $\varnothing \neq I \subset [m]$. We remark that φ_{K_I} is nullhomotopic if and only if $j_{K_I} \circ \varphi_{K_I}$ is, because $\mathbb{R}Z_{K_I}^{|I|-1}$ is a retract of $\mathbb{R}Z_K^{|I|-1}$. The fat-wedge filtration is useful for desuspending the BBCG decomposition because we have the following criterion.

Theorem 5.4 [17, Theorem 1.2] *If the fat-wedge filtration of $\mathbb{R}Z_K$ is trivial, then for any $\underline{X}$, there is a homotopy equivalence*

$$Z_K(C\underline{X}, \underline{X}) \simeq \bigvee_{\varnothing \neq I \subset [m]} |\Sigma K_I| \wedge \hat{X}^I.$$

For $\varnothing \neq I \subset [m]$, define a map $\alpha_I : \mathbb{R}Z_{K_I}^{|I|-1} \to \mathbb{R}Z_K^{m-1}$ by $\alpha_I(x_i \mid i \in I) = (y_1, \dots, y_m)$ such that

$$y_i = \begin{cases} x_i & \text{if } i \in I, \\ +1 & \text{if } i \notin I, \end{cases}$$

for $(x_i \mid i \in I) \in \mathbb{R}Z_{K_I}^{|I|-1}$. Note that α_I is not the natural inclusion because the basepoint of $S^0 = \{-1, +1\}$ is taken to be -1 as mentioned above. For $\varnothing \neq J \subset I \subset [m]$ and $|J| \leq i \leq |I|$, let π denote the composite of projections

$$\mathbb{R}Z_{K_I}^i \to \mathbb{R}Z_{K_J} \to \mathbb{R}Z_{K_J}/\mathbb{R}Z_{K_J}^{|J|-1} = |\Sigma K_J|.$$

By the construction of φ_K, we have:

Lemma 5.5 *For $\varnothing \neq J \subsetneq I \subset [m]$, there is a commutative diagram*

$$\begin{tikzcd} {|K_I|} \arrow[r, "\varphi_{K_I}"] \arrow[d] & \mathbb{R}Z_{K_I}^{|I|-1} \arrow[r, "\pi"] \arrow[d, "\alpha_I"] & {|\Sigma K_J|} \arrow[d, "|\Sigma j|"] \\ {|K|} \arrow[r, "\varphi_K"] & \mathbb{R}Z_K^{m-1} \arrow[r, "\pi"] & {|\Sigma K_{J \sqcup ([m]-I)}|} \end{tikzcd}$$

where $j : K_J \to K_{J \sqcup ([m]-I)}$ is the inclusion.

The following two lemmas, proved in [17, Proof of Theorem 7.2] and [17, Lemma 10.1] respectively, are quite useful in detecting the triviality of φ_K.

Lemma 5.6 *Let $\overline{K}$ be a simplicial complex obtained by filling all minimal nonfaces into K. Then φ_K factors through the inclusion $|K| \to |\overline{K}|$.*

Lemma 5.7 *If $\varphi_{K_I} \simeq *$ for each $\varnothing \neq I \subsetneq [m]$, then the composite*

$$|K| \xrightarrow{\varphi_K} \mathbb{R}Z_K^{m-1} \to \mathbb{R}Z_{K_J} \xrightarrow{\pi} |\Sigma K_J|$$

is nullhomotopic for each $\varnothing \neq J \subsetneq [m]$.

Finally, we estimate the connectivity of $\mathbb{R}Z_K$.

Lemma 5.8 *If K is k–neighborly, then $\mathbb{R}Z_K$ is k–connected.*

Proof The proof can be done by the same calculation as [17, Proposition 5.3]. Here, we give an alternative proof. By definition, $\pi_*(\mathbb{R}Z_K)$ is isomorphic to $\pi_*(\mathbb{R}Z_{K_k})$ for $* \le k$, where K_k denotes the k–skeleton of K. Since K is k–neighborly, $K_k = \Delta_k^{m-1}$. Since Δ_k^{m-1} is shifted, it follows from [14] that there is a homotopy equivalence

$$\mathbb{R}Z_{\Delta_k^{m-1}} \simeq \bigvee_{\varnothing \neq I \subset [m]} |\Sigma(\Delta_k^{m-1})_I|.$$

Since each $|\Sigma(\Delta_k^{m-1})_I|$ is k–connected, the proof is done. □

6 Proof of Theorem 1.5

Throughout this section, let M be a tight-neighborly triangulation of a closed connected $\mathbb{F}$–orientable d–manifold with vertex set $[m]$, unless otherwise is specified. We aim to prove that the fat-wedge filtration of $\mathbb{R}Z_M$ is trivial. First, we compute the fundamental group of $|F(M)_I|$ for $\varnothing \neq I \subset [m]$.

Lemma 6.1 *For each $\varnothing \neq I \subset [m]$, $\pi_1(|F(M)_I|)$ is a free group.*

Proof Since the fundamental group of a suspension is a free group, we prove $|F(M)_I|$ is a suspension by induction on I. For $|I| = 1$, $|F(M)_I|$ is obviously a suspension. Suppose that $|F(M)_{I-v}|$ is a suspension for $v \in I$. Note that

$$F(M)_I = F(M)_{I-v} \cup (\mathrm{lk}_{F(M)_I}(v) * v) \tag{3}$$

where $F(M)_{I-v} \cap (\mathrm{lk}_{F(M)_I}(v) * v) = \mathrm{lk}_{F(M)_I}(v)$. Since $\mathrm{lk}_{F(M)_I}(v) = \mathrm{lk}_{F(M)}(v)_{I-v}$, it follows from Proposition 4.3 that there are inclusions

$$\mathrm{lk}_{F(M)_I}(v) \to (\Delta(V(v,1)) \circ \cdots \circ \Delta(V(v,n_v)))_{I-v} \to F(M)_{I-v}.$$

Since M is neighborly by Theorem 2.3, so is M_{I-v}, implying $F(M)_{I-v}$ is connected. On the other hand, each component of $(\Delta(V(v,1)) \circ \cdots \circ \Delta(V(v,n_v)))_{I-v}$ is contractible. Then the inclusion $|(\Delta(V(v,1)) \circ \cdots \circ \Delta(V(v,n_v)))_{I-v}| \to |F(M)_{I-v}|$ is nullhomotopic, and so the inclusion $|\mathrm{lk}_{F(M)_I}(v)| \to |F(M)_{I-v}|$ is nullhomotopic too. Thus by (3), we get a homotopy equivalence

$$|F(M)_I| \simeq |F(M)_{I-v}| \vee |\Sigma\, \mathrm{lk}_{F(M)_I}(v)|.$$

Since $|F(M)_{I-v}|$ is a suspension by the induction hypothesis, $|F(M)_I|$ turns out to be a suspension, completing the proof. □

Let $\varnothing \neq I \subset [m]$. By Lemma 5.6, the map φ_{M_I} decomposes as

$$|M_I| \to |F(M)_I| \to \mathbb{R}Z_{M_I}^{|I|-1}. \tag{4}$$

By Lemma 6.1, there is a map $f_I : B_I \to |F(M)_I|$, where B_I is a wedge of circles, such that f_I is an isomorphism in π_1. Let $\widehat{F}(M)_I$ denote the cofiber of f_I, where $\widehat{F}(M)_{[m]}$ coincides with $\widehat{F}(M)$ in Section 4. On the other hand, since M is neighborly by Lemma 2.1, so is M_J for any $\varnothing \neq J \subset [m]$. Then by (2) and Lemma 5.8, we can see that $\mathbb{R}Z_{M_I}^{|I|-1}$ is simply connected. In particular, there is a commutative diagram

$$\begin{tikzcd} |F(M)_I| \arrow[r] \arrow[d] & \widehat{F}(M)_I \arrow[d] \\ \mathbb{R}Z_{M_I}^{|I|-1} \arrow[r, equal] & \mathbb{R}Z_{M_I}^{|I|-1} \end{tikzcd} \tag{5}$$

Then by combining (4) and (5), we get:

Lemma 6.2 *For each $\varnothing \neq I \subset [m]$, the map φ_{M_I} factors through the inclusion $|M_I| \to \widehat{F}(M)_I$.*

Proposition 6.3 *For each $\varnothing \neq I \subsetneq [m]$, the map φ_{M_I} is nullhomotopic.*

Proof As is computed in the proof of Proposition 4.5, $\widetilde{H}_*(F(M)_I;\mathbb{Z}) = 0$ unless $* = 1, d$. Thus as well as $\widehat{F}(M)$, we can see that $\widehat{F}(M)_I$ is $(d-1)$–connected. Since $I \neq [m]$, $|M_I|$ is homotopy equivalent to a CW complex of dimension $\leq d-1$. Then we obtain that the inclusion $|M_I| \to \widehat{F}(M)_I$ is nullhomotopic. Thus by Lemma 6.2, the proof is complete. ◻

It remains to show that φ_M is nullhomotopic. By Lemma 5.5, there is a commutative diagram

$$\begin{tikzcd} \bigvee_{I\in S(M)} |M_I| \arrow[r] \arrow[d, "\bigvee_{I\in S(M)} \varphi_{M_I}"'] & |M| \arrow[d, "\varphi_M"] \\ \bigvee_{I\in S(M)} \mathbb{R}Z_{M_I}^{d+1} \arrow[r, "\bigvee_{I\in S(M)} \alpha_I"] & \mathbb{R}Z_M^{m-1} \end{tikzcd}$$

Then since $F(M)_I = \partial\Delta(I)$ for $I \in S(M)$ by Proposition 4.3, we get a commutative diagram

$$\begin{tikzcd} \bigvee_{I\in S(M)} |\partial\Delta(I)| \arrow[r, "\bigvee_{I\in S(M)} g_I"] \arrow[d] & |F(M)| \arrow[d] \\ \bigvee_{I\in S(M)} \mathbb{R}Z_{M_I}^{d+1} \arrow[r, "\bigvee_{I\in S(M)} \alpha_I"] & \mathbb{R}Z_M^{m-1} \end{tikzcd} \tag{6}$$

Juxtaposing the commutative diagrams (5) and (6), we get a commutative diagram

$$\begin{tikzcd} \bigvee_{I\in S(M)} |\partial\Delta(I)| \arrow[r, "g"] \arrow[d] & \hat{F}(M) \arrow[d] \\ \bigvee_{I\in S(M)} \mathbb{R}Z_{M_I}^{d+1} \arrow[r, "\bigvee_{I\in S(M)}\alpha_I"] & \mathbb{R}Z_M^{m-1} \end{tikzcd}$$

and by Corollary 4.7 and Lemma 6.2, we obtain:

Lemma 6.4 *The map $\varphi_M\colon |M| \to \mathbb{R}Z_M^{m-1}$ is homotopic to the composite*

$$|M| \to \hat{F}(M) \xrightarrow{g^{-1}} \bigvee_{I\in S(M)} |\partial\Delta(I)| \to \bigvee_{I\in S(M)} \mathbb{R}Z_{M_I}^{d+1} \xrightarrow{\bigvee_{I\in S(M)}\alpha_I} \mathbb{R}Z_M^{m-1}.$$

We will investigate the composition of maps in Lemma 6.4 by identifying a homotopy set with a homology.

Lemma 6.5 *Let W be a finite wedge of S^d. Then there is an isomorphism of sets*

$$[|M|, W] \cong H^d(M;\mathbb{Z}) \otimes H_d(W;\mathbb{Z})$$

which is natural with respect to maps among finite wedges of S^d.

Proof Since $\dim M = d$, the statement follows from the Hopf degree theorem. □

Lemma 6.6 *For each $v \in I \in S(M)$, the natural map*

$$H^d(M;\mathbb{Z}) \otimes H_{d-1}(M_{I-v};\mathbb{Z}) \to H^d(M;\mathbb{Z}) \otimes H_{d-1}(M_{[m]-v};\mathbb{Z})$$

is injective.

Proof By Lemma 4.2, $|M_{I-v}|$ is contractible or S^{d-1}. In particular, $H_{d-1}(M_{I-v};\mathbb{Z})$ is a free abelian group, and so there is a natural isomorphism

$$H_{d-1}(M_{I-v};\mathbb{F}) \cong H_{d-1}(M_{I-v};\mathbb{Z}) \otimes \mathbb{F}. \tag{7}$$

By definition, $|M_{[m]-v}|$ is $|M|$ removed the open star of v, which is homotopy equivalent to $|M| - v$ by [25, Lemma 70.1]. Then by Theorem 2.3, $|M_{[m]-v}|$ is homotopy equivalent to a wedge of finitely many, possibly zero, copies of S^1 and S^{d-1}. Then $H_*(M_{[m]-v};\mathbb{Z})$ is a free abelian group, and so there is a natural isomorphism

$$H_{d-1}(M_{[m]-v};\mathbb{F}) \cong H_{d-1}(M_{[m]-v};\mathbb{Z}) \otimes \mathbb{F}. \tag{8}$$

Since M is $\mathbb{F}$–tight by Theorem 2.3, the natural map

$$H_{d-1}(M_{I-v};\mathbb{F}) \to H_{d-1}(M_{[m]-v};\mathbb{F})$$

is injective. Then by (7) and (8), the natural map

$$H_{d-1}(M_{I-v};\mathbb{Z})\otimes\mathbb{F} \to H_{d-1}(M_{[m]-v};\mathbb{Z})\otimes\mathbb{F}$$

is injective too. Since both $H_{d-1}(M_{I-v};\mathbb{Z})$ and $H_{d-1}(M_{[m]-v};\mathbb{Z})$ are free abelian groups, the case that M is orientable is proved because $H^d(M;\mathbb{Z})\cong\mathbb{Z}$. If M is nonorientable, then $H^d(M;\mathbb{Z})\cong\mathbb{F}_2$ and the base field $\mathbb{F}$ is of characteristic 2, where $\mathbb{F}_2$ is the field of two elements. Thus the case that M is not orientable is proved too. □

Proposition 6.7 *The map $\varphi_M\colon |M|\to\mathbb{R}Z_M^{m-1}$ is nullhomotopic.*

Proof Note that $m\ge d+2$. Let $\varnothing\ne J\subset I\in S(M)$. By Lemma 4.2, $|M_J|$ is contractible for $|J|\le d$, and $|M_J|$ is contractible or S^{d-1} for $|J|=d+1$. Then by Proposition 6.3, there is a homotopy equivalence

$$\mathbb{R}Z_{M_I}^{d+1}\simeq\bigvee_{v\in I}|\Sigma M_{I-v}|, \tag{9}$$

where $|\Sigma M_{I-v}|$ is contractible or S^d as mentioned above. Let

$$A=\bigvee_{I\in S(M)}\bigvee_{v\in I}|\Sigma M_{I-v}| \quad\text{and}\quad B=\bigvee_{I\in S(M)}\bigvee_{v\in I}|\Sigma M_{[m]-v}|,$$

where $A\simeq\bigvee_{I\in S(M)}\mathbb{R}Z_{M_I}^{d+1}$ by (9). Let $f\colon|M|\to A$ denote the composition of the first three maps in Lemma 6.4. Then it suffices to show f is nullhomotopic. By Lemma 6.5, f is identified with some element ϕ of $H_d(M;\mathbb{Z})\otimes H_d(A;\mathbb{Z})$, so f is nullhomotopic if and only if $\phi=0$.

As in the proof of Lemma 6.6, $|\Sigma M_{[m]-v}|$ is a wedge of finitely many copies of S^2 and S^d for each vertex v of M. Let C_v denote the S^d–wedge part of $|\Sigma M_{[m]-v}|$. Then there is a projection $q_v\colon B\to C_v$. By Lemmas 5.5, 5.7 and 6.4, the composite

$$|M|\xrightarrow{f}A\to|\Sigma M_{I-v}|\to|\Sigma M_{[m]-v}| \tag{10}$$

is nullhomotopic for each $v\in I\in S(M)$. Then by Lemma 6.5, ϕ is mapped to 0 by

$$1\otimes(q_v\circ j)_*\colon H^d(M;\mathbb{Z})\otimes H_d(A;\mathbb{Z})\to H^d(M;\mathbb{Z})\otimes H_d(C_v;\mathbb{Z})$$

for each $v\in I\in S(M)$, where $j\colon A\to B$ denotes the inclusion. Since the map

$$\bigoplus_{v\in I\in S(M)}(q_v)_*\colon H_d(B;\mathbb{Z})\to\bigoplus_{v\in I\in S(M)}H_d(C_v;\mathbb{Z})$$

is an isomorphism, we get $(1 \otimes j_*)(\phi) = 0$. Thus we obtain $\phi = 0$ by Lemma 6.6, completing the proof. □

Now we are ready to prove Theorem 1.5.

Proof of Theorem 1.5 The implications (1) $\Longrightarrow$ (2) $\Longleftarrow$ (3) are proved by Theorems 2.3 and 3.4. The implication (3) $\Longrightarrow$ (4) is proved by Propositions 6.3 and 6.7. If (4) holds, then by Theorem 5.4, Z_M is a suspension. So by the fact that K is $\mathbb{F}$–Golod whenever Z_K is a suspension, as mentioned in Section 1, we obtain the implication (4) $\Longrightarrow$ (1), completing the proof. □

7 A further problem

So far, we have been studying a relationship between Golodness and tightness through tight-neighborliness which perfectly works in dimension 3. However, in dimensions ≥ 4, tight-neighborliness does not work well because it is not equivalent to tightness as mentioned in Section 1. So we pose:

Problem 7.1 What condition on closed connected d–manifold triangulations with $d \geq 4$ guarantees $\mathbb{F}$–Golodness and $\mathbb{F}$–tightness being equivalent?

One approach is to put a topological condition on manifolds. For example, the condition on the Betti number is stated in Section 1. We also have the following theorem, in which manifold triangulations are not tight-neighborly.

Theorem 7.2 *Let M be a triangulation of a closed $(d-1)$–connected $2d$–manifold for $d \geq 2$. Then the following are equivalent*:

1. *M is $\mathbb{F}$–Golod for any field $\mathbb{F}$.*
2. *M is $\mathbb{F}$–tight for any field $\mathbb{F}$.*
3. *M is d–neighborly.*
4. *the fat-wedge filtration of $\mathbb{R}Z_M$ is trivial.*

Proof The implication (1) $\Longrightarrow$ (2) holds by Theorem 3.4 because M is orientable. Suppose M has a minimal nonface I with $|I| \leq d+1$. Then $M_I = \partial\Delta(I)$, implying $H_{|I|-2}(M_I;\mathbb{F}) \neq 0$. Since M is $\mathbb{F}$–tight, the natural map

$$H_{|I|-2}(M_I;\mathbb{F}) \to H_{|I|-2}(M;\mathbb{F})$$

is injective, and since M is $(d-1)$–connected, $\widetilde{H}_*(M;\mathbb{F})=0$ for $*<d$. Then we get a contradiction, so we obtain the implication (2) $\Longrightarrow$ (3). The implication (3) $\Longrightarrow$ (4) follows from [17, Theorem 1.6]. The implication (4) $\Longrightarrow$ (1) holds by the fact that K is $\mathbb{F}$–Golod over any field $\mathbb{F}$ whenever Z_K is a suspension, as mentioned in Section 1. Therefore, the proof is complete. □

In closing the paper, we consider a relation between weak $\mathbb{F}$–Golodness and $\mathbb{F}$–tightness. As proved in Theorem 3.4, weak $\mathbb{F}$–Golodness implies $\mathbb{F}$–tightness for a closed connected $\mathbb{F}$–orientable manifold triangulations. So one might ask whether or not this implication holds for simplicial complexes which are not manifolds. The answer is no. For example, if K is the join of a vertex and the boundary of a simplex, then it is $\mathbb{F}$–Golod for any field $\mathbb{F}$ as the fat-wedge filtration of $\mathbb{R}Z_K$ is trivial but it is not $\mathbb{F}$–tight as in the proof of Lemma 4.2. However, the opposite implication always holds as follows, which shows that Theorem 3.4 is thought of as a "wrong way" implication.

Proposition 7.3 *Let K be a simplicial complex with vertex set $[m]$. If K is $\mathbb{F}$–tight, then it is weakly $\mathbb{F}$–Golod.*

Proof Take any disjoint subsets $\varnothing \neq I, J \subset [m]$. Then there is a map

$$\iota_{I,J}\colon K_{I\sqcup J} \to K_I * K_J$$

as in Section 3. By Lemma 3.2, K is weakly $\mathbb{F}$–Golod if and only if the map $\iota_{I,J}$ is trivial in homology with coefficients in $\mathbb{F}$. Now we suppose K is $\mathbb{F}$–tight. Then $K_{I\sqcup J}$ is $\mathbb{F}$–tight too, and so we only need to consider the case $I\sqcup J=[m]$. By the Künneth theorem, the map

$$(j_I * j_J)_*\colon \widetilde{H}_*(K_I * K_J;\mathbb{F}) \to \widetilde{H}_*(K * K;\mathbb{F})$$

is injective, where $j_I\colon K_I \to K$ denotes the inclusion. Then it suffices to show the composite $(j_I * j_J)\circ \iota_{I,J}$ is nullhomotopic.

Now we may assume $|K|\subset \mathbb{R}^m$ by identifying a simplex $\{i_1,\dots,i_k\}\in K$ with

$$\{t_1e_{i_1}+\cdots+t_ke_{i_k} \mid t_1+\cdots+t_k=1,\ t_1,\dots,t_k\geq 0\},$$

where $e_1,\dots,e_m$ is the standard basis of $\mathbb{R}^m$. We may assume $|K*K|\subset\mathbb{R}^{2m}$ in the same way. Consider a homotopy $h_t^i\colon \mathbb{R}^{2m}\times[0,1]\to\mathbb{R}^{2m}$ defined by

$$\begin{aligned} h_t^i&(x_1,\dots,x_m,y_1,\dots,y_m)\\ &=(x_1,\dots,(1-t)x_i+ty_i,\dots,x_m,y_1,\dots,tx_i+(1-t)y_i,\dots,y_m)\end{aligned}$$

for $(x_1,\dots,x_m,y_1,\dots,y_m)\in\mathbb{R}^{2m}$. Then h_t^i restricts to a homotopy

$$h_t^i\colon |K*K|\times[0,1]\to|K*K|$$

such that for $i\in I$,

$$(j_I*j_J)\circ\iota_{I,J}=h_0^i\circ(j_I*j_J)\circ\iota_{I,J}\simeq h_1^i\circ(j_I*j_J)\circ\iota_{I,J}=(j_{I-i}*j_{J\cup i})\circ\iota_{I-i,J\cup i}.$$

Thus for $v\in[m]$, $(j_I*j_J)\circ\iota_{I,J}\simeq(j_v*j_{[m]-v})\circ\iota_{v,[m]-v}$. Since $|v*K_{[m]-v}|$ is contractible, we get $(j_I*j_J)\circ\iota_{I,J}\simeq *$, completing the proof. □

References

[1] **B Bagchi**, **B Datta**, *On k–stellated and k–stacked spheres*, Discrete Math. 313 (2013) 2318–2329 MR Zbl

[2] **B Bagchi**, **B Datta**, *On stellated spheres and a tightness criterion for combinatorial manifolds*, European J. Combin. 36 (2014) 294–313 MR Zbl

[3] **B Bagchi**, **B Datta**, **J Spreer**, *A characterization of tightly triangulated 3–manifolds*, European J. Combin. 61 (2017) 133–137 MR Zbl

[4] **A Bahri**, **M Bendersky**, **F R Cohen**, *Polyhedral products and features of their homotopy theory*, from "Handbook of homotopy theory" (H Miller, editor), CRC Press, Boca Raton, FL (2020) 103–144 MR Zbl

[5] **A Bahri**, **M Bendersky**, **F R Cohen**, **S Gitler**, *The polyhedral product functor: A method of decomposition for moment-angle complexes, arrangements and related spaces*, Adv. Math. 225 (2010) 1634–1668 MR Zbl

[6] **I V Baskakov**, **V M Buchstaber**, **T E Panov**, *Algebras of cellular cochains, and torus actions*, Uspekhi Mat. Nauk 59 (2004) 159–160 MR Zbl In Russian; translated in Russian Math. Surveys 59 (2004) 562–563

[7] **A Berglund**, **M Jöllenbeck**, *On the Golod property of Stanley–Reisner rings*, J. Algebra 315 (2007) 249–273 MR Zbl

[8] **B A Burton**, **B Datta**, **N Singh**, **J Spreer**, *A construction principle for tight and minimal triangulations of manifolds*, Exp. Math. 27 (2018) 22–36 MR Zbl

[9] **B Datta**, **S Murai**, *On stacked triangulated manifolds*, Electron. J. Combin. 24 (2017) art. no. P4.12 MR Zbl

[10] **J Grbić**, **T Panov**, **S Theriault**, **J Wu**, *The homotopy types of moment-angle complexes for flag complexes*, Trans. Amer. Math. Soc. 368 (2016) 6663–6682 MR Zbl

[11] **J Grbić**, **S Theriault**, *The homotopy type of the polyhedral product for shifted complexes*, Adv. Math. 245 (2013) 690–715 MR Zbl

[12] **V Grujić**, **V Welker**, *Moment-angle complexes of pairs (D^n, S^{n-1}) and simplicial complexes with vertex-decomposable duals*, Monatsh. Math. 176 (2015) 255–273 MR Zbl

[13] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[14] **K Iriye**, **D Kishimoto**, *Decompositions of polyhedral products for shifted complexes*, Adv. Math. 245 (2013) 716–736 MR Zbl

[15] **K Iriye**, **D Kishimoto**, *Golodness and polyhedral products for two-dimensional simplicial complexes*, Forum Math. 30 (2018) 527–532 MR Zbl

[16] **K Iriye**, **D Kishimoto**, *Golodness and polyhedral products of simplicial complexes with minimal Taylor resolutions*, Homology Homotopy Appl. 20 (2018) 69–78 MR Zbl

[17] **K Iriye**, **D Kishimoto**, *Fat-wedge filtration and decomposition of polyhedral products*, Kyoto J. Math. 59 (2019) 1–51 MR Zbl

[18] **K Iriye**, **D Kishimoto**, *Whitehead products in moment-angle complexes*, J. Math. Soc. Japan 72 (2020) 1239–1257 MR Zbl

[19] **K Iriye**, **D Kishimoto**, *Two-dimensional Golod complexes*, Homology Homotopy Appl. 23 (2021) 215–226 MR Zbl

[20] **K Iriye**, **T Yano**, *A Golod complex with non-suspension moment-angle complex*, Topology Appl. 225 (2017) 145–163 MR Zbl

[21] **L Katthän**, *A non-Golod ring with a trivial product on its Koszul homology*, J. Algebra 479 (2017) 244–262 MR Zbl

[22] **W Kühnel**, **F H Lutz**, *A census of tight triangulations*, Period. Math. Hungar. 39 (1999) 161–183 MR Zbl

[23] **N H Kuiper**, *Minimal total absolute curvature for immersions*, Invent. Math. 10 (1970) 209–238 MR Zbl

[24] **F H Lutz**, **T Sulanke**, **E Swartz**, *f–vectors of 3–manifolds*, Electron. J. Combin. 16 (2009) art. no. R13 MR Zbl

[25] **J R Munkres**, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, CA (1984) MR Zbl

[26] **J-P Serre**, *Algèbre locale. Multiplicités*, Lecture Notes in Math. 11, Springer (1965) MR Zbl

Department of Mathematics, Osaka Metropolitan University
Osaka, Japan
Faculty of Mathematics, Kyushu University
Fukuoka, Japan

kiriye@omu.ac.jp, kishimoto@math.kyushu-u.ac.jp

Received: 8 February 2021 Revised: 8 March 2022

Algebraic & Geometric Topology **23**:5 (2023) 2213–2219
DOI: 10.2140/agt.2023.23.2213
Published: 25 July 2023

A remark on the finiteness of purely cosmetic surgeries

TETSUYA ITO

By estimating the knot Floer thickness in terms of the genus and the braid index, we show that a knot K in S^3 does not admit purely cosmetic surgery whenever $g(K) \geq \frac{3}{2}b(K)$, where $g(K)$ and $b(K)$ denote the genus and the braid index, respectively. In particular, this establishes the finiteness of purely cosmetic surgeries; for a fixed b, all but finitely many knots with braid index b satisfies the cosmetic surgery conjecture.

57K10; 57K30

For a knot K in the 3–sphere S^3 and $r \in \mathbb{Q}$, let $S^3_K(r)$ be the r–surgery on K. Two Dehn surgeries $S^3_K(r)$ and $S^3_K(r')$ on the same knot K are *purely cosmetic* if $r \neq r'$ but $S^3_K(r) \cong S^3_K(r')$. Here we write $M \cong N$ if M and N are *orientation-preservingly* homeomorphic.

Conjecture 1 (cosmetic surgery conjecture) *A nontrivial knot does not admit purely cosmetic surgeries.*

One must be careful to take account of orientations; there are several examples of *chirally cosmetic surgery*, a pair of Dehn surgeries on the same knot, that yields orientation-reversingly homeomorphic 3–manifolds. For example, for the trefoil knot K, $S^3_K(9) \cong -S^3_K\left(\frac{9}{2}\right)$; see Mathieu [7]. Here $-M$ is the 3–manifold M with opposite orientation.

For a knot K in S^3, let $g(K)$ be the genus and $b(K)$ be the braid index of K. The aim of this note is to point out the following finiteness result on purely cosmetic surgeries, which gives strong supporting evidence for Conjecture 1:

Theorem 1 *If $g(K) \geq \frac{3}{2}b(K)$, then K does not admit a purely cosmetic surgery. In particular, for given $b > 0$, there are only finitely many knots with braid index b that admit purely cosmetic surgeries.*

Here, the latter finiteness assertion follows from Birman and Menasco's finiteness theorem [2]: for given $g, b > 0$ there are only finitely many knots with genus g and braid index b.

Our proof of Theorem 1 is based on a quantitative refinement of Birman and Menasco's finiteness theorem [5] and the following quite strong constraint for purely cosmetic surgeries:

Theorem 2 (Hanselman [4]) *Let K be a nontrivial knot and $\mathrm{th}(K)$ be the Heegaard Floer thickness of K. If $S^3_K(r) \cong S^3_K(r')$ for $r \neq r'$, then either*

- $\{r, r'\} = \{2, -2\}$ *and* $g(K) = 2$, *or*
- $\{r, r'\} = \{1/q, -1/q\}$ *for some* $0 < q \leq (\mathrm{th}(K) + 2g(K))/2g(K)(g(K)-1)$.

Here, $\mathrm{th}(K)$ is the thickness of the knot Floer homology.

Thus, if $g(K) \neq 2$ and $\mathrm{th}(K)$ is small compared with $g(K)$, then K does not admit purely cosmetic surgery. This motivates us to study a relation between $g(K)$ and $\mathrm{th}(K)$, in particular the (upper) bound of $\mathrm{th}(K)/g(K)$. Here, we give an upper bound of the thickness $\mathrm{th}(K)$ in terms of $g(K)$ and $b(K)$.

Although our argument applies in the cases $b(K) = 2, 3$, we restrict our attention to the case $b(K) \geq 4$.

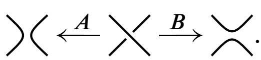
Lemma 3 *If $b(K) \geq 4$,*

$$\mathrm{th}(K) \leq \tfrac{1}{2}(2b(K) - 5)(2g(K) - 1 + b(K)).$$

Proof For a knot diagram D, the *Turaev genus* $g_T(D)$ is defined by

$$g_T(D) = \tfrac{1}{2}(c(D) + 2 - |s_A| - |s_B|),$$

where $c(D)$ is the crossing number of D and $|s_A|$ and $|s_B|$ are the number of circles obtained by A– and B–smoothing, respectively, of crossings of D given by



The Turaev genus $g_T(K)$ of a knot K is the minimum of $g_T(D)$ among diagrams D of K. In [6], Lowrance showed the inequality

$$\mathrm{th}(K) \leq g_T(K).$$

For any diagram D, $|s_A|, |s_B| \geq 1$, so $g_T(D) \leq \frac{1}{2}c(D)$. Hence, we have a canonical upper bound of the Turaev genus,

$$g_T(K) \leq \tfrac{1}{2}c(K). \tag{1}$$

Finally, by the quantitative Birman–Menasco finiteness theorem[1] [5], if $b(K) \geq 4$, we get

$$c(K) \leq (2b(K)-5)(2g(K)-1+b(K)).$$

These three inequalities prove the desired inequality. □

Proof of Theorem 1 In the following we assume that $b(K) \geq 4$ since Varvarezos [8] proved the cosmetic surgery conjecture for the case $b(K)=3$. Also, we assume that $g(K) \neq 2$.

Assume to the contrary that K admits a purely cosmetic surgery. By Theorem 2, such a knot must satisfy

$$1 \leq \frac{\mathrm{th}(K)+2g(K)}{2g(K)(g(K)-1)} \iff 2g(K)(g(K)-2) \leq \mathrm{th}(K),$$

so, by Lemma 3, we conclude that, when a knot K admits a purely cosmetic surgery, it satisfies

$$2g(K)(g(K)-2) \leq \tfrac{1}{2}(2b(K)-5)(2g(K)-1+b(K)).$$

That is, we get a constraint for a knot K to admit a purely cosmetic surgery:

$$4g(K)^2+(2-4b(K))g(K)+(2b(K)-5)(1-b(K)) \leq 0. \tag{2}$$

Now the assertion of the theorem follows from an easy computation that, if $g(K) \geq \frac{3}{2}b(K)$, then (2) is never satisfied. □

As the proof indicates, our sufficient condition $g(K) \geq \frac{3}{2}b(K)$ can be improved if one can improve on the estimate of $\mathrm{th}(K)$ in Lemma 3.

Remark 4 Instead of using an obvious bound (1) of the Turaev genus, by using a different upper bound [3, Corollary 7.3]

$$g_T(K) \leq c(K) - \mathrm{span}\, V_K(t),$$

where $V_K(t)$ denotes the Jones polynomial, we get a different constraint: if K admits a purely cosmetic surgery, then

$$2g(K)^2+(6-4b(K))g(K)+(2b(K)-5)(1-b(K))+\mathrm{span}\, V_K(t) \leq 0. \tag{3}$$

[1] When $b(K)=2,3$, a similar inequality holds but the coefficient $2b(K)-5$ is 1 or $\frac{5}{3}$, respectively.

Here, we give a mild improvement of Lemma 3. For a diagram D of a knot K, the *dealternation number* $\mathrm{dalt}(D)$ of D is the minimum number of crossing change needed to make D into an alternating diagram. The *dealternation number* of a knot K is the minimum of $\mathrm{dalt}(D)$ among diagrams D of K. It is known that $g_T(K) \le \mathrm{dalt}(K)$ [1], so evaluating the dealternation number also gives an upper bound on the thickness.

We prove the following estimate of the dealternation number (and hence the Turaev genus and the thickness) in terms of the genus and braid index, which is interesting in its own right:

Theorem 5 *If $b(K) \ge 4$, then*

$$\mathrm{th}(K) \le g_T(K) \le \mathrm{dalt}(K) \le \left(b(K) - 3 + \frac{1}{b(K)}\right)(2g(K) - 1 + b(K)).$$

Proof Let $n = b(K)$ and let B_n be the braid group of n strands. We denote the standard generators of B_n by $\sigma_1, \dots, \sigma_{n-1}$. We say that a braid is *alternating* if it is a product of $\{\sigma_1, \sigma_2^{-1}, \sigma_3, \sigma_4^{-1}, \dots, \sigma_{2i-1}, \sigma_{2i}^{-1}, \dots\}$. Obviously, the closure of an alternating braid is an alternating diagram.

For $1 \le i < j \le n$, let $a_{i,j}$ be the band generator given by

$$a_{i,j} = (\sigma_i \sigma_{i+1} \cdots \sigma_{j-2})\sigma_{j-1}(\sigma_i \sigma_{i+1} \cdots \sigma_{j-2})^{-1}.$$

A band generator $a_{i,j}$ can be seen as the boundary of a twisted band connecting the i^{th} and j^{th} strands of the braid. Thus, when K is represented as the closure of a braid $\beta \in B_n$, by giving β as a product of band generators, we get a Seifert surface F_β of K, called the *Bennequin surface* associated to the braid (word) β.

First we treat the case that K bounds a minimum genus Bennequin surface of minimum braid index. That is, K is represented by a closed n–braid β such that its Bennequin surface F_β is a minimum genus Seifert surface of K.

Thanks to the relation

$$\sigma_{j-1}\sigma_j^{\pm 1}\sigma_{j-1}^{-1} = \sigma_j^{-1}\sigma_{j-1}^{\pm 1}\sigma_j,$$

by taking suitable word representatives of the $a_{i,j}^{\pm 1}$, each band generator $a_{i,j}$ except $a_{1,n}$ can be made so that it is alternating by changing at most $n-3$ crossings. The exceptional case $a_{1,n}^{\pm 1}$ can be made so that it is alternating by changing $n-2$ crossings. Thus,

$$\mathrm{dalt}(K) \le \sum_{\substack{1\le i<j\le n\\(i,j)\neq(1,n)}} (n-3)r_{i,j} + (n-2)r_{1,n} = \sum_{1\le i<j\le n} (n-3)r_{i,j} + r_{1,n},$$

where $r_{i,j}$ is the number of $a_{i,j}^{\pm 1}$ in the braid β.

On the other hand, since we assume that the Bennequin surface F_β associated with the n–braid β has genus $g(K)$,

$$\sum_{1\le i<j\le n} r_{i,j} = 2g(K)-1+n.$$

Let $\delta=\sigma_1\sigma_2\sigma_3\cdots\sigma_{n-1}$. Since $\delta a_{i,j}\delta^{-1} = a_{i+1,j+1}$ (here we regard indices modulo n; for example, $\delta a_{1,n}\delta^{-1} = a_{2,n+1}$ is understood as $a_{1,2}$), by taking conjugates of δ if necessary, we may assume that

$$r_{1,n} \le \frac{1}{n}(r_{1,2}+r_{2,3}+r_{3,4}+\cdots+r_{n-1,n}+r_{1,n}) \le \frac{1}{n}(2g(K)-1+n).$$

Thus, we conclude

$$\mathrm{dalt}(K) \le \sum_{1\le i<j\le n} (n-3)r_{i,j} + r_{1,n} \le \left(n-3+\frac{1}{n}\right)(2g(K)-1+n),$$

as desired.

Next, we assume that K does not bound a minimum genus Bennequin surface of the minimum braid index. To treat this case we quickly review a main strategy of the proof of the quantitative Birman–Menasco theorem [5], namely how to relate the genus, braid index and crossing number (although we do not need to use or know the details).

We put a minimum genus Seifert surface F of K so that it admits a braid foliation. Let R_{aa} and R_{ab} be the number of *aa* tiles and *ab* tiles of the braid foliation. What we showed in [5] is two inequalities:

$$c(K) \le (2n-5)R_{aa} + (n-3)R_{ab} \tag{4}$$

and

$$2R_{aa} + R_{ab} \le 2(2g(K)-1+b(K)). \tag{5}$$

More precisely, the inequality (4) is obtained by observing that the braid foliation gives rise to an explicit closed n–braid representative β such that one *aa* tile provides a braid which is a band generator,

$$a_{i,j}^{\pm 1}, \quad (i,j)\neq(1,n),$$

and that one ab tile provides a braid of the form

$$\gamma_{i,j}^{\pm 1}, \quad |i-j| \le n-3.$$

Here, $\gamma_{i,j}$ denotes the braid

$$\gamma_{i,j} = \begin{cases} \sigma_i\sigma_{i+1}\cdots\sigma_{j-1} & \text{if } i<j, \\ \sigma_i\sigma_{i-1}\cdots\sigma_{j-1} & \text{if } i>j. \end{cases}$$

(when $i=j$, we regard $\gamma_{i,j}$ as the trivial braid).

If n is odd, then each braid $\gamma_{i,j}$ can be made into an alternating braid by at most $\frac{1}{2}(n-3)$ crossing changes. Each band generator $a_{i,j}$ coming from an aa tile can be made into an alternating braid by at most $n-3$ changes since $a_{1,n}$ does not appear. Therefore,

$$\begin{aligned} \operatorname{dalt}(K) &\le (n-3)R_{aa} + \tfrac{1}{2}(n-3)R_{ab} = \tfrac{1}{2}(n-3)(2R_{aa}+R_{ab}) \\ &\le (n-3)(2g(K)-1+n). \end{aligned}$$

If n is even, let M be the number of the $\gamma_{i,j}$ produced by ab tiles such that $\gamma_{i,j}$ is made into an alternating braid by $\frac{1}{2}(n-2)$ crossing changes. By taking the mirror image of β if necessary, we may assume that $M \le \frac{1}{2}R_{ab}$. Since other braids $\gamma_{i,j}$ from ab tiles can be made into an alternating braid by at most $\frac{1}{2}(n-4)$ crossing changes,

$$\begin{aligned} \operatorname{dalt}(K) &\le (n-3)R_{aa} + \tfrac{1}{2}(n-4)(R_{ab}-M) + \tfrac{1}{2}(n-2)M \\ &\le (n-3)R_{aa} + \tfrac{1}{2}(n-3)R_{ab} = \tfrac{1}{2}(n-3)(2R_{aa}+R_{ab}) \\ &\le (n-3)(2g(K)-1+n). \end{aligned} \qquad \square$$

Using this refinement we can improve a sufficient condition in Theorem 1. For example, for the case $b(K)=4$, a direct computation shows that:

Corollary 6 *A knot K with braid index 4 does not admit purely cosmetic surgery if $g(K) \ne 2,3$.*

Acknowledgement

The author has been partially supported by JSPS KAKENHI grants 19K03490 and 21H04428. He would like to thank the referee for many helpful comments.

References

[1] **T Abe**, **K Kishimoto**, *The dealternating number and the alternation number of a closed 3–braid*, J. Knot Theory Ramifications 19 (2010) 1157–1181 MR Zbl

[2] **J S Birman**, **W W Menasco**, *Studying links via closed braids, VI: A nonfiniteness theorem*, Pacific J. Math. 156 (1992) 265–285 MR Zbl

[3] **O T Dasbach**, **D Futer**, **E Kalfagianni**, **X-S Lin**, **N W Stoltzfus**, *The Jones polynomial and graphs on surfaces*, J. Combin. Theory Ser. B 98 (2008) 384–399 MR Zbl

[4] **J Hanselman**, *Heegaard Floer homology and cosmetic surgeries in S^3*, J. Eur. Math. Soc. 25 (2023) 1627–1670 MR

[5] **T Ito**, *A quantitative Birman–Menasco finiteness theorem and its application to crossing number*, J. Topol. 15 (2022) 1794–1806 MR

[6] **A M Lowrance**, *On knot Floer width and Turaev genus*, Algebr. Geom. Topol. 8 (2008) 1141–1162 MR Zbl

[7] **Y Mathieu**, *Closed 3–manifolds unchanged by Dehn surgery*, J. Knot Theory Ramifications 1 (1992) 279–296 MR Zbl

[8] **K Varvarezos**, *3–Braid knots do not admit purely cosmetic surgeries*, Acta Math. Hungar. 164 (2021) 451–457 MR Zbl

Department of Mathematics, Graduate School of Science, Kyoto University
Kyoto, Japan

tetitoh@math.kyoto-u.ac.jp

Received: 25 April 2021 Revised: 24 November 2021

Algebraic & Geometric Topology **23**:5 (2023) 2221–2270
DOI: 10.2140/agt.2023.23.2221
Published: 25 July 2023

Geodesic complexity of homogeneous Riemannian manifolds

STEPHAN MESCHER
MAXIMILIAN STEGEMEYER

We study the geodesic motion planning problem for complete Riemannian manifolds and investigate their geodesic complexity, an integer-valued isometry invariant introduced by D Recio-Mitter. Using methods from Riemannian geometry, we establish new lower and upper bounds on geodesic complexity and compute its value for certain classes of examples with a focus on homogeneous Riemannian manifolds. To achieve this, we study properties of stratifications of cut loci and use results on their structures for certain homogeneous manifolds obtained by T Sakai and others.

55M30; 53C22

1 Introduction

A topological abstraction of the motion planning problem in robotics was introduced by M Farber [12]. The *topological complexity* of a path-connected space X is denoted by $\mathsf{TC}(X)$ and intuitively given by the minimal number of open sets needed to cover

$X \times X$ such that, on each of the open sets, there exists a continuous motion planner. Here, a continuous motion planner is a map associating with each pair of points a continuous path from the first point to the second point which varies continuously with the endpoints. Such maps are interpreted as algorithms telling an autonomous robot in the workspace X how it is supposed to move from its position to a desired endpoint. Unfortunately, the topological complexity of a space does not tell us anything about the feasibility or efficiency of the paths taken by motion planners having $\mathsf{TC}(X)$ domains of continuity; see the discussion of Z Błaszczyk and J Carrasquel-Vera [3, Introduction]. For example, the explicitly constructed motion planners for configuration spaces of Euclidean spaces by H Mas-Ku and E Torres-Giese [29] and Farber [16, Section 8] require few domains of continuity, but have paths among their values which are far from being length-minimizing. Considering a general metric space, paths taken by the motion planners might become arbitrarily long and thus be unsuited for practical motion planning problems.

Recently, D Recio-Mitter [34] has introduced the notion of *geodesic complexity* of metric spaces. There, the paths taken by motion planners are additionally required to be length-minimizing between their endpoints. Intuitively, this is seen as the complexity of *efficient* motion planning in metric spaces. Recio-Mitter's seminal article has already triggered research in geodesic complexity, especially computations of geodesic complexity for interesting classes of examples; see Davis, Harrison and Recio-Mitter [8; 9; 10].

In this article we study the geodesic complexity of complete Riemannian manifolds and derive new lower and upper bounds for their geodesic complexities by methods from Riemannian geometry.

Before continuing, we recall the definition of geodesic complexity of geodesic spaces from [34, Definition 1.7] for the special case of a complete Riemannian manifold. Let (M, g) be a complete connected Riemannian manifold and let $PM := C^0([0,1], M)$ be equipped with the compact–open topology. We recall that a geodesic segment $\gamma\colon [0,1] \to M$ is called *minimal* if it minimizes the length compared to all rectifiable paths from $\gamma(0)$ to $\gamma(1)$. For simplicity, we shall call a minimal geodesic segment simply a *minimal geodesic*. Consider

$$GM := \{\gamma \in PM \mid \gamma \text{ is a minimal geodesic in } (M, g)\}$$

as a subspace of PM and let

$$\pi\colon GM \to M \times M, \quad \pi(\gamma) = (\gamma(0), \gamma(1)).$$

By standard results from Riemannian geometry, π is surjective since (M,g) is complete; see Petersen [33, Corollary 5.8.5]. The *geodesic complexity of* (M,g) is given by $\mathsf{GC}(M,g)=r$, where $r\in\mathbb{N}$ is the smallest integer with the following property: there are r pairwise disjoint locally compact subsets $E_1,\dots,E_r\subset M\times M$ with $\bigcup_{i=1}^r E_i=M\times M$ such that, for each $i\in\{1,2,\dots,r\}$, there exists a continuous geodesic motion planner $s_i\colon E_i\to GM$, ie a continuous local section of the map π. If there is no such r, we let $\mathsf{GC}(M,g)=+\infty$. Since it is not at all evident how to compute this number explicitly, one is interested in establishing lower and upper bounds for $\mathsf{GC}(M,g)$. This approach is also common in studies of Lusternik–Schnirelmann category or, more generally, sectional categories of fibrations. Given a fibration $p\colon E\to B$, the *sectional category of* p is given by $\mathsf{secat}(p)=k$, where $k\in\mathbb{N}$ is the minimal number with the following property: there exists an open cover of B consisting of k open subsets such that p admits a continuous local section over each of these sets. This notion was introduced under the name *genus of a fibration* by A Schwarz [39]. The topological complexity of a topological space X is for example given as the sectional category of the fibration

$$PX\to X\times X,\quad \gamma\mapsto(\gamma(0),\gamma(1)).$$

Schwarz worked out several ways of obtaining lower and upper bounds for sectional categories which have direct consequences for topological complexity; see eg Farber [14; 15, Chapter 4] for an overview.

However, the restriction $\pi\colon GM\to M\times M$ of this fibration to minimal geodesics is in general *not* a fibration. For example, if $M=S^n$ is an n–sphere with $n\in\mathbb{N}$, equipped with a round metric, then $\pi^{-1}(\{(p,q)\})$ consists of one element if $q\neq -p$, while it is homeomorphic to S^{n-1} if $q=-p$. In particular, not all preimages are homotopy-equivalent, so π is not a fibration in this case. Therefore, Schwarz's results are not applicable to the setting of geodesic complexity. Instead we will derive several lower and upper bounds for the geodesic complexity of Riemannian manifolds using methods from Riemannian geometry. By [34, Remark 1.9], every complete Riemannian manifold satisfies $\mathsf{TC}(M)\le\mathsf{GC}(M,g)$. This formalizes the observation that requiring the paths a robot takes to be as short as possible can increase the complexity of the problem. For example, as shown in [34, Theorem 1.11], for each $n\ge 3$ there exists a Riemannian metric g_n on the sphere S^n for which $\mathsf{GC}(S^n,g_n)-\mathsf{TC}(S^n)\ge n-3$. In practical applications, a person designing robotic systems that are supposed to move autonomously might not mind a higher complexity. In fact, such a person might accept more instabilities in the motions of robots as a downside if the upside is that the robots move fast and efficiently.

An important observation is that the difficulties of geodesic motion planning lie in the cut loci of (M, g), as was pointed out by Recio-Mitter [34, page 144] in the more general framework of metric spaces. Let $\mathrm{Cut}_p(M)$ denote the cut locus of $p \in M$ in (M, g). We refer to Lee [27, page 308], Petersen [33, page 219] or Definition 2.5 below for its definition. If $A \subset M \times M$ satisfies $q \notin \mathrm{Cut}_p(M)$ for each $(p, q) \in A$, then there is a unique minimal geodesic from p to q for each $(p, q) \in A$. The corresponding geodesic motion planner $A \to GM$ is continuous; see also the observations of Błaszczyk and Carrasquel-Vera [3]. Thus, to compute the geodesic complexity of a manifold, we need to understand its cut loci. While the cut locus of a point in a Riemannian manifold is always closed and of measure zero — see [27, Theorem 10.34(a)] — little else is known about cut loci in general.

In [34, Corollary 3.14], Recio-Mitter establishes a lower bound on the geodesic complexity of metric spaces given in terms of the structure of their cut loci. He considers cut loci which possess stratifications admitting finite coverings. For this purpose, Recio-Mitter introduces the notion of a *levelwise stratified covering* in [34, Definition 3.8]. He then defines a notion of *inconsistency*, which is roughly a condition on the relations between the coverings of the different strata of cut loci by minimal geodesics. It formalizes certain incompatibility properties of families of geodesics connecting a point with points in its cut locus.

Focusing on complete Riemannian manifolds, we will use Riemannian exponential maps to establish a similar inconsistency condition on cut loci, which is more concise than the one from [34]. Given a complete Riemannian manifold M and a point $p \in M$ for which $\mathrm{Cut}_p(M)$ admits a stratification, we study the preimages of the different strata of $\mathrm{Cut}_p(M)$ under the Riemannian exponential map $\exp_p \colon T_pM \to M$. Assume that some $x \in M$ lies in the closure of multiple connected components of the same stratum of $\mathrm{Cut}_p(M)$. We then study the closures of the preimages of all of these components under $\exp_p$ as subsets of T_pM. The inconsistency condition demands that these closures have no point in common that is mapped to x by $\exp_p$. We will see that this condition excludes the existence of an open neighborhood U of x with a single continuous geodesic motion planner which connects p to all points of $\mathrm{Cut}_p(M)$ that lie in U.

Note that our definition is only applicable to Riemannian manifolds and not to arbitrary geodesic spaces. One of its benefits in the Riemannian setting is the fact that we can deduce an easier condition than the one introduced by Recio-Mitter. More precisely, we do not require anymore that any point in a cut locus of another point is connected to that point by only finitely many minimal geodesics. Moreover, our inconsistency condition is explicitly stated as an intersection condition on certain subsets of a tangent cut locus, instead of using the notion of levelwise stratified coverings as in [34].

Our main result on inconsistent stratifications is the following theorem. This result is similar to [34, Corollary 3.14] and our proof is inspired by Recio-Mitter's proof as well.

Theorem 4.8 *Let (M, g) be a closed Riemannian manifold. Assume that there exists a point $p \in M$ for which $\mathrm{Cut}_p(M)$ admits an inconsistent stratification of depth $N \in \mathbb{N}$. Then*

$$\mathsf{GC}(M) \geq N + 1.$$

There is more to say about cut loci of *homogeneous* Riemannian manifolds, ie Riemannian manifolds (M, g) whose isometry groups act transitively on M. An isometry $\phi\colon M \to M$ maps the cut locus of $p \in M$ onto that of $\phi(p)$. Hence, the cut locus of a point is identified with that of another point by an isometry. This translation property of the cut loci allows us to estimate the geodesic complexity of M from above, once we understand how we can decompose one single cut locus into domains of continuous geodesic motion planners. The following result provides an upper bound for geodesic complexity in terms of a sectional category and the subspace geodesic complexities of considerably smaller subsets of $M \times M$. Here, the subspace geodesic complexity of $A \subset M \times M$ is defined in terms of covers of A by domains of continuous geodesic motion planners.

Corollary 5.8 *Let (M, g) be a homogeneous Riemannian manifold with isometry group $\mathrm{Isom}(M, g)$. Let $p \in M$ and assume that $\mathrm{Cut}_p(M)$ has a stratification $(S_1, \dots, S_k)$ of depth k. Then*

$$\mathsf{GC}(M) \leq \mathsf{secat}(\mathrm{ev}_p\colon \mathrm{Isom}(M, g) \to M) \cdot \sum_{i=1}^{k} \max_{Z_i \in \pi_0(S_i)} \mathsf{GC}_p(Z_i) + 1,$$

where $\mathrm{ev}_p(\phi) = \phi(p)$ for all $\phi \in \mathrm{Isom}(M, g)$ and where $\mathsf{GC}_p(Z_i)$ is the subspace geodesic complexity of $\{p\} \times Z_i \subset M \times M$.

In the case of compact, simply connected, irreducible symmetric spaces, we are able to further estimate this upper bound from above in terms of certain sectional categories. This means that for such symmetric spaces we obtain an upper bound on $\mathsf{GC}(M)$ which does not involve any geodesic complexities.

Note that this result produces the first upper bound for geodesic complexity in terms of categorical invariants. Indeed, the only previously known upper bounds were derived by Recio-Mitter [34] either from explicit constructions of geodesic motion planners or from the existence of particularly simple coverings of cut loci. We pick up Recio-Mitter's so-called trivially covered stratifications in this article in the setting of Riemannian manifolds as well.

In addition to establishing new lower and upper bounds for geodesic complexity, we compute the geodesic complexities of some Riemannian manifolds whose cut loci are well understood. We will show that every three-dimensional Berger sphere (S^3, g_α) satisfies $\mathsf{GC}(S^3, g_\alpha) = 2$ and that $\mathsf{GC}(T^2, g_f) = 3$ for every flat metric g_f on the two-dimensional torus. This extends the two-dimensional case of Recio-Mitter's computation of the geodesic complexity of the standard flat n–torus from [34, Theorem 4.4].

The article is structured as follows: In Section 2 we introduce some additional terminology and recall elementary facts about geodesic complexity and cut loci. Section 3 contains some basic nonexistence results on continuous geodesic motion planners. These results illustrate the difficulties for motion planning that cut loci can create. In Section 4 we establish lower bounds on geodesic complexity by two different approaches. On the one hand, this is done in terms of principal bundles over the manifold and the topological complexities of their total spaces. On the other hand, we study manifolds with stratified cut loci whose stratifications satisfy the above-mentioned inconsistency property. We focus on homogeneous Riemannian manifolds in Section 5. More precisely, we show that their geodesic complexities can be estimated from above in terms of the subspace complexities of a single cut locus. In Section 6 we consider Riemannian manifolds whose cut loci admit trivially covered stratifications. For such stratifications the relations between a cut locus and its corresponding tangent cut locus are particularly simple. Section 7 deals with examples of geodesic complexities. Combining results from the previous sections with new observations, we reobtain Recio-Mitter's computation of geodesic complexity of the standard flat n–torus and determine the geodesic complexity of arbitrary flat 2–tori. As another class of examples, we explicitly compute the geodesic complexity of three-dimensional Berger spheres. In the final Section 8 we consider consequences of the previous results for compact simply connected symmetric spaces. In both situations, the considered cut loci have been studied by T Sakai. Using the estimates from Section 5, we derive an upper bound for geodesic complexity that is given in terms of the Lie groups from which the symmetric space is built. We further make explicit computations for two examples of symmetric spaces.

Acknowledgements

The authors thank the referee for their careful and thoughtful reading of our manuscript. Their suggestions highly improved the exposition and the clarity of the article.

Throughout this article we assume all manifolds to be smooth and connected and all Riemannian metrics to be smooth.

2 Basic notions and definitions

We begin by introducing subspace versions of geodesic complexity for Riemannian manifolds. Afterwards, we recall some basic computations from [34] and several facts about cut loci in Riemannian manifolds.

Definition 2.1 Let (M, g) be a complete Riemannian manifold and $\pi : GM \to M \times M$, $\pi(\gamma) = (\gamma(0), \gamma(1))$. Let GM be equipped with the subspace topology of $C^0([0,1], M)$ with the compact–open topology.

(a) Let $X \subset M \times M$. A *geodesic motion planner on* X is a section $s\colon X \to GM$ of π.

(b) Given $A \subset M \times M$ we let $\mathsf{GC}_{(M,g)}(A)$ be the minimum $r \in \mathbb{N}$ for which there are r pairwise disjoint locally compact subsets $E_1, \dots, E_r \subset M \times M$ such that $A \subset \bigcup_{i=1}^r E_i$ and, for each $i \in \{1, 2, \dots, r\}$, there exists a continuous geodesic motion planner $s_i\colon E_i \to GM$. If no such r exists, then we put $\mathsf{GC}_{(M,g)}(A) := +\infty$. We call $\mathsf{GC}_{(M,g)}(A)$ the *subspace geodesic complexity of* A.

We recall that the map π is surjective for complete Riemannian manifolds. This is a consequence of the Hopf–Rinow theorem; see [33, Corollary 5.8.5].

Remarks 2.2 (1) If it is obvious which Riemannian metric we are referring to, we occasionally suppress it from the notation and write

$$\mathsf{GC}(M) := \mathsf{GC}(M, g) \quad \text{and} \quad \mathsf{GC}_M(A) := \mathsf{GC}_{(M,g)}(A).$$

Note that, in particular, $\mathsf{GC}(M) = \mathsf{GC}_M(M \times M)$.

(2) Given $p \in M$ and $B \subset M$, we further put

$$\mathsf{GC}_p(B) := \mathsf{GC}_{(M,g)}(\{p\} \times B).$$

(3) Our definition differs from Recio-Mitter's original definition by 1 in the sense that for us $\mathsf{GC}(\{\mathrm{pt}\}) = 1$, while it would be 0 in the sense of [34, Definition 1.7].

Examples 2.3 (1) As proven in [34, Proposition 4.1], if g_r is a round metric on the sphere S^n with $n \in \mathbb{N}$, then

$$\mathsf{GC}(S^n, g_r) = \mathsf{TC}(S^n) = \begin{cases} 2 & \text{if } n \text{ is odd,} \\ 3 & \text{if } n \text{ is even.} \end{cases}$$

(2) Let g_f be the standard flat metric on T^2 and let g_{emb} denote the metric induced by the standard embedding $T^2 \hookrightarrow \mathbb{R}^3$ and the Euclidean metric on $\mathbb{R}^3$. By [34, Theorems 4.4 and 5.1],

$$\mathsf{GC}(T^2, g_f) = 3, \quad \mathsf{GC}(T^2, g_{\mathrm{emb}}) = 4.$$

(3) It was further shown in [34, Theorem 1.11] that, for each $k \in \mathbb{N}$ with $k \geq 3$, there exists a Riemannian metric g_k on S^k with $\mathsf{GC}(S^k, g_k) \geq k$.

Remarks 2.4 Let (M, g) be a complete Riemannian manifold.

(1) For all $A \subset M \times M$, it holds that $\mathsf{TC}_M(A) \leq \mathsf{GC}_M(A)$, where $\mathsf{TC}_M(A)$ is the relative topological complexity of A in $M \times M$; see [15, Section 4.3]. Here, we made use of the characterization of topological complexity by locally compact subsets shown in [15, Proposition 4.9].

(2) It is easy to see that

$$\mathsf{GC}_M(A \cup B) \leq \mathsf{GC}_M(A) + \mathsf{GC}_M(B) \quad \text{for all } A, B \subset M \times M. \tag{2-1}$$

This is shown in analogy with [15, Proposition 4.24].

As pointed out by Recio-Mitter, the crucial ingredients for the discussion of geodesic complexity are the cut loci of points in the space under consideration. The notions of cut loci in metric and in Riemannian geometry are slightly different from each other. While Recio-Mitter used the former notion in his work — see [34, Definition 3.1] — we will use the latter throughout this manuscript. We next recall the notion of cut loci from Riemannian geometry. The relation between the two will be discussed in Remark 2.7(3) below. See also [27, page 308] or [33, page 219] for the following definition:

Definition 2.5 Let (M,g) be a complete Riemannian manifold and let $p \in M$.

(a) Let $\gamma\colon [0,+\infty) \to M$ be a unit-speed geodesic with $\gamma(0)=p$ and $\dot\gamma(0) \in T_pM$. The *cut time of* γ is given by

$$t_{\mathrm{cut}}(\gamma) = \sup\{t>0 : \gamma|_{[0,t]} \text{ is minimal}\}.$$

If $t_{\mathrm{cut}}(\gamma)$ is finite, then $t_{\mathrm{cut}}(\gamma)\dot\gamma(0) \in T_pM$ is a *tangent cut point of* p and $\gamma(t_{\mathrm{cut}}(\gamma)) \in M$ is a *cut point of* p *along* γ. Note that

$$\gamma(t_{\mathrm{cut}}(\gamma)) = \exp_p(t_{\mathrm{cut}}(\gamma)\dot\gamma(0)).$$

(b) The set of all cut points of p is called the *cut locus of* p and denoted by $\mathrm{Cut}_p(M)$. The set of all tangent cut points of p is called the *tangent cut locus of* p and denoted by $\widetilde{\mathrm{Cut}}_p(M)$.

(c) The *total cut locus of* M is given by

$$\mathrm{Cut}(M) := \bigcup_{p\in M} (\{p\}\times \mathrm{Cut}_p(M)) \subset M\times M.$$

Example 2.6 Let $n \in \mathbb{N}$ and let g be a round metric on the sphere S^n. Then, by [27, Example 10.30(a)], $\mathrm{Cut}_p(S^n) = \{-p\}$ for every $p \in S^n$.

Further examples of cut loci will appear in the upcoming sections.

Remarks 2.7 Let (M,g) be a complete Riemannian manifold.

(1) In general, $\mathrm{Cut}_p(M)$ does not need to be a submanifold of M. H Gluck and D Singer [20, Theorem A] have shown that, if $\dim M \geq 2$, then there exists a Riemannian metric on M and a point $p \in M$ for which $\mathrm{Cut}_p(M)$is not triangulable.

(2) By [34, Theorem 3.3], there exists a continuous geodesic motion planner

$$(M\times M)\smallsetminus \mathrm{Cut}(M) \to GM,$$

from which Recio-Mitter derived that $\mathsf{GC}(M)=1$ if $\mathrm{Cut}(M)=\varnothing$. By [3, Lemma 4.2], $(M\times M)\smallsetminus\mathrm{Cut}(M)$ is open and therefore locally compact. Using (2-1), this shows that

$$\mathsf{GC}(M,g) \leq \mathsf{GC}_M(\mathrm{Cut}(M)) + 1.$$

(3) Let $p \in M$. By [2, page 133], the set of points $q \in M$ such that there is more than one minimal geodesic from p to q is a dense subset of $\mathrm{Cut}_p(M)$. This set is also called the *ordinary cut locus* of p. In metric geometry — in particular in

[34, Definition 3.1] — the ordinary cut locus of a point is called its cut locus. The reader should thus keep in mind that the cut locus of a point, as considered in [34], is *not* the cut locus of a point in the sense of this article, but a dense subset of the cut locus.

3 Nonexistence results for geodesic motion planners

We begin our study by discussing two nonexistence results showing that certain subsets of a Riemannian manifold never admit continuous geodesic motion planners. First, we will study complete oriented Riemannian manifolds and see that the Euler class obstructs the existence of some geodesic motion planners. Then we will show that a complete Riemannian manifold (M,g) has the following property: if a subset $A \subset M \times M$ contains an element of the total cut locus in its interior, then there will be no continuous geodesic motion planner on A. Before doing so, we first want to establish a technical proposition that we will make frequent use of throughout the article.

Definition 3.1 Let (M,g) be a complete Riemannian manifold. We call the map

$$v\colon GM \to TM, \quad v(\gamma) = \dot{\gamma}(0),$$

the *velocity map of GM*.

Proposition 3.2 *Let (M,g) be a complete Riemannian manifold. The velocity map $v\colon GM \to TM$ is continuous.*

Proof Let $(\gamma_n)_{n\in\mathbb{N}}$ be a convergent sequence in GM and let $\gamma := \lim_{n\to\infty} \gamma_n \in GM$. By our choice of topology on GM, this means that

$$\lim_{n\to\infty} \gamma_n(t) = \gamma(t) \quad \text{for all } t \in [0,1]. \tag{3-1}$$

We need to show that $\lim_{n\to\infty} v(\gamma_n) = v(\gamma)$. Let $\mathcal{L}_g : GM \to \mathbb{R}$ denote the length of a minimal geodesic with respect to g. From the minimality property of the curves, we derive that

$$\lim_{n\to\infty} \mathcal{L}_g(\gamma_n) = \lim_{n\to\infty} \mathrm{d}_M(\gamma_n(0),\gamma_n(1)) = \mathrm{d}_M(\gamma(0),\gamma(1)) = \mathcal{L}_g(\gamma),$$

where $\mathrm{d}_M : M \times M \to \mathbb{R}$ is the distance function induced by g. Let $|\cdot| : TM \to \mathbb{R}$ denote the fiberwise norm induced by g. Since $\mathcal{L}_g(\alpha) = |\dot{\alpha}(0)| = |v(\alpha)|$ for each $\alpha \in GM$, it follows that

$$\lim_{n\to\infty} |v(\gamma_n)| = |v(\gamma)|. \tag{3-2}$$

To show the continuity of v, we need to derive that $\lim_{n\to\infty} v(\gamma_n) = v(\gamma)$. Let

$$\mathrm{Exp}\colon TM \to M \times M, \quad \mathrm{Exp}(p, v) = (p, \exp_p(v)),$$

be the extended exponential map. Let $K \subset M$ be a compact neighborhood of $\gamma(0)$ and let

$$\rho_0 := \sup\{r > 0 : \exp_x|_{B_r(0)} \text{ is injective for all } x \in K\},$$

where $B_r(0)$ denotes the open n–ball around the origin in the respective tangent space. Since K is compact, $\rho_0 > 0$ by [27, Lemma 6.16]. For $r \in (0, \rho_0)$ we put

$$D_r K := \{(p, v) \in TM \mid p \in K, \|v\| \le r\},$$

ie $D_r K$ is the closed disk bundle over K of radius r. Then Exp maps $D_r K$ diffeomorphically onto its image

$$V_r K := \mathrm{Exp}(D_r K) = \{(p, q) \in K \times M \mid \mathrm{d}_M(p, q) \le r\}.$$

Let $\mathrm{Exp}_K \colon D_r K \to V_r K$ be the corresponding restriction of Exp. Since $\mathrm{Exp}_K \colon D_r K \to V_r K$ is a diffeomorphism, its inverse $\mathrm{Exp}_K^{-1} \colon V_r K \to D_r K$ is a diffeomorphism as well. Thus, if we choose and fix a distance function $\mathrm{d}_{TM} \colon TM \times TM \to \mathbb{R}$ which induces the topology of TM, then $\mathrm{Exp}_K^{-1} \colon V_r K \to D_r K$ is locally Lipschitz-continuous with respect to $\mathrm{d}_M \times \mathrm{d}_M$ and d_{TM}. We further observe that, for all $\alpha \in GM$ with $\alpha(0) \in K$ and $\mathrm{d}_M(\alpha(0), \alpha(1)) \le r$,

$$\mathrm{Exp}_K^{-1}(\alpha(0), \alpha(1)) = v(\alpha).$$

We consider two different cases:

Case 1 Assume that $|v(\gamma)| < r$. This implies that $(\gamma(0), \gamma(1)) \in V_r K$. Then, by (3-2), there exists $n_0 \in \mathbb{N}$ with

$$\gamma_n(0) \in K \quad \text{and} \quad |v(\gamma_n)| \le r \quad \text{for all } n \ge n_0.$$

Thus, $(\gamma_n(0), \gamma_n(1)) \in V_r K$ for all $n \ge n_0$. Let C be a local Lipschitz constant for Exp_K^{-1} in a neighborhood of $(\gamma(0), \gamma(1))$. Then, for sufficiently big $n \in \mathbb{N}$,

$$\begin{aligned} \mathrm{d}_{TM}(v(\gamma_n), v(\gamma)) &= \mathrm{d}_{TM}\big(\mathrm{Exp}_K^{-1}(\gamma_n(0), \gamma_n(1)), \mathrm{Exp}_K^{-1}(\gamma(0), \gamma(1))\big) \\ &\le C\big(\mathrm{d}_M(\gamma_n(0), \gamma(0)) + \mathrm{d}_M(\gamma_n(1), \gamma(1))\big). \end{aligned}$$

By (3-1), this yields $\lim_{n\to\infty} \mathrm{d}_{TM}(v(\gamma_n), v(\gamma)) = 0$, which we wanted to show.

Case 2 Consider the case that $|v(\gamma)| \ge r$. By (3-2), there exists $n_1 \in \mathbb{N}$ such that

$$\gamma_n(0) \in K \quad \text{and} \quad |v(\gamma_n)| < |v(\gamma)| + 1 \quad \text{for all } n \ge n_1.$$

Let $c := r/\bigl(|v(\gamma)|+1\bigr) \in (0,1)$ and put $\xi_n := c \cdot v(\gamma_n)$ for each $n \in \mathbb{N}$ and $\xi := c \cdot v(\gamma)$. Then $\xi \in D_r K$ and $\xi_n \in D_r K$ for all $n \geq n_1$. If we define

$$\widetilde{\gamma}_n, \widetilde{\gamma}\colon [0,1] \to M, \qquad \widetilde{\gamma}_n(t) := \gamma_n(ct), \quad \widetilde{\gamma}(t) := \gamma(ct) \quad \text{for all } t \in [0,1],$$

then $\widetilde{\gamma}, \widetilde{\gamma}_n \in GM$ with $v(\widetilde{\gamma}) = \xi$ and $v(\widetilde{\gamma}_n) = \xi_n$ for each $n \geq n_1$. Since $(\gamma_n)_{n\in\mathbb{N}}$ converges to γ in the C^0–topology, it easily follows that $\lim_{n\to\infty} \widetilde{\gamma}_n = \widetilde{\gamma}$ in the C^0–topology as well. Thus, it follows from Case 1 that $\lim_{n\to\infty} \xi_n = \xi$, which obviously yields $\lim_{n\to\infty} v(\gamma_n) = v(\gamma)$. □

In the following proposition, we observe that the Euler class of an oriented manifold can obstruct the existence of geodesic motion planners:

Proposition 3.3 *Let (M,g) be a complete, oriented Riemannian manifold whose Euler class is nonvanishing. Let $f\colon M \to M$ be a continuous map with $f(p) \neq p$ for all $p \in M$. If $A \subset M \times M$ satisfies* graph $f \subset A$, *then there will be no continuous geodesic motion planner on A.*

Proof Assume by contradiction that there exists a continuous geodesic motion planner $s\colon A \to GM$. Then, by Proposition 3.2, the map

$$g\colon M \to TM, \quad g(p) = (v \circ s)(p, f(p)),$$

is a continuous vector field, where v is the velocity map. Since $f(p) \neq p$ for each p, the geodesic $s(p, f(p))$ is nonconstant for all $p \in M$. Hence, $g(p) \neq 0$ for all $p \in M$. But such a vector field cannot exist since the Euler class of M is nonvanishing. □

Corollary 3.4 *Let (M,g) be a complete, oriented manifold whose Euler class is nonvanishing. Let $f\colon M \to M$ be continuous and fixed-point-free. Then, for every Riemannian metric g on M, there exists $p \in M$ with $f(p) \in \mathrm{Cut}_p(M,g)$.*

Proof Assume by contradiction that there is such a metric g for which $f(p) \notin \mathrm{Cut}_p(M,g)$ for all $p \in M$. Then graph f lies in $(M \times M) \setminus \mathrm{Cut}(M,g)$. But, since there exists a continuous geodesic motion planner on $(M \times M) \setminus \mathrm{Cut}(M,g)$ — see Remark 2.7(3) — this contradicts Proposition 3.3. Hence, such a metric does not exist. □

Corollary 3.5 *Let $n \in \mathbb{N}$. For every Riemannian metric g on S^{2n} there exists $p \in S^{2n}$, such that $-p \in \mathrm{Cut}_p(S^{2n}, g)$.*

Proof Apply Corollary 3.4 to the case of $M = S^{2n}$ and $f(x) = -x$. □

Remark 3.6 Our Corollary 3.4 is complementary to results of M Frumosu and S Rosenberg from [17, page 338], who studied far-point sets, ie sets of points mapped to their cut loci under self-maps of a Riemannian manifold, in a very general way. Frumosu and Rosenberg focused on self-maps whose far-point sets are infinite and established connections to the Lefschetz numbers of such maps.

In [34, Remark 3.17], Recio-Mitter mentioned that, whenever a subset of $M \times M$ contains a point of the total cut locus in its interior, there is no continuous geodesic motion planner defined on that subset. For the sake of completeness, we report here a proof in the case of Riemannian manifolds.

Proposition 3.7 *Let (M, g) be a complete Riemannian manifold, $p \in M$, $q \in \mathrm{Cut}_p(M)$ and let $U \subset M$ be an open neighborhood of q. Then there is no continuous geodesic motion planner on $\{p\} \times U$.*

Proof As discussed in Remark 2.7(3), the set of points $r \in M$ for which there is more than one minimal geodesic from p to r is dense in $\mathrm{Cut}_p(M)$. Hence, U contains a point q_0 such that there are at least two minimal geodesics from p to q_0. In the following, we thus assume without loss of generality that q itself has this property. Assume that a continuous geodesic motion planner $s\colon \{p\} \times U \to GM$ existed. By our choice of q, there are $\gamma_1, \gamma_2 \in GM$ with

$$\gamma_1 \neq \gamma_2, \quad \gamma_1(0) = \gamma_2(0) = p \quad \text{and} \quad \gamma_1(1) = \gamma_2(1) = q.$$

Let $(t_n)_{n\in\mathbb{N}}$ be a sequence in $(0, 1)$ with $\lim_{n\to\infty} t_n = 1$ and $\gamma_1(t_n), \gamma_2(t_n) \in U$ for all $n \in \mathbb{N}$. One checks without difficulties that $\gamma_1(t) \neq \gamma_2(t)$ for all $t \in (0, 1)$, so that, in particular, $\gamma_1(t_n) \neq \gamma_2(t_n)$ for all $n \in \mathbb{N}$.

By definition of a cut locus, it follows for all $r \in (0, 1)$ and $i \in \{1, 2\}$ that

$$\gamma_{i,r} \in GM, \quad \gamma_{i,r}(t) := \gamma_i(rt),$$

is the unique minimal geodesic from p to $\gamma_i(r)$. In particular, this shows that necessarily

$$s(p, \gamma_i(t_n)) = \gamma_{i,t_n} \quad \text{for all } n \in \mathbb{N},\ i \in \{1, 2\}. \tag{3-3}$$

Let $v\colon GM \to TM$ be the velocity map. It follows from Proposition 3.2 that

$$v \circ s\colon \{p\} \times U \to TM$$

is continuous. Since $\gamma_1 \neq \gamma_2$, there are $\xi_1, \xi_2 \in T_pM$ with $\xi_1 \neq \xi_2$ such that $\gamma_1(t) = \exp_p(t\xi_1)$ and $\gamma_2(t) = \exp_p(t\xi_2)$ for all $t \in [0, 1]$. By (3-3) and the fact that the differential of $\exp_p$ in 0 is id_{T_pM}, we thus obtain that

$$\lim_{n\to\infty} (v \circ s)(p, \gamma_i(t_n)) = \lim_{n\to\infty} \dot{\gamma}_{i,t_n}(0) = \lim_{n\to\infty} t_n\xi_i = \xi_i.$$

In particular, $\lim_{n\to\infty}(v\circ s)(p,\gamma_1(t_n)) \neq \lim_{n\to\infty}(v\circ s)(p,\gamma_2(t_n))$. This contradicts the continuity of s, since by assumption $(v\circ s)(p,\gamma_1(1)) = (v\circ s)(p,\gamma_2(1))$. Thus, such a continuous s does not exist. □

This proposition has an immediate consequence in terms of geodesic complexity.

Corollary 3.8 *Let (M,g) be a complete Riemannian manifold and let $A\subset M\times M$ be a locally compact subset with*

$$\mathrm{int}(A)\cap \mathrm{Cut}(M)\neq\varnothing,$$

where $\mathrm{int}(A)$ is the interior of A as a subset of $M\times M$. Then $\mathsf{GC}_M(A)\geq 2$.

Proof Assume that there was a continuous geodesic motion planner $s\colon A\to GM$. Let $(p,q)\in\mathrm{int}(A)\cap\mathrm{Cut}(M)$. By definition of the product topology, there are open neighborhoods U of p and V of q with $U\times V\subset\mathrm{int}(A)$, so, in particular, $s|_{\{p\}\times V}$ would be a continuous geodesic motion planner. Since $q\in\mathrm{Cut}_p(M)$, this contradicts Proposition 3.7, so there is no such motion planner. This shows that $\mathsf{GC}_M(A)\geq 2$. □

Remark 3.9 There is another connection between cut loci and another numerical invariant, namely the Lusternik–Schnirelmann category of a Riemannian manifold M, which we denote by $\mathsf{cat}(M)$. Here, we use the convention that $\mathsf{cat}(X)=1$ if X is contractible. One observes that $M\smallsetminus\mathrm{Cut}_p(M)$ is contractible for all $p\in M$, which follows from [27, Theorem 10.34(c)]. If $p_1,\dots,p_k\in M$ satisfy $\bigcap_{i=1}^k\mathrm{Cut}_{p_i}(M)=\varnothing$, then

$$\{M\smallsetminus\mathrm{Cut}_{p_1}(M),\dots,M\smallsetminus\mathrm{Cut}_{p_k}(M)\}$$

will be an open cover of M by contractible subsets, and hence $\mathsf{cat}(M)\leq k$. By contraposition this shows that, if $\mathsf{cat}(M)\geq k+1$ for some $k\in\mathbb{N}$, then, for every choice of $p_1,\dots,p_k\in M$,

$$\bigcap_{i=1}^k\mathrm{Cut}_{p_i}(M)\neq\varnothing.$$

4 Lower bounds for geodesic complexity

Lower bounds on topological complexity are mostly derived from the cohomology rings of a space. In this section, we derive lower bounds on geodesic complexity from the Riemannian structures of manifolds. We first establish a result involving a principal bundle over the manifold under consideration. By explicitly constructing motion planners, we will establish a lower bound on geodesic complexity in terms of categorical invariants of total space and fiber of the bundle. Afterwards, we will establish the notion of inconsistent stratification that we lined out in the introduction. Then we will go on to prove the second theorem stated in that introduction.

We first establish a technical lemma, whose proof follows that of [14, Theorem 13.1].

Lemma 4.1 *Let E and X be topological spaces, let $p\colon E \to X$ be a fibration with $r := \mathsf{secat}(p) < +\infty$ and assume that X is normal. Then there are pairwise disjoint locally compact subsets $A_1, \dots, A_r \subset X$ with $X = \bigcup_{i=1}^r A_i$ such that, for each $i \in \{1, 2, \dots, r\}$, there exists a continuous local section $A_i \to E$ of p.*

Proof Let $\{U_1, \dots, U_r\}$ be an open cover of X such that, for each $i \in \{1, 2, \dots, r\}$, there exists a continuous local section $s_i : U_i \to E$ of p. Since X is normal, there exists a partition of unity $\{f_1, \dots, f_r\}$ subordinate to this finite open cover by Theorem 36.1 of [30]. Let $c_1, \dots, c_r \in (0, +\infty)$ with $c_1 + \dots + c_r = 1$. For each $i \in \{1, 2, \dots, r\}$, we put

$$A_i := \{x \in X \mid f_i(x) \geq c_i,\ f_j(x) < c_j \text{ for all } j < i\}.$$

Each A_i is the intersection of a closed and an open subset of X, and hence is locally compact. One checks without difficulties that the A_i are pairwise disjoint and that $X = \bigcup_{i=1}^r A_i$. Moreover, $A_i \subset U_i$ for each i, so $s_i|_{A_i} : A_i \to E$ is a continuous local section of p for each $i \in \{1, 2, \dots, r\}$. □

The following proposition establishes a lower bound on $\mathsf{GC}(M, g)$ in terms of a principal G–bundle over M that is a Riemannian submersion. This submersion property will be used in its proof to ensure the existence of horizontal lifts of curves. For each orientable M, its orthonormal frame bundle is an example of such a bundle with $G = \mathrm{SO}(\dim M)$; see eg [25, Example I.5.7].

Proposition 4.2 *Let (M, g) be a complete Riemannian manifold and let $\pi : E \to M$ be a smooth principal G–bundle, where G is a connected Lie group. Assume that E is equipped with a Riemannian metric for which π is a Riemannian submersion. Then*

$$\mathsf{GC}(M, g) \geq \frac{\mathsf{TC}(E)}{\mathsf{cat}(G)}.$$

Proof Let $\mathsf{GC}(M)=k$ and choose pairwise disjoint and locally compact subsets $A_1,\dots,A_k\subset M\times M$ with $\bigcup_{i=1}^k A_i=M\times M$ such that, for each $i\in\{1,2,\dots,k\}$, there exists a continuous geodesic motion planner $s_i\colon A_i\to GM$. Let $v\colon GM\to TM$ be the velocity map and put

$$v_i\colon A_i\to TM,\quad v_i:=v\circ s_i\quad\text{for all } i\in\{1,2,\dots,k\}.$$

The v_i are continuous by Proposition 3.2. For each i we put

$$B_i:=(\pi\times\mathrm{id})^{-1}(A_i)=\{(u,q)\in E\times M\mid(\pi(u),q)\in A_i\}.$$

Clearly the B_i are again pairwise disjoint with $\bigcup_{i=1}^k B_i=E\times M$. Let $\mathrm{Hor}(E)\subset TE$ denote the horizontal subbundle with respect to π. Since $d\pi|_{\mathrm{Hor}(E)}\colon\mathrm{Hor}(E)\to TM$ maps $\mathrm{Hor}_u(E)$ isomorphically onto $T_{\pi(u)}M$ for each $u\in E$, we obtain continuous lifts of the v_i by

$$w_i\colon B_i\to\mathrm{Hor}(E),\quad w_i(u,q)=(d\pi|_{\mathrm{Hor}_u(E)})^{-1}v_i(\pi(u),q)\quad\text{for all } i\in\{1,2,\dots,k\}.$$

For each $u\in E$ we let $\exp_u\colon T_uE\to E$ be the exponential map of the given Riemannian metric on E. With $PE=C^0([0,1],E)$, we define continuous maps

$$\eta_i\colon B_i\to PE$$

by

$$(\eta_i(u,q))(t)=\exp_u(t\,w_i(u,q))\quad\text{for all } (u,q)\in B_i,\ t\in[0,1],\ i\in\{1,2,\dots,k\}.$$

Each η_i induces a continuous map

$$\alpha_i\colon B_i\to E\times E,\quad \alpha_i(u,q)=\big((\eta_i(u,q))(0),(\eta_i(u,q))(1)\big)=\big(u,\exp_u(w_i(u,q))\big).$$

Since horizontal geodesics in E project to geodesics in M, we compute that

$$(\mathrm{id}\times\pi)(\alpha_i(u,q))=(\mathrm{id}\times\pi)(u,\eta_i(u,q)(1))=\big(u,\big(s_i(\pi(u),q)\big)(1)\big)=(u,q)$$

for all $(u,q)\in B_i$. Here we used that $\pi(\eta_i(u,q))=s_i(u,q)$ for all $(u,q)\in B_i$. Hence, for each $i\in\{1,2,\dots,k\}$, the map α_i is a continuous local section of $\mathrm{id}\times\pi\colon E\times E\to E\times M$, which is again a principal G–bundle. The right G–action on $E\times E$ is given by $E\times E\times G\to E\times E$, $(u,v,h)\mapsto(u,vh)$, where we consider the right G–action on E given by the bundle structure. Thus, we get a local trivialization of $\mathrm{id}\times\pi$ over each B_i, given explicitly by the homeomorphism

$$\Phi_i\colon B_i\times G\to E\times E|_{B_i},\quad \Phi_i(u,q,h)=\alpha_i(u,q)h=\big(u,\exp_u(w_i(u,q))h\big).$$

Put $l = \mathsf{cat}(G)$. Let $e \in G$ be the unit, $P_eG = \{\gamma \in PG \mid \gamma(0) = e\}$ and

$$q\colon P_eG \to G, \quad q(\gamma) = \gamma(1).$$

Since P_eG is contractible, by [39, Theorem 18], $\mathsf{cat}(G) = \mathsf{secat}(q\colon P_eG \to G)$. By Lemma 4.1, there are pairwise disjoint and locally compact subsets $C_1, \dots, C_l \subset G$ with $\bigcup_{j=1}^{l} C_j = G$ such that, for each $j \in \{1, 2, \dots, l\}$, there is a continuous local section $r_j : C_j \to P_eG$ of q.

If we put $D_{i,j} := \Phi_i(B_i \times C_j) \subset E \times E$ for all $i \in \{1, 2, \dots, k\}$ and $j \in \{1, 2, \dots, l\}$, then the $D_{i,j}$ are pairwise disjoint, locally compact and satisfy $\bigcup_{i=1}^{k} \bigcup_{j=1}^{l} D_{i,j} = E \times E$. For all i and j we further consider the map

$$\sigma_{i,j} : B_i \times C_j \to PE$$

given by

$$(\sigma_{i,j}(u,q,h))(t) = (\eta_i(u,q))(t) \cdot (r_j(h))(t) \quad \text{for all } (u,q) \in B_i,\ h \in C_j.$$

Then

$$\begin{aligned}
(\sigma_{i,j}(u,q,h))(0) &= (\eta_i(u,q))(0) = u, \\
(\sigma_{i,j}(u,g,h))(1) &= (\eta_i(u,q))(1)(r_j(h))(1) = \exp_u(w_i(u,q))h
\end{aligned}$$

and thus

$$(\sigma_{i,j}(u,q,h)(0), \sigma_{i,j}(u,q,h)(1)) = \Phi_i(u,q,h) \quad \text{for all } (u,q) \in B_i,\ h \in C_j.$$

This shows that $\sigma_{i,j} \circ \Phi_i^{-1}|_{D_{i,j}} : D_{i,j} \to PE$ is a continuous motion planner for all $i \in \{1, 2, \dots, k\}$ and $j \in \{1, 2, \dots, l\}$. As a smooth manifold, E is a Euclidean neighborhood retract (ENR). Since the $D_{i,j}$ are locally compact subsets of an ENR, they are ENRs themselves. Hence, it follows from [13, Theorem 6.1] that

$$\mathsf{TC}(E) \le k \cdot l = \mathsf{GC}(M) \cdot \mathsf{cat}(G),$$

which proves the claimed inequality. □

Remark 4.3 Since $\mathsf{GC}(M) \ge \mathsf{TC}(M)$ for all complete Riemannian manifolds M, the lower bound from Proposition 4.2 improves this basic inequality if and only if

$$\frac{\mathsf{TC}(E)}{\mathsf{cat}(G)} > \mathsf{TC}(M) \iff \mathsf{TC}(E) > \mathsf{cat}(G)\,\mathsf{TC}(M) = \mathsf{TC}(G)\,\mathsf{TC}(M),$$

where we used [13, Lemma 8.2]. Note that the assumption on the bundle to be *principal* in the previous result is necessary, as the following example shows. Consider the Klein bottle K, which is given as a fiber bundle over S^1 with fiber S^1 and satisfies $\mathsf{TC}(K) = 5$ by [5], while $\mathsf{TC}(S^1) = 2$. Since the round metric g_r on S^1 satisfies

$$\mathsf{GC}(S^1, g_r) = \mathsf{TC}(S^1) = 2 < \tfrac{5}{2} = \frac{\mathsf{TC}(K)}{\mathsf{cat}(S^1)}$$

by [34, Proposition 4.1], the inequality from Proposition 4.2 would indeed be false in this situation. However, K is *not* given as a *principal* S^1–bundle over S^1, so Proposition 4.2 is not applicable to this setting. By the classification theorem for principal bundles — see [11, Theorem 14.4.1] — the set of isomorphism classes of principal S^1–bundles over S^1 is in bijection with the set of homotopy classes $[S^1, BS^1] = [S^1, \mathbb{C}P^\infty]$. But $\mathbb{C}P^\infty$ is simply connected, so it follows that $[S^1, \mathbb{C}P^\infty]$ has only one element. Thus, every principal S^1–bundle over S^1 is trivial. Since $\pi_1(K) \ncong \mathbb{Z}^2 = \pi_1(S^1 \times S^1)$, the bundle K is a nontrivial S^1–bundle. Hence, it cannot be principal.

Our next aim is to derive a lower bound on geodesic complexity from the structure of the cut locus of a point in the manifold. We first introduce the notion of stratification that we are using.

Definition 4.4 Let M be a manifold and let $B \subset M$ be a subset. A *stratification of B of depth $N \in \mathbb{N}$* is a family $(S_1, \dots, S_N)$ of locally closed and pairwise disjoint subsets of M such that the following conditions hold:

(i) $B = \bigcup_{i=1}^N S_i$ and $\overline{S}_i = \bigcup_{j=i}^N S_j$ for all $i \in \{1, 2, \dots, N\}$.

(ii) Let $i, j \in \{1, 2, \dots, N\}$. If Z_j is a connected component of S_j and Z_i is a connected component of S_i with $Z_j \cap \overline{Z}_i \neq \varnothing$, then $Z_j \subset \overline{Z}_i$.

Example 4.5 Let $M = \mathbb{R}^2$ and let $B = [-1, 1]^2$. Consider

$$\begin{aligned} S_1 &= (-1, 1) \times (-1, 1), \\ S_2 &= ((-1, 1) \times \{-1, 1\}) \cup (\{-1, 1\} \times (-1, 1)), \\ S_3 &= \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}. \end{aligned}$$

One checks without difficulties that (S_1, S_2, S_3) has properties (i) and (ii) from Definition 4.4. Hence, (S_1, S_2, S_3) is a stratification of B.

Given a stratification of the cut locus of a point, we want to introduce an additional condition on those parts of the corresponding tangent cut locus that are mapped to the same stratum. This will be the crucial step for finding a lower bound for geodesic complexity. The following notion is an analogue of [34, Definition 3.10]; see our introduction and Remark 4.7(2) below for a comparison of the two notions. The terms from Riemannian geometry that are used are to be found, for example, in [27, page 310].

Definition 4.6 Let (M, g) be a complete Riemannian manifold, $p \in M$ and let $\mathcal{S} = (S_1, \dots, S_N)$ be a stratification of $\mathrm{Cut}_p(M)$. Let $K \subset T_pM$ denote the union of the tangent cut locus $\widetilde{\mathrm{Cut}}_p(M)$ with the domain of injectivity of $\exp_p$ and let

$$\exp_K := \exp_p|_K \colon K \to M \tag{4-1}$$

denote the restriction. We call $\mathcal{S}$ *inconsistent* if, for all $i \in \{2, 3, \dots, N\}$ and $x \in S_i$, there exists an open neighborhood $U \subset M$ of x with the following property: Let $Z_1, \dots, Z_s$ be the connected components of $U \cap S_{i-1}$. Then $x \in \overline{Z}_j$ for all $j \in \{1, 2, \dots, s\}$ and

$$\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{x\}) \cap \bigcap_{j=1}^{s} \overline{\exp_K^{-1}(Z_j)} = \varnothing.$$

In Section 7.1, we will encounter explicit examples of inconsistent stratifications when we consider flat tori. Examples for cut loci with nontrivial stratifications which are not inconsistent are Berger spheres, as we shall see in Section 7.2.

Remarks 4.7 Let (M, g) be a complete Riemannian manifold.

(1) If M is a closed manifold, then the set K from Definition 4.6 will be homeomorphic to a closed ball — see [27, Corollary 10.35] — and the map $\exp_K$ from (4-1) is a surjection. As an example, consider the round n–dimensional sphere S^n of radius 1. If $p \in S^n$ is a point, then the domain of injectivity of $\exp_p$ is an open ball of radius π in the tangent space T_pS^n. The tangent cut locus $\widetilde{\mathrm{Cut}}_p(S^n)$ is the $(n-1)$–sphere of radius π in T_pM. Consequently, the set K in this example is the closed ball of radius π in T_pM.

(2) Recio-Mitter [34, Definition 3.8] introduced the concept of a levelwise stratified covering for arbitrary surjective maps. He then applied this concept to the restriction of the path fibration

$$\pi \colon GX \to X \times X,$$

where X is a geodesic space and GX is the space of geodesic paths in X.

To work with this notion, one must study a stratification of the total cut locus of X and explore covering properties of the restrictions of π to its preimage. In contrast, the above Definition 4.6 for Riemannian manifolds only requires a stratification of the cut locus of a single point p in a Riemannian manifold M as well as properties of the Riemannian exponential map $\exp_p$. Thus, for complete Riemannian manifolds the above definition seems easier to verify than the corresponding notion from [34].

The following result is an analogue of the corresponding result of Recio-Mitter; see [34, Corollary 3.14]. The proof requires M to be compact, since we will use the property mentioned in Remark 4.7(1). We recall the notation $\mathsf{GC}_p(A) = \mathsf{GC}_M(\{p\} \times A)$ for all $A \subset M$.

Theorem 4.8 *Let (M, g) be a closed Riemannian manifold. Assume that there exists $p \in M$ for which $\mathrm{Cut}_p(M)$ admits an inconsistent stratification of depth $N \in \mathbb{N}$. Then*

$$\mathsf{GC}(M) \geq \mathsf{GC}_p(M) \geq N + 1.$$

Proof Let $(S_1, \dots, S_N)$ be an inconsistent stratification of $\mathrm{Cut}_p(M)$. Assume that there are pairwise disjoint locally compact sets $E_1, E_2, \dots, E_r \subset M$ with $\bigcup_{i=1}^r E_i = M$ such that, for each $i \in \{1, 2, \dots, r\}$, there exists a continuous geodesic motion planner $s_i : \{p\} \times E_i \to GM$.

We want to show by induction that, for all $k \in \{1, 2, \dots, N\}$ and all $x \in S_k$,

$$\#\bigl\{i \in \{1, 2, \dots, r\} \mid x \in \overline{E}_i\bigr\} \geq k + 1. \tag{4-2}$$

Consider the base case of $k = 1$ and assume by contradiction that there is an $i \in \{1, 2, \dots, r\}$ with $x \in \overline{E}_i$, but $x \notin \overline{E}_j$ for all $j \neq i$. Then x has an open neighborhood $U \subset M$ such that $U \subset E_i$ and the restriction $s_i|_{\{p\}\times U}$ is a continuous geodesic motion planner on $\{p\} \times U$. But, since $x \in \mathrm{Cut}_p(M)$, this contradicts Proposition 3.7. Hence, $\#\bigl\{i \in \{1, 2, \dots, r\} \mid x \in \overline{E}_i\bigr\} \geq 2$, which we wanted to show.

Assume as induction hypothesis that, for some $k \in \{2, 3, \dots, N\}$, we have shown that

$$\#\bigl\{i \in \{1, 2, \dots, r\} \mid y \in \overline{E}_i\bigr\} \geq k \quad \text{for all } y \in S_{k-1}.$$

Let $x \in S_k$. Assume that (4-2) is false and assume up to reordering that $x \notin \overline{E}_i$ for all $i > k$. Then there exists an open neighborhood U of x with $U \subset \bigcup_{i=1}^k E_i$. By the induction hypothesis, this yields

$$U \cap S_{k-1} \subset \overline{E}_i \quad \text{for all } i \in \{1, 2, \dots, k\}. \tag{4-3}$$

We assume without loss of generality that U is chosen as in Definition 4.6, since this can be achieved by shrinking U. We further assume that $x \in E_1$. Let $Z_1, \dots, Z_s$ be the connected components of $U \cap S_{k-1}$, where $s \in \mathbb{N}$ is suitably chosen.

Let $j \in \{1, 2, \ldots, s\}$ and let $(a_n)_{n\in\mathbb{N}}$ be a sequence in Z_j with $\lim_{n\to\infty} a_n = x$, which exists by our choice of U. For all $n \in \mathbb{N}$ it further holds by (4-3) that $a_n \in \overline{E}_1$. Thus, for each n, there exists a sequence $(b^n_m)_{m\in\mathbb{N}}$ in $U \cap E_1$ with $\lim_{m\to\infty} b^n_m = a_n$. Put

$$v_1\colon E_1 \to T_pM, \quad v_1(y) := (v \circ s_1)(p, y),$$

where v is the velocity map. By Proposition 3.2, v_1 is continuous. Let $\exp_K\colon K \to M$ be given as in (4-1). The set K is homeomorphic to a closed ball in T_pM; see Remark 4.7(1). By construction, $v_1(y) \in K$ for each $y \in E_1$; hence, $(v_1(b^n_m))_{m\in\mathbb{N}}$ is a sequence in K for each $n \in \mathbb{N}$. Since K is compact, it has a convergent subsequence $(v_1(b^n_{m_k}))_{k\in\mathbb{N}}$ for each $n \in \mathbb{N}$. Put $\xi_n := \lim_{k\to\infty} v_1(b^n_{m_k})$ for all $n \in \mathbb{N}$. By continuity of the exponential map,

$$\exp_K(\xi_n) = \exp_p(\xi_n) = \lim_{k\to\infty} \exp_p(v_1(b^n_{m_k})) = \lim_{k\to\infty} b^n_{m_k} = a_n \quad \text{for all } n \in \mathbb{N}.$$

Thus,

$$\xi_n \in K \cap \exp_K^{-1}(\{a_n\}) \subset K \cap \exp_K^{-1}(Z_j) \quad \text{for all } n \in \mathbb{N}.$$

Now $(\xi_n)_{n\in\mathbb{N}}$ is a sequence in K, so it has a convergent subsequence $(\xi_{n_l})_{l\in\mathbb{N}}$. With $\xi_0 := \lim_{l\to\infty} \xi_{n_l}$, we obtain

$$\exp_p(\xi_0) = \lim_{l\to\infty} \exp_p(\xi_{n_l}) = \lim_{l\to\infty} a_{n_l} = x.$$

In particular, it follows from $x \in \mathrm{Cut}_p(M)$ that $\xi_0 \in \widetilde{\mathrm{Cut}}_p(M)$. Since $\xi_n \in \exp_K^{-1}(\{a_n\})$ for each $n \in \mathbb{N}$, we conclude that

$$\xi_0 \in \widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{x\}) \cap \overline{\exp_K^{-1}(Z_j)}.$$

Note that ξ_0 depends on the choice of j. To conclude, we still need to show that the same ξ_0 can be chosen for each $j \in \{1, 2, \ldots, s\}$. We will do so by showing next that $\xi_0 = v_1(x)$, which does not depend on j.

Let $d_M\colon M \times M \to \mathbb{R}$ be the distance function induced by the Riemannian metric. By definition of the ξ_{n_l}, for each $l \in \mathbb{N}$ there exists $k_l \in \mathbb{N}$ such that

$$d_M(a_{n_l}, b^{n_l}_{m_k}) < \frac{1}{l} \quad \text{and} \quad \|\xi_{n_l} - v_1(b^{n_l}_{m_k})\| < \frac{1}{l} \quad \text{for all } k \geq k_l.$$

We can further choose the k_l in such a way that $\lim_{l\to\infty} k_l = \infty$. By a diagonal argument, $\lim_{l\to\infty} b^{n_l}_{m_{k_l}} = x$. This in particular shows, by continuity of v_1, that

$$\xi_0 = \lim_{l\to\infty} \xi_{n_l} = \lim_{l\to\infty} v_1(b^{n_l}_{m_{k_l}}) = v_1(x).$$

Thus, $v_1(x) \in \exp_p^{-1}(\{x\}) \cap \overline{\exp_K^{-1}(Z_j)}$. Since j was chosen arbitrarily, it follows that

$$v_1(x) \in \widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{x\}) \cap \bigcap_{j=1}^{s} \overline{\exp_K^{-1}(Z_j)}.$$

This contradicts the inconsistency of the stratification $(S_1, \dots, S_N)$. Hence, there is no such U, which concludes the proof of the induction step. For $k = N$, it in particular follows from (4-2) that $r \geq N+1$. Thus, $\mathsf{GC}_p(M) \geq N+1$. □

We will see in Section 7.1 that flat tori are indeed examples for Riemannian manifolds whose cut loci admit inconsistent stratifications. Next we will discuss a more tangible criterion on a cut locus that implies the existence of an inconsistent stratification. For this purpose, we will use results and constructions of J-I Itoh and Sakai [24]. Large parts of these methods are extensions of those applied by V Ozols [32].

Definition 4.9 [24, page 68 and Definition 2.1] Let (M, g) be a complete Riemannian manifold and let $p \in M$.

(a) We say that $q \in \mathrm{Cut}_p(M)$ is *of order* $k+1$, where $k \in \mathbb{N}$, if there are precisely $k+1$ minimal geodesics $\gamma_0, \gamma_1, \dots, \gamma_k \in GM$ with $\gamma_i \neq \gamma_j$ if $i \neq j$ and with $\gamma_i(0) = p$ and $\gamma_i(1) = q$ for all $i \in \{0, 1, 2, \dots, k\}$.

(b) We call q *nondegenerate* if the vectors $\dot\gamma_0(1), \dot\gamma_1(1), \dots, \dot\gamma_k(1) \in T_qM$ are in general position, ie if $\{\dot\gamma_i(1) - \dot\gamma_0(1) \mid i \in \{1, 2, \dots, k\}\}$ is linearly independent.

As carried out by Itoh and Sakai [24, Remark 2.2], a large class of two-dimensional flat tori provides an example for manifolds with nondegenerate cut points. However, our study of flat tori in Section 7.1 will not rely on this notion of nondegeneracy, but will employ the above inconsistency condition directly.

We recall that a *conjugate point* of a point p in a Riemannian manifold (M, g) is a point $q \in M$ such that there is a geodesic segment from p to q along which there exists a nontrivial Jacobi field which vanishes in p and q; see [27, page 298].

Remarks 4.10 (1) As shown by A Weinstein [40, page 29], every closed manifold M with $\dim M \geq 2$ and not homeomorphic to S^2 admits a Riemannian metric for which there exists $p \in M$ such that $\mathrm{Cut}_p(M)$ does not contain any conjugate points. Itoh and Sakai conjectured in [24, Remark 2.9] that the set of all such metrics on M contains as a dense subset the set of those metrics for which all points in $\mathrm{Cut}_p(M)$ are nondegenerate.

(2) It is evident from the definition of nondegeneracy that the order of a nondegenerate cut point is at most $\dim M + 1$.

Theorem 4.11 *Let (M, g) be a closed Riemannian manifold and assume that there exists $p \in M$ for which $\mathrm{Cut}_p(M)$ does not contain any conjugate points of p and for which all points in $\mathrm{Cut}_p(M)$ are nondegenerate. Let*

$$N := \max\{k \in \mathbb{N} \mid \text{there is } q \in \mathrm{Cut}_p(M) \text{ of order } k+1\}.$$

Then $\mathrm{Cut}_p(M)$ admits an inconsistent stratification of depth N.

Proof Let $\mathcal{C} := (C_1, \dots, C_N)$ be given by

$$C_k := \{q \in \mathrm{Cut}_p(M) \mid q \text{ is of order } k+1\} \quad \text{for all } k \in \{1, 2, \dots, N\}.$$

It is shown in [24, Proposition 2.4] that, under the nondegeneracy assumption on the points in $\mathrm{Cut}_p(M)$, $\mathcal{C}$ is a Whitney stratification of $\mathrm{Cut}_p(M)$, as defined in [21, page 37]. Hence, $\mathcal{C}$ is in particular an $\mathscr{S}$–decomposition in the sense of Goresky and MacPherson; see [21, page 36]. One checks immediately that the two conditions defining such an $\mathscr{S}$–decomposition imply that $\mathcal{C}$ is a stratification of $\mathrm{Cut}_p(M)$ in the sense of Definition 4.4. It remains to show that $\mathcal{C}$ is inconsistent. Fix $k \in \{1, 2, \dots, N\}$, let $q \in C_k$ and let $\gamma_0, \gamma_1, \dots, \gamma_k : [0,1] \to M$ be geodesics from p to q with $\gamma_i \neq \gamma_j$ whenever $i \neq j$. For each $i \in \{0, 1, \dots, k\}$, put $v_i := \dot\gamma_i(0) \in T_pM$, so that

$$\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{q\}) = \{v_0, v_1, \dots, v_k\}.$$

Choose an open neighborhood U of q such that $U \cap C_k$ is connected and such that

$$\mathrm{Cut}_p(M) \cap U = \bigcup_{i=1}^{k} C_i \cap U. \tag{4-4}$$

Such a neighborhood exists by the stratification properties. As discussed in [24, page 68], since q is nondegenerate, we can choose an open neighborhood $V_i \subset T_pM$ of v_i for each $i \in \{0, 1, \dots, k\}$ such that $\exp_p$ maps V_i diffeomorphically onto U. Put $F_i := (\exp|_{V_i})^{-1} : U \to V_i$. As explained in [32, pages 220–221], up to shrinking U we can assume that every minimal geodesic γ from p to an element of U has $\dot\gamma(0) \in \bigcup_{i=0}^{k} V_i$. We further assume that $\overline{V}_i \cap \overline{V}_j = \varnothing$ whenever $i \neq j$. For $i \in \{1, 2, \dots, k\}$, we define

$$f_i : U \to \mathbb{R}, \quad f_i(x) = \|F_i(x)\| - \|F_0(x)\|,$$

where $\|\cdot\|$ denotes the norm on T_pM defined by the Riemannian metric. With $f\colon U \to \mathbb{R}^k$, $f = (f_1, f_2, \dots, f_k)$, it follows that $f^{-1}(\{0\}) = C_k \cap U$. For $i \in \{1, 2, \dots, k\}$ we further let

$$g_i\colon U \to \mathbb{R}^{k-1}, \quad g_i = (f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k)$$

and put

$$g_0\colon U \to \mathbb{R}^{k-1},$$
$$g_0(x) = \big(\|F_2(x)\| - \|F_1(x)\|, \|F_3(x)\| - \|F_1(x)\|, \dots, \|F_k(x)\| - \|F_1(x)\|\big).$$

Then, by assumption on U,

$$C_{k-1} \cap U = \bigcup_{i=0}^{k} g_i^{-1}(\{0\}) \setminus C_k = \bigcup_{i=0}^{k} g_i^{-1}(\{0\}) \setminus f^{-1}(\{0\}).$$

The connected components of $C_{k-1} \cap U$ are the sets $Z_0, Z_1, \dots, Z_k$, where

$$Z_i := g_i^{-1}(\{0\}) \cap f_i^{-1}(0, +\infty) \quad \text{for all } i \in \{1, 2, \dots, k\},$$
$$Z_0 := g_0^{-1}(\{0\}) \cap f_1^{-1}(-\infty, 0).$$

By construction of the sets,

$$\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(Z_i) \subset \bigcup_{j \neq i} V_j \quad \text{for all } i \in \{0, 1, \dots, k\}.$$

A closer investigation, using that $\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{q\}) = \{v_0, v_1, \dots, v_k\}$ and that the closures of the V_i are pairwise disjoint, shows that

$$\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{q\}) \cap \overline{\exp_p^{-1}(Z_i)} = \{v_0, v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_k\}$$

for all $i \in \{0, 1, \dots, k\}$. This implies

$$\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(\{q\}) \cap \bigcap_{i=0}^{k} \overline{\exp_p^{-1}(Z_i)} = \varnothing.$$

Since k and q were chosen arbitrarily, this shows the inconsistency of $\mathcal{C}$. □

Combining the previous theorem with our lower bound from Theorem 4.8 yields:

Corollary 4.12 *Let (M, g) be a closed Riemannian manifold and assume that there exists $p \in M$ such that $\mathrm{Cut}_p(M)$ does not contain any conjugate points of p and such that all points in $\mathrm{Cut}_p(M)$ are nondegenerate. If $\mathrm{Cut}_p(M)$ contains a point of order $k+1$, where $k \in \mathbb{N}$, then*

$$\mathsf{GC}(M, g) \geq k + 1.$$

5 An upper bound for homogeneous Riemannian manifolds

From this section on, we will mostly consider homogeneous Riemannian manifolds and exploit their symmetry properties. Given a Riemannian manifold (M, g), we let $\mathrm{Isom}(M) := \mathrm{Isom}(M, g)$ denote its group of isometries and consider it as a subspace of $C^0(M, M)$ with the compact–open topology. We recall that (M, g) is called *homogeneous* if $\mathrm{Isom}(M)$ acts transitively on M. Note that every homogeneous Riemannian manifold is necessarily complete; see [25, Theorem IV.4.5].

Having derived lower bounds for geodesic complexity in the previous section, we next want to find upper bounds. After some preparatory lemmas, we will establish an upper bound on $\mathsf{GC}(M)$ for a homogeneous Riemannian manifold M in terms of the subspace complexity $\mathsf{GC}_M(\{p\} \times \mathrm{Cut}_p(M))$ and a categorical invariant determined by its isometry action. Intuitively, the transitivity of the isometry action implies that continuous geodesic motion planners on subsets of cut loci of single points can be continuously extended to larger subsets of the total cut locus. We will then go on to study further upper bounds on $\mathsf{GC}(M)$ in the case that $\mathrm{Cut}_p(M)$ admits a stratification. The following is a folklore result from Riemannian geometry:

Lemma 5.1 *Let (M, g) be a homogeneous Riemannian manifold and let $p \in M$. Then*

$$\mathrm{ev}_p \colon \mathrm{Isom}(M) \to M, \quad \mathrm{ev}_p(\phi) = \phi(p),$$

is a principal $\mathrm{Isom}_p(M)$–bundle, where $\mathrm{Isom}_p(M)$ denotes the isotropy group of the isometry action on M in p.

Proof By [26, Theorem 21.17], ev_p induces an $\mathrm{Isom}(M)$–equivariant diffeomorphism $f \colon \mathrm{Isom}(M)/\mathrm{Isom}_p(M) \to M$. Moreover, the projection $q \colon \mathrm{Isom}(M) \to \mathrm{Isom}(M)/\mathrm{Isom}_p(M)$ is a principal $\mathrm{Isom}_p(M)$–bundle by [25, Example I.5.1]. One easily shows that $\mathrm{ev}_p = f \circ q$, which implies that ev_p is a principal $\mathrm{Isom}_p(M)$–bundle as well. □

Example 5.2 Given a Lie group G with a left-invariant Riemannian metric, the left multiplication $l_g \colon G \to G$, $l_g(h) = gh$, is an isometry for each $g \in G$. With $e \in G$ denoting the unit, one further derives from $l_g(e) = g$ for each $g \in G$ that the map $s \colon G \to \mathrm{Isom}(G)$, $s(g) = l_g$, is a continuous section of the bundle $\mathrm{ev}_e \colon \mathrm{Isom}(G) \to G$.

Lemma 5.3 *Let $A, B \subset M$ and $p \in M$. Assume that there are a continuous geodesic motion planner $\sigma_B \colon \{p\} \times B \to GM$ and a continuous local section $s \colon A \to \mathrm{Isom}(M)$*

of ev_p. *Then there exists a continuous geodesic motion planner* $\sigma\colon F \to GM$, *where*

$$F := \{(x,y) \in M \times M \mid x \in A,\ y \in s(x)\cdot B\}.$$

Proof We define $\sigma\colon F \to GM$ by

$$\sigma(x,y) = s(x) \circ \sigma_B(p, s(x)^{-1}\cdot y) \quad \text{for all } (x,y) \in F.$$

By construction, $\sigma_B(p, s(x)^{-1}\cdot y)$ is a minimal geodesic from p to $s(x)^{-1}\cdot y$. Since $s(x)$ is an isometry for each x, $\sigma(x,y)$ is indeed a minimal geodesic from

$$s(x)\cdot p = \mathrm{ev}_p(s(x)) = x \quad \text{to} \quad s(x)\cdot(s(x)^{-1}\cdot y) = y.$$

So σ is a geodesic motion planner and it only remains to show its continuity.

Let $\rho\colon \mathrm{Isom}(M)\times M \to M$ denote the action of the isometry group by evaluation and again let $PM = C^0([0,1], M)$. By [4, Theorem VII.2.10], the composition map

$$\varphi\colon C^0(M,M)\times PM \to PM, \quad \varphi(f,\gamma) = f\circ\gamma,$$

is continuous with respect to the compact–open topologies. Thus, the restriction of φ to

$$\mathrm{Isom}(M)\times GM \subset C^0(M,M)\times PM$$

defines a continuous action

$$\tilde{\rho}\colon \mathrm{Isom}(M)\times GM \to GM, \quad \tilde{\rho} = \varphi|_{\mathrm{Isom}(M)\times GM}.$$

The inversion $i\colon \mathrm{Isom}(M)\to\mathrm{Isom}(M)$, $i(g) = g^{-1}$, is continuous since $\mathrm{Isom}(M)$ is a topological group. We can express σ as

$$\sigma(x,y) = \tilde{\rho}\big(s(x), \sigma_B\big(p, \rho\big(i(s(x)), y\big)\big)\big) \quad \text{for all } (x,y)\in F.$$

All maps on the right-hand side are continuous, so σ is continuous as well. □

The previous lemma allows us to make a useful estimate between the subspace geodesic complexity of the total cut locus and that of one single cut locus in the homogeneous case.

Theorem 5.4 *Let* (M,g) *be a homogeneous Riemannian manifold and let* $p \in M$. *Then*

$$\mathsf{GC}(M) \le \mathsf{secat}(\mathrm{ev}_p\colon \mathrm{Isom}(M)\to M)\cdot \mathsf{GC}_p(\mathrm{Cut}_p(M)) + 1.$$

Proof As seen in Remark 2.7(2), it holds that $\mathsf{GC}(M) \le \mathsf{GC}_M(\mathrm{Cut}(M)) + 1$, so it suffices to show that

$$\mathsf{GC}_M(\mathrm{Cut}(M)) \le \mathsf{secat}(\mathrm{ev}_p)\cdot\mathsf{GC}_p(\mathrm{Cut}_p(M)).$$

Let $k := \mathsf{secat}(\mathrm{ev}_p)$ and $r := \mathsf{GC}_p(\mathrm{Cut}_p(M))$. By Lemma 4.1, there are pairwise disjoint locally compact $A_1, \dots, A_k \subset M$ with $M = \bigcup_{i=1}^k A_i$ for which there is a continuous local section $s_i : A_i \to \mathrm{Isom}(M)$ of ev_p for each $i \in \{1, 2, \dots, k\}$. Let $B_1, \dots, B_r \subset M$ be pairwise disjoint and locally compact with $\mathrm{Cut}_p(M) \subset \bigcup_{j=1}^r B_j$ such that, for each j, there exists a continuous geodesic motion planner $\sigma_j : \{p\} \times B_j \to GM$. Put

$$F_{i,j} := \{(x, y) \in M \times M \mid x \in A_i,\ y \in s_i(x) \cdot B_j\}$$

for all $i \in \{1, 2, \dots, k\}$ and $j \in \{1, 2, \dots, r\}$. By construction, the elements of

$$\bigl\{F_{i,j} \mid i \in \{1, 2, \dots, k\},\ j \in \{1, 2, \dots, r\}\bigr\}$$

are pairwise disjoint. Furthermore, for all $i \in \{1, 2, \dots, k\}$ and $j \in \{1, 2, \dots, r\}$,

$$\psi_{i,j} : A_i \times B_j \to F_{i,j}, \quad \psi_{i,j}(x, y) = (x, s_i(x) \cdot y),$$

is a homeomorphism. Consequently, the $F_{i,j}$ are locally compact. If $(x, y) \in \mathrm{Cut}(M)$, then $x \in A_i$ for some $i \in \{1, 2, \dots, k\}$. Since $s_i(x)^{-1}$ is an isometry, it holds that $s_i(x)^{-1} \cdot y \in \mathrm{Cut}_p(M)$. Hence, there is a $j \in \{1, 2, \dots, r\}$ with $s_i(x)^{-1} \cdot y \in B_j$ and therefore $(x, y) \in F_{i,j}$ by definition. This shows that

$$\mathrm{Cut}(M) \subset \bigcup_{i=1}^{k} \bigcup_{j=1}^{r} F_{i,j}.$$

Moreover, by Lemma 5.3 we can find a continuous geodesic motion planner $F_{i,j} \to GM$ of p for all i and j. Thus, $\mathsf{GC}_M(\mathrm{Cut}(M)) \le kr$, which shows the claim. $\square$

The previous upper bound has a particularly strong consequence for connected Lie groups.

Corollary 5.5 *Let G be a connected Lie group equipped with a left-invariant Riemannian metric and let $e \in G$ denote the unit element. Then*

$$\mathsf{GC}(G) \le \mathsf{GC}_e(\mathrm{Cut}_e(G)) + 1.$$

Proof This is an immediate consequence of Theorem 5.4. Since $\mathrm{ev}_e : \mathrm{Isom}(G) \to G$ admits a continuous section — see Example 5.2 — it follows that $\mathsf{secat}(\mathrm{ev}_e) = 1$. $\square$

Sectional categories of fibrations are in general hard to compute. A common way of estimating their values from above is by the Lusternik–Schnirelmann categories of their base spaces using [39, Theorem 18]. In our situation, this leads to the following estimate:

Corollary 5.6 *Let (M,g) be a homogeneous Riemannian manifold and let $p \in M$. Then*

$$\mathsf{GC}(M) \le \mathsf{cat}(M) \cdot \mathsf{GC}_p(\mathrm{Cut}_p(M)) + 1.$$

Proof This is an immediate consequence of Theorem 5.4 and the fact that every fibration $p\colon E \to B$ satisfies $\mathsf{secat}(p) \le \mathsf{cat}(B)$ by [39, Theorem 18]. □

We want to further estimate geodesic complexity from above by finding upper bounds for subspace geodesic complexities of cut loci. When $\mathrm{Cut}_p(M)$ admits a stratification, we can compare $\mathsf{GC}_p(\mathrm{Cut}_p(M))$ to the subspace geodesic complexities of its strata.

Proposition 5.7 *Let (M,g) be a complete Riemannian manifold, let $p \in M$ and assume that $\mathrm{Cut}_p(M)$ has a stratification $(S_1,\dots,S_k)$ of depth k. Then*

$$\mathsf{GC}_p(\mathrm{Cut}_p(M)) \le \sum_{i=1}^{k} \max_{Z_i \in \pi_0(S_i)} \mathsf{GC}_p(Z_i),$$

where $\pi_0(X)$ denotes the set of connected components of a space X.

Proof Since $\mathrm{Cut}_p(M) = S_1 \cup \dots \cup S_k$, it follows from Remark 2.2(3) that

$$\mathsf{GC}_p(\mathrm{Cut}_p(M)) \le \sum_{i=1}^{k} \mathsf{GC}_p(S_i).$$

Now fix $i \in \{1,2,\dots,k\}$ and let $Z_1,\dots,Z_r$ be the connected components of S_i. Put

$$s_i := \max_{j \in \{1,2,\dots,r\}} \mathsf{GC}_p(Z_j).$$

For each $j \in \{1,2,\dots,r\}$, let $A_1^j,\dots,A_{s_i}^j \subset Z_j$ be pairwise disjoint and locally compact, such that, for each $j \in \{1,2,\dots,r\}$ and $l \in \{1,2,\dots,s_i\}$, either $A_l^j = \varnothing$ or there exists a continuous geodesic motion planner $\sigma_{j,l}\colon \{p\} \times A_l^j \to GM$. Put $A_l := \bigcup_{j=1}^r A_l^j$ for each $l \in \{1,2,\dots,s_i\}$. Then the A_l are pairwise disjoint and locally compact with $S_i = \bigcup_{l=1}^{s_i} A_l$. Moreover, since, by definition of a stratification, $Z_i \cap \overline{Z}_j = \varnothing$ for all $i \neq j$, the maps

$$\sigma_l\colon \{p\} \times A_l \to GM, \qquad \sigma_l(p,x) = \sigma_{j,l}(p,x) \quad \text{for all } x \in A_l^j,\ j \in \{1,2,\dots,r\},$$

are well-defined continuous geodesic motion planners. This shows $\mathsf{GC}_p(S_i) \le s_i$ for each $i \in \{1,2,\dots,k\}$, which implies the claim. □

Corollary 5.8 *Let (M,g) be a homogeneous Riemannian manifold, let $p \in M$ and assume that $\mathrm{Cut}_p(M)$ has a stratification $(S_1,\dots,S_k)$ of depth k. Then*

$$\mathsf{GC}(M) \le \mathsf{secat}(\mathrm{ev}_p\colon \mathrm{Isom}(M) \to M) \cdot \sum_{i=1}^{k} \max_{Z_i \in \pi_0(S_i)} \mathsf{GC}_p(Z_i) + 1.$$

Proof This follows from Theorem 5.4 and Proposition 5.7. □

6 Trivially covered stratifications

Recio-Mitter [34] considered total cut loci with stratifications whose strata are finitely covered by the path fibration. As a part of [34, Corollary 3.14], he showed that, if such a stratification is inconsistent and trivially covered, this knowledge about the total cut locus suffices to compute the geodesic complexity of the space.

In this section, we will revisit the notion of trivially covered stratifications in the setting of Riemannian manifolds, but, in contrast to [34], we will put a covering condition on the cut locus of a single point instead of the total cut locus. We will then derive an upper bound for the numbers $\mathsf{GC}_p(M)$ that we have studied in the previous section. From this estimate we will derive an upper bound for the geodesic complexity of homogeneous Riemannian manifolds for which the cut locus of a point admits a trivially covered stratification.

Definition 6.1 Let M be a complete Riemannian manifold, let $p \in M$ and let $\mathcal{S} = (S_1,\dots,S_N)$ be a stratification of $\mathrm{Cut}_p(M)$. We call $\mathcal{S}$ *trivially covered* if, for all $k \in \{1,2,\dots,N\}$ and for all connected components Z of S_k, the restriction

$$\exp_p|_{\widetilde{\mathrm{Cut}}_p(M)\cap \exp_p^{-1}(Z)}\colon \widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(Z) \to Z$$

is a trivial covering. Here, a trivial covering is understood to be a covering $q\colon X \to Y$ for which there is a discrete set D and a homeomorphism $f\colon X \to Y \times D$ such that $q = \mathrm{pr} \circ f$, where $\mathrm{pr}\colon Y \times D \to Y$ is the projection onto the first factor.

Theorem 6.2 *Let M be a complete Riemannian manifold, let $p \in M$ and assume that $\mathrm{Cut}_p(M)$ admits a trivially covered stratification of depth $N \in \mathbb{N}$. Then*

$$\mathsf{GC}_p(\mathrm{Cut}_p(M)) \le N.$$

Proof Let $\mathcal{S} = (S_1,\dots,S_N)$ be a trivially covered stratification of $\mathrm{Cut}_p(M)$. We want to show that $\{p\} \times S_k$ admits a continuous geodesic motion planner for each $k \in \{1,2,\dots,N\}$. For a fixed $k \in \{1,2,\dots,N\}$, let $Z_1,\dots,Z_r$ be the connected components of S_k for suitable $r \in \mathbb{N}$. For $i \in \{1, 2, \ldots, r\}$, let B_i be an arbitrary sheet of the trivial covering

$$\exp_p|_{\widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(Z_i)} : \widetilde{\mathrm{Cut}}_p(M) \cap \exp_p^{-1}(Z_i) \to Z_i.$$

Then $\exp_p|_{B_i} : B_i \to Z_i$ is a homeomorphism. With $\varphi_i := (\exp_p|_{B_i})^{-1} : Z_i \to B_i$, one checks without difficulties that

$$s_i : \{p\} \times Z_i \to GM, \quad (s_i(p,q))(t) = \exp_p(t\varphi_i^{-1}(q)),$$

is a continuous geodesic motion planner and thus $\mathsf{GC}_p(Z_i) = 1$. Since $k \in \{1, 2, \ldots, N\}$ was chosen arbitrarily, the claim follows from Proposition 5.7. □

With the additional hypotheses that M is compact and that the stratification in Theorem 6.2 is inconsistent, one can derive an equality from Theorem 6.2. The following result is analogous to the corresponding part of [34, Corollary 3.14]:

Corollary 6.3 *Let M be a closed Riemannian manifold, let $p \in M$ and assume that $\mathrm{Cut}_p(M)$ admits a trivially covered inconsistent stratification of depth $N \in \mathbb{N}$. Then $\mathsf{GC}_p(M) = N + 1$.*

Proof By restricting the motion planner from Remark 2.7(2), one obtains a continuous geodesic motion planner on $\{p\} \times (M \smallsetminus \mathrm{Cut}_p(M))$. It follows from Theorem 6.2 that

$$\mathsf{GC}_p(M) \le \mathsf{GC}_p(\mathrm{Cut}_p(M)) + 1 \le N + 1.$$

But, by Theorem 4.8, it also holds that $\mathsf{GC}_p(M) \ge N + 1$, which proves the equality. □

Corollary 6.4 *Let G be a compact connected Lie group equipped with a left-invariant Riemannian metric and let $e \in G$ denote the unit element. If $\mathrm{Cut}_e(G)$ admits a trivially covered inconsistent stratification of depth N, then*

$$\mathsf{GC}(G) = N + 1.$$

Proof Combining Theorem 6.2 with Corollary 5.5 yields

$$\mathsf{GC}(G) \le \mathsf{GC}_e(\mathrm{Cut}_e(G)) + 1 \le N + 1.$$

But, by Theorem 4.8, $\mathsf{GC}(G) \ge N + 1$ as well, so the claim follows. □

7 Examples: flat tori and Berger spheres

We want to use the results of Sections 5 and 6 to compute the geodesic complexities of two classes of examples: two-dimensional flat tori and three-dimensional Berger spheres.

The cut loci of points in such spaces are well understood and admit stratifications of a well-behaved kind.

7.1 Geodesic complexity of flat tori

Recio-Mitter has computed the geodesic complexity of a standard flat n–dimensional torus in [34, Theorem 4.4]. More precisely, he has shown that the standard flat metric g_f on the n–torus T^n satisfies $\mathsf{GC}(T^n, g_f) = n+1$ for each $n \in \mathbb{N}$.

In the course of this subsection, we will extend the two-dimensional case of Recio-Mitter's result to *arbitrary* flat metrics on two-dimensional tori. The cut loci of such metrics are well understood.

Before we do so, we will reobtain Recio-Mitter's computation for standard flat tori using the methods of this article. This example is particularly instructive and illustrates the use of inconsistent stratifications. Moreover, in contrast to [34, Theorem 4.4], we only need to consider the cut locus of a single point, while in the proof of [34, Theorem 4.4] a stratification of $T^n \times T^n$ is required and the structure of the space of geodesic paths in T^n needs to be examined.

Example 7.1 Let $n \in \mathbb{N}$ and consider the n–torus T^n with the standard flat metric g_f, ie the quotient metric induced by the standard metric on $\mathbb{R}^n$ and by identifying $T^n = \mathbb{R}^n/(2\mathbb{Z})^n$. Equivalently, T^n is obtained from $\mathbb{R}^n$ by collapsing the lattice defined by an arbitrary family of n pairwise orthogonal vectors of length two. Let $\pi\colon \mathbb{R}^n \to T^n$ be the projection and put $o := \pi(0)$ and $M := (T^n, g_f)$. We identify $\mathbb{R}^n$ with T_oM in the obvious way.

Note that T^n is isometric to the Riemannian product $(\mathbb{R}/(2\mathbb{Z}))^n$. For $N := \mathbb{R}/(2\mathbb{Z})$ let $\mathrm{pr}\colon \mathbb{R} \to N$ be the obvious Riemannian covering and put $p_0 := \mathrm{pr}(0) \in N$. Then $\mathrm{Cut}_{p_0}(N) = \{\mathrm{pr}(1)\}$ and the tangent cut locus is given by

$$\widetilde{\mathrm{Cut}}_{p_0}(N) = \{-1, 1\}$$

under the identification $T_{p_0}N \cong \mathbb{R}$.

Given the Riemannian product of two Riemannian manifolds (M_1, g_1) and (M_2, g_2), the cut locus of a point $(p_1, p_2) \in M_1 \times M_2$ is easily seen to be

$$\mathrm{Cut}_{(p_1,p_2)}(M_1 \times M_2) = (\mathrm{Cut}_{p_1}(M_1) \times M_2) \cup (M_1 \times \mathrm{Cut}_{p_2}(M_2));$$

see [7, page 328]. For $i \in \{1, 2\}$, let K_i be the union of the injectivity domain in $T_{p_i}M_i$ with the tangent cut locus $\widetilde{\mathrm{Cut}}_{p_i}(M_i)$. Similar to the cut locus, the tangent cut locus of

(p_1, p_2) is given by

$$\widetilde{\mathrm{Cut}}_{(p_1,p_2)}(M_1 \times M_2) = (\widetilde{\mathrm{Cut}}_{p_1}(M_1) \times K_2) \cup (K_1 \times \widetilde{\mathrm{Cut}}_{p_2}(M_2))$$

under the identification $T_{(p_1,p_2)}(M_1 \times M_2) \cong T_{p_1}M_1 \times T_{p_2}M_2$. For products of finitely many manifolds, one iteratively derives analogous results for cut loci and tangent cut loci.

We conclude that, if $I^n := [-1,1]^n$, then the tangent cut locus of o in $M = (T^n, g_f)$ is

$$\widetilde{\mathrm{Cut}}_o(M) = \partial I^n.$$

See also [19, page 107] for the case $n = 2$. The boundary ∂I^n admits a stratification $\partial I^n = \bigcup_{k=1}^n \mathcal{A}_k$ of depth n, given as follows: For each $k \in \{1, 2, \dots, n\}$, we put

$$J_k := \{(i_1, \dots, i_k) \in \mathbb{N}^k \mid 1 \le i_1 < \dots < i_k \le n\}.$$

Then each $\mathcal{A}_k$ is given as the disjoint union $\mathcal{A}_k = \bigcup_{(i_1,\dots,i_k)\in J_k} A_{i_1,\dots,i_k}$, where

$$A_{i_1,\dots,i_k} := \bigl\{(x_1,\dots,x_n) \in I^n : |x_l| = 1 \text{ if } l \in \{i_1, i_2, \dots, i_k\}, |x_l| < 1 \text{ if } l \notin \{i_1,\dots,i_k\}\bigr\}.$$

For $(i_1, \dots, i_k) \in J_k$ and $j_1, \dots, j_k \in \{-1, 1\}$, we put

$$\begin{aligned} &A_{i_1,\dots,i_k,j_1,\dots,j_k} \\ &\qquad = \bigl\{(x_1,\dots,x_n) \in \mathbb{R}^n : x_{i_1} = j_1, \dots, x_{i_k} = j_k,\ |x_l| < 1 \text{ if } l \notin \{i_1,\dots,i_k\}\bigr\}. \end{aligned}$$

Then the sets $A_{i_1,\dots,i_k,j_1,\dots,j_k}$, where $j_1, \dots, j_k \in \{-1, 1\}$, are precisely the connected components of $A_{i_1,\dots,i_k}$.

Put $\mathcal{B}_k := \exp_o(\mathcal{A}_k)$. We claim that $(\mathcal{B}_1, \dots, \mathcal{B}_n)$ is a trivially covered stratification of $\mathrm{Cut}_o(M)$. One checks that the connected components of each of the $\mathcal{B}_k$ are precisely the sets

$$B_{i_1,\dots,i_k} := \exp_o(A_{i_1,\dots,i_k}) \quad \text{for } (i_1,\dots,i_k) \in J_k.$$

Moreover, for all $(i_1, \dots, i_k) \in J_k$ and all $j_1, \dots, j_k \in \{-1, 1\}$, the restriction

$$\exp_o|_{A_{i_1,\dots,i_k,j_1,\dots,j_k}} : A_{i_1,\dots,i_k,j_1,\dots,j_k} \to B_{i_1,\dots,i_k}$$

is a homeomorphism. From the explicit description of the $A_{i_1,\dots,i_k}$, one derives that $(\mathcal{B}_1, \dots, \mathcal{B}_n)$ is a stratification. It further follows from the above observations that

$$\exp_o|_{A_{i_1,\dots,i_k}} : A_{i_1,\dots,i_k} \to B_{i_1,\dots,i_k}$$

is a trivial covering map for all $(i_1, \dots, i_k) \in J_k$. Since $k \in \{1, 2, \dots, n\}$ was chosen arbitrarily, this shows that $(\mathcal{B}_1, \dots, \mathcal{B}_n)$ is trivially covered.

We now want to prove that $(\mathcal{B}_1,\dots,\mathcal{B}_n)$ is indeed an inconsistent stratification of $\mathrm{Cut}_o(M)$. For this purpose, let $k\in\{2,3,\dots,n\}$, $(i_1,\dots,i_k)\in J_k$ and $x\in B_{i_1,\dots,i_k}$. We assume without loss of generality that $(i_1,i_2,\dots,i_k)=(1,2,\dots,k)$. Then there are $y_1,\dots,y_{n-k}\in(-1,1)$ such that

$$x=\exp_o(1,1,\dots,1,y_1,\dots,y_{n-k}).$$

It further holds that

$$\exp_K^{-1}(\{x\})=\bigl\{(j_1,\dots,j_k,y_1,\dots,y_{n-k})\in T_oM\mid j_1,\dots,j_k\in\{-1,1\}\bigr\},\tag{7-1}$$

where $K:=I^n$ and $\exp_K:=\exp_o|_K\colon K\to M$, which is a special case of the map defined in (4-1). Let $i\in\{1,2,\dots,k\}$ and let $\widehat{B}_i:=B_{1,\dots,i-1,i+1,\dots,k}\subset\mathcal{B}_{k-1}$. Given $\varepsilon>0$, put

$$U_\varepsilon:=\exp_o\biggl((1-\varepsilon,1+\varepsilon)^k\times\prod_{j=1}^{n-k}(y_j-\varepsilon,y_j+\varepsilon)\biggr)\subset M.$$

Then U_ε is an open neighborhood of x and, for sufficiently small $\varepsilon>0$, it holds that $\widehat{B}_i\cap U_\varepsilon$ has two components C_i^+ and C_i^-. With $I_+:=(1-\varepsilon,1)$ and $I_-:=(-1,-1+\varepsilon)$, we put, for all $j_1,\dots,j_{i-1},j_{i+1},\dots,j_k\in\{-1,1\}$,

$$A^{\pm}_{j_1,\dots,j_{i-1},j_{i+1},\dots,j_k}:=\biggl\{(j_1,\dots,j_{i-1},t,j_{i+1},\dots,j_k,q)\biggm| t\in I_\pm,\ q\in\prod_{l=1}^{n-k}(y_l-\varepsilon,y_l+\varepsilon)\biggr\}.$$

The two components C_i^+ and C_i^- then satisfy

$$\exp_K^{-1}(C_i^\pm)=\bigcup_{j_1,\dots,j_{i-1},j_{i+1},\dots,j_k\in\{-1,1\}}A^{\pm}_{j_1,\dots,j_{i-1},j_{i+1},\dots,j_k}.$$

Combining this observation with (7-1) yields

$$\widetilde{\mathrm{Cut}}_o(M)\cap\exp_o^{-1}(\{x\})\cap\overline{\exp_K^{-1}(C_i^\pm)}$$
$$=\bigl\{(j_1,\dots,j_{i-1},\pm1,j_{i+1},\dots,j_k,y_1,\dots,y_{n-k})\mid j_1,\dots,j_{i-1},j_{i+1},\dots,j_k\in\{-1,1\}\bigr\},$$

In particular, $\exp_o^{-1}(\{x\})\cap\overline{\exp_K^{-1}(C_i^+)}\cap\overline{\exp_K^{-1}(C_i^-)}=\varnothing$, implying that $(\mathcal{B}_1,\dots,\mathcal{B}_n)$ satisfies the inconsistency condition at x. Since $x\in\mathrm{Cut}_o(M)$ was chosen arbitrarily, this shows that $(\mathcal{B}_1,\dots,\mathcal{B}_n)$ is inconsistent. Note that in general $\mathcal{B}_{k-1}\cap U_\varepsilon$ has more connected components than C_i^+ and C_i^-, but considering these two components is sufficient for proving the inconsistency condition.

Since T^n is a Lie group and g_f is left-invariant, it follows from Corollary 6.4 that

$$\mathsf{GC}(T^n,g_f)=n+1.$$

Next we will compute the geodesic complexity of arbitrary two-dimensional flat tori. The reader should note that, in general, the geodesic complexity of (T^2, g) will vary with the metric g; see Example 2.3(2). For arbitrary flat tori of higher dimensions, the cut loci of points are not as well understood as in the two-dimensional case. While it might be possible to extend our result to flat tori of higher dimensions, we are not aware of any systematic study of cut loci of flat higher-dimensional tori in the literature.

For $p, q \in \mathbb{R}^2$, we let $[p, q] = \{(1-t)p + tq \in \mathbb{R}^2 \mid t \in [0, 1]\}$ denote the line segment from p to q.

Theorem 7.2 *Let g be an arbitrary flat metric on T^2. Then $\mathsf{GC}(T^2, g) = 3$.*

Proof By elementary Riemannian geometry, (T^2, g) is isometric to T^2 with a quotient metric induced by the standard metric on $\mathbb{R}^2$ and a projection $\pi : \mathbb{R}^2 \to \mathbb{R}^2/\Gamma = T^2$, where $\Gamma \subset \mathbb{R}^2$ is a lattice. We thus assume that g itself is such a quotient metric. Put $o := \pi(0, 0)$. We are going to describe $\widetilde{\mathrm{Cut}}_o(T^2)$, following [19, page 108]. The case that Γ is generated by two orthogonal vectors is covered in Example 7.1, so we assume in the following that Γ is generated by two vectors $a_1, a_2 \in \mathbb{R}^2$ such that the angle between a_1 and a_2 is acute.

If we identify T_oM with $\mathbb{R}^2$, then $\widetilde{\mathrm{Cut}}_o(M)$ is given by a hexagon whose construction we will describe next. Consider the perpendicular bisectors of the line segments

$$[0, a_1], \quad [0, a_2], \quad [0, -a_1], \quad [0, -a_2], \quad [0, a_1 - a_2], \quad [0, a_2 - a_1].$$

These perpendicular bisectors enclose a hexagon in $\mathbb{R}^2$; see Figure 1. The tangent cut locus $\widetilde{\mathrm{Cut}}_o(M)$ consists of the boundary curve of the hexagon, while the domain of injectivity of $\exp_o$ is given by the interior of the hexagon. Let the segments and the corner points of the hexagon be labeled as in Figure 1. Then there are $p, q \in M$ with $p \neq q$ such that $p = \exp_o(p_1) = \exp_o(p_2) = \exp_o(p_3)$ and $q = \exp_o(q_1) = \exp_o(q_2) = \exp_o(q_3)$.

For $x, y \in \mathbb{R}^2$ we put $[\![x, y]\!] := [x, y] \setminus \{x, y\}$. With p and q as above, the set $\mathrm{Cut}_o(M) \setminus \{p, q\}$ has three connected components

$$\begin{aligned} A_1 &:= \exp_o([\![p_1, q_1]\!]) = \exp_o([\![q_2, p_3]\!]), \\ A_2 &:= \exp_o([\![q_1, p_2]\!]) = \exp_o([\![p_3, q_3]\!]), \\ A_3 &:= \exp_o([\![p_2, q_2]\!]) = \exp_o([\![q_3, p_1]\!]). \end{aligned}$$

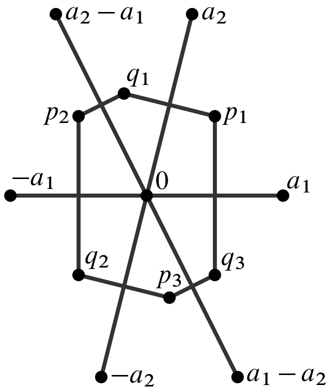


Figure 1: Tangent cut loci of flat two-dimensional tori.

More precisely, $\exp_o$ maps both $[\![p_1, q_1]\!]$ and $[\![q_2, p_3]\!]$ homeomorphically onto A_1, both $[\![q_1, p_2]\!]$ and $[\![p_3, q_3]\!]$ homeomorphically onto A_2, and both $[\![p_2, q_2]\!]$ and $[\![q_3, p_1]\!]$ homeomorphically onto A_3.

Let $S_2 := \{p, q\}$ and $S_1 := A_1 \cup A_2 \cup A_3$. By construction, (S_1, S_2) is a trivially covered stratification of $\mathrm{Cut}_o(T_2, g)$. We want to show that (S_1, S_2) is inconsistent as well. Let $K \subset T_oT^2$ denote the union of $\widetilde{\mathrm{Cut}}_o(T^2, g)$ with the domain of injectivity of $\exp_o$ and let $\exp_K \colon K \to T^2$ be the restriction of $\exp_o$ to K. This is again a special case of the map defined in (4-1). Let $U \subset T^2$ be an open neighborhood of p and put $Z_i := A_i \cap U$ for all $i \in \{1, 2, 3\}$. If U is chosen sufficiently small, then, by the above description of $\widetilde{\mathrm{Cut}}_o(T^2, g)$, there are $x_1 \in [\![p_1, q_1]\!]$ and $x_1' \in [\![q_2, p_3]\!]$ such that

$$\exp_K^{-1}(Z_1) = [\![x_1', p_3]\!] \cup [\![p_1, x_1]\!].$$

Analogously, one shows that there are

$$x_2 \in [\![q_1, p_2]\!], \quad x_2' \in [\![p_3, q_3]\!], \quad x_3 \in [\![p_2, q_2]\!] \quad \text{and} \quad x_3' \in [\![q_3, p_1]\!]$$

such that

$$\exp_K^{-1}(Z_2) = [\![x_2, p_2]\!] \cup [\![p_3, x_2']\!], \quad \exp_K^{-1}(Z_3) = [\![x_3', p_1]\!] \cup [\![p_2, x_3]\!].$$

Since $\exp_K^{-1}(\{p\}) = \{p_1, p_2, p_3\}$, this shows that

$$\begin{aligned}
\exp_K^{-1}(\{p\}) \cap \overline{\exp_K^{-1}(Z_1)} &= \{p_2, p_3\},\\
\exp_K^{-1}(\{p\}) \cap \overline{\exp_K^{-1}(Z_2)} &= \{p_1, p_3\},\\
\exp_K^{-1}(\{p\}) \cap \overline{\exp_K^{-1}(Z_3)} &= \{p_1, p_2\}.
\end{aligned}$$

Consequently,

$$\widetilde{\mathrm{Cut}}_o(T^2) \cap \exp_o^{-1}(\{p\}) \cap \bigcap_{i=1}^{3} \overline{\exp_K^{-1}(Z_i)} = \varnothing,$$

which shows that (S_1, S_2) satisfies the inconsistency condition at p. In complete analogy, one shows that the condition is satisfied at q as well, implying that (S_1, S_2) is inconsistent. Since g is, by construction, left-invariant, it follows from Corollary 6.4 that $\mathsf{GC}(T^2, g) = 3$. □

7.2 Geodesic complexity of Berger spheres

In this subsection we consider a class of homogeneous Riemannian manifolds whose geodesic complexity can be computed explicitly without making use of the upper and lower bounds we previously studied. In [1], M Berger has constructed a one-parameter family of homogeneous metrics g_α for $0 < \alpha \le \frac{\pi}{2}$ on the three-dimensional sphere S^3, whose cut loci have been described by Sakai [38].

In the following, we will first recall a particularly interesting class of homogeneous Riemannian manifolds, namely naturally reductive spaces. Berger spheres are special cases of them and we will outline the construction of Berger's metrics following [38].

Given a Lie group G, we always let $e \in G$ denote its unit element. Let $\mathfrak{g}$ denote the Lie algebra of G and assume that H is a closed subgroup of G. Then the Lie algebra $\mathfrak{h}$ of the Lie group H is a Lie subalgebra of $\mathfrak{g}$. If there is an Ad_H–invariant subspace $\mathfrak{m}$ of the Lie algebra $\mathfrak{g}$ which is complementary to $\mathfrak{h}$ then there is a bijective correspondence between Ad_H–invariant inner products on $\mathfrak{m}$ and G–invariant metrics on the homogeneous space G/H. See [31, Proposition 11.22(2)] for details.

Definition 7.3 [31, page 317] Let G be a Lie group with a closed subgroup H. Let $\mathfrak{g}$ be the Lie algebra of G and $\mathfrak{h}$ be the Lie algebra of H. Assume that there is a subspace $\mathfrak{m} \subset \mathfrak{g}$ which is complementary to $\mathfrak{h}$ and such that $\mathrm{Ad}_H(\mathfrak{m}) \subset \mathfrak{m}$, where Ad_H denotes the adjoint representation of H. Suppose there is an Ad_H–invariant inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{m}$ such that

$$\langle [X, Y]_{\mathfrak{m}}, Z \rangle = \langle X, [Y, Z]_{\mathfrak{m}} \rangle$$

for all $X, Y, Z \in \mathfrak{m}$, where the subscript $\mathfrak{m}$ of an element of $\mathfrak{g}$ denotes its projection onto $\mathfrak{m}$. Then G/H together with the G–invariant Riemannian metric corresponding to this inner product is called a *naturally reductive space*.

Example 7.4 All symmetric spaces are examples of naturally reductive spaces as discussed in [31, page 317]. The real Stiefel manifolds $V_k(\mathbb{R}^n)$ for $n \geq 4$ and $2 \leq k \leq n-2$ are examples of naturally reductive spaces which are not symmetric spaces; see [18, page 748].

For our purposes, the crucial property of naturally reductive spaces is the observation made in the following proposition. We refer to [31, Proposition 11.25] for its proof.

Proposition 7.5 *Let G be a Lie group and $H \subset G$ be a closed subgroup. If $M = G/H$ is a naturally reductive space and $\pi\colon G \to M$ is the projection, then the geodesics starting at $o = \pi(e)$ are precisely the curves of the form $\gamma(t) = \pi(\exp(t\xi))$ for $\xi \in \mathfrak{m}$, where $\exp\colon \mathfrak{g} \to G$ is the Lie group exponential of G.*

We proceed by constructing Berger spheres as naturally reductive spaces following the exposition of [38]. Let $G = \mathrm{SU}(2) \times \mathbb{R}$ and let $\mathfrak{g} = \mathfrak{su}(2) \oplus \mathbb{R}$ be its Lie algebra. We consider the Ad_G–invariant inner product on $\mathfrak{g}$ given by

$$\langle (A,x),(B,y)\rangle = -\tfrac{1}{2}\operatorname{Tr}(AB) + xy \quad \text{for all } (A,x),(B,y) \in \mathfrak{g}.$$

For $\alpha \in \left(0, \frac{\pi}{2}\right]$, we define a linear subspace of $\mathfrak{g}$ as

$$\mathfrak{h}_\alpha = \left\{ \left(\begin{pmatrix} il\cos\alpha & 0 \\ 0 & -il\cos\alpha \end{pmatrix}, l\sin\alpha \right) \in \mathfrak{g} \,\middle|\, l \in \mathbb{R} \right\}.$$

Consider the closed subgroup $H_\alpha \subset G$, $H_\alpha = \exp(\mathfrak{h}_\alpha)$, where exp again denotes the Lie group exponential of G. Explicitly, H_α is given as

$$H_\alpha = \left\{ \left(\begin{pmatrix} e^{il\cos\alpha} & 0 \\ 0 & e^{-il\cos\alpha} \end{pmatrix}, l\sin\alpha \right) \,\middle|\, l \in \mathbb{R} \right\}.$$

One checks that G/H_α is diffeomorphic to S^3. The orthogonal complement to $\mathfrak{h}_\alpha$ in $\mathfrak{g}$ with respect to $\langle\cdot,\cdot\rangle$ is the space

$$\mathfrak{m}_\alpha = \left\{ \left(\begin{pmatrix} ir\sin\alpha & a+ib \\ -a+ib & -ir\sin\alpha \end{pmatrix}, -r\cos\alpha \right) \in \mathfrak{g} \,\middle|\, a,b,r \in \mathbb{R} \right\}.$$

A direct computation shows that $\mathfrak{m}_\alpha$ is Ad_{H_α}–invariant. The restriction of the inner product $\langle\cdot,\cdot\rangle$ to $\mathfrak{m}_\alpha \times \mathfrak{m}_\alpha$ defines an Ad_{H_α}–invariant inner product on $\mathfrak{m}_\alpha$. We equip the homogeneous space G/H_α with the G–invariant metric that is defined by this inner product and the abovementioned correspondence between G–invariant Riemannian metrics on G/H_α and Ad_{H_α}–invariant inner products on $\mathfrak{m}_\alpha$.

Since $\mathfrak{m}_\alpha \perp \mathfrak{h}_\alpha$ by construction, the space $M_\alpha = G/H_\alpha$ equipped with the described homogeneous metric is a naturally reductive space; see [18, Proposition 23.29]. Thus, by Proposition 7.5, the geodesics in M_α emanating from o are precisely the images of the one-parameter groups in G under π of elements of $\mathfrak{m}_\alpha$. For $\alpha = \frac{\pi}{2}$, one further observes that G/H_α is isometric to the round sphere S^3 of sectional curvature one; see [28, page 77].

The following observation gives us a strong upper bound on $\mathsf{GC}(M_\alpha)$. We refer to [28, Section 3] for its proof.

Proposition 7.6 *For each $\alpha \in \left(0, \frac{\pi}{2}\right)$ the Berger sphere M_α is isometric to* SU(2) *equipped with a left-invariant metric.*

Combining Proposition 7.6 with Corollary 5.5 yields

$$\mathsf{GC}(M_\alpha) \leq \mathsf{GC}_o(\mathrm{Cut}_o(M_\alpha)) + 1, \tag{7-2}$$

where $o = \pi(1)$. To compute $\mathsf{GC}_o(\mathrm{Cut}_o(M_\alpha))$, we will outline the results from [38] about the cut loci of M_α. For $\alpha = \frac{\pi}{2}$, we already know that $\mathsf{GC}(M_{\pi/2}) = 2$; see Examples 2.3(1). Thus, in the following we fix $\alpha \in \left(0, \frac{\pi}{2}\right)$.

Let $S \subset \mathfrak{m}_\alpha$ denote the unit sphere in $\mathfrak{m}_\alpha$ with respect to the norm induced by $\langle \cdot , \cdot \rangle$ and let $D\pi_e \colon \mathfrak{g} \to T_o M_\alpha$ denote the differential of π in the unit $e \in G$. Consider the isometric isomorphism of vector spaces

$$\varphi := D\pi_e|_{\mathfrak{m}_\alpha} \colon \mathfrak{m}_\alpha \xrightarrow{\cong} T_o M_\alpha .$$

Then φ maps S to the unit sphere in $T_o M_\alpha$. Let $\psi \colon T_o M_\alpha \to T_o M_\alpha$ be the radial homeomorphism which maps the unit sphere homeomorphically onto the tangent cut locus $\widetilde{\mathrm{Cut}}_o(M_\alpha)$ of o in M_α. Then the map $F \colon S \to \widetilde{\mathrm{Cut}}_o(M_\alpha)$, $F := \psi \circ \varphi|_S$ is a homeomorphism. We consider $e_1, e_2 \in \mathfrak{m}_\alpha$ given by

$$e_1 = \left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, 0 \right), \quad e_2 = \left(\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, 0 \right).$$

Let $U := \mathrm{span}_{\mathbb{R}}(\{e_1, e_2\})$ and let $r_U \colon \mathfrak{m}_\alpha \to \mathfrak{m}_\alpha$ denote the reflection through U. Let $\mathfrak{m}_1$ and $\mathfrak{m}_2$ denote the two components of $\mathfrak{m}_\alpha \setminus U$ and put $D_i := S \cap \mathfrak{m}_i$ for $i \in \{1, 2\}$. By the results of [38, page 151]:

- $\mathrm{Cut}_o(M_\alpha) = (\exp_o \circ F)(S)$.
- $\exp_o \circ F|_{\overline{D}_1}$ and $\exp_o \circ F|_{\overline{D}_2}$ are injective.
- $(\exp_o \circ F)(v) = (\exp_o \circ F)(r_U(v))$ for all $v \in D_1 \cup D_2$.

Hence, the map $\exp_o \circ F|_{\overline{D}_1}$ is a bijective continuous map from a closed disk onto the cut locus $\mathrm{Cut}_o(M_\alpha)$. Since the disk is compact and $\mathrm{Cut}_o(M_\alpha)$ is a Hausdorff space, this shows that $\mathrm{Cut}_o(M_\alpha)$ is homeomorphic to a closed disk. Moreover, $\exp_o \circ F|_{\overline{D}_i} : \overline{D}_i \to M_\alpha$ is an embedding of $\overline{D}_i$ onto $\mathrm{Cut}_o(M_\alpha)$ for $i \in \{1,2\}$.

Theorem 7.7 *For all $\alpha \in \left(0, \frac{\pi}{2}\right]$, it holds that $\mathsf{GC}(M_\alpha) = 2$.*

Proof For $\alpha = \frac{\pi}{2}$, ie the case of a round metric, this is observed in [34, Proposition 4.1], so we will only consider the case of $\alpha \in \left(0, \frac{\pi}{2}\right)$. In the notation from above, we put $E := F(\overline{D}_1)$ and let $f : \mathrm{Cut}_o(M_\alpha) \to E$, $f := (\exp_o \circ F|_E)^{-1}$. Define

$$s : \{o\} \times \mathrm{Cut}_o(M_\alpha) \to GM_\alpha, \qquad (s(o,q))(t) = \exp_o(t \cdot f(q)) \quad \text{for all } t \in [0,1].$$

By Proposition 7.5, the map s is a continuous geodesic motion planner, which shows

$$\mathsf{GC}_o(\mathrm{Cut}_o(M_\alpha)) = 1$$

and thus $\mathsf{GC}(M_\alpha) \le 2$ by (7-2). Since $\mathsf{GC}(M_\alpha) \ge \mathsf{TC}(S^3) = 2$, this shows the claim. □

Remarks 7.8 (1) As Recio-Mitter has shown [34, Example 2.4], there exists a Riemannian metric g_m on S^3 for which $\mathsf{GC}(S^3, g_m) = 3$. This shows that also in the case of S^3, the value of GC depends on the chosen metric.

(2) The cut locus of a point in the Berger sphere M_α for $0 < \alpha < \frac{\pi}{2}$ is a closed disk. It therefore seems tempting to determine the geodesic complexity of M_α via a stratification of this cut locus similarly to what we have done in previous sections. More precisely, an obvious stratification of a closed disk is given by taking one stratum as its interior and another stratum as its boundary. However, this is not an inconsistent stratification as in Definition 4.6 since we would then obtain $\mathsf{GC}(M_\alpha) \ge 3$, whereas we have shown that $\mathsf{GC}(M_\alpha) = 2$.

(3) As this example is particularly instructive, we want to sketch briefly how to show directly that the stratification from the previous paragraph is not inconsistent. Let $K \subset T_o M_\alpha$ be the union of the injectivity domain with the tangent cut locus $\widetilde{\mathrm{Cut}}_o(M_\alpha)$. Using the same notation as in the exposition above, put

$$S_1 := (\exp_o \circ F)(\mathring{D}_1) \subset \mathrm{Cut}_o(M_\alpha) \quad \text{and} \quad S_2 := (\exp_o \circ F)(\partial D_1) \subset \mathrm{Cut}_o(M_\alpha).$$

Under the identification of $\mathrm{Cut}_o(M_\alpha)$ and a closed disk, this is the decomposition from part (2) of this remark. Evidently, this is a stratification in the sense of Definition 4.4.

Let $x \in S_2$ and let $U \subset M_\alpha$ be a neighborhood of x. For sufficiently small U, the intersection $U \cap S_1$ has only one connected component, which we call Z. We claim that

$$\widetilde{\mathrm{Cut}}_o(M_\alpha) \cap \exp_o^{-1}(\{x\}) \cap \overline{\exp_K^{-1}(Z)} \neq \varnothing.$$

By the above discussion of $\mathrm{Cut}_o(M_\alpha)$, the intersection $\widetilde{\mathrm{Cut}}_o(M_\alpha) \cap \exp_o^{-1}(\{x\})$ consists of a single point $v \in T_o M_\alpha$. By choosing a sequence in Z converging to x and recalling that $\exp_o \circ F|_{\overline{D}_i}$ is a homeomorphism for $i \in \{1,2\}$, we see that

$$v \in \overline{\exp_K^{-1}(Z)}.$$

This shows that

$$\widetilde{\mathrm{Cut}}_o(M_\alpha) \cap \exp_o^{-1}(\{x\}) \cap \overline{\exp_K^{-1}(Z)} = \{v\} \neq \varnothing.$$

Hence, the stratification (S_1, S_2) of $\mathrm{Cut}_o(M_\alpha)$ is not inconsistent.

8 Explicit upper bounds for symmetric spaces

In [36, Theorem 5.3], Sakai has determined the cut loci of compact simply connected irreducible symmetric spaces. He showed that their cut loci always allow for stratifications for which each stratum is a submanifold. Since every symmetric space is a Riemannian product of irreducible symmetric spaces, Sakai's results are enough to determine the cut loci of compact simply connected symmetric spaces in general; see our explanations on cut loci of product manifolds in Example 7.1.

In this section, we will first apply the results from Section 5 to find an upper bound for the geodesic complexity of a compact, simply connected, irreducible symmetric space. From Sakai's results, in particular [36, Proposition 4.10], we will further derive estimates on the subspace geodesic complexities of the strata of a cut locus. These numbers appeared on the right-hand side of the inequality in Corollary 5.8 and we will show that they can be estimated from above by certain sectional categories. As a result, we will obtain an upper bound for the geodesic complexity of compact, simply connected, irreducible symmetric spaces given purely in terms of categorical invariants.

We begin by summarizing the main results of [36], stated here in the form of [37, Section 4]. We assume basic knowledge on symmetric spaces that is provided by textbooks in Riemannian geometry like [23] or [33]. In the following, we always let Df_x denote the differential of a differentiable map f in the point x.

Let $M = G/K$ be a compact, simply connected and irreducible symmetric space, where (G, K) is a Riemannian symmetric pair. Explicitly, G is a compact, connected Lie group, K is a closed, connected Lie subgroup of G and G admits an involutive automorphism $\sigma\colon G \to G$ whose fixed-point set satisfies $(\mathrm{Fix}(\sigma))_0 \subset K \subset \mathrm{Fix}(\sigma)$, where $(\mathrm{Fix}(\sigma))_0$ is the identity component of $\mathrm{Fix}(\sigma)$.

Let $\pi\colon G \to M$ denote the orbit space projection, let $e \in G$ denote the unit element and put $o := \pi(e) \in M$. Let $\mathfrak{g}$ and $\mathfrak{k}$ denote the Lie algebras of G and K, respectively, and let $\mathfrak{m} \subset \mathfrak{g}$ be the -1 eigenspace of $D\sigma_e$. Then, since $\mathfrak{k}$ is the $+1$ eigenspace of $D\sigma_e$, there is a vector space decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{m}$. Furthermore, the restriction $D\pi_e|_{\mathfrak{m}}\colon \mathfrak{m} \to T_oM$ is a linear isomorphism; see [23, Theorem IV.3.3].

In the following we give a concise overview of the most important notions related to root systems of symmetric spaces:

- Let $\mathfrak{g}_{\mathbb{C}}$ denote the complexification of $\mathfrak{g}$. By [23, page 284], there exists a Cartan subalgebra $\mathfrak{h} \subset \mathfrak{g}_{\mathbb{C}}$. We recall that a *root* α of the Lie algebra $\mathfrak{g}_{\mathbb{C}}$ is an element of the dual space $\mathfrak{h}^*$ such that there is a nonzero vector $X \in \mathfrak{g}_{\mathbb{C}}$ satisfying

$$[H, X] = \alpha(H)X \quad \text{for all } H \in \mathfrak{h}.$$

The set of nonzero roots of the Lie algebra $\mathfrak{g}_{\mathbb{C}}$ will be called R.

- Let $\mathfrak{a}$ be a maximal abelian subalgebra of $\mathfrak{m}$, which again exists by [23, page 284]. We will call $\mathfrak{a}$ a *Cartan subalgebra of* (G, K).

A root $\alpha \in R$ with $\alpha|_{\mathfrak{a}} \neq 0$ will be called a *root of the symmetric pair* (G, K). The set of roots of the symmetric pair (G, K) will be denoted by $R(G, K)$.

- By choosing a certain real subspace $\mathfrak{h}_{\mathbb{R}}$ of the Cartan subalgebra $\mathfrak{h}$ and defining a lexicographic ordering on $\mathfrak{h}_{\mathbb{R}}$, one defines an ordering on the set of roots R; see [23, page 173]. This defines a set of positive roots $R^+ \subset R$ of $\mathfrak{g}_{\mathbb{C}}$. The set of *positive roots of* (G, K) is then defined as

$$R(G, K)^+ := R^+ \cap R(G, K).$$

There is a maximal element of $R(G, K)^+$ with respect to this ordering, which we denote by δ and call the *highest root of* (G, K).

- Let k be the rank of the symmetric space $M = G/K$. A *simple root of* (G, K) is a positive root α which cannot be written as a sum $\alpha = \beta + \gamma$ with $\beta, \gamma \in R(G, K)^+$. There are precisely k simple roots and one finds that every positive root can be written as a linear combination of the simple roots with nonnegative integer coefficients; see [23, Theorem VII.2.19]. Denote the system of simple roots of (G, K) by $\pi(G, K)$.

- By virtue of the chosen Ad_K–invariant inner product on $\mathfrak{g}$, we will from now on consider the roots to be vectors in $\mathfrak{a}$ in order to follow [36, Section 2].

Based on this terminology, we next recall Sakai's results on the structure of cut loci of symmetric spaces. In the case that there are two or more positive roots of (G, K), we define a subset $\mathcal{D}$ of the power set of $\pi(G, K)$ by

$$\mathcal{D} := \{\Delta \subset \pi(G, K) \mid \Delta \neq \varnothing, \delta \notin \Delta\}.$$

If there is only one positive root γ, which is therefore also the only simple root and also the highest root, define

$$\mathcal{D} := \{\{\gamma\}\}.$$

Let $\langle\cdot,\cdot\rangle$ denote the chosen Ad_K–invariant inner product on $\mathfrak{g}$ and consider the Weyl chamber of $\pi(G, K)$ that is given by

$$W := \{X \in \mathfrak{a} \mid \langle\gamma, X\rangle > 0 \text{ for all } \gamma \in \pi(G, K)\}.$$

See [23, Section VII.2] for further details on Weyl chambers and their connection to root systems. If there is more than one positive root, let

$$S_\Delta := \{X \in \overline{W} \mid \langle\gamma, X\rangle > 0 \text{ for all } \gamma \in \Delta,\ \langle\gamma, X\rangle = 0 \text{ for all } \gamma \in \pi(G, K) \smallsetminus \Delta,\ 2\langle\delta, X\rangle = 1\}$$

for each $\Delta \in \mathcal{D}$. If there is just one positive root γ, then set

$$S_{\{\gamma\}} := \{X \in \mathfrak{a} \mid 2\langle\gamma, X\rangle = 1\}.$$

Since $\mathfrak{a}$ is one-dimensional in that case, $S_{\{\gamma\}}$ consists of a single point.

Let $\exp\colon \mathfrak{g} \to G$ be the exponential map of G and put

$$\mathrm{Exp}\colon \mathfrak{m} \to M, \quad \mathrm{Exp} := \pi \circ \exp|_{\mathfrak{m}}.$$

For $\Delta \in \mathcal{D}$, we let

$$\widetilde{\Phi}_\Delta\colon K \times S_\Delta \to M, \quad \widetilde{\Phi}_\Delta(k, X) = \mathrm{Exp}(\mathrm{Ad}(k)(X)),$$

and put $Z_\Delta := \{k \in K \mid \mathrm{Exp}(\mathrm{Ad}(k)(X)) = \mathrm{Exp}(X) \text{ for all } X \in S_\Delta\}$. One checks without difficulties that Z_Δ is a closed subgroup of K. As shown in [36, Proposition 4.10], each $\widetilde{\Phi}_\Delta$ induces a differentiable embedding

$$\Phi_\Delta\colon K/Z_\Delta \times S_\Delta \to M.$$

Put $C_\Delta := \operatorname{im} \Phi_\Delta$ for each $\Delta \in \mathcal{D}$. By [36, Theorem 5.3], the cut locus of M at o is then given by

$$\operatorname{Cut}_o(M) = \bigcup_{\Delta \in \mathcal{D}} C_\Delta$$

and the C_Δ satisfy

$$\begin{aligned} C_\Delta \cap C_{\Delta'} &= \varnothing && \text{for all } \Delta, \Delta' \in \mathcal{D},\ \Delta \neq \Delta', \\ \overline{C}_\Delta &= \bigcup_{\Delta' \subset \Delta} C_{\Delta'} && \text{for all } \Delta \in \mathcal{D}. \end{aligned} \tag{8-1}$$

Let k be the rank of M. For $i \in \{1, 2, \dots, k\}$, we put

$$\mathcal{D}_i := \{\Delta \in \mathcal{D} \mid \#\Delta = i\} \quad \text{and} \quad C_i := \bigcup_{\Delta \in \mathcal{D}_i} C_\Delta.$$

It follows from (8-1) that $(C_k, C_{k-1}, \dots, C_1)$ is a stratification of $\operatorname{Cut}_o(M)$ and that the C_Δ for $\Delta \in \mathcal{D}_i$ are precisely the connected components of C_i. Since M is a homogeneous Riemannian manifold, we thus obtain from Corollary 5.8 that

$$\mathsf{GC}(M) \leq \mathsf{secat}(\operatorname{ev}_o \colon \operatorname{Isom}(M) \to M) \cdot \sum_{i=1}^{k} \max_{\Delta \in \mathcal{D}_i} \mathsf{GC}_o(C_\Delta) + 1. \tag{8-2}$$

It remains to find upper bounds on the numbers $\mathsf{GC}_o(C_\Delta)$.

Proposition 8.1 *For each $\Delta \in \mathcal{D}$, it holds that*

$$\mathsf{GC}_o(C_\Delta) \leq \mathsf{secat}(q_\Delta \colon K \to K/Z_\Delta),$$

where q_Δ denotes the orbit space projection.

Proof Let $r := \mathsf{secat}(q_\Delta)$. Then, by Lemma 4.1, there are pairwise disjoint and locally compact subsets $B_1, \dots, B_r \subset K/Z_\Delta$ such that, for each $i \in \{1, 2, \dots, r\}$, there is a continuous local section $s_i \colon B_i \to K$ of q_Δ. Using these s_i, we define

$$\sigma_i \colon \{o\} \times \Phi_\Delta(B_i \times S_\Delta) \to GM, \quad \big(\sigma_i(o, \Phi_\Delta(x, X))\big)(t) = \operatorname{Exp}\big(t \cdot \operatorname{Ad}(s_i(x))(X)\big),$$

for every $i \in \{1, 2, \dots, r\}$. By construction, each σ_i is continuous and $\sigma_i(o, \Phi_\Delta(x, X))$ is a geodesic segment for all $(x, X) \in B_i \times S_\Delta$ and each $i \in \{1, 2, \dots, r\}$. Moreover,

$$\begin{aligned} \big(\sigma_i(o, \Phi_\Delta(x, X))\big)(0) &= \operatorname{Exp}(0) = o, \\ \big(\sigma_i(o, \Phi_\Delta(x, X))\big)(1) &= \operatorname{Exp}\big(\operatorname{Ad}(s(x))(X)\big) = \Phi_\Delta(x, X), \end{aligned}$$

by definition of Φ_Δ. Thus, the σ_i are continuous geodesic motion planners. Since the sets $\Phi_\Delta(B_1 \times S_\Delta), \dots, \Phi_\Delta(B_r \times S_\Delta)$ are pairwise disjoint, locally compact and cover $\Phi(K/Z_\Delta \times S_\Delta) = C_\Delta$, this shows that $\mathsf{GC}_o(C_\Delta) \leq r$. □

Combining Proposition 8.1 with (8-2) yields the following upper bound:

Theorem 8.2 *Let (G,K) be a Riemannian symmetric pair and let $M=G/K$ be the corresponding symmetric space. Assume that M is compact, simply connected and irreducible. Then, with $\mathcal{D}_i$ and Z_Δ given as above,*

$$\mathsf{GC}(M)\le \mathsf{secat}(\mathrm{ev}_o\colon \mathrm{Isom}(M)\to M)\cdot \sum_{i=1}^{\mathrm{rk}(M)} \max_{\Delta\in\mathcal{D}_i} \mathsf{secat}(q_\Delta\colon K\to K/Z_\Delta)+1,$$

where $\mathrm{rk}(M)$ denotes the rank of M.

Corollary 8.3 *Let (G,K) be a Riemannian symmetric pair and let $M=G/K$ be the corresponding symmetric space. Assume that M is compact, simply connected and irreducible. Then, with $\mathcal{D}_i$ and Z_Δ given as above,*

$$\mathsf{GC}(M)\le \mathsf{cat}(M)\cdot \sum_{i=1}^{\mathrm{rk}(M)} \max_{\Delta\in\mathcal{D}_i} \mathsf{cat}(K/Z_\Delta)+1.$$

Proof This is an immediate consequence of Theorem 8.2 and [39, Theorem 18]. □

We want to conclude by applying the upper bounds to two examples of compact symmetric spaces whose cut loci have already been discussed in the works of Sakai, more precisely in [36, Example 5.4; 35, Section 4.2].

Example 8.4 Consider the complex projective space $\mathbb{C}P^n=U(n+1)/(U(1)\times U(n))$ with the Fubini–Study metric. This is a compact and simply connected symmetric space of rank one. Its cut locus is studied in detail in [36, Example 5.4]. Let $G=U(n+1)$, let $\mathfrak{g}=\mathfrak{u}(n+1)$ be its Lie algebra and let $K=U(1)\times U(n)$. Let $\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{m}$ denote the decomposition of $\mathfrak{g}$ with respect to the symmetric pair (G,K). By the same methods as in [23, page 452], which treats the Lie algebra of $\mathrm{SU}(n)$, one computes that

$$\mathfrak{k}=\left\{\begin{pmatrix} ia & 0\\ 0 & \xi\end{pmatrix}\;\middle|\; a\in\mathbb{R},\ \xi\in\mathfrak{u}(n)\right\} \quad\text{and}\quad \mathfrak{m}=\left\{\begin{pmatrix} 0 & \bar{u}^T\\ -u & 0\end{pmatrix}\;\middle|\; u\in\mathbb{C}^n\right\}.$$

Then $\mathfrak{a}=\mathrm{span}_{\mathbb{R}}(\{H_0\})$ is a Cartan subalgebra of (G,K), where $H_0=(h_{ij})\in\mathfrak{g}$ is given by

$$h_{ij}=\begin{cases} \frac{\pi}{2} & \text{if } (i,j)=(1,n+1),\\ -\frac{\pi}{2} & \text{if } (i,j)=(n+1,1),\\ 0 & \text{otherwise.}\end{cases}$$

In particular, every system of simple roots of (G,K) consists of a unique element. Let o be the equivalence class of the neutral element of G in $G/K=\mathbb{C}P^n$. Then $\mathrm{Cut}_o(M)$

consists of a unique submanifold, given by

$$\mathrm{Cut}_o(M) = \{\mathrm{Exp}(\mathrm{Ad}(k)(H_0)) \mid k \in U(1) \times U(n)\}.$$

Sakai further showed that

$$Z_0 := \{k \in U(1) \times U(n) \mid \mathrm{Exp}(\mathrm{Ad}(k)(H_0)) = \mathrm{Exp}(H_0))\}$$

can be identified with $Z_0 = U(1) \times U(n-1) \times U(1)$. Hence, by Proposition 8.1,

$$\begin{aligned}(8\text{-}3)\quad & \mathsf{GC}_M(\mathrm{Cut}_o(M)) \\ & \qquad \le \mathsf{secat}\big(U(1) \times U(n) \to (U(1) \times U(n))/(U(1) \times U(n-1) \times U(1))\big).\end{aligned}$$

One easily checks that the map

$$\varphi\colon (U(1) \times U(n))/(U(1) \times U(n-1) \times U(1)) \to U(n)/(U(n-1) \times U(1)) = \mathbb{C}P^{n-1},$$
$$\varphi([z, A]) = [A],$$

where $(z, A) \in U(1) \times U(n)$, is a well-defined homeomorphism. Let $p\colon U(n) \to \mathbb{C}P^{n-1}$ denote the principal $U(1)$–bundle over the homogeneous space $\mathbb{C}P^{n-1}$. Assume that $s\colon V \to U(n)$ is a continuous local section of p over a subset $V \subset \mathbb{C}P^{n-1}$. Then we obtain a continuous local section $\tilde{s}\colon V \to U(1) \times U(n)$ of the principal fiber bundle

$$U(1) \times U(n) \to (U(1) \times U(n))/(U(1) \times U(n-1) \times U(1))$$

by setting $\tilde{s}(\varphi^{-1}(p)) = (z_0, s(p))$ for $p \in V$, where $z_0 \in U(1)$ is a fixed element. This shows that

$$\mathsf{secat}\big(U(1)\times U(n)\to(U(1)\times U(n))/(U(1)\times U(n-1)\times U(1))\big)\le\mathsf{secat}(U(n)\to\mathbb{C}P^{n-1}).$$

Hence, we derive from (8-3) that

$$\mathsf{GC}_o(\mathrm{Cut}_o(M)) \le \mathsf{secat}(U(n) \to \mathbb{C}P^{n-1}) \le \mathsf{cat}(\mathbb{C}P^{n-1}) = n,$$

where we used [39, Theorem 18] for the second inequality. The fact that $\mathsf{cat}(\mathbb{C}P^{n-1}) = n$ is shown in [6, Example 1.51]. Eventually, by Theorem 5.4 and the same references,

$$\begin{aligned}\mathsf{GC}(\mathbb{C}P^n) &\le \mathsf{secat}(U(n+1) \to \mathbb{C}P^n)\, \mathsf{GC}_o(\mathrm{Cut}_o(M)) + 1 \\ &\le \mathsf{cat}(\mathbb{C}P^n)\, \mathsf{cat}(\mathbb{C}P^{n-1}) + 1 \\ &= (n+1)n + 1.\end{aligned}$$

Since $\mathsf{TC}(\mathbb{C}P^n) = 2n + 1$, as computed in [14, Lemma 28.1], we derive using Remark 2.4(1) that

$$2n + 1 \le \mathsf{GC}(\mathbb{C}P^n) \le (n+1)n + 1 \quad \text{for all } n \in \mathbb{N}.$$

For $n = 2$, this shows that $\mathsf{GC}(\mathbb{C}P^2) \in \{5, 6, 7\}$.

Example 8.5 Consider the complex Grassmann manifold

$$G_2(\mathbb{C}^4) = U(4)/(U(2)\times U(2)).$$

As a quotient of a compact Lie group by a closed subgroup, $G_2(\mathbb{C}^4)$ is compact. Let $V_2(\mathbb{C}^4)$ be the corresponding complex Stiefel manifold. As shown in Example 4.54 of [22], $V_2(\mathbb{C}^4)$ is simply connected. Since the fiber $U(2)$ of the fibration $V_2(\mathbb{C}^4)\to G_2(\mathbb{C}^4)$ is connected, it follows from the long exact homotopy sequence of this fibration that $G_2(\mathbb{C}^4)$ is simply connected as well. The cut loci of $G_2(\mathbb{C}^4)$ are discussed in [35, Section 4.2; 36, Example 5.5]. The corresponding decomposition of the Lie algebra $\mathfrak{g}=\mathfrak{u}(4)$ of $G=U(4)$ is given by $\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{m}$, where

$$\mathfrak{k}=\left\{\begin{pmatrix}\alpha & 0\\ 0 & \beta\end{pmatrix}\;\middle|\;\alpha,\beta\in\mathfrak{u}(2)\right\}\quad\text{and}\quad\mathfrak{m}=\left\{\begin{pmatrix}0 & \xi\\ -\bar{\xi}^T & 0\end{pmatrix}\;\middle|\;\xi\in M_2(\mathbb{C})\right\}.$$

Here, $\mathfrak{k}$ is the Lie algebra of $K=U(2)\times U(2)$. A Cartan subalgebra $\mathfrak{a}\subset\mathfrak{m}$ is spanned by

$$e_1 := \frac{1}{2\pi}\begin{pmatrix}0&0&1&0\\0&0&0&0\\-1&0&0&0\\0&0&0&0\end{pmatrix}\quad\text{and}\quad e_2 := \frac{1}{2\pi}\begin{pmatrix}0&0&0&0\\0&0&0&1\\0&0&0&0\\0&-1&0&0\end{pmatrix}.$$

By [36, page 143], one can define positive roots and simple roots of $(U(4),U(2)\times U(2))$ in such a way that $2e_1=\delta$ is the highest root and that a system of simple roots is given by

$$\pi(G,K)=\{\gamma_1,\gamma_2\},\quad\text{where }\gamma_1:=2e_2,\ \gamma_2:=e_1-e_2.$$

Thus, in the notation from above, $\mathcal{D}=\{\Delta_0,\Delta_1,\Delta_2\}$, where $\Delta_0=\{\gamma_1,\gamma_2\}$, $\Delta_1=\{\gamma_1\}$ and $\Delta_2=\{\gamma_2\}$. With $S_i:=S_{\Delta_i}$ for $i\in\{0,1,2\}$, one computes that

$$S_0=\{\pi^2e_1+\lambda e_2\in\mathfrak{a}\mid\lambda\in(0,\pi^2)\},\quad S_1=\{\pi^2e_1\},\quad S_2=\{\pi^2(e_1+e_2)\}.$$

We further put $Z_i:=Z_{\Delta_i}$ for each i. By computing the corresponding matrix exponentials, we obtain

$$Z_1=\{\operatorname{diag}(a,b,c,d)\in U(2)\times U(2)\mid a,b,c,d\in U(1)\}\cong U(1)^4.$$

Since $U(2)/(U(1)\times U(1))$ is diffeomorphic to $\mathbb{C}P^1\cong S^2$, it follows that $K/Z_1\cong S^2\times S^2$. Hence, $\mathsf{cat}(K/Z_1)=\mathsf{cat}(S^2\times S^2)\le 3$ by the product inequality for cat; see [6, Theorem 1.37]. One further computes by matrix exponentials that $Z_2=K$, so K/Z_2 consists of a single point, which yields $\mathsf{cat}(K/Z_2)=1$. By [36, Lemma 4.9], for every fixed $X \in S_0$ we obtain $Z_0 = \{k \in K \mid \mathrm{Exp}(\mathrm{Ad}(k)(X)) = \mathrm{Exp}(X)\}$. We choose $X = \pi^2 e_1 + \frac{1}{2}\pi^2 e_2$ and claim that

$$Z_0 = \{\mathrm{diag}(a,b,c,b) \in U(2) \times U(2) \mid a,b,c \in U(1)\}.$$

To see this, we compute that

$$\exp(X) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -1 & 0 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

The condition $\mathrm{Exp}(\mathrm{Ad}_k(X)) = \mathrm{Exp}(X)$ is equivalent to $\exp(-X)k\exp(X) \in K$. One then checks by an explicit computation that $k \in K$ satisfies this condition if and only if

$$k = \mathrm{diag}(a,b,c,b) \quad \text{with } a,b,c \in U(1).$$

Hence, $K \to K/Z_0$ is a bundle with typical fiber $U(1)^3$, where an inclusion of the fiber is given by

$$f : U(1)^3 \to U(2) \times U(2), \quad f(a,b,c) = \mathrm{diag}(a,b,c,b).$$

We want to show that K/Z_0 is simply connected. By the long exact sequence of homotopy groups of that bundle, it suffices to show that $f_* : \pi_1(U(1)^3) \to \pi_1(U(2)^2)$ is surjective. Let $\gamma : [0,1] \to U(1)$, $\gamma(t) = e^{2\pi i t}$. We observe that $\pi_1(U(1)^3) \cong \mathbb{Z}^3$. A set of generators of $\pi_1(U(1)^3)$ is given by the homotopy classes of the loops $\gamma_1, \gamma_2, \gamma_3 : [0,1] \to U(1)^3$ defined as

$$\gamma_1(t) = (\gamma(t), 1, 1), \quad \gamma_2(t) = (1, \gamma(t), 1), \quad \gamma_3(t) = (1, 1, \gamma(t)).$$

We further observe that $\pi_1(U(2)^2) \cong \mathbb{Z}^2$, where a set of generators is given by the homotopy classes of

$$\beta_1, \beta_2 : [0,1] \to U(2) \times U(2), \quad \beta_1(t) = \mathrm{diag}(\gamma(t), 1, 1, 1), \quad \beta_2(t) = \mathrm{diag}(1, 1, \gamma(t), 1).$$

Here we used [4, Example VII.8.1]. One immediately sees that $f \circ \gamma_1 = \beta_1$ and $f \circ \gamma_3 = \beta_2$. This shows that the image f_* contains a set of generators, and hence f_* is surjective. Thus, $\pi_1(K/Z_0)$ is the trivial group, which implies by [6, Theorem 1.50] that

$$\mathsf{cat}(K/Z_0) \le \tfrac{1}{2}\dim(K/Z_0) + 1 = \tfrac{5}{2} + 1 = \tfrac{7}{2}.$$

Since cat is integer-valued, we obtain $\mathsf{cat}(K/Z_0) \le 3$. To employ Corollary 8.3, we still need to estimate $\mathsf{cat}(G_2(\mathbb{C}^4))$ from above. Another use of [6, Theorem 1.50] shows that

$$\mathsf{cat}(G_2(\mathbb{C}^4)) \le \tfrac{1}{2}\dim(G_2(\mathbb{C}^4)) + 1 = 5.$$

Inserting the results of our computations into Corollary 8.3, we derive

$$\begin{aligned}\mathsf{GC}(G_2(\mathbb{C}^4)) &\le \mathsf{cat}(G_2(\mathbb{C}^4))\big(\mathsf{cat}(K/Z_0)+\max\{\mathsf{cat}(K/Z_1),\mathsf{cat}(K/Z_2)\}\big)+1\\ &\le 5(3+3)+1=31.\end{aligned}$$

By [14, Lemma 28.1], it further holds that $\mathsf{TC}(G_2(\mathbb{C}^4)) = \dim(G_2(\mathbb{C}^4))+1 = 9$. Thus, by the previous inequality and Remarks 2.4, we obtain

$$9 \le \mathsf{GC}(G_2(\mathbb{C}^4)) \le 31.$$

References

[1] **M Berger**, *Les variétés riemanniennes homogènes normales simplement connexes à courbure strictement positive*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 15 (1961) 179–246 MR Zbl

[2] **R L Bishop**, *Decomposition of cut loci*, Proc. Amer. Math. Soc. 65 (1977) 133–136 MR Zbl

[3] **Z Błaszczyk**, **J G Carrasquel-Vera**, *Topological complexity and efficiency of motion planning algorithms*, Rev. Mat. Iberoam. 34 (2018) 1679–1684 MR Zbl

[4] **G E Bredon**, *Topology and geometry*, Graduate Texts in Math. 139, Springer (1993) MR Zbl

[5] **D C Cohen**, **L Vandembroucq**, *Topological complexity of the Klein bottle*, J. Appl. Comput. Topol. 1 (2017) 199–213 MR Zbl

[6] **O Cornea**, **G Lupton**, **J Oprea**, **D Tanré**, *Lusternik–Schnirelmann category*, Mathematical Surveys and Monographs 103, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[7] **R J Crittenden**, *Minimum and conjugate points in symmetric spaces*, Canadian J. Math. 14 (1962) 320–328 MR Zbl

[8] **D M Davis**, *Geodesics in the configuration spaces of two points in $\mathbb{R}^n$*, Tbilisi Math. J. 14 (2021) 149–162 MR Zbl

[9] **D M Davis**, **M Harrison**, **D Recio-Mitter**, *Two robots moving geodesically on a tree*, Algebr. Geom. Topol. 22 (2022) 785–814 MR Zbl

[10] **D M Davis**, **D Recio-Mitter**, *The geodesic complexity of n–dimensional Klein bottles*, New York J. Math. 27 (2021) 296–318 MR Zbl

[11] **T tom Dieck**, *Algebraic topology*, Eur. Math. Soc., Zürich (2008) MR Zbl

[12] **M Farber**, *Topological complexity of motion planning*, Discrete Comput. Geom. 29 (2003) 211–221 MR Zbl

[13] **M Farber**, *Instabilities of robot motion*, Topology Appl. 140 (2004) 245–266 MR Zbl

[14] **M Farber**, *Topology of robot motion planning*, from "Morse theoretic methods in nonlinear analysis and in symplectic topology" (P Biran, O Cornea, F Lalonde, editors), NATO Sci. Ser. II Math. Phys. Chem. 217, Springer (2006) 185–230 MR Zbl

[15] **M Farber**, *Invitation to topological robotics*, Eur. Math. Soc., Zürich (2008) MR Zbl

[16] **M Farber**, *Configuration spaces and robot motion planning algorithms*, from "Combinatorial and toric homotopy" (A Darby, J Grbic, Z Lü, J Wu, editors), Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. 35, World Sci., Hackensack, NJ (2018) 263–303 MR Zbl

[17] **M Frumosu**, **S Rosenberg**, *Lefschetz theory, geometric Thom forms and the far point set*, Tokyo J. Math. 27 (2004) 337–355 MR Zbl

[18] **J Gallier**, **J Quaintance**, *Differential geometry and Lie groups: a computational perspective*, Geometry and Computing 12, Springer (2020) MR Zbl

[19] **S Gallot**, **D Hulin**, **J Lafontaine**, *Riemannian geometry*, 3rd edition, Springer (2004) MR Zbl

[20] **H Gluck**, **D Singer**, *Scattering of geodesic fields, I*, Ann. of Math. 108 (1978) 347–372 MR Zbl

[21] **M Goresky**, **R MacPherson**, *Stratified Morse theory*, Ergebnisse der Math. (3) 14, Springer (1988) MR Zbl

[22] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[23] **S Helgason**, *Differential geometry*, *Lie groups*, *and symmetric spaces*, Pure and Applied Mathematics 80, Academic, New York (1978) MR Zbl

[24] **J-I Itoh**, **T Sakai**, *Cut loci and distance functions*, Math. J. Okayama Univ. 49 (2007) 65–92 MR Zbl

[25] **S Kobayashi**, **K Nomizu**, *Foundations of differential geometry*, *I*, Interscience, New York (1963) MR Zbl

[26] **J M Lee**, *Introduction to smooth manifolds*, 2nd edition, Graduate Texts in Math. 218, Springer (2013) MR Zbl

[27] **J M Lee**, *Introduction to Riemannian manifolds*, 2nd edition, Graduate Texts in Math. 176, Springer (2018) MR Zbl

[28] **M Mainkar**, **B Schmidt**, *Metric foliations of homogeneous three-spheres*, Geom. Dedicata 203 (2019) 73–84 MR Zbl

[29] **H Mas-Ku**, **E Torres-Giese**, *Motion planning algorithms for configuration spaces*, Bol. Soc. Mat. Mex. 21 (2015) 265–274 MR Zbl

[30] **J R Munkres**, *Topology*: *a first course*, Prentice-Hall, Englewood Cliffs, NJ (1975) MR Zbl

[31] **B O'Neill**, *Semi-Riemannian geometry*: *with applications to relativity*, Pure and Applied Mathematics 103, Academic, New York (1983) MR Zbl

[32] **V Ozols**, *Cut loci in Riemannian manifolds*, Tohoku Math. J. 26 (1974) 219–227 MR Zbl

[33] **P Petersen**, *Riemannian geometry*, 3rd edition, Graduate Texts in Math. 171, Springer (2016) MR Zbl

[34] **D Recio-Mitter**, *Geodesic complexity of motion planning*, J. Appl. Comput. Topol. 5 (2021) 141–178 MR Zbl

[35] **T Sakai**, *On cut loci of compact symmetric spaces*, Hokkaido Math. J. 6 (1977) 136–161 MR Zbl

[36] **T Sakai**, *On the structure of cut loci in compact Riemannian symmetric spaces*, Math. Ann. 235 (1978) 129–148 MR Zbl

[37] **T Sakai**, *Cut loci of compact symmetric spaces*, from "Minimal submanifolds and geodesics" (M Obata, editor), North-Holland, Amsterdam (1979) 193–207 MR Zbl

[38] **T Sakai**, *Cut loci of Berger's spheres*, Hokkaido Math. J. 10 (1981) 143–155 MR Zbl

[39] **A S Schwarz**, *The genus of a fiber space*, Trudy Moskov. Mat. Obšč. 10 (1961) 217–272 MR In Russian; translated in "Eleven papers on topology and algebra", Amer. Math. Soc. Transl. Ser. (2) 55, Amer. Math. Soc., Providence, RI (1966) 49–140

[40] **A D Weinstein**, *The cut locus and conjugate locus of a riemannian manifold*, Ann. of Math. 87 (1968) 29–41 MR Zbl

Institut für Mathematik, Martin-Luther-Universität Halle-Wittenberg
Halle, Germany

Mathematisches Institut, Universität Leipzig
Leipzig, Germany

stephan.mescher@mathematik.uni-halle.de,
maximilian.stegemeyer@math.uni-leipzig.de

Received: 25 May 2021 Revised: 29 November 2021

Algebraic & Geometric Topology **23**:5 (2023) 2271–2308
DOI: 10.2140/agt.2023.23.2271
Published: 25 July 2023

Adequate links in thickened surfaces and the generalized Tait conjectures

Hans U Boden
Homayun Karimi
Adam S Sikora

We apply Kauffman bracket skein algebras to develop a theory of skein adequate links in thickened surfaces. We show that any alternating link diagram on a surface is skein adequate. We apply our theory to establish the first and second Tait conjectures for adequate links in thickened surfaces. Our notion of skein adequacy is broader and more powerful than the corresponding notions of adequacy previously considered for link diagrams in surfaces.

For a link diagram D on a surface Σ of minimal genus $g(\Sigma)$, we show that

$$\operatorname{span}([D]_\Sigma) \le 4c(D) + 4|D| - 4g(\Sigma),$$

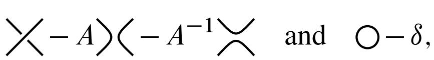
where $[D]_\Sigma$ is its skein bracket, $|D|$ is the number of connected components of D, and $c(D)$ is the number of crossings. This extends a classical result of Kauffman, Murasugi and Thistlethwaite. We further show that the above inequality is an equality if and only if D is weakly alternating. This is a generalization of a well-known result for classical links due to Thistlethwaite. Thus, the skein bracket detects the crossing number for weakly alternating links. As an application, we show that the crossing number is additive under connected sum for adequate links in thickened surfaces.

57K10, 57K12; 57K14, 57K31

1 Introduction

The Kauffman bracket is a $\mathbb{Z}[A^{\pm 1}]$–valued invariant of framed links in $\mathbb{R}^3$ determined by the skein relations

(1)



where $\delta = -A^2 - A^{-2}$.

It naturally extends to an invariant of framed links in an arbitrary oriented 3–manifold M (possibly with boundary), via the skein module construction: Let $\mathscr{L}(M)$ be the set of all unoriented, framed links in M, including the empty link $\varnothing$. The *skein module* $\mathscr{S}(M)$ of M is the quotient of the free $\mathbb{Z}[A^{\pm 1}]$–module spanned by $\mathscr{L}(M)$ by the submodule generated by the Kauffman bracket skein relations (1); see Przytycki [44] and Turaev [50; 51].

By this construction, the bracket

$$[\cdot]\colon \mathscr{L}(M) \to \mathscr{S}(M),$$

sending framed links to their equivalence classes in $\mathscr{S}(M)$, called the *skein bracket*, is the universal invariant of framed links in M satisfying (1).

Independently of this initial motivation, skein modules quickly began to play a much broader role in the development of quantum topology, for example in connection with $\mathrm{SL}(2,\mathbb{C})$ character varieties (see Bullock [12], Przytycki and Sikora [45], Frohman, Kania-Bartoszynska and Lê [25], Turaev [51] and Bullock, Frohman and Kania-Bartoszynska [13]), topological quantum field theory (see Blanchet, Habegger, Masbaum and Vogel [7] and Turaev [52]), (quantum) Teichmüller spaces and (quantum) cluster algebras (see Bonahon and Wong [11], Costantino [17], Fock and Goncharov [22], Fomin, Shapiro and Thurston [23] and Muller [42]), the AJ conjecture (see Frohman, Gelca and Lofaro [24] and Lê [36]), and many more.

Here we develop a general theory of skein adequacy (called adequacy, for short) for links in thickened surfaces with the aid of skein modules.

Let Σ be an oriented surface and $I=[0,1]$ be the unit interval. The skein module of the thickened surface $\Sigma \times I$ comes naturally equipped with a product structure given by stacking, ie the product $L_1 \cdot L_2$ is defined by placing L_1 on top of L_2 in $\Sigma \times I$. With this product structure, the skein module $\mathscr{S}(\Sigma \times I)$ becomes an algebra over $\mathbb{Z}[A^{\pm 1}]$.

Let $\mathscr{C}(\Sigma)$ denote the set of all nontrivial unoriented simple loops on Σ up to isotopy and $\mathscr{MC}(\Sigma)$ denote the set of all nontrivial unoriented multiloops on Σ, ie collections of pairwise disjoint simple noncontractible loops, including $\varnothing$, up to isotopy. Then, by [44] (cf Sikora and Westbury [46]), the skein module $\mathscr{S}(\Sigma \times I)$ is a free $\mathbb{Z}[A^{\pm 1}]$–module with basis $\mathscr{MC}(\Sigma)$. Consequently, via this identification, the skein bracket gives a map

$$[\cdot]_\Sigma\colon \mathscr{L}(\Sigma \times I) \to \mathscr{S}(\Sigma \times I) = \mathbb{Z}[A^{\pm 1}]\mathscr{MC}(\Sigma). \tag{2}$$

We use the association (2) to develop a theory of skein adequacy for links in $\Sigma \times I$ which extends that for classical links. This theory is broader and more powerful than the corresponding notions of simple adequacy (see Lickorish and Thistlethwaite [38]) and homological adequacy (see Boden and Karimi [8]). For example, we will see that every weakly alternating link in $\Sigma \times I$ without removable nugatory crossings is skein adequate.

We will apply the skein bracket to establish the first and second Tait conjectures for skein adequate link diagrams on surfaces. The first one says that skein adequate diagrams have minimal crossing number, and the second one says that two skein adequate diagrams for the same oriented link have the same writhe. (The writhe of a link diagram D is denoted by $w(D)$ and is defined to be the sum of its crossing signs.) These results strengthen the earlier work of Adams, Fleming, Levin and Turner [3], who showed the minimal crossing number result for reduced alternating knot diagrams in surfaces. We also strengthen the minimality result of [8] for homologically adequate link diagrams in surfaces, and further show that any connected sum of two skein adequate link diagrams on surfaces is again skein adequate. This implies that the crossing number and writhe are essentially additive under connected sum of skein adequate links in thickened surfaces.

For any link diagram D on a surface Σ of minimal genus, we prove that

$$\operatorname{span}([D]_\Sigma) \le 4c(D) + 4|D| - 4g(\Sigma),$$

where $|D|$ is the number of connected components of D, $c(D)$ is the number of crossings and $g(\Sigma)$ is the genus of Σ. This inequality generalizes a result proved by Kauffman, Murasugi and Thistlethwaite for link diagrams on $\mathbb{R}^2$ [31; 43; 48], extending their nice geometric application of the Kauffman bracket. It also extends and strengthens an analogous recent result proved in [8] using the homological Kauffman bracket.

Additionally, we prove that the above inequality is an equality if and only if D is weakly alternating. Therefore, the skein bracket, together with the crossing number, distinguishes weakly alternating links. That generalizes the analogous result of Thistlethwaite for classical links.

Broader context and motivation

While the results presented here are new only for links in noncontractible surfaces, generalized link theory is of growing interest and has many potential connections to classical links and 3–dimensional geometry. We take a moment to discuss some of them.

One motivation for our results is their connection to the theory of virtual knots and links, which can be viewed as links in thickened surfaces, considered up to homeomorphisms and stabilization; see Carter, Kamada and Saito [14]. By Kuperberg's theorem [33], minimal genus realizations of virtual links are unique up to homeomorphism. Our theory of adequate and alternating links in thickened surfaces is invariant under surface homeomorphisms and, therefore, many of the results given here can be restated in the language of virtual links.

A second motivation involves potentially novel applications to classical link theory. The Turaev surface construction associates to any classical link diagram an alternating link in a thickened surface; see Turaev [49], Dasbach, Futer, Kalfagianni, Lin and Stoltzfus [18] and Champanerkar and Kofman [15]. Menasco [40] famously proved hyperbolicity for prime alternating (nontorus) links in S^3, and his result has been extended to prime alternating links $L \subset \Sigma \times I$ by Adams, Albors-Riera, Haddock, Li, Nishida, Reinoso and Wang [1]. This result opens the door to using the hyperbolic geometry of alternating links in higher-genus surfaces to profitably study nonalternating classical links; eg see Adams, Eisenberg, Greenberg, Kapoor, Liang, O'Connor, Pachecho-Tallaj and Wang [2] and the many other papers cited below.

Dasbach and Lin [19] proved a remarkable result giving a bound on the volume of alternating link complements in terms of the second and penultimate coefficients of the Jones polynomial. Lackenby [34] established an equally remarkable bound on the volume of alternating link complements in terms of the diagrammatic *twist number*. For alternating hyperbolic links in S^3, the results of [19] imply that the twist number is essentially an isotopy invariant of L, but this is not true in general.

These methods have been generalized to nonalternating hyperbolic links in S^3 (see Blair, Allen and Rodriguez [5; 6]) and to hyperbolic links in arbitrary compact oriented 3–manifolds by Howie and Purcell [27]. In general, there is a notion of weakly generalized alternating link diagrams on surfaces due to Howie [26], extended to links in compact oriented 3–manifolds via "generalized projection surfaces" by Howie and Purcell [27].

The volume bounds have been extended to alternating links in thickened surfaces by Bavier and Kalfagianni [4] and Will [53] and also to virtual alternating links by Champanerkar and Kofman [16]. In [16; 53], the volume bounds are expressed in terms of the Jones–Krushkal polynomial — see Krushkal [32] and Boden and Karimi [8] — and in [4] they are expressed in terms of a skein invariant derived from fully contractible smoothings. In [4, Corollary 1.3], they deduce that, for certain alternating links in thickened surfaces, the twist number is an isotopy invariant. Interestingly, this result is consistent with the generalized Tait flyping conjecture.

2 State sum formula and the generalized Jones polynomial

We will assume throughout that Σ is an oriented surface with one or more connected components, which may also have boundary. Links in $\Sigma \times I$ will be represented as diagrams on Σ up to Reidemeister moves.

Every framed link in $\Sigma \times I$ can also be represented by a link diagram with framing given by the blackboard framing. Equivalence of framed links is given by regular isotopy, which includes the second and third Reidemeister moves, as well as the modified first Reidemeister move, which replaces or with .

Let D be a link diagram on a surface Σ. Given a crossing of D, we consider its A–type and B–type resolution, as in the Kauffman bracket construction. A choice of resolution for each crossing of D is called a *state*. Let $\mathfrak{S}(D)$ denote the set of all states of D. Thus, $|\mathfrak{S}(D)| = 2^{c(D)}$, where $c(D)$ is the crossing number of D.

For $S \in \mathfrak{S}(D)$, let $|S|$ denote the number of loops in S and $t(S)$ the number of contractible loops in S. Also let $\widehat{S}$ denote S with contractible loops removed. Hence, $\widehat{S} \in \mathscr{M}\mathscr{C}(\Sigma)$. Generalizing the usual formula for the classical Kauffman bracket, we get the following state sum formula as an immediate consequence of the definition:

$$[D]_\Sigma = \sum_{S \in \mathfrak{S}(D)} A^{a(S)-b(S)} \delta^{t(S)} \widehat{S} \in \mathbb{Z}[A^{\pm 1}]\mathscr{M}\mathscr{C}(\Sigma), \tag{3}$$

where $a(S)$ and $b(S)$ are the numbers of A– and B–smoothings in S and $\delta = -A^2 - A^{-2}$ as before. A similar formula appears in the paper of Dye and Kauffman on the surface bracket polynomial [21].

Any invariant of framed links in $\Sigma \times I$ satisfying (1) can be normalized to obtain a Jones-type polynomial invariant of oriented links. In the case of the skein bracket (2), one obtains the *generalized Jones polynomial*, an invariant for oriented links in $\Sigma \times I$ given by

$$J_\Sigma(D) = (-1)^{w(D)} t^{3w(D)/4} ([D]_\Sigma)_{A = t^{-1/4}}. \tag{4}$$

3 Adequate link diagrams in surfaces

Given a link diagram D, let S_A be the pure A state and let S_B be the pure B state. Then S_A and S_B are the states which theoretically give rise to the terms of maximal and minimal degree in (3). The notion of adequacy of a link diagram is designed to guarantee that the terms from S_A and S_B survive in the state sum formula. Therefore, when D is a skein adequate diagram, its skein bracket $[D]_\Sigma$ has maximal possible span.

Two states S and S' are said to be *adjacent* if their resolutions differ at exactly one crossing.

Definition 1 A link diagram D on a surface Σ is said to be *A–adequate* if $t(S) \le t(S_A)$ or $\widehat{S} \ne \widehat{S}_A$ in $\mathscr{MC}(\Sigma)$ for any state S adjacent to S_A. It is said to be *B–adequate* if $t(S) \le t(S_B)$ or $\widehat{S} \ne \widehat{S}_B$ for any state S adjacent to S_B. The diagram D is called *skein adequate* if it is both A– and B–adequate.

The notions of A– and B–adequacy are modeled on the notions of plus- and minus-adequacy for classical links [37]. Recall that a classical link diagram is said to be *plus-adequate* if $|S| = |S_A| - 1$ for any state S adjacent to S_A, and it is *minus-adequate* if $|S| = |S_B| - 1$ for any state S adjacent to S_B. This simpler notion of adequacy extends verbatim to link diagrams on surfaces. For link diagrams on surfaces, plus- and minus-adequacy is a special case of the notion of homological adequacy, which was introduced in [8] and will be reviewed in Section 4. We will see that adequacy as defined above is more general than simple or homological adequacy.

The following provides an alternative definition of adequacy:

Proposition 2 (1) *A link diagram D on Σ is A–adequate if and only if $t(S) \le t(S_A)$ or $|\widehat{S}| \ne |\widehat{S}_A|$ for any state S adjacent to S_A.*

(2) *A link diagram D on Σ is B–adequate if and only if $t(S) \le t(S_B)$ or $|\widehat{S}| \ne |\widehat{S}_B|$ for any state S adjacent to S_B.*

Proof We begin with some general comments. Given a link diagram D and two adjacent states S and S', the transition from S to S' is one of the following types:

(i) $|S'| = |S| + 1$, ie one cycle of S splits into two cycles of S'.

(ii) $|S'| = |S| - 1$, ie two cycles of S merge into one cycle of S'.

(iii) $|S'| = |S|$, ie one cycle C of S rearranges itself into a new cycle C' of S'.[1]

In cases (ii) and (iii), either $t(S') \le t(S)$ or $\widehat{S'} \ne \widehat{S}$. Specifically, in case (ii), $t(S') > t(S)$ only when two nontrivial parallel cycles in S merge to form one trivial cycle in S', which implies that $\widehat{S} \ne \widehat{S'}$. Likewise, in case (iii), we claim that neither C nor C' is

[1] The transition $S \to S'$ in this case is called a *single cycle bifurcation.*

trivial and, consequently, $t(S') = t(S)$. To see that, note that, if S' is obtained from S by a smoothing change of a crossing x, then there are two simple closed loops $\alpha, \beta \subset \Sigma$ intersecting at x only and such that the two different smoothings of x yield C and C'. Assigning some orientations to α and β, we see that C and C' with some orientations equal $\pm(\alpha+\beta)$ and $\pm(\alpha-\beta)$ in $H_1(\Sigma)$. Since the algebraic intersection number of α and β is 1, we know that $\alpha \neq \pm\beta$ and, consequently, neither C nor C' is trivial.

Therefore, to verify that a given diagram is A– or B–adequate, it is enough to check that the conditions of Definition 1 hold in case (i).

We will now prove part (1). Suppose S is a state adjacent to S_A with $t(S) = t(S_A)+1$. Then the transition from S_A to S must either be case (i) or (ii).

If it is case (i), then $|S| = |S_A|+1$ and $t(S) = t(S_A)+1$; therefore, $\widehat{S} = \widehat{S}_A$. Thus, D is not A–adequate and $|\widehat{S}| = |\widehat{S}_A|$. If it is case (ii), then $|S| = |S_A|-1$, and two nontrivial cycles of S_A must merge into a trivial cycle of S. In this case, the conditions for A–adequacy are satisfied and $|\widehat{S}| \neq |\widehat{S}_A|$.

The proof of part (2) is similar and is left to the reader. $\square$

For any diagram D, its bracket has a unique presentation

$$[D]_\Sigma = \sum_\mu p_\mu(D)\mu \in \mathscr{S}(\Sigma \times I),$$

where the sum is over all multiloops μ in Σ. Denote the maximal and minimal degrees (in the variable A) of the nonzero polynomials $p_\mu(D)$ in this expression by $d_{\max}([D]_\Sigma)$ and $d_{\min}([D]_\Sigma)$.

Proposition 3 *For any link diagram D on Σ,*

1. $d_{\max}([D]_\Sigma) \le c(D) + 2t(S_A)$, *with equality if D is A–adequate*;
2. $d_{\min}([D]_\Sigma) \ge -c(D) - 2t(S_B)$, *with equality if D is B–adequate.*

Proof (1) By (3), $[D]_\Sigma$ is given by a state sum with term $(-1)^{t(S_A)} A^{c(D)+2t(S_A)} \widehat{S}_A$ for the state S_A. Now the inequality of (1) follows from the fact that every change of a smoothing in S_A decreases $a(S) - b(S)$ by two and increases $t(S)$ by at most one.

The proof of equality in (1) when D is A–adequate follows immediately from part (1) of the lemma below.

The proof of (2) is analogous, and the proof of equality in (2) when D is B–adequate follows from part (2) of the lemma below. $\square$

Lemma 4 (1) *If D is A–adequate and S is a state with at least one B–smoothing, then either*

$$a(S)-b(S)+2t(S)<c(D)+2t(S_A) \quad \text{or} \quad \widehat{S}\neq\widehat{S}_A.$$

(2) *If D is B–adequate and S is a state with at least one A–smoothing, then either*

$$a(S)-b(S)+2t(S)>-c(D)-2t(S_B) \quad \text{or} \quad \widehat{S}\neq\widehat{S}_B.$$

Proof We prove (1) by contradiction: Suppose to the contrary that S is a state with at least one B–smoothing such that $\widehat{S}=\widehat{S}_A$ and

$$a(S)-b(S)+2t(S)=c(D)+2t(S_A).$$

Clearly, S can be obtained from S_A by a sequence of smoothing changes from A to B, $S_A=S_0\to S_1\to\cdots\to S_k=S$. Further, each smoothing change must increase $t(\cdot)$ by one, ie $t(S_{i+1})=t(S_i)+1$ for $i=0,\dots,k-1$. Since each smoothing change increases the number of cycles in a state by at most one, none of these smoothing changes can add a new cycle to $\widehat{S}_i$ for $i=0,\dots,k$. Therefore, $|\widehat{S}_{i+1}|\le|\widehat{S}_i|$ for $i=0,\dots,k-1$. However, since $\widehat{S}=\widehat{S}_A$, none of the smoothing changes can decrease $|\widehat{S}_i|$ either. It follows that $\widehat{S}_{i+1}=\widehat{S}_i$ for $i=0,\dots,k-1$. Thus, $|\widehat{S}_{i+1}|=|\widehat{S}_i|$ and

$$|S_{i+1}|=t(S_{i+1})+|\widehat{S}_{i+1}|=t(S_i)+1+|\widehat{S}_i|=|S_i|+1$$

for $i=0,\dots,k-1$. In particular, each transition $S_i\to S_{i+1}$ is of type (i), as discussed in the proof of Proposition 2, ie one where a cycle of S_i splits into two cycles of S_{i+1}.

However, since D is A–adequate, the first smoothing change $S_A=S_0\to S_1$ has either $t(S_1)\le t(S_A)$ or $\widehat{S}_1\neq\widehat{S}_A$, which is a contradiction.

This completes the proof of the first statement. The proof of the second one is similar and is left to the reader. □

The next result is an immediate consequence of Proposition 3. Below, $\operatorname{span}([D]_\Sigma)$ denotes the difference between the maximal and minimal A–degree of $[D]_\Sigma$.

Corollary 5 *If D is a link diagram on Σ, then*

$$\operatorname{span}([D]_\Sigma)\le 2c(D)+2t(S_A)+2t(S_B),$$

with equality if D is skein adequate.

The map $\Psi\colon\mathscr{MC}(\Sigma)\to\mathbb{Z}[z]$ sending S to $z^{|S|}$ extends linearly to the skein module,

$$\Psi\colon\mathscr{S}(\Sigma\times I)=\mathbb{Z}[A^{\pm1}]\mathscr{MC}(\Sigma)\to\mathbb{Z}[A^{\pm1},z].$$

The composition $\Psi([D]_\Sigma)$ is called the *reduced homotopy Kauffman bracket*. Obviously,

$$\operatorname{span}(\Psi([D]_\Sigma)) \le \operatorname{span}([D]_\Sigma),$$

where $\operatorname{span}(\cdot)$ refers to the span in the A–degree.

Proposition 6 *If D is a skein adequate link diagram on Σ, then*

$$\operatorname{span}(\Psi([D]_\Sigma)) = \operatorname{span}([D]_\Sigma).$$

Proof Let S be a state with at least one B–smoothing such that $|\widehat{S}| = |\widehat{S}_A|$ and $a(S) - b(S) + 2t(S) = c(D) + 2t(S_A)$. As before, S can be obtained from S_A by a sequence of smoothing changes from A to B, and each smoothing change can increase $t(\cdot)$ by at most one, ie $S_A = S_0 \to S_1 \to \cdots \to S_k = S$. As in the proof of Lemma 4, we must have $t(S_{i+1}) = t(S_i) + 1$. Further, since a smoothing change can increase the number of cycles in S_i by at most one, we have $|\widehat{S}_{i+1}| \le |\widehat{S}_i|$ for $i = 0, \dots, k-1$. Now the assumption that $|\widehat{S}| = |\widehat{S}_A|$ then implies that $|\widehat{S}_{i+1}| = |\widehat{S}_i|$ for $i = 0, \dots, k-1$. However, since D is adequate, for the first transition $S_A = S_0 \to S_1$, either $t(S_1) \ne t(S_0) + 1$ or $\widehat{S}_1 \ne \widehat{S}_0$. But $t(S_1) = t(S_0) + 1$ and $|\widehat{S}_1| = |\widehat{S}_0|$ imply that $\widehat{S}_1 = \widehat{S}_0$, which gives a contradiction.

Therefore, the term with maximum A–degree in $\Psi([D]_\Sigma)$ must survive. A similar argument applies to show that the term with minimum A–degree survives. It follows that

$$\operatorname{span}(\Psi([D]_\Sigma)) = 2c(D) + 2t(S_A) + 2t(S_B) = \operatorname{span}([D]_\Sigma). \qquad \square$$

The next proposition shows that skein adequacy is inherited under passing to subsurfaces $\Sigma' \subset \Sigma$.

Proposition 7 *If a link diagram D on a subsurface Σ' of Σ is A– or B–adequate in Σ, then it is A– or B–adequate, respectively, in Σ'.*

Proof In the following, let $t(S, \Sigma)$ be the value of $t(S)$ when S is regarded as a state in Σ, and let $t(S, \Sigma')$ be its value when S is regarded as a state in Σ'.

Suppose D is not A–adequate in Σ'. By Proposition 2, there exists a state S adjacent to S_A with $t(S, \Sigma') = t(S_A, \Sigma') + 1$ and $|\widehat{S}| = |\widehat{S}_A|$ in Σ'. In particular, $|S| = |S_A| + 1$, and the transition from S_A to S must involve one cycle C of S_A splitting into two cycles C_1 and C_2 of S. At least one of C_1 and C_2 must be trivial in Σ', for otherwise $t(S, \Sigma') \le t(S_A, \Sigma')$. If, say, C_1 is trivial in Σ', then it must also be trivial in Σ, because $\Sigma' \subset \Sigma$ is a subsurface.

As a cycle in Σ, C is either trivial or nontrivial. If it is trivial, then C_2 must also be trivial in Σ, and so in fact all three of C, C_1 and C_2 are trivial. This implies that $t(S,\Sigma) = t(S_A,\Sigma)+1$ and $|\widehat{S}| = |\widehat{S}_A|$ in Σ, contradicting the assumption of A–adequacy of D.

If, on the other hand, C is nontrivial in Σ, then C_2 must also be nontrivial in Σ. This again implies that $t(S,\Sigma) = t(S_A,\Sigma)+1$ and $|\widehat{S}| = |\widehat{S}_A|$ in Σ, leading to the same contradiction. Therefore, D must be A–adequate on Σ'.

The proof of B–adequacy of D is identical. □

4 Skein and homological adequacy

For completeness of discussion, in this section we compare Definition 1 of skein adequacy to two legacy versions, namely simple and homological adequacy. We will see that our notion of adequacy is broader and that the statements of Lemma 4 and Corollary 5 are strictly stronger than the corresponding statements for simple and homological adequacy. Henceforth, we will say a link diagram on a surface is adequate if it is skein adequate.

For any state $S \subset \Sigma$, let us denote the ranks of the kernel and the image of

$$i_*\colon H_1(S;\mathbb{Z}/2) \to H_1(\Sigma;\mathbb{Z}/2),$$

by $k(S)$ and $r(S)$, respectively.

The *homological Kauffman bracket*,

$$\langle D\rangle_\Sigma = \sum_{S\in\mathfrak{S}(D)} A^{a(S)-b(S)}\delta^{k(S)}z^{r(S)},$$

was introduced by Krushkal [32] and studied in [8].

Based on this invariant, Boden and Karimi [8] introduced the notion of homological adequacy for link diagrams in surfaces. A diagram D on Σ is *homologically A–adequate* if $k(S) \le k(S_A)$ for any state S adjacent to S_A, and it is *homologically B–adequate* if $k(S) \le k(S_B)$ for any state S adjacent to S_B. A diagram D is *homologically adequate* if it is both homologically A– and B–adequate.

It is not difficult to show that a diagram that is plus-adequate is homologically A–adequate, and one that is minus-adequate is homologically B–adequate. (For further details, see [8, Section 2.2].)

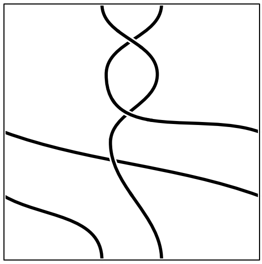


Figure 1: An alternating diagram on the torus.

Proposition 8 *Every homologically A–adequate link diagram is A–adequate and every homologically B–adequate link diagram is B–adequate.*

Proof Recall from the discussion at the beginning of the proof of Proposition 2 that there are the three possible cases and, to verify that a given diagram is A– or B–adequate, it is enough to check that the conditions of Definition 1 hold in case (i). Hence, it is enough to focus on states S adjacent to S_A or S_B with $|S| = |S_A| + 1$ or $|S| = |S_B| + 1$, respectively.

If D is not A–adequate, then there exists a state S adjacent to S_A with $|S| = |S_A| + 1$, $t(S) = t(S_A) + 1$ and $\widehat{S} = \widehat{S}_A$. (Notice that if $|S| = |S_A| + 1$ and $t(S) = t(S_A) + 1$, then $\widehat{S} = \widehat{S}_A$ automatically holds.) In this case, we have $k(S) = k(S_A) + 1$, and it follows that D is not homologically A–adequate. This proves the first statement in the proposition, and the proof of the second statement on B–adequacy is similar. □

In summary, then, for a link diagram D on a surface Σ, it follows that

$$\text{plus-adequacy} \implies \text{homological } A\text{–adequacy} \implies A\text{–adequacy}, \tag{5}$$

with similar statements relating minus-adequacy, homological B–adequacy, and B–adequacy.

In Example 20, we will see a knot diagram in a genus two surface which is adequate but not homologically adequate. On the other hand, it is easy to construct examples which are homologically adequate but not simply adequate. For instance, consider the alternating diagram D with three crossings on the torus in Figure 1. A straightforward calculation shows that it is homologically adequate but not simply adequate. These examples show that none of the reverse implications in (5) hold; therefore, the notion of adequacy in Definition 1 is strictly more general than either homological or simple adequacy.

In general, notice that

$$\operatorname{span}(\langle D\rangle_\Sigma) \le \operatorname{span}([D]_\Sigma) \le 2c(D)+2t(S_A)+2t(S_B) \le 2c(D)+2k(S_A)+2k(S_B),$$

where $\operatorname{span}(\,\cdot\,)$ is the span in the A–degree. Therefore, Corollary 5 immediately implies an analogous inequality holds for homological adequacy; see [8, Corollary 2.7].

5 Alternating links and the Tait conjectures

When tabulating knots, Tait formulated three conjectures on alternating links. The first one states that any reduced alternating diagram of a classical link has minimal crossing number. The second one asserts that any two such diagrams representing the same link have the same writhe. The third one states that any two reduced alternating diagrams of the same link are related by flype moves. The first two conjectures were resolved almost 100 years later, independently by Kauffman [31], Murasugi [43] and Thistlethwaite [48], using the newly discovered Jones polynomial. The third conjecture was established shortly after by Menasco and Thistlethwaite [41]. The first two Tait conjectures actually hold more generally for adequate links [38], and their proofs have been generalized to homologically adequate links in thickened surfaces in [8]. Here, we generalize these results even further to adequate links in thickened surfaces.

Henceforth, all links in thickened surfaces will be unframed, unless stated otherwise. Given an oriented link diagram D, let $c_+(D)$ be the numbers of crossings of type , and let $c_-(D)$ be the number of crossings of type . The proof of the following theorem can be found in Section 7.1:

Theorem 9 *Let D and E be oriented link diagrams on Σ representing the same oriented unframed link in $\Sigma\times I$.*

(i) *If D is A–adequate, then $c_-(D)\le c_-(E)$.*

(ii) *If D is B–adequate, then $c_+(D)\le c_+(E)$.*

The *crossing number* $c(L)$ of a link $L\subset\Sigma\times I$ is defined as the minimal crossing number among all diagram representatives of L. A link $L\subset\Sigma\times I$ is said to be *adequate* if it admits an adequate diagram on Σ.

Using Theorem 9, one can deduce the first and second Tait conjectures for adequate links in surfaces.

Corollary 10 (i) *Any adequate diagram of a link in $\Sigma \times I$ has $c(L)$ crossings.*

(ii) *Any two adequate diagrams of the same oriented link in $\Sigma \times I$ have the same writhe.*

Proof Statements (i) and (ii) are immediate consequences of Theorem 9. In the case of (ii), if adequate diagrams D and E represent the same oriented link, then $c_+(D) = c_+(E)$ and $c_-(D) = c_-(E)$ by the above theorem and, hence,

$$w(D) = c_+(D) - c_-(D) = c_+(E) - c_-(E) = w(E). \qquad \square$$

Corollary 10 implies that for an adequate link $L \subset \Sigma \times I$, the writhe is a well-defined invariant of its oriented link type.

Let $g(\Sigma)$ be the sum of the genera of the connected components of Σ. A link diagram D on Σ is *minimally embedded* if it does not lie on a subsurface of Σ of smaller genus. In other words, the complement of D on Σ has no nonseparating loops. Let N_D be a neighborhood of D in Σ small enough that it is a ribbon surface retractable onto D. A diagram D is minimally embedded if and only if $g(N_D) = g(\Sigma)$.

Furthermore, note that, if D is connected and Σ is closed, then D is minimally embedded if and only if $\Sigma \smallsetminus D$ is composed of disks. In that case, we say that D is *cellularly embedded*.

A link diagram D on a closed surface Σ is said to have *minimal genus* if it is minimally embedded within its isotopy class.

In [39], it is proved that any cellularly embedded knot diagram with minimal crossing number has minimal genus. This result was recently extended to link diagrams, and the following is a restatement of [10, Theorem 1]:

Theorem 11 *Any cellularly embedded link diagram with minimal crossing number has minimal genus.*

A link diagram D on Σ is *alternating* if, when traveling along any of its components, its crossings alternate between over and under. A link $L \subset \Sigma \times I$ is *alternating* if it can be represented by an alternating link diagram.

A crossing x of D is *nugatory* if there is a simple loop in Σ which separates Σ and intersects D only at x.

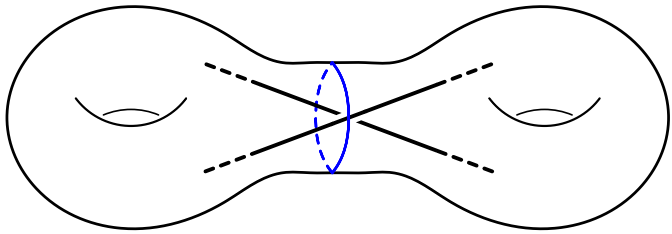


Figure 2: An essential nugatory crossing.

As observed in [8], although nugatory crossings in diagrams in $\Sigma = \mathbb{R}^2$ can always be removed by rotating one side of the diagram 180° relative to the other, this is not always true for diagrams in noncontractible surfaces Σ; see Figure 2. A nugatory crossing is said to be *removable* if the simple loop can be chosen to bound a disk, otherwise it is called *essential*. A link diagram is *reduced* if it does not contain any removable nugatory crossings. For example, the knot in Figure 6 contains an essential nugatory crossing.

The following strengthens [8, Proposition 2.8]. Its proof is given in Section 7.2.

Theorem 12 *Any reduced alternating diagram is adequate.*

Note that, unlike [8, Proposition 2.8], we do not assume here that D is cellularly embedded or checkerboard colorable, nor that D has no nugatory crossings.

A link diagram on Σ is called *weakly alternating* if it is a connected sum $D_0 \# D_1 \# \cdots \# D_k$ of an alternating diagram D_0 in Σ and alternating diagrams $D_1, \dots, D_k$ in S^2 (see Lemma 16). Theorem 12 can be generalized to show that weakly alternating diagrams are adequate. In fact, in the next section we will prove Proposition 17, showing that any diagram on a surface obtained as the connected sum of two adequate link diagrams is itself adequate.

Let us return to Tait conjectures now. By Corollary 10, any reduced alternating diagram D has the minimal crossing number for all diagrams representing the same unframed link L in $\Sigma \times I$. Furthermore, all such oriented diagrams representing the same link L have the same writhe.

The results of Kauffman, Murasugi and Thistlethwaite [31; 43; 48] imply that the span of the Kauffman bracket of any diagram $D \subset S^2$ satisfies

$$\operatorname{span}([D]_{S^2}) \le 4c(D) + 4,$$

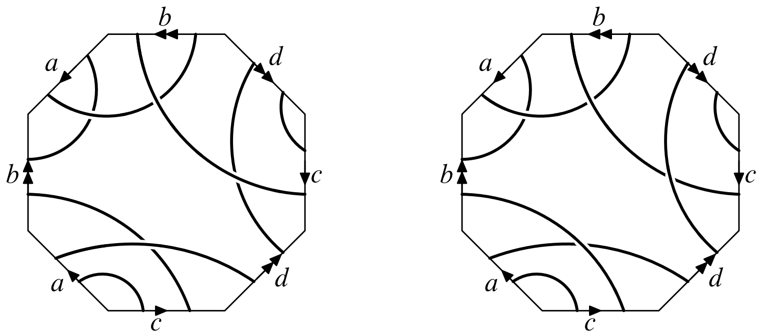


Figure 3: Two knots in a genus two surface with the same homological Kauffman bracket.

or, equivalently, for the Jones polynomial, that $\operatorname{span}(V_D(t)) \le c(D)$, with equality if D is alternating. Furthermore, in [48], Thistlethwaite proved that, if $D \subset S^2$ is prime and nonalternating, then

$$\operatorname{span}([D]_{S^2}) < 4c(D) + 4.$$

In [49], it is observed that the above results hold if $D \subset S^2$ is weakly alternating, namely if D is a connected sum of alternating diagrams. Thus, the Kauffman bracket $[D]_{S^2}$, together with $c(D)$, *detects* weakly alternating classical links.

The homological Kauffman bracket of [8] is not sufficiently strong to prove an analogous statement for links in thickened surfaces. Consider the two knots in the genus two surface in Figure 3. These knots have the same homological Kauffman bracket, namely

$$\langle D_1 \rangle_\Sigma = \langle D_2 \rangle_\Sigma = 3\delta z^2 - 4\delta^2 z + (A^4 + 3 + A^{-4})\delta,$$

but one of them is alternating and the other is not. Consequently, the homological Kauffman bracket does not detect alternating knots in thickened surfaces.

However, we are going to show that the Kauffman, Murasugi and Thistlethwaite statements hold for the Kauffman bracket $[\cdot]_\Sigma$ of diagrams in closed surfaces Σ after replacing 4 by $4|D| - 4g(\Sigma)$ on the right.

Let $|D|$ denote the number of connected components of D (which may be smaller than the number of connected components of the link in $\Sigma \times I$ represented by D).

Let $r(D)$ be the rank of the image of $i_*\colon H_1(D; \mathbb{Z}/2) \to H_1(\Sigma; \mathbb{Z}/2)$. If $D \subset \Sigma$ is minimally embedded, then i_* is surjective and $r(D) = 2g$.

The proof of the next result is given in Section 7.4.

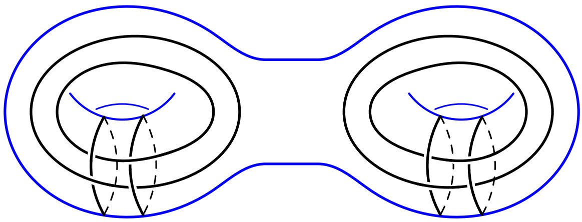


Figure 4: Minimally embedded alternating diagram for which the equality of Theorem 13(ii) does not hold.

Theorem 13 (i) *For any link diagram $D \subset \Sigma$,*

$$\operatorname{span}([D]_\Sigma) \le 4c(D) + 4|D| - 2r(D).$$

(ii) *If D is cellularly embedded, reduced, and weakly alternating, then*

$$\operatorname{span}([D]_\Sigma) = 4c(D) + 4|D| - 4g(\Sigma).$$

(iii) *If D is not weakly alternating then*

$$\operatorname{span}([D]_\Sigma) < 4c(D) + 4|D| - 2r(D).$$

The assumptions of Theorem 13(ii) are necessary:

If D has a removable nugatory crossing, then eliminating it decreases the right-hand side of the above equality but not the left-hand side. Therefore, (ii) does not hold for diagrams with removable crossings.

It can also fail when D is not cellularly embedded. For example, consider the alternating link in Figure 4. It has $t(S_A) = 4$ and $t(S_B) = 2$. Therefore, by Corollary 5, we have $\operatorname{span}([D]_\Sigma) \le 16 + 12 = 28$, whereas $4c(D) + 4|D| - 4g(\Sigma) = 32$. Note that this diagram is minimally embedded but not cellularly embedded.

Although (ii) holds for weakly alternating diagrams, in the next section we will see that it does not hold generally for connected sums of alternating diagrams in arbitrary surfaces (see Example 19).

Corollary 14 *Let L be a link in $\Sigma \times I$ with a reduced, weakly alternating diagram D which is cellularly embedded. Then any other cellularly embedded diagram E for L satisfies $c(D) \le c(E)$. If E is not weakly alternating, then $c(D) < c(E)$.*

Proof The first part is a direct consequence of Tait conjecture, Corollary 10. Let us prove the full statement now: Any cellularly embedded link diagram on a connected surface is itself connected. Therefore, it is enough to prove the statement under the assumption that Σ and D are both connected. Theorem 13(ii) then implies that $c(D) = \frac{1}{4}\operatorname{span}([D]_\Sigma) + g(\Sigma) - 1$. If E is a second link diagram for L on Σ, then, since E is cellularly embedded, it must also be connected. Theorem 13(i) implies that

$$c(D) = \tfrac{1}{4}\operatorname{span}([D]_\Sigma) + g(\Sigma) - 1 = \tfrac{1}{4}\operatorname{span}([E]_\Sigma) + g(\Sigma) - 1 \le c(E).$$

If E is not weakly alternating, then Theorem 13(iii) shows the last inequality is strict; therefore, it follows that $c(D) < c(E)$. $\square$

Remark 15 The corollary gives an alternative proof of Theorem 11 for nonsplit alternating links as follows. Let L be a nonsplit alternating link in $\Sigma \times I$, where Σ is closed oriented surface, and let $D \subset \Sigma$ a minimal crossing cellularly embedded diagram for L. Then Corollary 14 implies that D is an alternating diagram. The argument is completed by appealing to [9, Proposition 6], which shows that alternating link diagrams have minimal genus.

6 Crossing number and connected sums

In this section, we will study the behavior of the crossing number under connected sum of links in thickened surfaces. This problem is closely related to an old and famous conjecture for classical links, which asserts that, for any two links L_1 and L_2,

$$c(L_1 \# L_2) = c(L_1) + c(L_2). \tag{6}$$

This conjecture has been verified for a wide class of links, including alternating links, adequate links, and torus links [20]. Clearly, $c(L_1 \# L_2) \le c(L_1) + c(L_2)$. In addition, in [35], Lackenby has proved that, in general, one has a lower bound of the form

$$c(L_1 \# L_2) \ge \tfrac{1}{152}(c(L_1) + c(L_2)).$$

The operation of connected sum is not so well behaved for arbitrary links in thickened surfaces.

Just as for classical links, it depends on the choice of components which are joined as well as their orientations. However, unless one of the links is in $S^2 \times I$, it also depends on the diagram representatives as well as the choice of basepoints $x_i \in D_i$ where the link components are joined. The issue is the fact that a Reidemeister move applied to either of the link diagrams may change the link type of their connected sum. We take a moment to quickly review its construction.

Suppose Σ_1 and Σ_2 are oriented surfaces and let $\Sigma_1 \# \Sigma_2$ denote their connected sum. It is obtained from the union $(\Sigma_1 \smallsetminus \operatorname{int} B_1) \cup (\Sigma_2 \smallsetminus \operatorname{int} B_2)$ by gluing $\partial B_1 \subset \Sigma_1$ to $\partial B_2 \subset \Sigma_2$ by an orientation-reversing homeomorphism $g\colon \partial B_1 \to \partial B_2$. For connected surfaces, $\Sigma_1 \# \Sigma_2$ is independent of the choice of disks $B_i \subset \Sigma_i$ and gluing map.

If $D_1 \subset \Sigma_1$ and $D_2 \subset \Sigma_2$ are link diagrams, we can choose cutting points $x_i \in D_i$ and disk neighborhoods B_i from Σ_i such that $B_i \cap D_i$ is an interval for $i = 1, 2$. Then the surface $\Sigma_1 \# \Sigma_2$ can be formed in such a way that $D = (D_1 \smallsetminus \operatorname{int} B_1) \cup (D_2 \smallsetminus \operatorname{int} B_2)$ is a link diagram in $\Sigma_1 \# \Sigma_2$. If D_1 and D_2 are oriented link diagrams, then we require the gluing to respect the orientations of the arcs. The resulting diagram is called a *connected sum* of D_1 and D_2. In general, it depends on the choice of link diagrams D_1 and D_2, components being joined, and points $x_i \in D_i$. However, it is independent of the choice of disk neighborhoods B_i containing x_i.

The next result shows that, when one of the diagrams lies in $S^2 \times I$, the operation of connected sum is well behaved.

Lemma 16 *Let $D_1 \subset \Sigma \times I$ and $D_2 \subset S^2 \times I$ be oriented diagrams, where Σ is an arbitrary surface. Then the connected sum of D_1 and D_2 is independent of the choice of the cutting points x_1 and x_2 on the selected components of D_1 and of D_2.*

We will denote the connected sum in this case by $D_1 \# D_2$. The oriented link type of $D_1 \# D_2$ depends only on the link types of D_1 and D_2 and a choice of which components are joined.

Proof One can shrink the image of D_2 in the connected sum so that all its crossings lie in a small 3–ball B^3 in $\Sigma \times I$. By an isotopy, we can move the ball along arcs of D_1 representing the component to which D_1 is joined, and moving over or under the other arcs at any crossing that we encounter.

This shows that the connected sum is independent of the choice of the cut point x_1 on D_1. The independence on the cut point x_2 on D_2 follows from the well-known fact that all long knots — or rather $(1,1)$ tangles — obtained by cutting D_2 at different points x_2 of its specified component are isotopic (as $(1,1)$ tangles). Shrinking D_2 into a small 3–ball also allows one to translate any Reidemeister move of D_1 or D_2 into a Reidemeister move on the connected sum $D_1 \# D_2$. This proves the last statement. □

Proposition 17 *Any connected sum of two A– or B–adequate diagrams is itself A– or B–adequate, respectively.*

Proof Let D be a link diagram in $\Sigma_1 \# \Sigma_2$ obtained as the connected sum of A–adequate diagrams $D_1 \subset \Sigma_1$ and $D_2 \subset \Sigma_2$, and suppose to the contrary that D is not A–adequate. By Proposition 2, there is a state S for D adjacent to S_A with $t(S, \Sigma_1 \# \Sigma_2) = t(S_A, \Sigma_1 \# \Sigma_2) + 1$ and $|\widehat{S}| = |\widehat{S}_A|$ in $\Sigma_1 \# \Sigma_2$. In particular, $|S| = |S_A| + 1$, and the transition from S_A to S involves one cycle of S_A splitting into two cycles.

Let x be the crossing of D where the smoothing is changed in the transition from S_A to S. We can assume, without loss of generality, that x is a crossing from D_1. Let C be the cycle of S_A that splits into two cycles, C' and C'', under this transition. Since $t(S, \Sigma_1 \# \Sigma_2) = t(S_A, \Sigma_1 \# \Sigma_2) + 1$, one of the cycles C' and C'', say C', must be trivial.

If C is a cycle contained in $S_A(D_1)$, then the same is true for C' and C''. However, this contradicts the assumption that D_1 is A–adequate.

Otherwise, $C = C_1 \# C_2$ must be a connected sum of a cycle C_1 in $S_A(D_1)$ with a cycle C_2 in $S_A(D_2)$. In the transition from S_A to S, by the previous argument, we may assume the cycle $C_1 \# C_2$ splits into $C_1' \# C_2$ and C_1''. Further, since $C' = C_1' \# C_2$ is trivial, it follows that C_1' must be trivial in Σ_1 and C_2 must be trivial in Σ_2.

If $C_1 \# C_2$ is trivial, then $C_1'' \# C_2$ must also be trivial. That would imply that all three of C_1, C_1' and C_1'' are trivial in Σ_1. This again contradicts the assumption that D_1 is A–adequate, and we take a moment to explain this point.

Let $S(D_1)$ be the corresponding state for D_1. It is obtained from $S_A(D_1)$ by switching the smoothing at x. The transition from $S_A(D_1)$ to $S(D_1)$ involves C_1 splitting into C_1' and C_1''. Since all three of C_1, C_1' and C_1'' are trivial in Σ_1, we have $t(S(D_1)) = t(S_A(D_1)) + 1$ and $|\widehat{S}(D_1)| = \widehat{S}_A(D_1)$ in Σ_1, which contradicts the assumption of A–adequacy of D_1.

The other possibility is that $C_1 \# C_2$ is nontrivial. Since C_2 is trivial in Σ_2, the cycles C_1 and C_1'' must both be nontrivial in Σ_1. The transition from $S_A(D_1)$ to $S(D_1)$ still involves C_1 splitting into C_1' and C_1'', only now C_1 and C_1'' are nontrivial and C_1' is trivial in Σ_1. Thus, $t(S(D_1)) = t(S_A(D_1)) + 1$ and $|\widehat{S}(D_1)| = \widehat{S}_A(D_1)|$ in Σ_1, which again contradicts the assumption of A–adequacy of D_1. Therefore, $D = D_1 \# D_2$ must be A–adequate.

The proof of B–adequacy of D is similar. □

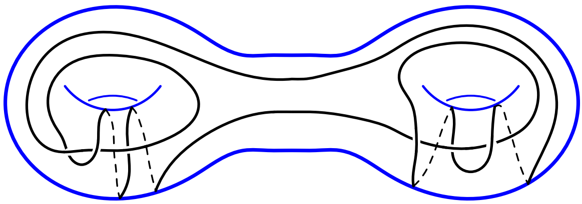


Figure 5: A connected sum of alternating diagrams.

Corollary 18 *Suppose $L_1 \subset \Sigma_1 \times I$ and $L_2 \subset \Sigma_2 \times I$ are links represented by adequate diagrams $D_1 \subset \Sigma_1$ and $D_2 \subset \Sigma_2$. Then any link L in $(\Sigma_1 \# \Sigma_2) \times I$ admitting a diagram which is a connected sum of D_1 and D_2 is itself adequate. Further, the crossing number and writhe satisfy $c(L) = c(L_1) + c(L_2)$ and $w(L) = w(L_1) + w(L_2)$.*

Proof Suppose L is represented by $D = D_1 \# D_2 \subset \Sigma_1 \# \Sigma_2$. Then D is adequate by Proposition 17. Further, by parts (i) and (ii) of Corollary 10, we see that

$$c(L) = c(D) = c(D_1) + c(D_2) = c(L_1) + c(L_2),$$
$$w(L) = w(D) = w(D_1) + w(D_2) = w(L_1) + w(L_2). \qquad \square$$

Example 19 Figure 5 shows a knot diagram D in the genus two surface obtained as the connected sum of two alternating diagrams of the same knot in the torus. One can easily verify that D is reduced and cellularly embedded, but not alternating. Further, Proposition 17 implies that this diagram is adequate, and therefore a minimal crossing diagram for the knot type. Direct calculation reveals that $t(S_A) = 2$, $t(S_B) = 0$ and $|\widehat{S}_A| = |\widehat{S}_B| = 1$. Therefore, $\operatorname{span}([D]_\Sigma) = 16$. On the other hand, since $4(c(D) + |D| - g(\Sigma)) = 20$, by Theorem 13(ii), it follows that D is not weakly alternating and, in fact, not equivalent to any weakly alternating knot in $\Sigma \times I$.

Example 20 Figure 6 shows a knot in a genus two surface with an essential nugatory crossing. Since it is reduced and alternating, Theorem 12 shows that it is adequate.

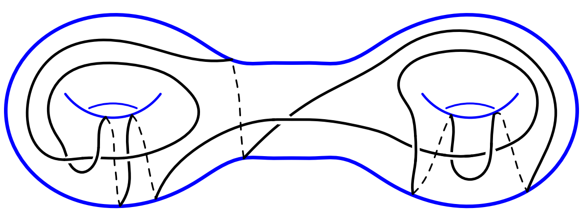


Figure 6: An alternating diagram with an essential nugatory crossing.

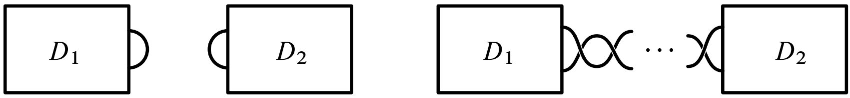


Figure 7: Adding twists to a connected sum to create essential nugatory crossings.

Note that this diagram is not homologically adequate. In fact, if S is the state with a B–smoothing at the nugatory crossing and A–smoothings at all the other crossings, then one can show that $|S| = |S_A| + 1$ and $k(S) > k(S_A)$.

Notice that this knot can also be obtained as the connected sum of two alternating knots K_1 and K_2 in $T^2 \times I$ with $c(K_i) = 3$, but after performing a Reidemeister one move on one of them to obtain a diagram with four crossings. In particular, this example shows that a connected sum of two diagrams $D_1 \subset \Sigma_1$ and $D_2 \subset \Sigma_2$ can be adequate even when one of them is not adequate.

Suppose $L_1 \subset \Sigma_1 \times I$ and $L_2 \subset \Sigma_2 \times I$ are two alternating links in thickened surfaces with $g(\Sigma_i) > 0$ for $i = 1, 2$. Suppose further that D_i is a link diagram on Σ_i representing L_i for $i = 1, 2$, and that D_1 and D_2 are both reduced and alternating.

Instead of forming the connected sum of D_1 and D_2, take one of the diagrams and insert an arbitrary number (say n) of twists before forming the connected sum. See Figure 7.

The result will be a diagram D which is similar to a connected sum of D_1 and D_2, but with n essential nugatory crossings in between. This construction can be carried out so that D is reduced and alternating. In particular, it will have crossing number $c(D) = c(D_1) + c(D_2) + n$. If L denotes the link type of D, and since D_1 and D_2 are alternating and have minimal crossing number, this shows that the analogue of (6) can fail arbitrarily badly for links in thickened surfaces other than $S^2 \times I$.

The reason (6) fails in general for connected sums of links in thickened surfaces is due to the use of nonminimal diagrams in forming the connected sum. However, if one restricts the connected sum operation to minimal crossing diagrams, then one gets a plausible generalization:

Conjecture 21 *Suppose $L_1 \subset \Sigma_1 \times I$ and $L_2 \subset \Sigma_2 \times I$ are links in thickened surfaces with minimal crossing representatives D_1 and D_2, respectively. Then any link L in the thickening of $\Sigma_1 \# \Sigma_2$ arising as a connected sum of D_1 and D_2 satisfies*

$$c(L) = c(L_1) + c(L_2).$$

Note that the assumption that D_1 and D_2 are minimal crossing representatives implies immediately that

$$c(L) \le c(L_1) + c(L_2).$$

In fact, the inequality may fail without that assumption. This is related to the fact that crossing number is not additive under connected sum for virtual knots. For example, the Kishino knot is the connected sum of two virtual unknots. As evidence, notice that Corollary 18 confirms that the conjecture is true if L_1 and L_2 are adequate links in thickened surfaces. In particular, it holds for alternating and weakly alternating links.

7 Proofs of Theorems 9, 12 and 13

7.1 Proof of Theorem 9

Given a link diagram D on Σ and positive integer r, the r^{th} *parallel* of D is the link diagram D^r on Σ in which each link component of D is replaced by r parallel copies, with each one repeating the same "over" and "under" behavior of the original component.

Lemma 22 *If D is A–adequate, then D^r is also A–adequate. If D is B–adequate, then D^r is also B–adequate.*

Proof Let $S_A(D)$ and $S_A(D^r)$ be the pure A–smoothings of D and the pure A–smoothings of D^r, respectively. It is straightforward to check that $S_A(D^r)$ is the r–parallel of $S_A(D)$.

Suppose D^r is not A–adequate. Then there is a state S' obtained by switching one A–smoothing in $S_A(D^r)$ to a B–smoothing such that $t(S_A(D^r)) < t(S')$ and $\widehat{S}_A(D^r) = \widehat{S}'$. In the terminology of the proof of Proposition 2, that can only happen for a smoothing change of type (i), more specifically when the smoothing change involves one of the innermost cycles in $S_A(D^r)$ which is self-abutting and which, when split, creates a new trivial cycle in S'. That is only possible if there is a self-abutting cycle in $S_A(D)$ which, when split, creates a new trivial cycle. Since D is A–adequate, this cannot happen.

An analogous argument proves the statement for B–adequate diagrams. □

Proof of Theorem 9 (i) Since

$$c(D) - w(D) = c_+(D) + c_-(D) - (c_+(D) - c_-(D)) = 2c_-(D),$$

we will prove that

$$c(D) - w(D) \le c(E) - w(E).$$

Our argument is an adaptation of Stong's proof [47] (see also [37, Theorem 5.13]).

Let $L_1,\dots,L_m$ be the components of L and let D_i and E_i be the subdiagrams of D and E corresponding to L_i. For each $i=1,\dots,m$, choose nonnegative integers μ_i and ν_i such that $w(D_i)+\mu_i = w(E_i)+\nu_i$. Let D' be composed of components $D'_1,\dots,D'_m$, where each D'_i is obtained from D_i by adding μ_i positive kinks to it. (These kinks do not cross with other components). Similarly, let E' be composed of components $E'_1,\dots,E'_m$, where each E'_i is obtained from E_i by adding ν_i positive kinks to it. Notice that D' is still A–adequate.

The writhes of the individual components satisfy

$$w(D'_i) = w(D_i)+\mu_i = w(E_i)+\nu_i = w(E'_i).$$

Further, the sum of the signs of the crossings of $D'_i \cap D'_j$ coincides with the sum of the signs of the crossings of $E'_i \cap E'_j$, since both are equal to the linking number of L_i and L_j. Thus, $w(D') = w(E')$.

For any r, consider the r^{th} parallels $(D')^r$ and $(E')^r$ now. Then $w((D')^r) = r^2 w(D')$, because each crossing of D' corresponds to r^2 crossings in $(D')^r$ of the same sign. The diagrams $(D')^r$ and $(E')^r$ are equivalent and have the same writhe; thus, their Kauffman brackets must be equal. In particular, we have $d_{\max}\big([(D')^r]_\Sigma\big) = d_{\max}\big([(E')^r]_\Sigma\big)$. Proposition 3 implies now that

$$d_{\max}\big([(D')^r]_\Sigma\big) = \Big(c(D)+\sum_{i=1}^m \mu_i\Big)r^2 + 2\Big(t(S_A(D))+\sum_{i=1}^m \mu_i\Big)r,$$

$$d_{\max}\big([(E')^r]_\Sigma\big) \le \Big(c(E)+\sum_{i=1}^m \nu_i\Big)r^2 + 2\Big(t(S_A(E))+\sum_{i=1}^m \nu_i\Big)r.$$

Since this is true for all r, by comparing coefficients of the r^2 terms, we find that

$$c(D)+\sum_{i=1}^m \mu_i \le c(E)+\sum_{i=1}^m \nu_i. \tag{7}$$

Subtracting $\sum_{i=1}^m(\mu_i + w(D_i)) = \sum_{i=1}^m(\nu_i + w(E_i))$ from both sides of (7), we get that

$$c(D)-\sum_{i=1}^m w(D_i) \le c(E)-\sum_{i=1}^m w(E_i). \tag{8}$$

Subtracting the total linking number of L from both sides of (8) gives the desired inequality.

The proof of (ii) is analogous. One adds negative kinks to D and E in this case. □

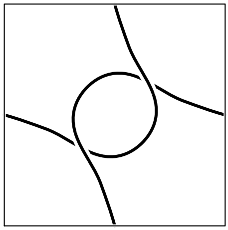


Figure 8: A knot diagram in the torus which is not alternable.

7.2 Proof of Theorem 12

A link diagram D on Σ is *alternable* if it can be made alternating by inverting some of its crossings. Every classical link diagram is alternable, but the same is not true for link diagrams in arbitrary surfaces. For example, the knot diagram in the torus in Figure 8 is not alternable.

A link diagram D on Σ is *checkerboard colorable* if the components of $\Sigma \smallsetminus D$ can be colored by two colors such that any two components of $\Sigma \smallsetminus D$ that share an edge have opposite colors.

Proposition 23 *Any minimal embedding D on Σ is alternable if and only if it is checkerboard colorable.*

Proof Observe that filling the boundaries of Σ with disks does not affect alternability or checkerboard colorability. Likewise, removing disks from $\Sigma \smallsetminus D$ also does not affect alternability or checkerboard colorability. This has two consequences:

(a) It is enough to prove this statement for surfaces Σ with all boundary components capped, ie for closed surfaces.

(b) Since Kamada proved that, if a diagram D is a deformation retract of Σ, then it is alternable if and only if it is checkerboard colorable [28, Lemma 7], our statement holds for cellularly embedded diagrams.

Our strategy is to reduce the proof to this case of cellular embeddings. Suppose that C is a nondisk component of $\Sigma \smallsetminus D$. Then it contains a noncontractible simple closed loop α. Let Σ' be obtained by cutting Σ along α and by capping the boundary components. The loop α must be separating Σ, since otherwise $D \hookrightarrow \Sigma'$ would be a lower-genus embedding of D. Observe now that, since Σ is a connected sum of two surfaces $\Sigma_1 \# \Sigma_2$, where $\Sigma_1 \cup \Sigma_2 = \Sigma'$ and D is a disjoint union of $D \cap \Sigma_1$ and of $D \cap \Sigma_2$, it is enough to prove that $D \subset \Sigma_i$ is checkerboard colored for $i = 1, 2$.

By repeating this process as long as possible, we reduce the statement to cellularly embedded diagrams, which is covered by (b) above. □

Lemma 24 *Any alternable diagram can be extended by disjoint simple closed loops to a checkerboard colorable one.*

Proof The surface $N_D \subset \Sigma$, being a regular neighborhood of D, is checkerboard colorable by the earlier mentioned result of Kamada [28, Lemma 7]. The only reason that coloring does not extend to $D \subset \Sigma$ is that some connected components C of $\Sigma \smallsetminus \operatorname{int} N_D$ may have multiple connected components of their boundary whose neighborhoods are colored differently. However, that issue can be resolved by adding simple closed loops around those boundary components of C which are white. □

Proof of Theorem 12 Let D be alternating diagram without removable crossings. By Lemma 24, by adding disjoint simple closed loops to D, we obtain a diagram D' which is alternating and checkerboard colorable. Hence, it is enough to prove that D' is adequate. Let us assume for simplicity of notation that D is checkerboard colorable.

We will prove the A–adequacy of D only, as the proof of B–adequacy is identical. Let S be a state with all A–smoothings except for a B–smoothing at a crossing x of D. We will prove that D is A–adequate "at x", meaning that $t(S) \le t(S_A)$ or $\widehat{S} \neq \widehat{S}_A$ in $\mathscr{S}(\Sigma \times I)$.

As in the proof of Proposition 2, there are three cases and, to check adequacy, it is enough to check that the conditions of Definition 1 hold in the first case, namely when $|S| = |S_A| + 1$. Therefore, S_A must contain a self-abutting cycle C and, in the transition from S_A to S, the cycle C splits into two cycles C_1 and C_2 of S. Since D is alternating and checkerboard colorable, S_A bounds a subsurface Σ' of Σ of a certain color, say white, which contains no crossings of D.

We claim that neither C_1 nor C_2 is trivial. Indeed, if, say, C_1 were trivial, then there would be a loop γ parallel to C_1 totally inside Σ' except for a little neighborhood of x, in which it would cross x. Such a curve would imply that the crossing x is removable, (see for example Figure 9), which is a contradiction. Therefore, neither C_1 nor C_2 is trivial, and it follows that $t(S) = t(S_A)$. Therefore, D is A–adequate at x, and this completes the proof of the theorem. □

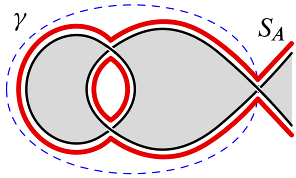


Figure 9: A trivial cycle, resulting in a removable crossing.

7.3 Link diagrams and shadows

A *link shadow* in Σ is a 4–valent graph in Σ, possibly with loop components. In other words, a shadow is a link diagram with crossing types ignored. For that reason we refer to shadow vertices as crossings and the components of any link realization of a shadow as its link components. (Not to be confused with connected components of a shadow.)

Some properties of link diagrams are entirely determined by its link shadow. For example, we will say that a link shadow D on Σ is *checkerboard colorable* if the components of $\Sigma \smallsetminus D$ can be colored by two colors such that any two components of $\Sigma \smallsetminus D$ that share an edge have opposite colors. Clearly, a link diagram is checkerboard colorable if and only if its link shadow is. Similarly, a link shadow is *minimally embedded* if it does not lie in a subsurface of Σ of smaller genus, and it is immediate that a link diagram on Σ is minimally embedded if and only if its link shadow is.

Each shadow crossing has two smoothings, which cannot be differentiated as A– and B–type, as in the case of link diagrams. For that reason, for link shadows it is customary to place markers at the crossings indicating the smoothing as in Figure 10.

Two consecutive crossings can have identical or opposite smoothings; see Figure 11. An *alternating state* of a shadow is one with alternating crossing smoothings along all of its link components. In other words, a state is alternating if the smoothings at every pair of consecutive crossings are opposite.

Not all link shadows admit alternating smoothings, for example the shadow of the nonalternable knot in the torus in Figure 8. On the other hand, any link shadow of

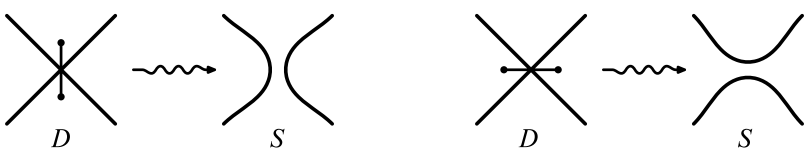


Figure 10: Two types of markers for a state of a link shadow.

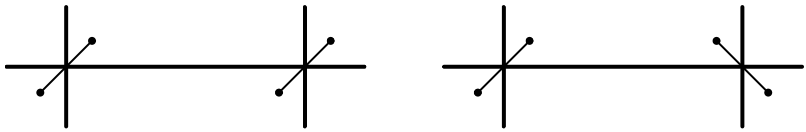


Figure 11: Two consecutive crossings with identical markers (left) and opposite markers (right).

an alternating link diagram admits two alternating smoothings, namely the shadow smoothings coming from S_A and S_B.

Given a state S for a link shadow D, the dual state is denoted by $S^\vee$ and has opposite smoothing to S at each crossing of D. Notice that a state S is alternating if and only if its dual state $S^\vee$ is alternating.

We say that a 2–disk D^2 is 2–*cutting* or, simply, *cutting* a shadow D if its boundary intersects D transversely at two points (which are not crossings) and $D^2 \cap D$ contains some but not all the crossings of D. A connected shadow D is said to be *strongly prime* if it has no cutting disk. More generally, a shadow D is strongly prime if all of its connected components are.

Lemma 25 *Every crossing of every strongly prime shadow $D \subset \Sigma$ has at least one smoothing producing a shadow which is again strongly prime. If D is connected, then the smoothing can be chosen so the resulting shadow is connected and strongly prime.*

For classical links, a proof of this statement can be found in [37]. That proof relies on checkerboard colorability of the diagram, which is of course true for classical links. Below, we give a proof that does not require the shadow to be checkerboard colorable.

Proof For the first part, it is enough to prove it for each of the connected components of D. Assume now that the smoothings of a crossing v in a strongly prime D produce diagrams D_1 and D_2 neither of which is strongly prime. Let B_1 and B_2 be cutting disks for D_1 and D_2. Since D is strongly prime, we can assume that $v \in \partial B_i$ for $i = 1, 2$. We can also assume that ∂B_1 and ∂B_2 are in transversal position. Let C be the connected component of $B_1 \cap B_2$ containing v, as in Figure 12, left. The circles ∂B_1 and ∂B_2 are broken because they may intersect each other many times.

By modifying B_1 or B_2 slightly if necessary we can assume that D does not contain the second intersection point, w, of $\partial B_1 \cap \partial B_2$ in C.

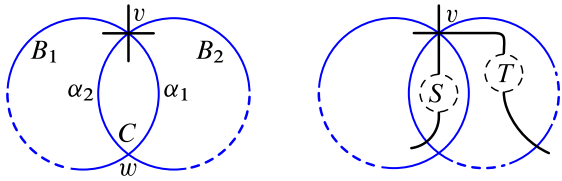


Figure 12: The cutting disks B_1 and B_2.

Let $\alpha_1 = \mathrm{int}(C \cap \partial B_1)$ and $\alpha_2 = \mathrm{int}(C \cap \partial B_2)$. (Note that $v \notin \alpha_1 \cup \alpha_2$.) Since D intersects $\partial B_i - \{v\}$ twice for $i = 1, 2$ and D intersects $\alpha_1 \cup \alpha_2$ at an odd number of points, we have the following possibilities:

(1) $|D \cap \alpha_2| = 1$ and $D \cap \alpha_1 = \varnothing$.

(2) $|D \cap \alpha_2| = 2$ and $|D \cap \alpha_1| = 1$.

(3) One of the two cases above with α_1 interchanged with α_2. We will ignore this case without loss of generality.

In the first case, D looks like in Figure 12, center, where S and T (in dashed circles) are shadow tangles. In that case, since neighborhoods of S and T are not cutting disks for D, the tangles S and T are crossingless. That means that B_2 is not a cutting disk for D_2 — a contradiction.

In the second case, D looks like in Figure 12, right, where R, S and T are shadow tangles. Note that all crossings of D, other than v, are contained in R, S or T, since otherwise a disk containing v, R, S and T but no other crossings of D would be cutting for D. Note also that, as in the first case, T is crossingless. That means that all crossings of D_1 are in R and S. Hence, B_1 is not cutting for D_1 — a contradiction.

For the second part, assume that D is connected. Then one of the smoothings of D at v will be connected. Let D' denote the connected shadow obtained from smoothing D, and assume the other smoothing is disconnected. We claim that D' is strongly prime.

Assume to the contrary that D' is not strongly prime. Then there is a cutting disk B containing some but not all the crossings of D' (see Figure 13). We can assume that

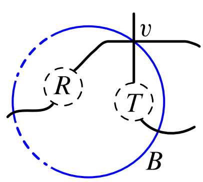


Figure 13: A cutting disk for D'.

$v \in \partial B$ and that D' is obtained by the smoothing of v tangential to ∂B. However, since the other smoothing of D at v is disconnected, the strands from the tangles R and T cannot cross each other. The neighborhoods of R and T give cutting disks for D unless the tangles R and T are crossingless, but then B would not be a cutting disk for D', which is a contradiction. □

Suppose $D \subset \Sigma$ is a link shadow. Let N_D denote a neighborhood of D in Σ small enough that it is a ribbon surface retractable onto D. A *local checkerboard coloring* of D is a checkerboard coloring of $D \subset N_D$. If one exists, we say that D is *locally checkerboard colorable*. (The pair (D, N_D) is the shadow of an abstract link diagram, or ALD for short [29]. This condition is equivalent to saying that (D, N_D) is the shadow of a checkerboard colorable ALD.)

Obviously, if $D \subset \Sigma$ is checkerboard colorable, then it is locally checkerboard colorable. The converse holds if $D \subset \Sigma$ is cellularly embedded, but, in general, a shadow can be locally checkerboard colorable without being checkerboard colorable.

Lemma 26 *Suppose $D \subset \Sigma$ is a link shadow. Then D is locally checkerboard colorable if and only if it admits an alternating state.*

Proof If D is locally checkerboard colorable, then let S be the state whose smoothings at each crossing join the white regions. Then S is an alternating state.

Conversely, suppose S is an alternating state. Let $\widehat{\Sigma}$ be the surface obtained from N_D by attaching disks to each of its boundary component. Then $D \subset \widehat{\Sigma}$ is cellularly embedded. We can color $\widehat{\Sigma} \smallsetminus D$ so that each cycle in S bounds a black disk and each cycle in $S^\vee$ bounds a white disk. To see this, notice that, at each smoothing of S, two local regions are joined. We can color the joined regions white and extend the coloring to the rest of $\widehat{\Sigma} \smallsetminus D$. This determines a local checkerboard coloring of D. □

If S and S' are adjacent states on a shadow D with $|S'| = |S|$, then the transition from S to S' is called a *single cycle bifurcation*.

Lemma 27 *A connected shadow D is locally checkerboard colorable if and only if there is no single cycle bifurcation in its cube of resolutions.*

Proof For one implication, we apply [30, Proposition 5.11] to see that, if D is locally checkerboard colorable, then its cube of resolutions does not contain any single cycle bifurcations.

The other implication is proved by induction on the crossing number. To start, we verify it for 1–crossing shadows, which can be classified into the *first type* or the *second type*. The shadows of the first type are locally checkerboard colorable and of the second type are not. The cubes of resolutions for these shadows are $\bullet \to \bullet$, and they have just one edge, which is a split/join for the shadow of the first type and a single cycle bifurcation for the shadow of the second type.

Now assume the lemma has been proved for all connected shadows with fewer than n crossings. Let D be a connected shadow with n crossings. We will show that, if D is not locally checkerboard colorable, then there is a single cycle bifurcation in its cube of resolutions. Pick a crossing x and let D' be the diagram obtained by smoothing D at x. (It does not matter which smoothing is chosen.)

Assume first that D' is not locally checkerboard colorable. By induction, the cube of resolutions for D' contains a single cycle bifurcation. Since the cube of resolutions of D' is a face of the cube of resolutions of D, the result follows.

On the other hand, if D' is locally checkerboard colorable, then, by Lemma 26, it admits an alternating state S'. We color $N_{D'} \smallsetminus D'$ consistently, so that the smoothings of S' join white regions. Let S be a state of D which coincides with S', and $S^\vee$ its dual state. Switching the smoothing of x in $S^\vee$, we obtain $S'^\vee$, considered as a state of D.

The ribbon surface N_D is obtained by adding a 2–dimensional 1–handle (a band) to $N_{D'}$. Unless the transition from $S'^\vee$ to $S^\vee$ is a single cycle bifurcation, we can extend the coloring of $(N_{D'}, D')$ to (N_D, D). Since D is not locally checkerboard colorable, the transition from $S'^\vee$ to $S^\vee$ must be a single cycle bifurcation. □

Recall that $r(D)$ denotes the rank of the image of $i_*\colon H_1(D;\mathbb{Z}/2) \to H_1(\Sigma;\mathbb{Z}/2)$. Any connected shadow is homotopy equivalent to a bouquet of circles. If D has $c(D)$ crossings, then $\chi(D) = -c(D)$. It follows that $0 \le r(D) \le c(D)+1$ for connected shadows with $c(D)$ crossings.

Proposition 28 *Let D be a link shadow in Σ (not necessarily connected).*

(i) *If S is a state of D, then*

$$t(S)+t(S^\vee) \le c(D)+2|D|-r(D).$$

(ii) *If D is not locally checkerboard colorable, then, for any state S of D,*

$$t(S)+t(S^\vee) < c(D)+2|D|-r(D).$$

(iii) *If D is strongly prime and S is nonalternating, then*

$$t(S) + t(S^\vee) < c(D) + 2|D| - r(D).$$

Proof Let us write $\Sigma = \Sigma_1 \cup \cdots \cup \Sigma_n$ as a disjoint union of connected components. Any component disjoint from D does not contribute to the terms in (i), (ii) and (iii), so it can be discarded. Therefore, we can assume that $D_i = D \cap \Sigma_i \neq \varnothing$ for $i = 1, \ldots, n$.

Since all terms of the inequalities of the statements are additive under taking disjoint unions of surfaces, it is enough to prove the statement for Σ connected.

On the other hand, if $D = D_1 \cup D_2$ is disconnected, then $r(D) \leq r(D_1) + r(D_2)$. Thus, $r(D)$ is subadditive, and, since the other terms on the right-hand side of (i), (ii) and (iii) are additive, it is enough to prove the proposition for connected shadows in connected surfaces. Assume henceforth that Σ is a connected surface.

Let us prove the statement for single crossing abstract shadows D now. Recall from the proof of Lemma 27 that single crossing shadows D are of two types. For both of them, $r(D) \leq 2$. If $r(D) = 0$, then $t(S) + t(S^\vee) = 2$. If $r(D) = 1, 2$, then $t(S) + t(S^\vee) \leq 1$. Therefore, statement (i) holds for 1–crossing shadows. Since shadows of the first type are locally checkerboard colorable and $t(S) = t(S^\vee) = 0$ for shadows of the second type, statements (ii) and (iii) hold as well.

The proof of (i) proceeds by induction on the crossing number $c(D)$. Let D be a connected shadow in Σ with $c(D) \geq 2$ crossings. We assume that statement (i) has been established for all connected shadows in Σ with fewer than $c(D)$ crossings.

Let D' be the shadow resulting from smoothing at a crossing x of D. We choose the smoothing so that D' is connected. Notice that

(9) $$r(D) - 1 \leq r(D') \leq r(D).$$

Let S be a state of D. The chosen smoothing of x coincides with the smoothing of x either in S or in $S^\vee$ and, without loss of generality, we can assume that it coincides with the smoothing of x in S. Then S induces a state on D', denoted by S'. Clearly, $t(S') = t(S)$. The dual state $S'^\vee$ to S' differs from $S^\vee$ at x only. The states $S^\vee$ and $S'^\vee$ are adjacent in the cube of resolutions of D. Thus,

(10) $$t(S'^\vee) - 1 \leq t(S^\vee) \leq t(S'^\vee) + 1.$$

Lemma 29 *Either $r(D') = r(D)$ or $t(S^\vee) \leq t(S'^\vee)$.*

Proof Assume that $t(S^\vee) > t(S'^\vee)$. Then, either two trivial loops in $S^\vee$ join to make a trivial loop in $S'^\vee$, or a trivial and a nontrivial loop in $S^\vee$ join to make a nontrivial loop in $S'^\vee$, or a trivial loop in $S^\vee$ splits to make two nontrivial loops in $S'^\vee$. In each case, $r(D') = r(D)$. □

We prove the inductive step for part (i). By Lemma 29, there are the two possibilities. If $r(D') = r(D)$, then (10) and the inductive assumption imply that

$$t(S) + t(S^\vee) \le t(S') + t(S'^\vee) + 1 \le c(D') + 2 - r(D') + 1 = c(D) + 2 - r(D).$$

On the other hand, if $r(D') \ne r(D)$, then $t(S^\vee) \le t(S'^\vee)$, and (9) and the inductive assumption imply that

$$t(S) + t(S^\vee) \le t(S') + t(S'^\vee) \le c(D') + 2 - r(D') = c(D) + 2 - r(D).$$

This completes the proof in case (i).

We prove part (ii) also by induction on $c(D)$. Let D be a connected shadow in Σ with $c(D) \ge 2$ crossings, and assume D is not locally checkerboard colorable. We assume that statement (ii) has been established for all connected shadows in Σ with fewer than $c(D)$ crossings that are not locally checkerboard colorable. By Lemma 27, there is a single cycle bifurcation in the cube of resolutions of D.

Let D' be the shadow resulting from smoothing D at a crossing x, and we assume D' is connected and that the smoothing at x coincides with the smoothing of x in S.

If D' is locally checkerboard colorable, then the transition from $S^\vee$ to $S'^\vee$ must be a single cycle bifurcation, for otherwise the local checkerboard coloring would extend from D' to D.

Since the transition is a single cycle bifurcation, we have $t(S^\vee) = t(S'^\vee)$ and $r(D) = r(D')$. Therefore, applying part (i) to D', we see that

$$t(S) + t(S^\vee) = t(S) + t(S'^\vee) \le c(D') + 2 - r(D') < c(D) + 2 - r(D).$$

If D' is not locally checkerboard colorable, then we can apply the inductive hypothesis for part (ii) to D' and use it to deduce the desired strict inequality just as before. This completes the proof of (ii).

The last step is to prove statement (iii). We begin by verifying (iii) for connected shadows with one or two crossings. For a single crossing shadow D of the first type,

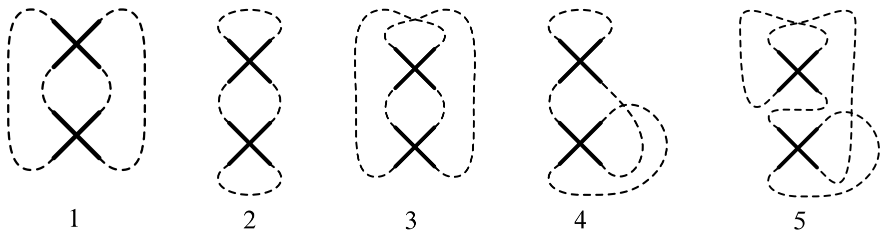


Figure 14: Connected shadow diagrams with 2–crossings.

both states are alternating, so (iii) is vacuously true. Single crossing shadow of the second type are not locally checkerboard colorable, and so the result follows from (ii).

All abstract connected 2–crossing shadows D are depicted in Figure 14. For a type 1 shadow D, its nonalternating states appear in Figure 15, left. Note that $0 \leq r(D) \leq 3$ and $0 \leq t(S), t(S^\vee) \leq 1$. If $r(D) = 0$ or 1, then $t(S)+t(S^\vee) \leq 2$ and $3 \leq c(D)+2-r(D)$. Thus, (iii) holds in this case. If $r(D) = 2$ or 3, then $t(S) = t(S^\vee) = 0$, and statement (iii) holds.

For a type 2 shadow D, its nonalternating states are shown in Figure 15, right. Note that $0 \leq r(D) \leq 3$ and $0 \leq t(S), t(S^\vee) \leq 2$. Since D is strongly prime, $r(D) > 0$ and $t(S), t(S^\vee) \leq 1$. If $r(D) = 1$, then $t(S)+t(S^\vee) \leq 2$; if $r(D) = 2$, then $t(S)+t(S^\vee) \leq 1$; and if $r(D) = 3$, then $t(S)+t(S^\vee) = 0$. In all three cases, statement (iii) is seen to hold.

Note that none of the shadows of the third, fourth and fifth types is locally checkerboard colorable. Therefore statement (iii) follows from (ii) in these cases.

The proof of (iii) proceeds by induction on the crossing number $c(D)$. Let D be a strongly prime connected shadow in Σ. By (ii), we can assume that D is locally

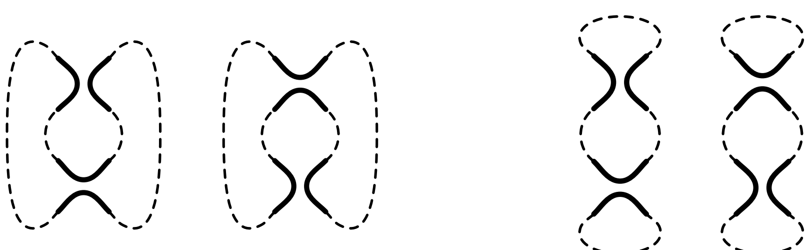


Figure 15: Nonalternating states on 2–crossing shadows of type 1 (left) and type 2 (right).

checkerboard colorable. We assume additionally that $c(D) \geq 3$ and that statement (iii) has been established for all strongly prime shadows in Σ with fewer than $c(D)$ crossings.

Let S be a nonalternating state for D. Then S has two consecutive smoothings that are identical, and we choose a third crossing x of D. By Lemma 25, one of the smoothings of x yields a shadow which is connected and strongly prime. Let D' be the resulting shadow. As before, we assume that the smoothing at x coincides with the smoothing of x in S. The state S induces a state on D', denoted by S', which is nonalternating. Since D' is connected, one can apply Lemma 29 as before and argue again by induction that (iii) holds for D. □

7.4 Proof of Theorem 13

Part (i) follows immediately by combining Corollary 5 and Proposition 28(i).

For parts (ii)–(iii), if D is a connected sum of $D_0 \subset \Sigma$ and $D_1, \dots, D_k \subset S^2$, then

$$[D]_\Sigma = \delta^{-k}[D_0]_\Sigma \cdot \prod_{i=1}^{k} [D_i]_{S^2}. \tag{11}$$

Therefore, it is enough to prove (ii)–(iii) for *prime* diagrams (alternating for (ii) and nonalternating for (iii)).

The condition that D is prime implies that it is not a nontrivial connected sum diagram as above. More precisely, a link diagram D on Σ is said to be *prime* if any contractible simple loop γ in Σ that meets D transversely at two points bounds a 2–disk that intersects D in an unknotted arc (possibly with self-crossings).

For the proof of part (iii), assume D is prime. If the shadow diagram of D is strongly prime, then the statement follows from Corollary 5 and Proposition 28(iii). If it is not strongly prime, then D must contain a self-crossing trivial arc. Let D' be obtained by replacing it by a simple trivial arc. Since $\mathrm{span}([D]_\Sigma)$ is invariant under Reidemeister moves and $r(D') = r(D)$,

$$\mathrm{span}([D]_\Sigma) = \mathrm{span}([D']_\Sigma) \leq 4c(D') + 4|D'| - 2r(D') < 4c(D) + 4|D| - 2r(D)$$

by part (i).

The proof of part (ii) follows that of [8, Theorem 2.9]. Since both sides of the equality in (ii) are additive under disjoint union of diagrams, it is enough to prove it for connected diagrams.

By Proposition 23, D is checkerboard colorable. Then all regions of one color, say white, are enclosed by the cycles in the state S_A of D, and all regions of the other color, ie black, are enclosed by the cycles in the state S_B. Therefore, the numbers of white and black regions are $t(S_A)$ and $t(S_B)$, respectively. Since D defines a cellular decomposition of Σ into $c(D)$ 0–cells, $2c(D)$ 1–cells, and $t(S_A)+t(S_B)$ 2–cells,

$$2-2g(\Sigma)=\chi(\Sigma)=c(D)-2c(D)+t(S_A)+t(S_B),$$

and

$$t(S_A)+t(S_B)=c(D)+2-2g(\Sigma).$$

By Proposition 3,

$$\begin{aligned}\operatorname{span}([D]_\Sigma)&=d_{\max}([D]_\Sigma)-d_{\min}([D]_\Sigma),\\&=2c(D)+2t(S_A)+2t(S_B),\\&=4c(D)+4-4g(\Sigma).\end{aligned}$$ □

Acknowledgements We would like to express our thanks to the referee for their many helpful comments and suggestions. HB would also like to acknowledge funding through the Natural Sciences and Engineering Research Council of Canada.

References

[1] **C Adams**, **C Albors-Riera**, **B Haddock**, **Z Li**, **D Nishida**, **B Reinoso**, **L Wang**, *Hyperbolicity of links in thickened surfaces*, Topology Appl. 256 (2019) 262–278 MR Zbl

[2] **C Adams**, **O Eisenberg**, **J Greenberg**, **K Kapoor**, **Z Liang**, **K O'Connor**, **N Pachecho-Tallaj**, **Y Wang**, *Turaev hyperbolicity of classical and virtual knots*, Algebr. Geom. Topol. 21 (2021) 3459–3482 MR Zbl

[3] **C Adams**, **T Fleming**, **M Levin**, **A M Turner**, *Crossing number of alternating knots in $S\times I$*, Pacific J. Math. 203 (2002) 1–22 MR Zbl

[4] **B Bavier**, **E Kalfagianni**, *Guts, volume and skein modules of 3–manifolds*, Fund. Math. 256 (2022) 195–220 MR Zbl

[5] **R C Blair**, *Alternating augmentations of links*, J. Knot Theory Ramifications 18 (2009) 67–73 MR Zbl

[6] **R Blair**, **H Allen**, **L Rodriguez**, *Twist number and the alternating volume of knots*, J. Knot Theory Ramifications 28 (2019) art. id. 1950016 MR Zbl

[7] **C Blanchet**, **N Habegger**, **G Masbaum**, **P Vogel**, *Topological quantum field theories derived from the Kauffman bracket*, Topology 34 (1995) 883–927 MR Zbl

[8] **H U Boden**, **H Karimi**, *The Jones–Krushkal polynomial and minimal diagrams of surface links*, Ann. Inst. Fourier (Grenoble) 72 (2022) 1437–1475 MR Zbl

[9] **H U Boden**, **H Karimi**, *A characterization of alternating links in thickened surfaces*, Proc. Roy. Soc. Edinburgh Sect. A 153 (2023) 177–195 MR Zbl

[10] **H U Boden**, **W Rushworth**, *Minimal crossing number implies minimal supporting genus*, Bull. Lond. Math. Soc. 53 (2021) 1174–1184 MR Zbl

[11] **F Bonahon**, **H Wong**, *Quantum traces for representations of surface groups in* $\mathrm{SL}_2(\mathbb{C})$, Geom. Topol. 15 (2011) 1569–1615 MR Zbl

[12] **D Bullock**, *Rings of* $\mathrm{SL}_2(\mathbb{C})$*–characters and the Kauffman bracket skein module*, Comment. Math. Helv. 72 (1997) 521–542 MR Zbl

[13] **D Bullock**, **C Frohman**, **J Kania-Bartoszyńska**, *Understanding the Kauffman bracket skein module*, J. Knot Theory Ramifications 8 (1999) 265–277 MR Zbl

[14] **J S Carter**, **S Kamada**, **M Saito**, *Stable equivalence of knots on surfaces and virtual knot cobordisms*, J. Knot Theory Ramifications 11 (2002) 311–322 MR Zbl

[15] **A Champanerkar**, **I Kofman**, *A survey on the Turaev genus of knots*, Acta Math. Vietnam. 39 (2014) 497–514 MR Zbl

[16] **A Champanerkar**, **I Kofman**, *A volumish theorem for alternating virtual links*, New York J. Math. 28 (2022) 337–356 MR Zbl

[17] **F Costantino**, **T T Q Lê**, *Stated skein algebras of surfaces*, J. Eur. Math. Soc. 24 (2022) 4063–4142 MR Zbl

[18] **O T Dasbach**, **D Futer**, **E Kalfagianni**, **X-S Lin**, **N W Stoltzfus**, *The Jones polynomial and graphs on surfaces*, J. Combin. Theory Ser. B 98 (2008) 384–399 MR Zbl

[19] **O T Dasbach**, **X-S Lin**, *A volumish theorem for the Jones polynomial of alternating knots*, Pacific J. Math. 231 (2007) 279–291 MR Zbl

[20] **Y Diao**, *The additivity of crossing numbers*, J. Knot Theory Ramifications 13 (2004) 857–866 MR Zbl

[21] **H A Dye**, **L H Kauffman**, *Minimal surface representations of virtual knots and links*, Algebr. Geom. Topol. 5 (2005) 509–535 MR Zbl

[22] **V Fock**, **A Goncharov**, *Moduli spaces of local systems and higher Teichmüller theory*, Publ. Math. Inst. Hautes Études Sci. 103 (2006) 1–211 MR Zbl

[23] **S Fomin**, **M Shapiro**, **D Thurston**, *Cluster algebras and triangulated surfaces, I: Cluster complexes*, Acta Math. 201 (2008) 83–146 MR Zbl

[24] **C Frohman**, **R Gelca**, **W Lofaro**, *The A–polynomial from the noncommutative viewpoint*, Trans. Amer. Math. Soc. 354 (2002) 735–747 MR Zbl

[25] **C Frohman**, **J Kania-Bartoszynska**, **T Lê**, *Unicity for representations of the Kauffman bracket skein algebra*, Invent. Math. 215 (2019) 609–650 MR Zbl

[26] **J A Howie**, *Surface-alternating knots and links*, PhD thesis, University of Melbourne (2015) Available at `http://hdl.handle.net/11343/58991`

[27] **J A Howie**, **J S Purcell**, *Geometry of alternating links on surfaces*, Trans. Amer. Math. Soc. 373 (2020) 2349–2397 MR Zbl

[28] **N Kamada**, *On the Jones polynomials of checkerboard colorable virtual links*, Osaka J. Math. 39 (2002) 325–333 MR Zbl

[29] **N Kamada**, **S Kamada**, *Abstract link diagrams and virtual knots*, J. Knot Theory Ramifications 9 (2000) 93–106 MR Zbl

[30] **H Karimi**, *The Khovanov homology of alternating virtual links*, Michigan Math. J. 70 (2021) 749–778 MR Zbl

[31] **L H Kauffman**, *State models and the Jones polynomial*, Topology 26 (1987) 395–407 MR Zbl

[32] **V Krushkal**, *Graphs, links, and duality on surfaces*, Combin. Probab. Comput. 20 (2011) 267–287 MR Zbl

[33] **G Kuperberg**, *What is a virtual link?*, Algebr. Geom. Topol. 3 (2003) 587–591 MR Zbl

[34] **M Lackenby**, *The volume of hyperbolic alternating link complements*, Proc. London Math. Soc. 88 (2004) 204–224 MR Zbl

[35] **M Lackenby**, *The crossing number of composite knots*, J. Topol. 2 (2009) 747–768 MR Zbl

[36] **T T Q Lê**, *The colored Jones polynomial and the A–polynomial of knots*, Adv. Math. 207 (2006) 782–804 MR Zbl

[37] **W B R Lickorish**, *An introduction to knot theory*, Graduate Texts in Math. 175, Springer (1997) MR Zbl

[38] **W B R Lickorish**, **M B Thistlethwaite**, *Some links with nontrivial polynomials and their crossing-numbers*, Comment. Math. Helv. 63 (1988) 527–539 MR Zbl

[39] **V O Manturov**, *Parity and projection from virtual knots to classical knots*, J. Knot Theory Ramifications 22 (2013) art. id. 1350044 MR Zbl

[40] **W Menasco**, *Closed incompressible surfaces in alternating knot and link complements*, Topology 23 (1984) 37–44 MR Zbl

[41] **W Menasco**, **M Thistlethwaite**, *The classification of alternating links*, Ann. of Math. 138 (1993) 113–171 MR Zbl

[42] **G Muller**, *Skein and cluster algebras of marked surfaces*, Quantum Topol. 7 (2016) 435–503 MR Zbl

[43] **K Murasugi**, *Jones polynomials and classical conjectures in knot theory*, Topology 26 (1987) 187–194 MR Zbl

[44] **J H Przytycki**, *Fundamentals of Kauffman bracket skein modules*, Kobe J. Math. 16 (1999) 45–66 MR Zbl

[45] **J H Przytycki**, **A S Sikora**, *On skein algebras and* $\mathrm{Sl}_2(\mathbb{C})$*–character varieties*, Topology 39 (2000) 115–148 MR Zbl

[46] **A S Sikora**, **B W Westbury**, *Confluence theory for graphs*, Algebr. Geom. Topol. 7 (2007) 439–478 MR Zbl

[47] **R Stong**, *The Jones polynomial of parallels and applications to crossing number*, Pacific J. Math. 164 (1994) 383–395 MR Zbl

[48] **M B Thistlethwaite**, *A spanning tree expansion of the Jones polynomial*, Topology 26 (1987) 297–309 MR Zbl

[49] **V G Turaev**, *A simple proof of the Murasugi and Kauffman theorems on alternating links*, Enseign. Math. 33 (1987) 203–225 MR Zbl

[50] **V G Turaev**, *The Conway and Kauffman modules of a solid torus*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov 167 (1988) 79–89 MR Zbl In Russian; translated in J. Soviet Math. 52 (1990) 2799–2805

[51] **V G Turaev**, *Skein quantization of Poisson algebras of loops on surfaces*, Ann. Sci. École Norm. Sup. 24 (1991) 635–704 MR Zbl

[52] **V G Turaev**, *Algebras of loops on surfaces, algebras of knots, and quantization*, from "Braid group, knot theory and statistical mechanics, II" (C N Yang, M L Ge, editors), Adv. Ser. Math. Phys. 17, World Sci., River Edge, NJ (1994) 324–360 MR Zbl

[53] **D A Will**, *Homological polynomial coefficients and the twist number of alternating surface links*, Algebr. Geom. Topol. 22 (2022) 3939–3963 MR Zbl

Department of Mathematics and Statistics, McMaster University
Hamilton, ON, Canada

Department of Mathematics and Statistics, McMaster University
Hamilton, ON, Canada

Department of Mathematics, University at Buffalo
Buffalo, NY, United States

boden@mcmaster.ca, karimih@mcmaster.ca, asikora@buffalo.edu

Received: 16 June 2021 Revised: 19 January 2022

Algebraic & Geometric Topology **23**:5 (2023) 2309–2327
DOI: 10.2140/agt.2023.23.2309
Published: 25 July 2023

Homotopy types of gauge groups over Riemann surfaces

Masaki Kameko
Daisuke Kishimoto
Masahiro Takeda

Let G be a compact connected Lie group with $\pi_1(G) \cong \mathbb{Z}$. We study the homotopy types of gauge groups of principal G–bundles over Riemann surfaces. This can be applied to an explicit computation of the homotopy groups of the moduli spaces of stable vector bundles over Riemann surfaces.

57S05; 55Q15

1 Introduction

Let G be a compact connected Lie group, and let P be a principal G–bundle over a finite complex X. The *gauge group* of P is defined to be the topological group of G–equivariant self-maps of P which fix X. There may be infinitely many distinct principal G–bundles over X. For example, there are infinitely many bundles when X is an orientable 4–manifold. Each bundle has a gauge group, so there may be potentially infinitely many gauge groups. However, Crabb and Sutherland [6] showed that these gauge groups have only finitely many homotopy types. Subsequently, the precise number of homotopy types of gauge groups for specific G and X has been intensely studied. The study began with simply connected Lie groups by Cutler [7], Hamanaka, Hasui, Kishimoto, Kono, So, Theriault and Tsutaya [10; 12; 15; 16; 18; 20; 30; 31], and recently, nonsimply connected cases are also studied by Hasui, Kamiyama, Kishimoto, Kono, Membrillo-Solis, Sato, Theriault and Tsukuda [11; 14; 17] and Rea [26].

In this paper, we study the homotopy types of gauge groups of principal G–bundles over a compact connected Riemann surface, where $\pi_1(G) \cong \mathbb{Z}$. This includes an important case: gauge groups of principal $\mathrm{U}(n)$–bundles over a Riemann surface, whose topology was first studied by Atiyah and Bott [2]. To state the results, we introduce a numerical invariant of G. Suppose $\pi_1(G) \cong \mathbb{Z}$. Then as in Mimura and Toda [24, Corollary 5.1, Chapter II], there is a compact connected simply connected Lie group H and a subgroup C of the center of $S^1 \times H$ such that

$$G \cong (S^1 \times H)/C. \tag{1-1}$$

In other words, G is locally isomorphic to $S^1 \times H$. Note that H is uniquely determined by G, but C is not. For example, if $G = S^1 \times H$, then C can be any finite subgroup of $S^1 \times 1 \subset S^1 \times H$. We define

$$s(G) = |p_2(C)|,$$

where $p_2 \colon S^1 \times H \to H$ is the projection. By Theorem 1.4 below, we can see that $s(G)$ is independent of the choice of C.

Example 1.1 Since $\mathrm{U}(n)$ is the quotient of $S^1 \times \mathrm{SU}(n)$ by the diagonal central subgroup isomorphic to $\mathbb{Z}/n$, we have $s(\mathrm{U}(n)) = n$.

Let X be a compact connected Riemann surface. Then there is a one-to-one correspondence between principal G–bundles over X and $\pi_2(BG) \cong \mathbb{Z}$. Let $\mathcal{G}_k(X, G)$ denote the gauge group of a principal G–bundle over X corresponding to $k \in \mathbb{Z}$. Now we state our results.

Theorem 1.2 *Let G be a compact connected Lie group with $\pi_1(G) \cong \mathbb{Z}$, and let X be a compact connected Riemann surface. If $(k, s(G)) = (l, s(G))$, then $\mathcal{G}_k(X, G)$ and $\mathcal{G}_l(X, G)$ are homotopy equivalent after localizing at any prime or zero.*

We remark that the p–localization of a disconnected space will mean the disjoint union of the p–localization of path-connected components. For a prime p, Theriault [29] gave a p–local homotopy decomposition of $\mathcal{G}_k(X, \mathrm{U}(p))$, which implies the converse implication of Theorem 1.2 holds for $G = \mathrm{U}(p)$. We will prove the converse implication of Theorem 1.2 holds for other Lie groups.

Theorem 1.3 *Let G be a compact connected Lie group with $\pi_1(G) \cong \mathbb{Z}$, and let X be a compact connected Riemann surface. If G is locally isomorphic to $S^1 \times \mathrm{SU}(n)^r$ or $S^1 \times \mathrm{SU}(4n-2)^s \times Sp(2n-1)^t$, then the following statements are equivalent:*

1. $(k, s(G)) = (l, s(G))$.
2. *$\mathcal{G}_k(X, G)$ and $\mathcal{G}_l(X, G)$ are homotopy equivalent after localizing at any prime or zero.*

Note that since $\mathrm{U}(n) = (S^1 \times \mathrm{SU}(n))/(\mathbb{Z}/n)$ as in Example 1.1, Theorem 1.3 applies to the case $G = \mathrm{U}(n)$.

The homotopy type of a gauge group $\mathcal{G}_k(X,G)$ is closely related with a Samelson product in G, as we will see in Section 2. In our context, the Samelson product of a generator of $\pi_1(G)\cong\mathbb{Z}$ and the identity map of G is of particular importance. We will prove the following theorem, which is of independent interest.

Theorem 1.4 *Let G be a compact connected Lie group with $\pi_1(G)\cong\mathbb{Z}$, and let ϵ denote a generator of $\pi_1(G)$. Then the Samelson product $\langle\epsilon,1_G\rangle$ in G is of order $s(G)$.*

Now we consider an application. Gauge groups over a Riemann surface are closely related to the moduli spaces of stable vector bundles over a Riemann surface as follows. Let X be a Riemann surface of genus g, and let $M(n,k)$ denote the moduli space of stable vector bundles over X of rank n and degree k. Daskalopoulos and Uhlenbeck [8] showed that there is an isomorphism

$$\pi_i(M(n,k))\cong\pi_{i-1}(\mathcal{G}_k(X,\mathrm{U}(n)))$$

for $2<i\le 2(g-1)(n-1)-2$ and $(n,k)\ne(2,2)$. There is a polystable Higgs bundle analog due to Bradlow, García-Prada and Gothen [5]. We can compute the homotopy groups of these moduli spaces in a range through the following homotopy decomposition.

Theorem 1.5 *Let G be a compact connected Lie group with $\pi_1(G)\cong\mathbb{Z}$, and let X be a compact connected Riemann surface of genus g. If $s(G)$ divides k, then*

$$\mathcal{G}_k(X,G)\simeq G\times(\Omega G)^{2g}\times\Omega^2G.$$

Moreover, the above homotopy equivalence also holds after localizing at p whenever p does not divide $s(G)$.

The paper is structured as follows. Section 2 recalls a connection between gauge groups and Samelson products, and then proves Theorems 1.2 and 1.5 by assuming Theorem 1.4 holds. Section 3 shows some general results on Samelson products in a Lie group, which will be used for a practical computation. Sections 4 and 5 compute the Samelson products in G when H is simple. Finally, Section 6 collects all results so far together to prove Theorems 1.3 and 1.4.

Acknowledgements

The authors are grateful to Jérôme Scherer and the referee for useful comments. The authors were partly supported by JSPS KAKENHI grant numbers 17K05263 (Kameko), 17K05248 and 19K03473 (Kishimoto) and 21J10117 (Takeda).

2 Gauge groups and Samelson products

This section recalls a connection between gauge groups and Samelson products, and then Theorems 1.2 and 1.5 are proved by assuming Theorem 1.4 holds. First, we recall a connection between gauge groups and mapping spaces. Let G be a topological group, and let P be a principal G–bundle over a base X, which is classified by a map $\alpha\colon X \to BG$. Recall that the gauge group of P, denoted by $\mathcal{G}(P)$, is the topological group of G–equivariant self-maps of P which fix X. Gottlieb [9] proved that there is a natural homotopy equivalence

$$B\mathcal{G}(P) \simeq \mathrm{map}(X, BG; \alpha),$$

where $\mathrm{map}(A, B; f)$ denotes the path component of the space of maps $\mathrm{map}(A, B)$ containing a map $f\colon A \to B$. Then evaluating at the basepoint of X yields a homotopy fibration

$$\mathrm{map}_*(X, BG; \alpha) \to B\mathcal{G}(P) \to BG, \tag{2-1}$$

where $\mathrm{map}_*(X, BG; \alpha)$ is the subspace of $\mathrm{map}(X, BG; \alpha)$ consisting of basepoint preserving maps. So the gauge group $\mathcal{G}(P)$ is homotopy equivalent to the homotopy fiber of the connecting map

$$\partial_\alpha\colon G \to \mathrm{map}_*(X, BG; \alpha)$$

of the above homotopy fibration.

Next, we assume $X = S^n$ for $n \geq 1$ and describe the connecting map ∂_α. Clearly, there is a homotopy equivalence $\mathrm{map}_*(S^n, BG; \alpha) \simeq \Omega_0^{n-1} G$, where $\Omega_0^{n-1} G$ denotes the path component of $\Omega^{n-1} G$ containing the constant map. Then by adjointing, the connecting map ∂_α corresponds to a map

$$d_\alpha\colon S^{n-1} \wedge G \to G.$$

The original definition of Whitehead products in [32] and adjointness of Whitehead products and Samelson products prove the following.

Lemma 2.1 *The map d_α is the Samelson product $\langle \bar{\alpha}, 1_G \rangle$ in G, where $\bar{\alpha}\colon S^{n-1} \to G$ is the adjoint of $\alpha\colon S^n \to BG$.*

The following lemma due to Theriault [27] shows how to identify the homotopy type of a gauge group $\mathcal{G}(P)$ from the order of a Samelson product $\langle \bar{\alpha}, 1_G \rangle$.

Lemma 2.2 *Suppose that a map $f\colon X \to Y$ into an H-space Y is of order $n < \infty$. Then $(n,k) = (n,l)$ implies $F_{k(p)} \simeq F_{l(p)}$ for any prime p, where F_k denotes the homotopy fiber of a map $k \circ f\colon X \to Y$.*

Finally, we recall a homotopy decomposition of a gauge group. Theriault [28] showed a homotopy decomposition of a gauge group over principal $\mathrm{U}(n)$–bundle over a Riemann surface. We can easily see that his proof works in verbatim for any compact connected Lie group G with $\pi_1(G) \cong \mathbb{Z}$. Then we get:

Proposition 2.3 *Let G be a compact connected Lie group with $\pi_1(G) \cong \mathbb{Z}$, and let X be a compact connected Riemann surface of genus g. Then there is a homotopy equivalence*

$$\mathcal{G}_k(X,G) \simeq (\Omega G)^{2g} \times \mathcal{G}_k(S^2,G).$$

Now we prove Theorems 1.2 and 1.5 by assuming Theorem 1.4 holds.

Proof of Theorem 1.2 Combine Lemmas 2.1 and 2.2, Proposition 2.3 and Theorem 1.4. □

Proof of Theorem 1.5 By Lemma 2.1 and Theorem 1.4, if k is divisible by $s(G)$, then $\mathcal{G}_k(S^2,G)$ is homotopy equivalent to the homotopy fiber of the constant map $G \to \Omega_0 G$. So since $\pi_2(G) = 0$, $\mathcal{G}_k(S^2,G) \simeq G \times \Omega^2 G$. Thus by Proposition 2.3, the proof is done. □

3 Samelson products in Lie groups

This section shows some criteria for computing Samelson products in a Lie group. For the rest of the paper, we will use the following notation:

- Let G be a compact connected Lie group with $\pi_1(G) \cong \mathbb{Z}$.
- Let ϵ_G denote a generator of $\pi_1(G) \cong \mathbb{Z}$.
- Let H and C be as in the decomposition (1-1).
- Let $j_H\colon \Sigma H \to BH$ denote the natural map.
- Let $p_G\colon S^1 \times H \to G$ denote the quotient map.
- Let $p_1\colon S^1 \times H \to S^1$ and $p_2\colon S^1 \times H \to H$ denote projections.
- Let $K = H/p_2(C)$.
- Let $q_G\colon G \to K$ and $\bar{q}_K\colon H \to K$ denote the quotient maps.

We will abbreviate ϵ_G, j_H, p_G, q_G and $\bar{q}_K$ to ϵ, j, p, q and $\bar{q}$, respectively, if G, H and K are clear from the context. First, we show two properties of the group C.

Lemma 3.1 *The abelian group $p_2(C)$ is cyclic.*

Proof There is a fibration

$$S^1 \to G \xrightarrow{q} K \tag{3-1}$$

and so by the homotopy exact sequence, we can see that $\pi_1(K) \cong p_2(C)$ is a quotient of $\pi_1(G) \cong \mathbb{Z}$. Then $p_2(C)$ is a cyclic group, as stated. □

Lemma 3.2 *We may choose a group C such that $|p_1(C)| = s(G)$.*

Proof Note that $p_2(C)$ is a cyclic group by (3-1). We prove that the inequality $|p_1(C)| \geq s(G)$ always holds. If $|p_1(C)| < s(G)$, then $C_1 = |p_1(C)|C$ is a nontrivial subgroup of the center of $1 \times H \subset S^1 \times H$. In particular, there is a covering

$$C/C_1 \to (S^1 \times H)/C_1 \to G.$$

Then $\pi_1(G) \cong \mathbb{Z}$ includes a nontrivial finite abelian group C_1, which is a contradiction. Thus $|p_1(C)| \geq s(G)$.

Suppose that $|p_1(C)| > s(G)$. Then $C_2 = s(G)C$ is a finite subgroup of $S^1 \times 1 \subset S^1 \times H$. Then $(S^1 \times H)/C_2 \cong S^1 \times H$, implying

$$G \cong (S^1 \times H)/C \cong ((S^1 \times H)/C_2)/(C/C_2) \cong (S^1 \times H)/(C/C_2).$$

Note that C is a subgroup of $p_1(C) \times p_2(C)$ generated by (g_1, g_2), where g_i is a generator of a cyclic group $p_i(C)$ for $i = 1, 2$. Then $C_2 = s(G)p_1(C) \times 0$, and so C/C_2 is identified with the diagonal subgroup of

$$(p_1(C)/s(G)p_1(C)) \times p_2(C) \cong \mathbb{Z}/s(G) \times Z/s(G).$$

Thus $|p_1(C/C_2)| = s(G)$, finishing the proof. □

By Lemma 3.1, $\pi_1(K) \cong p_2(C)$ is a cyclic group of order $s(G)$. For the rest of this section, we will also use the following notation:

- Let $\bar{\epsilon}_K$ denote a generator of $\pi_1(K)$.

We will abbreviate it by $\bar{\epsilon}$ if K is clear from the context.

Next, we show an upper bound and a lower bound for the order of $\langle \epsilon, 1_G \rangle$.

Lemma 3.3 *The order of $\langle \epsilon, 1_G \rangle$, hence $\langle \epsilon, p \rangle$, divides $s(G)$.*

Proof The proof of Lemma 3.1 implies $q \circ \epsilon = \bar{\epsilon}$. Then since q is a homomorphism, we get

$$q_*(s(G)\langle \epsilon, 1_G\rangle) = s(G)\langle q \circ \epsilon, q\rangle = \langle s(G)\bar{\epsilon}, q\rangle = 0.$$

So since there is a fibration (3-1), $s(G)\langle \epsilon, 1_G\rangle$ lifts to a map $S^1 \wedge G \to S^1$. Since $S^1 \wedge G$ is simply connected, this lift is trivial, and thus $s(G)\langle \epsilon, 1_G\rangle$ itself is trivial, completing the proof. □

Lemma 3.4 *The order of $\langle \bar{\epsilon}, \bar{q}\rangle$ divides the order of $\langle \epsilon, p\rangle$.*

Proof Let $i\colon H \to S^1 \times H$ denote the inclusion. By definition, $q \circ p \circ i = \bar{q}$, and the proof of Lemma 3.2 implies that $q \circ \epsilon = \bar{\epsilon}$. Then

$$(1 \wedge i)^* \circ q_*(\langle \epsilon, p\rangle) = q_*(\langle \epsilon, p \circ i\rangle) = \langle q \circ \epsilon, q \circ p \circ i\rangle = \langle \bar{\epsilon}, \bar{q}\rangle$$

and so the proof is done. □

Finally, we give a cohomological criterion for the Samelson product $\langle \bar{\epsilon}, \bar{q}\rangle$ being nontrivial. For an algebra A, let QA denote the module of indecomposables.

Lemma 3.5 *Suppose there are $x, y, z \in QH^*(BK; \mathbb{Z}/p)$ and a Steenrod operation θ satisfying the following conditions*:

1. $|y| = 2$ *and* $QH^n(BK; \mathbb{Z}/p) = \langle z\rangle$ *for* $n > 2$.
2. $\theta(x)$ *is decomposable and includes the term* $y \otimes z$.
3. $(\bar{q} \circ j)^*(z)$ *is nontrivial and not included in any element of* $\theta(H^*(\Sigma H; \mathbb{Z}/p))$.

Then the Samelson product $\langle \bar{\epsilon}, \bar{q}\rangle$ is nontrivial.

Proof Suppose that $\langle \bar{\epsilon}, \bar{q}\rangle$ is trivial. Let $\hat{\epsilon}\colon S^2 \to BK$ and $\hat{q}\colon \Sigma H \to BK$ denote the adjoint of $\bar{\epsilon}$ and $\bar{q}$, respectively. Then by adjointness of Samelson products and Whitehead products, the Whitehead product $[\hat{\epsilon}, \hat{q}]$ is trivial, so that there is a homotopy commutative diagram

$$\begin{tikzcd}
S^2 \vee \Sigma H \arrow[r, "\hat{\epsilon} \vee \hat{q}"] \arrow[d] & BK \arrow[d, equal] \\
S^2 \times \Sigma H \arrow[r, "\mu"] & BK
\end{tikzcd}$$

Since BK is simply connected, $H^1(BK; \mathbb{Z}/p) = 0$ and $H^2(BK; \mathbb{Z}/p) = \langle y\rangle$. Then by the Hurewicz theorem and the first condition in the statement, we may assume $\hat{\epsilon}^*(y) = u$, where u is a generator of $H^2(S^2; \mathbb{Z}/p) \cong \mathbb{Z}/p$. Hence by the first and the second conditions, $\mu^*(\theta(x))$ includes the term $u \otimes \hat{q}^*(z)$. Since $\hat{q} = \bar{q} \circ j$, the third condition implies $u \otimes \hat{q}^*(z) \neq 0$. On the other hand, by the third condition, $\theta(\mu^*(x))$ cannot include the term $u \otimes \hat{q}^*(z)$. Thus since $\mu^*(\theta(x)) = \theta(\mu^*(x))$, we obtain a contradiction. Therefore $\langle \bar{\epsilon}, \bar{q} \rangle$ is nontrivial, completing the proof. □

Recall that compact simply connected simple Lie groups with nontrivial center are

$$\mathrm{SU}(n), \qquad \mathrm{Sp}(n), \qquad \mathrm{Spin}(n) \quad (n \geq 7), \qquad E_6, \qquad E_7.$$

Then in the following two sections, we will compute the Samelson product $\langle \epsilon, p \rangle$ for H being one of the above Lie groups.

4 Classical case

This section determines the order of the Samelson product $\langle \epsilon, p \rangle$ for $H = \mathrm{SU}(n)$, $\mathrm{Sp}(n)$ and $\mathrm{Spin}(n)$.

4.1 The case $H = \mathrm{SU}(n)$

First we consider the case $H = \mathrm{SU}(n)$.

Proposition 4.1 *If $H = \mathrm{SU}(n)$, then $\langle \epsilon, p \rangle$ is of order $s(G)$.*

Proof By Lemma 3.3, it suffices to show that the order of $\langle \epsilon, p \rangle$ is a nonzero multiple of $s(G)$. The center of $\mathrm{SU}(n)$ is isomorphic to $\mathbb{Z}/n$. Then since $\mathrm{U}(n) = S^1 \times_{\mathbb{Z}/n} \mathrm{SU}(n)$, it follows from Lemma 3.2 that there is a homomorphism $\rho\colon G \to \mathrm{U}(n)$ which is a $n/s(G)$ sheeted covering. Let α_{2i-1} denote a generator of $\pi_{2i-1}(\mathrm{U}(n)) \cong \mathbb{Z}$ for $i = 1, 2, \ldots, n$. Then

$$\rho_*(\epsilon) = \frac{n}{s(G)} \alpha_1.$$

On the other hand, it is shown in [4] that the order of $\langle \alpha_1, \alpha_{2n-1} \rangle$ is a nonzero multiple of n. Since $\rho_*\colon \pi_{2n-1}(G) \to \pi_{2n-1}(\mathrm{U}(n))$ is an isomorphism, there is an $\tilde{\alpha} \in \pi_{2n-1}(G)$ such that $\rho_*(\tilde{\alpha}) = \alpha_{2n-1}$. Then since

$$\rho_*(\langle \epsilon, \tilde{\alpha} \rangle) = \langle \rho_*(\epsilon), \rho_*(\tilde{\alpha}) \rangle = \left\langle \frac{n}{s(G)} \alpha_1, \alpha_{2n-1} \right\rangle = \frac{n}{s(G)} \langle \alpha_1, \alpha_{2n-1} \rangle,$$

the order of $\rho_*(\langle \epsilon, \tilde{\alpha} \rangle)$ is a nonzero multiple of $s(G)$. Thus, since the map

$$\rho_*\colon \pi_{2n}(G) \to \pi_{2n}(\mathrm{U}(n))$$

is an isomorphism, the order of $\langle \epsilon, \tilde{\alpha} \rangle$ is a nonzero multiple of $s(G)$ too. Since

$$p_*\colon \pi_{2n-1}(S^1 \times \mathrm{SU}(n)) \to \pi_{2n-1}(G)$$

is an isomorphism, there is a $\beta \in \pi_{2n-1}(S^1 \times \mathrm{SU}(n))$ such that $p \circ \beta = \tilde{\alpha}$. Thus since $(1 \wedge \beta)^*(\langle \epsilon, p \rangle) = \langle \epsilon, \tilde{\alpha} \rangle$, the order of $\langle \epsilon, p \rangle$ is a nonzero multiple of $s(G)$, completing the proof. □

4.2 The case $H = \mathrm{Sp}(n)$

Next, we consider the case $H = \mathrm{Sp}(n)$. Recall that the center of $\mathrm{Sp}(n)$ is isomorphic to $\mathbb{Z}/2$, and the quotient of $\mathrm{Sp}(n)$ by its center is denoted by $\mathrm{PSp}(n)$. We apply Lemma 3.5 to the case $H = \mathrm{Sp}(n)$. To this end, we compute the mod 2 cohomology of $B\mathrm{PSp}(2n)$ in low dimensions.

Lemma 4.2 *Let $\Delta = \{\pm(1, \dots, 1) \in \mathrm{Sp}(2)^n\}$. Then for $* \le 7$,*

$$H^*(B(\mathrm{Sp}(2)^n/\Delta); \mathbb{Z}/2) = \mathbb{Z}/2[x_2, x_3, x_5] \otimes \bigotimes_{k=1}^{n} \mathbb{Z}/2[x_{4,k}], \quad \mathrm{Sq}^2 x_{4,k} = x_2 x_{4,k},$$

where $|x_i| = i$ and $|x_{4,k}| = 4$.

Proof Consider the Serre spectral sequence for a homotopy fibration

$$\mathbb{R}P^\infty \to B\mathrm{Sp}(2)^n \to B(\mathrm{Sp}(2)^n/\Delta).$$

Since $H^*(\mathbb{R}P^\infty; \mathbb{Z}/2) = \mathbb{Z}/2[w]$ with $|w| = 1$,

$$H^*(\mathbb{R}P^\infty; \mathbb{Z}/2) = \Delta(w, \mathrm{Sq}^1 w, \mathrm{Sq}^2 \mathrm{Sq}^1 w)$$

for $* \le 7$, where $\Delta(a_1, \dots, a_k)$ denotes the simple system of generators in $a_1, \dots, a_k$. Clearly, $\tau(w) = x_2$ for a generator x_2 of $H^2(B(\mathrm{Sp}(2)^n/\Delta); \mathbb{Z}/2) \cong \mathbb{Z}/2$, where τ denotes the transgression. Then by [23, Corollary 6.9], $\mathrm{Sq}^1 w$ and $\mathrm{Sq}^2 \mathrm{Sq}^1 w$ are also transgressive, and so we get $H^*(B(\mathrm{Sp}(2)^n/\Delta); \mathbb{Z}/2)$ for $* \le 7$ as stated. It remains to show $\mathrm{Sq}^2 x_{4,k} = x_2 x_{4,k}$. Recall that

$$H^*(B\mathrm{SO}(n); \mathbb{Z}/2) = \mathbb{Z}/2[w_2, w_3, \dots, w_n],$$

$$\mathrm{Sq}^i w_j = \sum_{k=0}^{i} \binom{j+k-i-1}{k} w_{i-k} w_{j+k}, \tag{4-1}$$

where w_i is the i^{th} Stiefel–Whitney class. Then since $\mathrm{PSp}(2) \cong \mathrm{SO}(5)$,

$$H^*(B\mathrm{PSp}(2); \mathbb{Z}/2) = \mathbb{Z}/2[y_2, y_3, y_4, y_5], \quad \mathrm{Sq}^2 y_4 = y_2 y_4,$$

where $|y_i| = i$. Let $q_k \colon B(\mathrm{Sp}(2)^n/\Delta) \to B\mathrm{PSp}(2)$ denote the induced map of the k^{th} projection for $k = 1, 2, \dots, n$. Then $q_k^*(y_2) = x_2$ and $q_k^*(y_4) = x_{4,k}$. Thus we obtain $\mathrm{Sq}^2 x_{4,k} = x_2 x_{4,k}$, completing the proof. □

Proposition 4.3 *For* $* \leq 7$,

$$H^*(B\mathrm{PSp}(n); \mathbb{Z}/2) = \mathbb{Z}/2[x_2, x_3, x_4, x_5], \quad \mathrm{Sq}^2 x_4 = x_4 x_2, \quad |x_i| = i.$$

Proof We can compute the mod 2 cohomology of $B\mathrm{PSp}(2n)$ in the same way as in the proof of Lemma 4.2 by considering a homotopy fibration

$$\mathbb{R}P^\infty \to B\mathrm{Sp}(2n) \to B\mathrm{PSp}(2n).$$

Then it remains to show $\mathrm{Sq}^2 x_4 = x_4 x_2$. Let Δ be as in Lemma 4.2. Then there is an inclusion $i : \mathrm{Sp}(2)^n/\Delta \to \mathrm{PSp}(2n)$. Clearly, $i^*(x_2) = x_2$ and $i^*(x_4) = x_{4,1} + \cdots + x_{4,n}$. Then we obtain $\mathrm{Sq}^2 x_4 = x_4 x_2$ by Lemma 4.2. □

Now we prove:

Proposition 4.4 *If* $H = \mathrm{Sp}(n)$, *then* $\langle \epsilon, p \rangle$ *is of order* $s(G)$.

Proof Since the center of $\mathrm{Sp}(n)$ is isomorphic to $\mathbb{Z}/2$, we only consider

$$G = S^1 \times_{\mathbb{Z}/2} \mathrm{Sp}(n).$$

In this case, $s(G) = 2$, so by Lemma 3.3, it suffices to show $\langle \epsilon, p \rangle$ is nontrivial. First, we consider the case $G = S^1 \times_{\mathbb{Z}/2} \mathrm{Sp}(2n-1)$. The natural inclusion

$$\mathrm{Sp}(2n-1) \to \mathrm{SU}(4n-2)$$

sends the center of $\mathrm{Sp}(2n-1)$ injectively into the center of $\mathrm{SU}(4n-2)$. Then we get a homomorphism $G \to S^1 \times_{\mathbb{Z}/2} \mathrm{SU}(4n-2)$ which is an isomorphism in π_1. It is well known that the induced map $\pi_{8n-5}(\mathrm{Sp}(2n-1)) \to \pi_{8n-5}(\mathrm{SU}(4n-2))$ is an isomorphism; hence so is $\pi_{8n-5}(G) \to \pi_{8n-5}(S^1 \times_{\mathbb{Z}/2} \mathrm{SU}(4n-2))$. Then the proof of Proposition 4.1 implies that the Samelson product $\langle \epsilon, p \rangle$ is nontrivial.

Next, we consider $G = S^1 \times_{\mathbb{Z}/2} \mathrm{Sp}(2n)$. We apply Lemma 3.5 to $K = \mathrm{PSp}(2n)$ by setting $x = z = x_4$, $y = x_2$ and $\theta = \mathrm{Sq}^2$. By Proposition 4.3, the first and the second conditions of Lemma 3.5 are satisfied. The proof of Proposition 4.3 implies $\bar{q}^*(x_4)$ is nontrivial, where $H^4(B\mathrm{Sp}(2n); \mathbb{Z}/2) \cong QH^4(B\mathrm{Sp}(2n); \mathbb{Z}/2) \cong \mathbb{Z}/2$. Since the map

$$j^* \colon QH^4(B\mathrm{Sp}(2n); \mathbb{Z}/2) \to \Sigma QH^3(\mathrm{Sp}(2n); \mathbb{Z}/2)$$

is an isomorphism, we have $(\bar{q} \circ j)^*(x_4) \neq 0$. Moreover, for degree reasons, $(\bar{q} \circ j)^*(x_4)$ is not included in any element of $\theta(H^*(\Sigma\mathrm{Sp}(2n); \mathbb{Z}/2))$. Then the third condition of Lemma 3.5 is also satisfied. Thus $\langle \bar{\epsilon}, \bar{q} \rangle$ is nontrivial, and so by Lemma 3.4, $\langle \epsilon, p \rangle$ is nontrivial too. □

4.3 The case $H = \mathrm{Spin}(n)$

Finally, we consider the case $H = \mathrm{Spin}(n)$. We show some properties of the mod 2 cohomology of $B\mathrm{Spin}(n)$ that we are going to use. Recall that the mod 2 cohomology of $B\mathrm{SO}(n)$ is given as in (4-1).

Lemma 4.5 (1) *The mod 2 cohomology of $B\mathrm{Spin}(n)$ is given by*

$$H^*(B\mathrm{Spin}(n);\mathbb{Z}/2) = \mathbb{Z}/2[u_2, u_3, \dots, u_n, z]/(u_2, \mathrm{Sq}^{2^k}\,\mathrm{Sq}^{2^{k-1}} \cdots \mathrm{Sq}^1 u_2 \mid k \geq 0),$$

where $\bar{q}^*_{\mathrm{SO}(n)}(w_j) = u_j$, $|z| = 2^h$ for some $h > 0$ and $\mathrm{Sq}^i u_j$ is computed by replacing w_j with u_j in (4-1).

(2) *For $2 \leq i \leq n$ with $i \neq 2^k + 1$, $j^*_{\mathrm{Spin}(n)}(u_i) \neq 0$.*

Proof Item (1) is a result of Quillen [25]. We prove statement (2). It is well known that $(j')^*(w_i) \neq 0$ for $i = 2, 3, \dots, n$, where $j'\colon \Sigma\mathrm{SO}(n) \to B\mathrm{SO}(n)$ is the natural map. On the other hand, it is shown in [13] that $(\Sigma\bar{q}_{\mathrm{SO}(n)})^* \circ (j')^*(w_i) \neq 0$. Then for $2 \leq i \leq n$ with $i \neq 2^k + 1$,

$$0 \neq (\Sigma\bar{q}_{\mathrm{SO}(n)})^* \circ (j')^*(w_i) = j^* \circ \bar{q}_{\mathrm{SO}(n)}(w_i) = j^*(u_i). \qquad \square$$

The following lemma is easily deduced from the formula (4-1).

Lemma 4.6 *In $H^*(B\mathrm{SO}(n);\mathbb{Z}/2)$, we have:*

1. *If $n \equiv 0, 1 \bmod 4$, then $\mathrm{Sq}^2 w_i$ for $i = n-3, n-1$ are decomposable and $\mathrm{Sq}^2 w_{n-1}$ includes the term $w_2 w_{n-1}$.*
2. *If $n \equiv 2 \bmod 8$, then $\mathrm{Sq}^5 w_i$ for $i = n-4, n-9$ are decomposable and $\mathrm{Sq}^5 w_{n-4}$ includes the term $w_2 w_{n-1}$.*
3. *If $n \equiv 6 \bmod 8$, then $\mathrm{Sq}^3 w_i$ for $i = n-2, n-4$ are decomposable and $\mathrm{Sq}^3 w_{n-2}$ includes the term $w_2 w_{n-1}$.*
4. *If $n \equiv 3 \bmod 4$, then $\mathrm{Sq}^2 w_i$ for $i = n-2, n$ are decomposable and $\mathrm{Sq}^2 w_n$ includes the term $w_2 w_n$.*

Let C_n denote the center of $\mathrm{Spin}(n)$. Then we have:

1. $C_{2n+1} \cong \mathbb{Z}/2$ and $\mathrm{Spin}(2n+1)/C_{2n+1} \cong \mathrm{SO}(2n+1)$.
2. $C_{4n+2} \cong \mathbb{Z}/4$ and $\mathrm{Spin}(4n+2)/(\mathbb{Z}/2) \cong \mathrm{SO}(4n+2)$.
3. $C_{4n} \cong \mathbb{Z}/2 \times \mathbb{Z}/2$, $\mathrm{Spin}(4n)/(\mathbb{Z}/2 \times 1) \cong \mathrm{SO}(4n)$ and $\mathrm{Spin}(4n)/(1 \times \mathbb{Z}/2) \cong Ss(4n)$.

Proposition 4.7 *If $H = \mathrm{Spin}(n)$ and $K = \mathrm{SO}(n)$, then $\langle \epsilon, p\rangle$ is of order $s(G)$.*

Proof We only give a proof for n odd because the case n even is quite similarly proved. We apply Lemma 3.5 by setting $x = z = w_{n-1}$, $y = w_2$ and $\theta = \mathrm{Sq}^2$. By Lemma 4.6, the first and the second conditions of Lemma 3.5 are satisfied. By Lemmas 4.5 and 4.6, $(\bar{q} \circ j)^*(w_{n-1})$ is nontrivial and not included in any element of $\mathrm{Sq}^2(H^*(\Sigma\mathrm{Spin}(n); \mathbb{Z}/2))$. Then the third condition of Lemma 3.5 is also satisfied, so $\langle \bar{\epsilon}, \bar{q} \rangle \neq 0$. Thus, since $s(G) = 2$, Lemmas 3.3 and 3.4 complete the proof. □

Let $\mathrm{PO}(n) = \mathrm{Spin}(n)/C_n$. Then we have:

Corollary 4.8 *If $H = \mathrm{Spin}(4n+2)$ and $K = \mathrm{PO}(4n+2)$, then $\langle \epsilon, p \rangle$ is of order $s(G)$.*

Proof Let $\bar{\rho}\colon \mathrm{SO}(4n+2) \to \mathrm{PO}(4n+2)$ denote the projection. Then $\bar{\rho}_*(\bar{\epsilon}_{\mathrm{SO}(4n+2)}) = 2\bar{\epsilon}_{\mathrm{PO}(4n+2)}$. Since $S^1 \wedge \mathrm{Spin}(4n+2)$ is simply connected, the map

$$\bar{\rho}_*\colon [S^1 \wedge \mathrm{Spin}(4n+2), \mathrm{SO}(4n+2)] \to [S^1 \wedge \mathrm{Spin}(4n+2), \mathrm{PO}(4n+2)]$$

is an isomorphism. By definition, $\bar{q}_{\mathrm{PO}(4n+2)} = \bar{\rho} \circ \bar{q}_{\mathrm{SO}(4n+2)}$. So by Proposition 4.7,

$$2\langle \bar{\epsilon}_{\mathrm{PO}(4n+2)}, \bar{q}_{\mathrm{PO}(4n+2)} \rangle = \bar{\rho}_*(\langle \bar{\epsilon}_{\mathrm{SO}(4n+2)}, \bar{q}_{\mathrm{SO}(4n+2)} \rangle) \neq 0.$$

Then by Lemma 3.3, the order of $\langle \bar{\epsilon}_{\mathrm{PO}(4n+2)}, \bar{q}_{\mathrm{PO}(4n+2)} \rangle$ is a nonzero multiple of $s(G) = 4$. Thus the proof is complete by Lemmas 3.3 and 3.4. □

Let Δ denote the diagonal subgroup of $\mathbb{Z}/2 \times \mathbb{Z}/2$.

Proposition 4.9 *If $H = \mathrm{Spin}(4n)$ and $p_2(C) = 1 \times \mathbb{Z}/2, \Delta$, then $\langle \epsilon, p \rangle$ is of order $s(G)$.*

Proof By triality of $\mathrm{Spin}(8)$, the case $H = \mathrm{Spin}(8)$ is proved by Proposition 4.7. Then we assume $n > 2$. The mod 2 cohomology of $\mathrm{PO}(4n)$ was determined by Baum and Browder [3] such that

$$H^*(\mathrm{PO}(4n); \mathbb{Z}/2) = \mathbb{Z}/2[v]/(v^{2^r}) \otimes \Delta(u_1, \dots, \hat{u}_{2^r-1}, \dots, u_{n-1}), \quad \bar{\rho}^*(u_i) = w_i,$$

where $4n = 2^r(2m+1)$, $|v| = 1$ and $|u_i| = i$. The elements v and u_1 correspond respectively to generators of subgroups $1 \times \mathbb{Z}/2$ and $\mathbb{Z}/2 \times 1$ of $C_{4n} \cong \mathbb{Z}/2 \times \mathbb{Z}/2$. The Hopf algebra structure of $H^*(\mathrm{PO}(4n); \mathbb{Z}/2)$ was also determined such that

$$\bar{\phi}(v) = 0 \quad \text{and} \quad \bar{\phi}(u_i) = \sum_{j=1}^{i-1} \binom{i}{j} u_j \otimes v^{i-j},$$

where $\bar{\phi}$ is the reduced diagonal map. Let $\gamma\colon \mathrm{PO}(4n)^2 \to \mathrm{PO}(4n)$ denote the commutator map. Since $\bar{\epsilon}(v) \neq 0$, it suffices to show $\gamma^*(x)$ includes the term $v \otimes y$

such that $\rho^*(y) \neq 0$, where $\rho\colon \mathrm{Spin}(4n) \to \mathrm{PO}(4n)$ denotes the projection. Let $\mu\colon \mathrm{PO}(4n)^2 \to \mathrm{PO}(4n)$ and $\Delta\colon \mathrm{PO}(4n) \to \mathrm{PO}(4n)^2$ denote the multiplication and the diagonal map, respectively. Let $\iota\colon \mathrm{PO}(4n) \to \mathrm{PO}(4n)$ be a map given by $\iota(x) = x^{-1}$, and let $T\colon \mathrm{PO}(4n)^2 \to \mathrm{PO}(4n)^2$ be the switching map. Then

$$\gamma = \mu \circ (\mu \times \mu) \circ (1 \times 1 \times \iota \times \iota) \circ (1 \times T \times 1) \circ (\Delta \times \Delta).$$

Let $I_k = \widetilde{H}^*(\mathrm{PO}(n)^k; \mathbb{Z}/2)$. Now we compute $\gamma^*(u_i)$:

$$\begin{aligned}
u_i &\xmapsto{\mu^*} u_i \otimes 1 + 1 \otimes u_i + i\, u_{i-1} \otimes v \bmod I_2^3 \\
&\xmapsto{(\mu\times\mu)^*} i\,(u_{i-1}\otimes v\otimes 1\otimes 1 + 1\otimes 1\otimes u_{i-1}\otimes v + u_{i-1}\otimes 1\otimes 1\otimes v + 1\otimes u_{i-1}\otimes v\otimes 1) \\
&\qquad\qquad \bmod I_1\otimes 1\otimes I_1\otimes 1 + 1\otimes I_1\otimes 1\otimes I_1 + I_4^3 \\
&\xmapsto{(1\times 1\times\iota\times\iota)^*} i\,(u_{i-1}\otimes v\otimes 1\otimes 1 + 1\otimes 1\otimes u_{i-1}\otimes v + u_{i-1}\otimes 1\otimes 1\otimes v + 1\otimes u_{i-1}\otimes v\otimes 1) \\
&\qquad\qquad \bmod I_1\otimes 1\otimes I_1\otimes 1 + 1\otimes I_1\otimes 1\otimes I_1 + I_4^3 \\
&\xmapsto{(1\times T\times 1)^*} i\,(u_{i-1}\otimes 1\otimes v\otimes 1 + 1\otimes u_{i-1}\otimes 1\otimes v + u_{i-1}\otimes 1\otimes 1\otimes v + 1\otimes v\otimes u_{i-1}\otimes 1) \\
&\qquad\qquad \bmod I_1\otimes I_1\otimes 1\otimes 1 + 1\otimes 1\otimes I_1\otimes I_1 + I_4^3 \\
&\xmapsto{(\Delta\times\Delta)^*} i\,(u_{i-1}\otimes v + v\otimes u_{i-1}) \bmod I_1\otimes 1 + 1\otimes I_1 + I_2^3.
\end{aligned}$$

Then for n odd, $\gamma^*(u_7)$ includes the term $v \otimes u_6$, where $\rho^*(u_6) \neq 0$ by Lemma 4.5, and for n even, $\gamma^*(u_{11})$ includes the term $v \otimes u_{10}$, where $\rho^*(u_{10}) \neq 0$ by Lemma 4.5. Thus the Samelson product $\langle \bar{\epsilon}, \bar{q} \rangle$ is nontrivial, completing the proof by Lemmas 3.3 and 3.4 because $s(G) = 2$. □

5 Exceptional case

First, we consider the case $H = E_6$.

Proposition 5.1 *If $H = E_6$, then $\langle \epsilon, p \rangle$ is of order $s(G)$.*

Proof Since the center of E_6 is isomorphic to $\mathbb{Z}/3$, we only need to consider the case $G = S^1 \times_{\mathbb{Z}/3} E_6$. The mod 3 cohomology of $\mathrm{Ad}(E_6)$, which is the quotient of E_6 by its center, was determined by Kono [19] as

$$H^*(\mathrm{Ad}(E_6); \mathbb{Z}/3) = \mathbb{Z}/3[x_2, x_8]/(x_2^9, x_8^3) \otimes \Lambda(x_1, x_3, x_7, x_9, x_{11}, x_{16})$$

such that

$$\bar{\phi}(x_9) = x_8 \otimes x_1 + x_2 \otimes x_7 - x_2^3 \otimes x_3 + x_2^4 \otimes x_1 \quad \text{and} \quad \bar{q}^*(x_8) \neq 0,$$

where $|x_i| = i$. Then by the same computation as in the proof of Proposition 4.9, we can see that $\langle \bar{\epsilon}, \bar{q} \rangle$ is nontrivial. Thus by Lemmas 3.3 and 3.4, $\langle \epsilon, 1_G \rangle$ is of order $s(G) = 3$. □

Next, we consider the case $H = E_7$. Because the center of E_7 is isomorphic to $\mathbb{Z}/2$, we only need to consider the case $G = S^1 \times_{\mathbb{Z}/2} E_7$. The Hopf algebra structure of $H^*(\mathrm{Ad}(E_7); \mathbb{Z}/2)$ was determined by Ishitoya, Kono and Toda [13], from which we can see that the same computation as $\mathrm{Ad}(E_6)$ does not apply to $\mathrm{Ad}(E_7)$. So we apply Lemma 3.5. Kono and Mimura [21] showed that the mod 2 cohomology of $B\mathrm{Ad}(E_7)$ is generated by elements x_i for $i \in \{2, 3, 6, 7, 10, 11, 18, 19, 34, 35, 64, 66, 67, 96, 112\}$, where $|x_i| = i$. We determine $\mathrm{Sq}^2 x_6$.

Let $e_1, e_2, \dots, e_n$ be the standard basis of $\mathbb{R}^n$. Elements of the spin group $\mathrm{Spin}(n)$ are expressed by using $e_1, e_2, \dots, e_n$. See [1, Chapter 3]. Recall from [1, Proposition 4.2] that there are two representations

$$\Delta_{2n}^+, \Delta_{2n}^- \colon \mathrm{Spin}(2n) \to \mathrm{SU}(2^{n-1})$$

such that Δ_n^+ has weights $\frac{1}{2}(\pm x_1 \pm x_2 \pm \cdots \pm x_n)$ with even numbers of minus signs and Δ_n^- has weights $\frac{1}{2}(\pm x_1 \pm x_2 \pm \cdots \pm x_n)$ with odd numbers of minus signs.

Proposition 5.2 *There is a natural isomorphism*

$$\mathrm{Spin}(4) \cong \mathrm{Ker}\, \Delta_4^+ \times \mathrm{Ker}\, \Delta_4^-.$$

Proof There is a product decomposition $\mathrm{Spin}(4) \cong \mathrm{SU}(2) \times \mathrm{SU}(2)$ such that

$$\Delta_4^\pm \colon \mathrm{Spin}(4) \to \mathrm{SU}(2)$$

are identified with projections $\mathrm{SU}(2) \times \mathrm{SU}(2) \to \mathrm{SU}(2)$. □

As in [1, Theorem 6.1], there is a homomorphism

$$\theta \colon \mathrm{Spin}(16) \to E_8$$

whose kernel is $\{1, e_1 e_2 \cdots e_{16}\}$. Let $\mu \colon \mathrm{Spin}(4) \times \mathrm{Spin}(12) \to \mathrm{Spin}(16)$ denote the homomorphism covering the inclusion

$$\mathrm{SO}(4) \times \mathrm{SO}(12) \to \mathrm{SO}(16), \quad (A, B) \mapsto \begin{pmatrix} A & O \\ O & B \end{pmatrix}.$$

Define $\bar{\mu} = \theta \circ \mu \colon \mathrm{Spin}(4) \times \mathrm{Spin}(12) \to E_8$. Then

$$\mathrm{Ker}\, \bar{\mu} = \{(1, 1), (-1, -1), (e_1 e_2 e_3 e_4, e_5 e_6 \cdots e_{16}), (-e_1 e_2 e_3 e_4, -e_5 e_6 \cdots e_{16})\}.$$

Recall from [1, Chapter 8] that E_7 is defined as the centralizer of $\bar{\mu}(\mathrm{Ker}\,\Delta_4^+ \times 1)$ in E_8. Then, by Proposition 5.2, there is a homomorphism

$$\hat{\mu}\colon \mathrm{Ker}\,\Delta_4^- \times \mathrm{Spin}(12) \to E_7.$$

Since $-e_1e_2e_3e_4 \in \mathrm{Ker}\,\Delta_4^+$, $\bar{\mu}(-e_1e_2e_3e_4, 1)$ commutes with every element of E_7 in E_8. Moreover, $\bar{\mu}(-e_1e_2e_3e_4, 1) = \bar{\mu}(e_1e_2e_3e_4, -1) = \hat{\mu}(e_1e_2e_3e_4, -1)$, which belongs to E_7 and is not the unit of E_7. Then we obtain:

Proposition 5.3 *The center of E_7 is $\{1, \hat{\mu}(e_1e_2e_3e_4, -1)\}$.*

Let $L = (\mathrm{Ker}\,\Delta_4^- \times \mathrm{Spin}(12))/\{(1,1), (e_1e_2e_3e_4, -1)\}$. Then by Proposition 5.3, there is a map

$$\rho\colon L \to \mathrm{Ad}(E_7),$$

which is an isomorphism in the second mod 2 cohomology.

Lemma 5.4 *In $H^*(B\mathrm{Ad}(E_7); \mathbb{Z}/2)$, $\mathrm{Sq}^2 x_6$ is decomposable and includes the term x_2x_6.*

Proof By [21; 22], $(\bar{\mu} \circ (1 \times \bar{q}))^*(x_6)$ includes the term $1 \otimes u_6$, where u_i is as in Lemma 4.5. Note that the composition

$$\mathrm{Spin}(12) \to \mathrm{Ker}\,\Delta_4^- \times \mathrm{Spin}(12) \to L \xrightarrow{q_2} \mathrm{SO}(12)$$

is the natural projection, where q_2 is the second projection. Then by degree reasons,

$$\rho^*(x_6) + a\rho^*(x_2)^3 + b\rho^*(x_3)^2 = q_2^*(w_6)$$

for some $a, b \in \mathbb{Z}/2$. On the other hand, $q_2^*\colon H^2(B\mathrm{SO}(12); \mathbb{Z}/2) \to H^2(BL; \mathbb{Z}/2)$ is an isomorphism, implying $\rho^*(x_2) = q_2^*(w_2)$. Then since $\mathrm{Sq}^2 w_6 = w_2w_6$ by (4-1) and $\mathrm{Sq}^2 x_6$ is decomposable by degree reasons, $\mathrm{Sq}^2 x_6$ is decomposable and includes the term x_2x_6, as stated. □

We are ready to prove:

Proposition 5.5 *If $H = E_7$, then $\langle \epsilon, p \rangle$ is of order $s(G)$.*

Proof As mentioned above, we only need to consider $G = S^1 \times_{\mathbb{Z}/2} E_7$. We apply Lemma 3.5 by setting $x = z = x_6$, $y = x_2$ and $\theta = \mathrm{Sq}^2$. By Lemma 5.4, the first and second conditions of Lemma 3.5 are satisfied. As in [22], $\bar{q}^*(x_6)$ is a generator of $H^6(BE_7; \mathbb{Z}/2)$ such that $(\bar{q} \circ j)^*(x_6)$ is nontrivial. Then by degree reasons, the third condition of Lemma 3.5 is also satisfied, implying $\langle \bar{\epsilon}, \bar{q} \rangle$ is nontrivial. Since $s(G) = 2$, the proof is complete by Lemmas 3.3 and 3.4. □

6 Proofs of Theorems 1.3 and 1.4

This section proves Theorems 1.3 and 1.4. First, we prove Theorem 1.4.

Proof of Theorem 1.4 Suppose $H \cong H_1 \times \cdots \times H_k$, where each H_i is a simple Lie group. Let $r_i \colon S^1 \times H \to S^1 \times H_i$ be the projection, and let $G_i = (S^1 \times H_i)/(r_i(C))$ for $i = 1, 2, \dots, k$. By definition, $s(G)$ is the least common multiple of $s(G_1), \dots, s(G_k)$. Let $\bar{r}_i \colon G \to G_i$ and $\iota_i \colon S^1 \times H_i \to S^1 \times H$ denote the projection and the inclusion, respectively. Then $\bar{r}_i \circ \epsilon_G = \epsilon_{G_i}$ and $\bar{r}_i \circ p_G \circ \iota_i = p_{G_i}$, so

$$(1 \wedge \iota_i)^* \circ (\bar{r}_i)_*(\langle \epsilon_G, p_G \rangle) = \langle \bar{r}_i \circ \epsilon_G, \bar{r}_i \circ p_G \circ \iota_i \rangle = \langle \epsilon_{G_i}, p_{G_i} \rangle.$$

Thus the order of $\langle \epsilon_G, p_G \rangle$ is a nonzero multiple of the order of $\langle \epsilon_{G_i}, p_{G_i} \rangle$. So by Propositions 4.1, 4.4, 4.7, 5.1 and 5.5, the order of $\langle \epsilon_G, p_G \rangle$ is a nonzero multiple of $s(G_i)$ for $i = 1, 2, \dots, k$; hence so is $\langle \epsilon_G, 1_G \rangle$. Therefore, by Lemma 3.3, the proof is complete. □

Next, we prove Theorem 1.3.

Proof of Theorem 1.3 First, we prove the case $H = \mathrm{SU}(n)^r$. The implication (1) $\Longrightarrow$ (2) follows from Theorem 1.2. We prove the implication (2) $\Longrightarrow$ (1). Let $\partial_k \colon G \to \mathrm{map}_*(S^2, BG; k) \simeq \Omega_0 G$ be as in Section 2, and let $q_i \colon H \to \mathrm{SU}(n)$ be the projection onto the i^{th} $\mathrm{SU}(n)$. Then by Lemma 2.1, the proof of Proposition 4.1 implies that the image of the map

$$(\partial_k)_* \colon \pi_{2n-1}(G) \to \pi_{2n-1}(\Omega_0 G)$$

is isomorphic to $\prod_{i=1}^r \mathbb{Z}/\frac{n!}{(k,|q_i(C)|)}$, where $\pi_{2n-1}(\Omega_0 G) \cong (\mathbb{Z}/n!)^r$. By (2-1), there is an exact sequence

$$0 \to \prod_{i=1}^r \mathbb{Z}\Big/\frac{n!}{(k, |q_i(C)|)} \to \pi_{2n-1}(B\mathcal{G}_k(S^2, G)) \to \pi_{2n-1}(BG) \cong \pi_{2n-1}(B\mathrm{SU}(n)^r) = 0.$$

Then since $\pi_{2n-1}(B\mathcal{G}_k(S^2, G)) \cong \pi_{2n-2}(\mathcal{G}_k(S^2, G))$,

$$\pi_{2n-2}(\mathcal{G}_k(S^2, G)) \cong \prod_{i=1}^r \mathbb{Z}/(k, |q_i(C)|).$$

So if $\mathcal{G}_k(X, G) \simeq \mathcal{G}_l(X, G)$, then $\pi_{2n-2}(\mathcal{G}_k(S^2, G)) \simeq \pi_{2n-2}(\mathcal{G}_l(S^2, G))$, implying

$$\big(k, |q_1(C)|\big) \cdots \big(k, |q_r(C)|\big) = \big(l, |q_1(C)|\big) \cdots \big(l, |q_r(C)|\big).$$

As in the proof of Theorem 1.4, $s(G)$ is the least common multiple of

$$|q_1(C)|, \dots, |q_r(C)|.$$

Then it is easy to see that the above equality implies $(k, s(G)) = (l, s(G))$.

Next, we prove the case $H = \mathrm{SU}(4n-2)^s \times \mathrm{Sp}(2n-1)^t$. Note that

$$\pi_{8n-4}(\mathrm{Sp}(2n-1)) \cong \mathbb{Z}/2.$$

Then similarly to the above case, the proofs of Propositions 4.1 and 4.4 imply that the image of the map

$$(\partial_k)_* \colon \pi_{8n-5}(G) \to \pi_{8n-5}(\Omega_0 G)$$

is isomorphic to

$$\prod_{i=1}^{s} \mathbb{Z}\Big/\frac{(4n-2)!}{(k, |q_i(C)|)} \times \prod_{i=1}^{t} \mathbb{Z}\Big/\frac{2}{(k, q_i(C))}.$$

So we also get an exact sequence

$$\begin{aligned} 0 \to \prod_{i=1}^{s} \mathbb{Z}\Big/\frac{(4n-2)!}{(k, |q_i(C)|)} \times \prod_{i=1}^{t} \mathbb{Z}\Big/\frac{2}{(k, |q_i(C)|)} &\to \pi_{8n-5}(B\mathcal{G}_k(S^2, G)) \\ &\to \pi_{2n-1}(BG) \cong \pi_{8n-5}(B\mathrm{SU}(4n-2)^s \times B\mathrm{Sp}(2n-1)^t) = 0. \end{aligned}$$

Thus, by arguing as above, we obtain $(k, s(G)) = (l, s(G))$ whenever $\mathcal{G}_k(X, G) \simeq \mathcal{G}_l(X, G)$. Therefore, the proof is complete. □

References

[1] **J F Adams**, *Lectures on exceptional Lie groups*, Univ. Chicago Press (1996) MR Zbl

[2] **M F Atiyah**, **R Bott**, *The Yang–Mills equations over Riemann surfaces*, Philos. Trans. Roy. Soc. London Ser. A 308 (1983) 523–615 MR Zbl

[3] **P F Baum**, **W Browder**, *The cohomology of quotients of classical groups*, Topology 3 (1965) 305–336 MR Zbl

[4] **R Bott**, *A note on the Samelson product in the classical groups*, Comment. Math. Helv. 34 (1960) 249–256 MR Zbl

[5] **S B Bradlow**, **O García-Prada**, **P B Gothen**, *Homotopy groups of moduli spaces of representations*, Topology 47 (2008) 203–224 MR Zbl

[6] **M C Crabb**, **W A Sutherland**, *Counting homotopy types of gauge groups*, Proc. London Math. Soc. 81 (2000) 747–768 MR Zbl

[7] **T Cutler**, *The homotopy types of* $\mathrm{Sp}(3)$*–gauge groups*, Topology Appl. 236 (2018) 44–58 MR Zbl

[8] **G D Daskalopoulos**, **K K Uhlenbeck**, *An application of transversality to the topology of the moduli space of stable bundles*, Topology 34 (1995) 203–215 MR Zbl

[9] **D H Gottlieb**, *Applications of bundle map theory*, Trans. Amer. Math. Soc. 171 (1972) 23–50 MR Zbl

[10] **H Hamanaka**, **A Kono**, *Unstable* K^1*–group and homotopy type of certain gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 136 (2006) 149–155 MR Zbl

[11] **S Hasui**, **D Kishimoto**, **A Kono**, **T Sato**, *The homotopy types of* $\mathrm{PU}(3)$*– and* $\mathrm{PSp}(2)$*–gauge groups*, Algebr. Geom. Topol. 16 (2016) 1813–1825 MR Zbl

[12] **S Hasui**, **D Kishimoto**, **T So**, **S Theriault**, *Odd primary homotopy types of the gauge groups of exceptional Lie groups*, Proc. Amer. Math. Soc. 147 (2019) 1751–1762 MR Zbl

[13] **K Ishitoya**, **A Kono**, **H Toda**, *Hopf algebra structure of mod 2 cohomology of simple Lie groups*, Publ. Res. Inst. Math. Sci. 12 (1976/77) 141–167 MR Zbl

[14] **Y Kamiyama**, **D Kishimoto**, **A Kono**, **S Tsukuda**, *Samelson products of* $\mathrm{SO}(3)$ *and applications*, Glasg. Math. J. 49 (2007) 405–409 MR Zbl

[15] **D Kishimoto**, **A Kono**, *On the homotopy types of* $\mathrm{Sp}(n)$ *gauge groups*, Algebr. Geom. Topol. 19 (2019) 491–502 MR Zbl

[16] **D Kishimoto**, **A Kono**, **M Tsutaya**, *On* p*–local homotopy types of gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 144 (2014) 149–160 MR Zbl

[17] **D Kishimoto**, **I Membrillo-Solis**, **S Theriault**, *The homotopy types of* $\mathrm{SO}(4)$*–gauge groups*, Eur. J. Math. 7 (2021) 1245–1252 MR Zbl

[18] **D Kishimoto**, **S Theriault**, **M Tsutaya**, *The homotopy types of* G_2*–gauge groups*, Topology Appl. 228 (2017) 92–107 MR Zbl

[19] **A Kono**, *Hopf algebra structure of simple Lie groups*, J. Math. Kyoto Univ. 17 (1977) 259–298 MR Zbl

[20] **A Kono**, *A note on the homotopy type of certain gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 117 (1991) 295–297 MR Zbl

[21] **A Kono**, **M Mimura**, *On the cohomology mod 2 of the classifying space of* $\mathrm{Ad}E_7$, J. Math. Kyoto Univ. 18 (1978) 535–541 MR Zbl

[22] **A Kono**, **M Mimura**, **N Shimada**, *On the cohomology mod 2 of the classifying space of the 1–connected exceptional Lie group* E_7, J. Pure Appl. Algebra 8 (1976) 267–283 MR Zbl

[23] **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Studies in Advanced Mathematics 58, Cambridge Univ. Press (2001) MR Zbl

[24] **M Mimura**, **H Toda**, *Topology of Lie groups, I, II*, Translations of Mathematical Monographs 91, Amer. Math. Soc., Providence, RI (1991) MR Zbl

[25] **D Quillen**, *The mod 2 cohomology rings of extra-special 2–groups and the spinor groups*, Math. Ann. 194 (1971) 197–212 MR Zbl

[26] **S Rea**, *Homotopy types of gauge groups of* $\mathrm{PU}(p)$*–bundles over spheres*, J. Homotopy Relat. Struct. 16 (2021) 61–74 MR Zbl

[27] **S D Theriault**, *The homotopy types of* $\mathrm{Sp}(2)$*–gauge groups*, Kyoto J. Math. 50 (2010) 591–605 MR Zbl

[28] **S D Theriault**, *Odd primary homotopy decompositions of gauge groups*, Algebr. Geom. Topol. 10 (2010) 535–564 MR Zbl

[29] **S D Theriault**, *Homotopy decompositions of gauge groups over Riemann surfaces and applications to moduli spaces*, Internat. J. Math. 22 (2011) 1711–1719 MR Zbl

[30] **S Theriault**, *The homotopy types of* $\mathrm{SU}(5)$*–gauge groups*, Osaka J. Math. 52 (2015) 15–29 MR Zbl

[31] **S Theriault**, *Odd primary homotopy types of* $\mathrm{SU}(n)$*–gauge groups*, Algebr. Geom. Topol. 17 (2017) 1131–1150 MR Zbl

[32] **G W Whitehead**, *On products in homotopy groups*, Ann. of Math. 47 (1946) 460–475 MR Zbl

Department of Mathematical Sciences, Shibaura Institute of Technology
Saitama, Japan

Faculty of Mathematics, Kyushu University
Fukuoka, Japan

Department of Mathematics, Kyoto University
Kyoto, Japan

kameko@shibaura-it.ac.jp, kishimoto@math.kyushu-u.ac.jp,
takeda.masahiro.87u@kyoto-u.jp

Received: 3 August 2021 Revised: 6 February 2022

msp

Algebraic & Geometric Topology **23**:5 (2023) 2329–2345
DOI: 10.2140/agt.2023.23.2329
Published: 25 July 2023

Diffeomorphisms of odd-dimensional discs, glued into a manifold

JOHANNES EBERT

Let $\mu_M : B\mathrm{Diff}_\partial(D^{2n+1}) \to B\mathrm{Diff}(M)$, for a compact $(2n+1)$–dimensional smooth manifold M, be the map defined by extending diffeomorphisms on an embedded disc by the identity. By a classical result of Farrell and Hsiang, the rational homotopy groups and the rational homology of $B\mathrm{Diff}_\partial(D^{2n+1})$ are known in the concordance stable range. We prove two results on the behaviour of the map μ_M in the concordance stable range. Firstly, it is *injective* on rational homotopy groups, and secondly, it is *trivial* on rational homology if M contains sufficiently many embedded copies of $S^n \times S^{n+1} \setminus \mathrm{int}(D^{2n+1})$. We also show that μ_M is generally not injective on homotopy groups outside the stable range.

The homotopical statement is probably not new and follows from the theory of smooth torsion invariants. The noninjectivity outside the stable range is based on recent work by Krannich and Randal-Williams. The homological statement relies on work by Botvinnik and Perlmutter on diffeomorphisms of odd-dimensional manifolds.

57S05

1 Introduction

For a smooth compact manifold M with boundary, we denote by $\mathrm{Diff}(M)$ the topological group of diffeomorphisms of M, and by $\mathrm{Diff}_\partial(M) \subset \mathrm{Diff}(M)$ the subgroup of those diffeomorphisms which agree with the identity near ∂M. A celebrated classical result by Farrell and Hsiang [10] states that

(1.1) $$\pi_k(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \cong \begin{cases} \mathbb{Q} & \text{if } k \equiv 0 \pmod 4, \\ 0 & \text{if } k \not\equiv 0 \pmod 4, \end{cases}$$

in a range of degrees which was originally given by $k < \frac{1}{3}n$, but (1.1) holds more generally if $k \leq \phi^{\mathbb{Q}}(D^{2n})$, where $\phi^{\mathbb{Q}}(D^{2n})$ is the *rational concordance stable range for* D^{2n}, which we briefly recall.

For a compact smooth manifold M, let

$$C(M) := \mathrm{Diff}(M \times [0,1], M \times \{0\} \cup \partial M \times [0,1])$$

be the concordance diffeomorphism group of M, and let $\sigma_+ : C(M) \to C(M \times [0,1])$ be the (positive) suspension map defined in eg Igusa [16, Section 6.2]. Define $\phi(M)$ to be the largest integer k such that the maps $\sigma_+ : C(M \times [0,1]^m) \to C(M \times [0,1]^{m+1})$, for $m \geq 0$, are all k–connected. Similarly, define $\phi^{\mathbb{Q}}(M) \geq \phi(M)$ using rational connectivity instead of connectivity (this makes sense if $\dim(M) \geq 6$ as $\pi_0(C(M))$ is abelian in that case, by Hatcher and Wagoner [14, Lemma 1.1]).

Igusa's stability theorem [15, page 6] states that

$$\phi(M^d) \geq \min\big(\tfrac{1}{2}(d-7), \tfrac{1}{3}(d-4)\big). \tag{1.2}$$

Recent work by Krannich and Randal-Williams [21] gives the optimal range in which (1.1) holds. Corollary B of [21] shows that

$$\phi^{\mathbb{Q}}(D^d) = d-4 \quad \text{if } d \geq 10, \tag{1.3}$$

and hence (1.1) holds if $k \leq 2n-4$, provided that $n \geq 5$. Theorem A of [21] improves this to $k \leq 2n-3$, again for $n \geq 5$. These results slightly exceed Krannich [20, Corollary B].

For an arbitrary smooth compact and nonempty manifold M of dimension $2n+1$, choose an embedding $D^{2n+1} \to \operatorname{int} M$. Extending diffeomorphisms by the identity gives a gluing map

$$\mu_M^{\partial} : \mathrm{Diff}_\partial(D^{2n+1}) \to \mathrm{Diff}_\partial(M).$$

We may also consider the composition

$$\mu_M : \mathrm{Diff}_\partial(D^{2n+1}) \xrightarrow{\mu_M^{\partial}} \mathrm{Diff}_\partial(M) \to \mathrm{Diff}(M).$$

The purpose of this note is to study the effect of the maps $B\mu_M$ and $B\mu_M^{\partial}$ on rational homotopy and homology. The precise choice of the embedding does not play a role for this question as long as M is connected. This is because the homotopy class of $B\mu_M$ only depends on the isotopy class of the embedding. If M is connected and not orientable there is only one isotopy class of embeddings, and if M is connected and orientable there are two such isotopy classes, which differ by the reflection automorphism of the group $\mathrm{Diff}_\partial(D^{2n+1})$.

Theorem 1.4 (homotopical theorem) *For every $(2n+1)$–dimensional manifold M, the maps*

$$(\mu_M)_* : \pi_k(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \to \pi_k(B\mathrm{Diff}(M)) \otimes \mathbb{Q}$$

and

$$(\mu_M^\partial)_*\colon \pi_k(B\mathrm{Diff}_\partial(D^{2n+1}))\otimes\mathbb{Q}\to\pi_k(B\mathrm{Diff}_\partial(M))\otimes\mathbb{Q}$$

are injective when $k\neq 1$ *and* $k\leq\phi^{\mathbb{Q}}(D^{2n})$.

Remark 1.5 Note that $\pi_1(B\mathrm{Diff}_\partial(D^{2n+1}))=\pi_0(\mathrm{Diff}_\partial(D^{2n+1}))$ is always a finite group; this is trivial when $n=0$ and follows from Cerf [7] for $n=1$. For $n\geq 2$, Cerf [8, corollaire 2] and the h-cobordism theorem identify $\pi_0(\mathrm{Diff}_\partial(D^{2n+1}))$ with the group of homotopy $(2n+2)$–spheres, which is finite by Kervaire and Milnor [19].

Remark 1.6 By [21, Corollary B], Theorem 1.4 holds for $k\leq 2n-4$ if $n\geq 5$. Theorem 1.4 is also true for $k=2n-3$ and $n\geq 5$, since $\pi_{2n-3}(B\mathrm{Diff}_\partial(D^{2n+1})\otimes\mathbb{Q}=0$ for such n by [21, Theorem A].

Theorem 1.4 could have been proven with little effort in Badzioch, Dorabiała, Klein and Williams [1] at latest. It was in fact known by experts and we learnt the statement from Mauricio Bustamante. The proof is given here for the sake of completeness and to contrast it with our main result (Theorem 1.7 below), which seemingly goes the opposite direction.

Our result concerns the effect of μ_M in rational homology. Since $B\mathrm{Diff}_\partial(D^{2n+1})$ is a connected E_{2n+1}–space (and hence a homotopy commutative H-space), we have that $H_*(B\mathrm{Diff}_\partial(D^{2n+1});\mathbb{Q})$ (with the Pontryagin product) is the free graded-commutative algebra generated by $\pi_*(B\mathrm{Diff}_\partial(D^{2n+1}))\otimes\mathbb{Q}$. Therefore, in the concordance stable range, $H_*(B\mathrm{Diff}_\partial(D^{2n+1});\mathbb{Q})$ is a polynomial algebra with one generator in each dimension divisible by 4.

Let

$$U_g^n:=\#^g(S^n\times S^{n+1})$$

be the connected sum of g copies of $S^n\times S^{n+1}$, and let

$$U_{g,1}^n:=U_g^n\setminus\mathrm{int}\,D^{2n+1}$$

be U_g^n with the interior of a disc removed.

Theorem 1.7 (homological theorem) *Let M be a connected manifold of dimension $2n+1\geq 9$ and suppose that M contains an embedded copy of $U_{g,1}^n$. Then the maps*

$$(\mu_M)_*\colon \widetilde{H}_k(B\mathrm{Diff}_\partial(D^{2n+1});\mathbb{Q})\to\widetilde{H}_k(B\mathrm{Diff}(M);\mathbb{Q})$$

and

$$(\mu_M^\partial)_*\colon \widetilde{H}_k(B\mathrm{Diff}_\partial(D^{2n+1});\mathbb{Q})\to\widetilde{H}_k(B\mathrm{Diff}_\partial(M);\mathbb{Q})$$

are trivial if $k\leq\phi^{\mathbb{Q}}(D^{2n})+1$ *and* $k\leq\frac{1}{2}(g-4)$.

Finally, using the recent work [21], we can show that the range for the validity of Theorem 1.4 given in Remark 1.6 is optimal.

Theorem 1.8 *For even $n \geq 6$, there is a closed $(2n+1)$–dimensional smooth manifold M such that the kernel of*

$$(\mu_M)_* \colon \pi_{2n-2}(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \to \pi_{2n-2}(B\mathrm{Diff}(M)) \otimes \mathbb{Q}$$

is nonzero.

Acknowledgements

I am grateful to Mauricio Bustamante for stating Theorem 1.4 as a fact to me, to Markus Land and Oscar Randal-Williams for some remarks, in particular concerning the proof of Corollary 3.4, and to Manuel Krannich for his detailed feedback on the entire paper.

I am particularly indebted to the referee for an exceptionally speedy and thorough report, and in particular the suggestion to improve the main results by using Krannich and Randal-Williams [21], and Theorem 1.8 together with its proof.

I was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project ID 427320536, SFB 1442, as well as Germany's Excellence Strategy EXC 2044 – 390685587, Mathematics Münster: Dynamics–Geometry–Structure.

2 Proof of the homotopical theorem

The proof of Theorem 1.4 relies on higher torsion invariants as axiomatized by Igusa [18], and we review some background beforehand. Let $K(\mathbb{Z})$ be the algebraic K–theory spectrum of $\mathbb{Z}$ and let

$$u \colon Q(S^0) \to \Omega^\infty K(\mathbb{Z})$$

be the unit map on infinite loop spaces.

Let M be a finite CW complex of dimension d, and let $\pi : E \to B$ be a fibration with fibres homotopy equivalent to M. Let

(2.1) $$\rho_k(\pi) \colon B \to B\mathrm{GL}(H_k(M;\mathbb{Z}))$$

be the map induced by the monodromy action of the fundamental group on the homology of the fibre. The $\mathbb{Z}$–module $H_k(M;\mathbb{Z})$ is finitely generated, and hence there is a canonical map

$$\iota \colon B\mathrm{GL}(H_k(M;\mathbb{Z})) \to \Omega^\infty K(\mathbb{Z}).$$

Such a map exists even if the homology groups are not free, essentially because $\mathbb{Z}$ is a regular ring (each finitely generated $\mathbb{Z}$–module has a finite length resolution by projective finitely generated $\mathbb{Z}$–modules). See the discussion leading up to [9, Proposition 6.7] for details.

The map ι hits the component of $[H_k(M;\mathbb{Z})] \in K_0(\mathbb{Z}) = \pi_0(\Omega^\infty K(\mathbb{Z})) \cong \mathbb{Z}$. The algebraic K–theory Euler characteristic of the fibration π is the alternating sum

$$\chi(\pi) := \sum_{k=0}^{d} (-1)^k \iota \circ \rho_k(\pi)\colon B \to \Omega^\infty K(\mathbb{Z}),$$

where $d = \dim(M)$ and we have used the H–space structure on $\Omega^\infty K(\mathbb{Z})$ to form the sum. Of course, $\chi(\pi)$ hits the component indexed by $\chi(M) \in \mathbb{Z} = K_0(\mathbb{Z})$.

The fibration π has an associated transfer map [3]

$$\mathrm{trf}_\pi\colon \Sigma^\infty B_+ \to \Sigma^\infty E_+$$

on the level of suspension spectra; we mostly consider its adjoint, also written

$$\mathrm{trf}_\pi\colon B \to Q(E_+).$$

The Dwyer–Weiss–Williams index theorem [9, Corollary 8.12] implies that if π is a *smooth* fibre bundle, the diagram

(2.2)
$$\begin{tikzcd} B \arrow[r, "\mathrm{trf}_\pi"] \arrow[dr, "\chi(\pi)"'] & Q(E_+) \arrow[d, "u\circ c"] \\ & \Omega^\infty K(\mathbb{Z}) \end{tikzcd}$$

commutes, up to a preferred homotopy (here $u \circ c$ is the composition of the unit map with the collapse map $c\colon Q(E_+) \to Q(S^0)$).

Remark 2.3 Actually, Theorem 8.5 of [9] proves a stronger version involving the algebraic K–theory $A(E)$ of the space E and a fibrewise version thereof. Raptis and Steimle gave a substantially simpler proof of the homotopy-commutativity of (2.2) in [24]; they also showed [9, Theorem 8.5] for smooth bundles in [25].

The diagram (2.2) can be used to define secondary invariants under additional hypotheses on the bundle π; we follow the approach of [2] here, with some modifications. The extra hypothesis to be made is that the monodromy action of $\pi_1(B)$ on $H_k(M;\mathbb{Z})$ is unipotent for all k (in [2] the authors consider homology with coefficients in a field, but the construction generalizes to regular rings such as $\mathbb{Z}$; see [2, Remark 6.11]). Moreover, we assume as in [2] that the base space B is a compact manifold, possibly with boundary.

Under these assumptions, the map $\chi(\pi)$ comes with a preferred homotopy to the constant map to the point

$$\chi(M) := \sum_{k=0}^{d} (-1)^k [H_k(M;\mathbb{Z})] \in \Omega^\infty K(\mathbb{Z});$$

see [2, Theorem 6.7]. Combining this homotopy with the preferred homotopy from (2.2) yields a map

(2.4) $$T(\pi)\colon B \to \mathrm{hofib}_{\chi(M)}(u) \simeq \mathrm{hofib}_0(u),$$

where we use the infinite loop space structures to identify the homotopy fibres. Using Borel's computation [5] of $\pi_*(\Omega^\infty K(\mathbb{Z})) \otimes \mathbb{R}$, we define characteristic classes of unipotent smooth bundles as follows. Note that

(2.5) $$H^*(\mathrm{hofib}_0(u);\mathbb{R}) = \mathbb{R}[a_4, a_8, \dots]$$

for certain generators a_{4k} of degree $4k$ (the transgression of a_k is the Borel class in $H^{4k+1}(\Omega^\infty K(\mathbb{Z});\mathbb{R})$). Following [2, Section 7] (but using the notation of [1]), define

(2.6) $$t^s_{4k}(\pi) := T(\pi)^* a_{4k} \in H^{4k}(B;\mathbb{R}).$$

It is convenient for us to replace the coefficient field by $\mathbb{Q}$, which can be done as follows. First, the Borel class comes from a spectrum cohomology class $b_k \in H^{4k+1}(K(\mathbb{Z});\mathbb{R})$. Second, Borel showed that $H^{4k+1}(K(\mathbb{Z});\mathbb{R})$ is 1–dimensional, and so there is $\alpha_k \in \mathbb{R}^\times$ such that $\alpha_k b_k$ lies in $H^*(K(\mathbb{Z});\mathbb{Q})$. We now define

(2.7) $$\bar{t}^s_{4k}(\pi) = \bar{t}^s_{4k}(E) := \alpha_k t^s_{4k}(\pi) \in H^{4k}(B;\mathbb{Q}).$$

The construction of (2.4) is given in [2] only for compact manifold bases; the definition of (2.7) can be generalized to arbitrary base spaces as follows. For an arbitrary unipotent bundle $E \to B$, we define $\bar{t}^s_{4k}(E) \in H^{4k}(B;\mathbb{Q})$ as the class corresponding to the homomorphism

(2.8) $$\Omega^{\mathrm{fr}}_{4k}(B) \to \mathbb{Q};\ [X, f] \mapsto \langle \bar{t}^s_{4k}(f^*E); [X] \rangle$$

from the framed bordism group of B under the isomorphism

$$H^{4k}(B;\mathbb{Q}) \cong \mathrm{Hom}(\Omega^{\mathrm{fr}}_{4k}(B);\mathbb{Q})$$

(we need [2, Proposition 7.3] to show that (2.8) is well defined).

Now let $\mathrm{Torr}(M) \subset \mathrm{Diff}(M)$ be the *Torelli diffeomorphism group*, ie the subgroup of those diffeomorphisms which act as the identity on $H_*(M;\mathbb{Z})$ and $H_*(M,\partial M;\mathbb{Z})$. The universal M–bundle over $B\mathrm{Torr}(M)$ is clearly unipotent, and combining the classes $\bar{t}^s_{4k}$, we obtain a map

$$\tau_M \colon B\mathrm{Torr}(M) \to \prod_{k\geq 1} K(\mathbb{Q},4k). \tag{2.9}$$

Observing that $\mathrm{Torr}(D^{2n+1}) = \mathrm{Diff}^+(D^{2n+1})$ is the group of orientation-preserving diffeomorphisms, we obtain in particular

$$\begin{aligned}\tau_{D^{2n+1}} \colon B\mathrm{Diff}_\partial(D^{2n+1}) &\to B\mathrm{Diff}^+(D^{2n+1}) \\ &= B\mathrm{Torr}(D^{2n+1}) \to \prod_{k\geq 1} K(\mathbb{Q},4k).\end{aligned} \tag{2.10}$$

Farrell and Hsiang's theorem may be restated as follows.

Theorem 2.11 *The map $\tau_{D^{2n+1}}$ induces an isomorphism on rational homotopy groups in degrees at most $\phi^{\mathbb{Q}}(D^{2n})$.*

Proof The map $\tau_{D^{2n+1}}$ factors through

$$B\mathrm{Diff}_\partial(D^{2n+1}) \to BC(D^{2n}) \to B\mathrm{Diff}^+(D^{2n+1}).$$

Consider the diagram

$$\begin{array}{ccc} BC(D^{2n}) & \longrightarrow & B\mathrm{Diff}^+(D^{2n+1}) \\ \Big\downarrow \sigma_+^{\circ m} & & \Big\downarrow \times\, \mathrm{id}_{D^m} \\ BC(D^{2n+m}) & \longrightarrow & B\mathrm{Diff}^+(D^{2n+m+1}) \end{array} \tag{2.12}$$

The left vertical map is a composition of the suspension map and the right vertical map is given by taking products with D^m and the identification $D^{2n+m+1} = D^{2n+1} \times D^m$. The square commutes up to homotopy by the definition of the suspension map σ_+. A special case of [1, Theorem 7.1] states that $\bar{t}^s_{4k}(E \times D^m) = \bar{t}^s_{4k}(E)$ for each unipotent bundle $E \to B$. It follows that $\tau_{D^{2n+1}} \colon B\mathrm{Diff}_\partial(D^{2n+1}) \to \prod_{k\geq 1} K(\mathbb{Q},4k)$ factors as

$$\begin{aligned}B\mathrm{Diff}_\partial(D^{2n+1}) \to BC(D^{2n}) &\to B\mathcal{C}(D^{2n}) \\ &:= \mathrm{hocolim}_m BC(D^{2n+m}) \to \prod_{m\geq 1} K(\mathbb{Q},4k).\end{aligned} \tag{2.13}$$

All three maps in (2.13) induce isomorphisms on rational homotopy up to degree $\phi^{\mathbb{Q}}(D^{2n})$. This is true for the second map by definition.

The third map is a rational equivalence. Since the Whitehead group of $\pi_1(D^{2n})$ is trivial, $B\mathcal{C}(D^{2n})$ is the stable h–cobordism space $\mathcal{H}(D^{2n})$. The stable h–cobordism theorem [27] states an equivalence $\mathcal{H}(D^{2n}) \simeq \mathrm{hofib}(Q(S^0) \to A(*))$, and the linearization map from $A(*)$ to $\Omega^\infty K(\mathbb{Z})$ induces a rational equivalence

$$\mathrm{hofib}(Q(S^0) \to A(*)) \to \mathrm{hofib}(Q(S^0) \to \Omega^\infty K(\mathbb{Z})).$$

Together with [5], this shows that the rational homotopy groups of $B\mathcal{C}(D^{2n})$ and of $\prod_{m\geq 1} K(\mathbb{Q}, 4k)$ are abstractly isomorphic and at most 1–dimensional.

To conclude that the third map in (2.13) is a rational isomorphism, it is therefore enough to prove that the induced map on rational homotopy groups is nontrivial whenever its target is nonzero, and this amounts to proving that for each $k \geq 1$, there is an m and an element in $\pi_{4k}(BC(D^{2n+m}))$ such that $\bar{t}^s_{D^{2n+m+1}}$ is nontrivial on that element. This was done by Igusa in [16, Theorem 6.4.2], but with the *higher Franz–Reidemeister torsion* classes $t^{\mathrm{FR}}_{4k} \in H^{4k}(BC(D^{2n+m}); \mathbb{R})$ in place of $\bar{t}^s_{4k}$. These were constructed using ideas from Morse theory in [16, Section 5.7.2] (for bundles with structure group $\mathrm{Torr}(M)$) and in [17, Section 2.11] for unipotent bundles. The main theorem of [1] shows that there is a universal constant $\lambda_{4k} \in \mathbb{R}^\times$ such that $\bar{t}^s_{4k}(\pi) = \lambda_{4k} t^{\mathrm{IK}}_{4k}(\pi)$ for all unipotent bundles over compact manifold bases, and so the third map in (2.13) is a rational equivalence.

It is shown in [16, Section 6.5] that the first map in (2.13) induces an isomorphism on rational homotopy groups up to degree $\phi^{\mathbb{Q}}(D^{2n})$. In loc. cit., the result is stated in terms of (1.2), so we give a few more details here. Consider the fibre sequence

$$\mathrm{Diff}_\partial(D^{2n+1}) \to C(D^{2n}) \to \mathrm{Diff}_\partial(D^{2n}).$$

The two maps are compatible with the following involutions on the spaces: the group inversion on $\mathrm{Diff}_\partial(D^{2n})$, an involution defined at the beginning of [16, Section 6.5] on $C(D^{2n})$, and the involution

$$I : \mathrm{Diff}_\partial(D^{2n+1}) \to \mathrm{Diff}_\partial(D^{2n+1}) \tag{2.14}$$

given by conjugation with the reflection map

$$r(x_1, \ldots, x_{2n+1}) := (x_1, \ldots, x_{2n}, -x_{2n+1}).$$

The rational homotopy sequence of the fibration splits into negative and positive eigenspaces of these involutions. An Eckmann–Hilton argument proves that

$$\pi_*^+(\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \to \pi_*^+(C(D^{2n})) \otimes \mathbb{Q}$$

is an isomorphism in all degrees. On the other hand, $I_* = \mathrm{id}$ on $\pi_k(\_) \otimes \mathbb{Q}$ for $k \le \phi^{\mathbb{Q}}(D^{2n})$; Corollary 6.5.3 of [16] states this when k is in the range given by (1.2), but the proof clearly works for $k \le \phi^{\mathbb{Q}}(D^{2n})$. Hence, for $k \le \phi^{\mathbb{Q}}(D^{2n}) + 1$,

$$\pi_k(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} = \pi_k^+(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \cong \pi_k^+(BC(D^{2n})) \otimes \mathbb{Q}.$$

Finally, $\pi_k^-(BC(D^{2n})) \otimes \mathbb{Q} = 0$ for $k \le \phi^{\mathbb{Q}}(D^{2n})$. To see this, observe that the stabilization map $BC(D^d)) \to BC(D^{d+1})$ switches the eigenspaces of the involutions by [16, Lemma 6.5.1]. Hence it is enough to check that $\pi_k^-(BC(D^{2n})) \otimes \mathbb{Q} = 0$ for very large n, and this follows from Theorem 6.4.2 and Lemma 6.5.4 of [16], using that the third map in (2.13) is a rational equivalence. □

Proof of Theorem 1.4 for closed M We first consider the case where M is closed. The cohomology classes $\bar{t}^s_{4k}$ have the following additivity property: for unipotent bundles $\pi_j : E_j \to B$ with common (vertical) boundary bundle $E_{01} = \partial E_0 = \partial E_1$, the boundary bundle $\pi_{01} : E_{01} \to B$ and the glued bundle $\pi : E = E_0 \cup_{\partial E_j} E_1 \to B$ are also unipotent, and

$$\bar{t}^s_{4k}(\pi) + \bar{t}^s_{4k}(\pi_{01}) = \bar{t}^s_{4k}(\pi_1) + \bar{t}^s_{4k}(\pi_2) \in H^{4k}(B;\mathbb{Q}). \tag{2.15}$$

This is proven for B a compact manifold in [1, Corollary 5.2], the case of a general base follows by using framed bordism as in the construction of $\bar{t}^s_{4k}$ for general base spaces.

Next, $\mu_M : B\mathrm{Diff}_\partial(D^{2n+1}) \to B\mathrm{Diff}(M)$ lifts to $\tilde{\mu}_M : B\mathrm{Diff}_\partial(D^{2n+1}) \to B\mathrm{Torr}(M)$, and (2.15) shows that

$$\tau_M \circ \tilde{\mu}_M \sim \tau_{D^{2n+1}}. \tag{2.16}$$

By Theorem 2.11, it follows that $\tilde{\mu}_M$ is injective on $\pi_k(\_) \otimes \mathbb{Q}$ when $k \le \phi^{\mathbb{Q}}(D^{2n})$. Because $\mathrm{Torr}(M) \subset \mathrm{Diff}(M)$ is a union of path components,

$$p_* : \pi_k(B\mathrm{Torr}(M)) \to \pi_k(B\mathrm{Diff}(M))$$

is injective when $k = 1$ and an isomorphism when $k \ge 2$, and $\mu_M = p \circ \tilde{\mu}_M$ is injective on rational homotopy groups up to degree $\phi^{\mathbb{Q}}(D^{2n})$. □

We have used that M is closed in order to apply (2.15), which in the quoted source is only covered for closed M. The case of a general M reduces to the closed case by "doubling". Let M be a manifold with boundary, let $A \subset \partial M$ be a compact codimension-0 submanifold, and form $M \cup_{\partial M - \mathrm{int}\, A} M$. This is a manifold with boundary $A \cup_{\partial A} A$. Let $\mathrm{Diff}_A(M)$ be the group of diffeomorphisms which fix A pointwise. There is a doubling map

$$d_A \colon B\mathrm{Diff}_A(M) \to B\mathrm{Diff}_\partial(M \cup_{\partial M - \mathrm{int}\, A} M)$$

given by extending a diffeomorphism with its reflection. The diagram

$$\begin{tikzcd} B\mathrm{Diff}_\partial(D^{2n+1}) \arrow[r, "\mu_M"] \arrow[d, "d_{D^{2n}}"] & B\mathrm{Diff}_A(M) \arrow[d, "d_A"] \\ B\mathrm{Diff}_\partial(D^{2n+1} \cup_{D^{2n}} D^{2n+1}) \arrow[r, "\mu_{M \cup_{\partial M - \mathrm{int}\, A} M}"] & B\mathrm{Diff}_\partial(M \cup_{\partial M - \mathrm{int}\, A} M) \end{tikzcd} \tag{2.17}$$

commutes up to homotopy, where $D^{2n} \subset \partial D^{2n+1}$ denotes a half disc in the boundary.

Lemma 2.18 *The doubling map induces an isomorphism*

$$(d_{D^{2n}})_* \colon \pi_k(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \to \pi_k(B\mathrm{Diff}_\partial(D^{2n+1} \cup_{D^{2n}} D^{2n+1})) \otimes \mathbb{Q}$$

when $k \leq \phi^{\mathbb{Q}}(D^{2n}) + 1$.

Proof By the Eckmann–Hilton argument, the effect of d on rational homotopy groups is the map

$$1 + (BI)_* \colon \pi_*(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q} \to \pi_*(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q}$$

(here I is the involution (2.14), and we identify $D^{2n+1} \cup_{D^{2n}} D^{2n+1} = D^{2n+1}$). The lemma now follows from the fact that $BI_* = \mathrm{id}$ on $\pi_k(\_) \otimes \mathbb{Q}$ for $k \leq \phi^{\mathbb{Q}}(D^{2n}) + 1$ (see the proof of Theorem 2.11 above for more details). □

Remark 2.19 The bound given in Lemma 2.18 is optimal: [21, Corollary 8.4] shows that the involution acts nontrivially on $\pi_{2n-2}(B\mathrm{Diff}_\partial(D^{2n+1})) \otimes \mathbb{Q}$ when $n \geq 5$, while $\phi^{\mathbb{Q}}(D^{2n}) + 1 = 2n - 3$ for such n.

Proof of Theorem 1.4 for general M To prove the statement for μ_M, use (2.17) with $A = \varnothing$ and apply Lemma 2.18. The statement for μ_M^∂ follows from that for μ_M in view of the definition of μ_M. □

Remark 2.20 From the proof of Theorem 1.4 given above, one can also deduce a statement about $H_*(B\mathrm{Torr}(M); \mathbb{Q})$, namely that $\tilde{\mu}_M$ is injective on rational homology in the concordance stable range, at least when M is closed or orientable.

In the case where M is closed, this follows from (2.16) and Theorem 2.11. For manifolds with nonempty boundary, a bit more care is needed to check that doubling really gives a map $\mathrm{Torr}(M) \to \mathrm{Torr}(M \cup_{\partial M} M)$. For oriented M, the argument goes as follows.

Mapping both halves of $M \cup_{\partial M} M$ to M and excision in homology gives a $\mathrm{Diff}(M)$–equivariant isomorphism $H_*(M \cup_{\partial M} M; \mathbb{Z}) \cong H_*(M; \mathbb{Z}) \oplus H_*(M, \partial M; \mathbb{Z})$, from which it follows that the double of f also induces the identity on homology.

We leave it to the reader to figure out statements in cohomology or a variant for nonorientable M.

3 Proof of the homological theorem

We now turn to the proof of Theorem 1.7, which relies on work by Botvinnik and Perlmutter [6; 23]. To state their results, let

$$V_g^n := \natural^g (S^n \times D^{n+1})$$

be the boundary connected sum of g copies of $S^n \times D^{n+1}$, and let $D = D^{2n} \subset \partial V_g^n$ be a disk in the boundary of V_g^n. Note that $V_0^n = D^{2n+1}$. There is a stabilization map

(3.1) $$B\mathrm{Diff}_D(V_g^n) \to B\mathrm{Diff}_D(V_{g+1}^n),$$

given by taking the boundary connected sum with $S^n \times D^{n+1}$ at D and extending diffeomorphisms by the identity. Perlmutter [23, Theorem 1.1] proved that the map (3.1) induces an isomorphism in homology in degrees $* \le \frac{1}{2}(g-4)$ provided that $n \ge 4$. Botvinnik and Perlmutter [6] computed the homology of $B\mathrm{Diff}_D(V_g^n)$ in the stable range. Let

$$\theta_n \colon BO(2n+1)\langle n \rangle \to BO(2n+1)$$

be the n–connected cover of $BO(2n+1)$. Let $\pi : E \to B$ be a bundle with fibre V_g^n and structure group $\mathrm{Diff}_D(V_g^n)$. The vertical tangent bundle $T_v E$ admits a θ_n–structure, ie a bundle map $\ell \colon T_v E \to \theta_n^* \gamma_{2n+1}$ to the pullback of the universal bundle over $BO(2n+1)$. This θ_n–structure is unique up to contractible choice once the following condition is imposed. Inside E, there is a trivial D–subbundle $B \times D$. The restriction of the vertical tangent bundle $T_v E$ to $B \times D$ has a canonical trivialization, and one requires that ℓ is compatible with that trivialization (see [6, Proposition 6.16] for all this).

Let $\underline{\ell}\colon E \to BO(2n+1)\langle n\rangle$ be the map of spaces underlying ℓ. Let

$$\alpha_\pi\colon B \xrightarrow{\mathrm{trf}_\pi} Q(E_+) \xrightarrow{Q(\underline{\ell})} Q(BO(2n+1)\langle n\rangle_+)$$

be the composition of the transfer with the map induced by $\underline{\ell}$. In particular, we can apply this construction to the universal bundle over $B\mathrm{Diff}_D(V_g^n)$ and obtain a map

(3.2) $$\alpha_g\colon B\mathrm{Diff}_D(V_g^n) \to Q_{1+(-1)^n g}(BO(2n+1)\langle n\rangle_+).$$

The target is the path component indexed by

$$\chi(V_g^n) = 1+(-1)^n g \in \mathbb{Z} = \pi_0\big(Q(BO(2n+1)\langle n\rangle_+)\big).$$

The following is essentially [6, Corollary B].

Theorem 3.3 (Botvinnik and Perlmutter) *Let $n \geq 4$. Then the map (3.2) induces an isomorphism in integral homology in degrees $* \leq \frac{1}{2}(g-4)$.*

Theorem 3.3 as stated above differs from the formulation given in [6] in so far as loc. cit. does not mention the transfer at all, so some remarks have to be made here. For a fibration $\theta\colon X \to BO(d)$, Genauer [13] introduced the cobordism category $\mathrm{Cob}_\theta^\partial$ of $(d-1)$–dimensional θ–manifolds with boundaries and their cobordisms (which have corners). He proved that there is a weak equivalence $B\mathrm{Cob}_\theta^\partial \simeq \Omega^{\infty-1}\Sigma^\infty X_+$, and the equivalence is given by a parametrized Pontryagin–Thom construction (this result is parallel to the well-known result [12] for the usual cobordism category). Given any bundle $\pi\colon E \to B$ of smooth compact d–manifolds with boundary equipped with a θ–structure ℓ on the vertical tangent bundle, one obtains a tautological map $B \to \Omega B\mathrm{Cob}_\theta^\partial$, and from the description of the transfer for smooth bundles, one sees that the composition of this tautological map with Genauer's equivalence agrees with the composition $B \xrightarrow{\mathrm{trf}_\pi} Q(E_+) \xrightarrow{Q(\ell_+)} Q(X_+)$. Using this observation, one derives Theorem 3.3 from the results of [6].

Corollary 3.4 *If $n \geq 4$, the iterated stabilization map*

$$B\mathrm{Diff}_D(V_0^n) \to B\mathrm{Diff}_D(V_g^n)$$

induces the zero map on integral reduced homology in degrees $* \leq \frac{1}{2}(g-4)$.

Proof The transfer has an additivity property [4] which implies that

$$\begin{tikzcd}
B\mathrm{Diff}_D(V_0^n) \arrow[r] \arrow[d, "\alpha_0"] & B\mathrm{Diff}_D(V_g^n) \arrow[d, "\alpha_g"] \\
Q_1(BO(2n+1)\langle n\rangle_+) \arrow[r] & Q_{1+(-1)^n g}(BO(2n+1)\langle n\rangle_+)
\end{tikzcd}$$

commutes up to homotopy. The lower map takes the sum with a fixed point in $Q_{(-1)^n g}(BO(2n+1)\langle n\rangle_+)$ and is a weak equivalence. We shall show that the left vertical map is trivial in reduced homology (in all degrees); this will imply the claim by Theorem 3.3.

The map α_0 factors as

$$B\mathrm{Diff}_D(V_0^n) \xrightarrow{\mathrm{trf}} Q_1((E\mathrm{Diff}_D(V_0^n) \times_{\mathrm{Diff}_D(V_0^n)} V_0^n)_+) \xrightarrow{Q(\underline{l})} Q_1(BO(2n+1)\langle n\rangle_+).$$

The map $Q(\underline{l})$ is induced from the vertical tangent bundle of the universal bundle $E\mathrm{Diff}_D(V_0^n) \times_{\mathrm{Diff}_D(V_0^n)} V_0^n \to B\mathrm{Diff}_D(V_0^n)$, which is trivial by the following argument: The choice of a point in $D \subset \partial V_0^n$ determines a section s of the bundle, which is a homotopy equivalence as V_0^n is a disc, and the pullback of the vertical tangent bundle along s is trivial. Hence $Q(\underline{l})$ factors through $Q_1(S^0)$, which is rationally acyclic. Therefore α_0 induces the zero map on rational homology. This finishes the proof for rational homology, which is all we need for the proof of Theorem 1.7.

For the integral version, we use that the transfer is defined more generally for fibrations with finite CW fibres. It follows that there is a commutative diagram

$$\begin{tikzcd}
B\mathrm{Diff}_D(V_0^n) \arrow[r, "\mathrm{trf}"] \arrow[d] & Q_1((E\mathrm{Diff}_D(V_0^n) \times_{\mathrm{Diff}_D(V_0^n)} V_0^n)_+) \arrow[d] \\
B\mathrm{Homeo}_D(V_0^n) \arrow[r, "\mathrm{trf}"] & Q_1((E\mathrm{Homeo}_D(V_0^n) \times_{\mathrm{Homeo}_D(V_0^n)} V_0^n)_+)
\end{tikzcd}$$

Because the map

$$\underline{\ell}\colon E\mathrm{Diff}_D(V_0^n) \times_{\mathrm{Diff}_D(V_0^n)} V_0^n \to BO(2n+1)\langle n\rangle$$

is nullhomotopic as we just argued, it extends to a map

$$\underline{\ell}'\colon E\mathrm{Homeo}_D(V_0^n) \times_{\mathrm{Homeo}_D(V_0^n)} V_0^n \to BO(2n+1)\langle n\rangle.$$

Therefore α_0 factors up to homotopy through $B\mathrm{Homeo}_D(V_0^n)$, which is contractible by the Alexander trick. □

Proof of Theorem 1.7 Since both μ_M and μ_M^∂ factor through

$$\mu_{U_{g,1}^n}^\partial : B\mathrm{Diff}_\partial(D^{2n+1}) \to B\mathrm{Diff}_\partial(U_{g,1}^n),$$

it suffices to show that this map induces the trivial map in rational homology in the indicated range of degrees. Note that

$$V_g^n \cup_{\partial V_g^n \setminus \mathrm{int}\, D} V_g^n = U_{g,1}^n.$$

Pick an embedding $f\colon D^{2n+1}\to V_0^n\subset V_g^n$ disjoint from the disc $D\subset V_0^n$ such that $f^{-1}(\partial V_0^n)=D^{2n}$ is a disc in ∂D^{2n+1}. This gives $\mu\colon B\mathrm{Diff}_\partial(D^{2n+1})\to B\mathrm{Diff}_D(V_0^n)$, and by Corollary 3.4 composition with the stabilization map to $B\mathrm{Diff}_D(V_g^n)$ is trivial in integral homology in degrees $*\le\frac{1}{2}(g-4)$.

Diagram (2.17) becomes

$$
\begin{tikzcd}
B\mathrm{Diff}_\partial(D^{2n+1}) \arrow[r] \arrow[d, "d"] & B\mathrm{Diff}_D(V_g^n) \arrow[d, "d"] \\
B\mathrm{Diff}_\partial(D^{2n+1}\cup_{D^{2n}} D^{2n+1}) \arrow[r, "\mu"] & B\mathrm{Diff}_\partial(U_{g,1}^n)
\end{tikzcd}
\tag{3.5}
$$

Lemma 2.18 shows that the left vertical in (3.5) induces an isomorphism in rational homotopy in degrees at most $\phi^{\mathbb{Q}}(D^{2n})+1$. The same is true in rational homology since both spaces are connected H–spaces and their rational homology is the free graded commutative algebra on the rational homotopy, and so the proof is complete. □

4 Optimality of the range in the homotopical theorem

Proof of Theorem 1.8 The composition

$$B\mathrm{Diff}_\partial(D^{2n+1})\xrightarrow{\mu_M} B\mathrm{Diff}(M)\to B\mathrm{Homeo}(M)$$

factors through $B\mathrm{Homeo}_\partial(D^{2n+1})\simeq *$. Hence μ_M factors through the space

$$\mathrm{hofib}(B\mathrm{Diff}(M)\to B\mathrm{Homeo}(M)).$$

By [21, Theorem A], $\pi_{2n-2}(B\mathrm{Diff}_\partial(D^{2n+1}))\otimes\mathbb{Q}\neq 0$ if $n\ge 6$. Therefore, it is enough to find a closed $(2n+1)$–manifold M such that

$$\pi_{2n-2}(\mathrm{hofib}(B\mathrm{Diff}(M)\to B\mathrm{Homeo}(M)))\otimes\mathbb{Q}=0.$$

Now, by smoothing theory [26], $\mathrm{hofib}(B\mathrm{Diff}(M)\to B\mathrm{Homeo}(M))$ is homotopy equivalent to a union of path components of the section space

$$\Gamma\Bigl(M;\mathrm{Fr}(M)\times_{O(2n+1)}\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\Bigr).$$

We now prove that

$$\pi_{2n-2}\Bigl(\Gamma\Bigl(\mathbb{RP}^2\times S^{2n-1};\mathrm{Fr}(\mathbb{RP}^2\times S^{2n-1})\times_{O(2n+1)}\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\Bigr)\Bigr)\otimes\mathbb{Q}=0 \tag{4.1}$$

when n is even. Homotopy groups of section spaces can be computed by means of the Federer spectral sequence [11]; see [22, Section 5.2] for the variant we shall be using. Let $E \to B$ be a fibration over a finite-dimensional CW-complex with simply connected fibre F and a fixed section s. Then there is a spectral sequence

$$E^2_{p,q} = H^{-p}(B; \pi_q(F)) \Rightarrow \pi_{p+q}(\Gamma(B; E), s)$$

(the coefficient systems in the E_2–term are twisted). Hence $\pi_m(\Gamma(B; E), s)$ admits a finite filtration whose filtration quotients are subquotients of $H^{-p}(B; \pi_q(F))$ with $p+q=m$, and so in order to prove that $\pi_m(\Gamma(B; E), s) \otimes \mathbb{Q} = 0$, it suffices to show that $H^{-p}(B; \pi_q(F)) \otimes \mathbb{Q} = 0$ if $p+q=m$.

Because $\mathrm{Top}(2n+1)/O(2n+1)$ is simply connected by [26, 5.0(4)–(5)], we can apply the Federer spectral in our case. Furthermore, by loc. cit., $\mathrm{Top}(2n+1)/O(2n+1)$ is rationally $(2n+2)$–connected. So the only entries in the E_2–page which could potentially be rationally nonzero and contribute to (4.1) are

$$H^{2n+1-i}\left(\mathbb{RP}^2 \times S^{2n-1}; \pi_{4n-1-i}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right) \otimes \mathbb{Q}\right)$$

for $0 \le i \le 2$. It remains to be shown that

(4.2) $$H^{2-i}\left(\mathbb{RP}^2; \pi_{4n-1-i}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right) \otimes \mathbb{Q}\right) = 0$$

when $0 \le i \le 2$.

The fundamental group $\pi_1(\mathbb{RP}^2) = C_2$ acts on $\pi_{4n-1-i}(\mathrm{Top}(2n+1)/O(2n+1))$ by conjugation with an isometry of determinant -1. There is an isomorphism

$$\pi_{4n-1-i}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right) = \pi_{2n-2-i}(B\mathrm{Diff}_\partial(D^{2n+1}))$$

coming from Morlet's theorem, [26, Section 3.4] which states that

$$B\mathrm{Diff}_\partial(D^{2n+1}) \simeq \Omega_0^{2n+1}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right).$$

By the discussion in [21, Section 8.2], we have that the action of the generator of C_2 on $\pi_{4n-1-i}(\mathrm{Top}(2n+1)/O(2n+1))$ corresponds under this isomorphism to *minus* the involution $(BI)_*$ considered in the proof of Lemma 2.18 above. By [21, Corollary 8.4], we therefore have, for *even* n, that

$$\pi_{4n-1}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right)\otimes\mathbb{Q}=\mathbb{Q}^+,$$
$$\pi_{4n-2}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right)\otimes\mathbb{Q}=0,$$
$$\pi_{4n-3}\left(\frac{\mathrm{Top}(2n+1)}{O(2n+1)}\right)\otimes\mathbb{Q}=\mathbb{Q}^-$$

as C_2–modules. Since $H^2(\mathbb{RP}^2;\mathbb{Q}^+)=H^0(\mathbb{RP}^2;\mathbb{Q}^-)=0$, we obtain (4.2), which concludes the proof. □

References

[1] **B Badzioch**, **W Dorabiała**, **J R Klein**, **B Williams**, *Equivalence of higher torsion invariants*, Adv. Math. 226 (2011) 2192–2232 MR Zbl

[2] **B Badzioch**, **W Dorabiała**, **B Williams**, *Smooth parametrized torsion: a manifold approach*, Adv. Math. 221 (2009) 660–680 MR Zbl

[3] **J C Becker**, **D H Gottlieb**, *Transfer maps for fibrations and duality*, Compositio Math. 33 (1976) 107–133 MR Zbl

[4] **J C Becker**, **R E Schultz**, *Axioms for bundle transfers and traces*, Math. Z. 227 (1998) 583–605 MR Zbl

[5] **A Borel**, *Stable real cohomology of arithmetic groups*, Ann. Sci. École Norm. Sup. 7 (1974) 235–272 MR Zbl

[6] **B Botvinnik**, **N Perlmutter**, *Stable moduli spaces of high-dimensional handlebodies*, J. Topol. 10 (2017) 101–163 MR Zbl

[7] **J Cerf**, *Sur les difféomorphismes de la sphère de dimension trois* ($\Gamma_4=0$), Lecture Notes in Math. 53, Springer (1968) MR Zbl

[8] **J Cerf**, *La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie*, Inst. Hautes Études Sci. Publ. Math. 39 (1970) 5–173 MR Zbl

[9] **W Dwyer**, **M Weiss**, **B Williams**, *A parametrized index theorem for the algebraic K–theory Euler class*, Acta Math. 190 (2003) 1–104 MR Zbl

[10] **F T Farrell**, **W C Hsiang**, *On the rational homotopy groups of the diffeomorphism groups of discs, spheres and aspherical manifolds*, from "Algebraic and geometric topology, I" (R J Milgram, editor), Proc. Sympos. Pure Math. 23, Amer. Math. Soc., Providence, RI (1978) 325–337 MR Zbl

[11] **H Federer**, *A study of function spaces by spectral sequences*, Trans. Amer. Math. Soc. 82 (1956) 340–361 MR Zbl

[12] **S r Galatius**, **U Tillmann**, **I Madsen**, **M Weiss**, *The homotopy type of the cobordism category*, Acta Math. 202 (2009) 195–239 MR Zbl

[13] **J Genauer**, *Cobordism categories of manifolds with corners*, Trans. Amer. Math. Soc. 364 (2012) 519–550 MR Zbl

[14] **A Hatcher**, **J Wagoner**, *Pseudo-isotopies of compact manifolds*, Astérisque 6, Soc. Math. France, Paris (1973) MR Zbl

[15] **K Igusa**, *The stability theorem for smooth pseudoisotopies*, K–Theory 2 (1988) 1–355 MR Zbl

[16] **K Igusa**, *Higher Franz–Reidemeister torsion*, AMS/IP Studies in Advanced Math. 31, Amer. Math. Soc., Providence, RI (2002) MR Zbl

[17] **K Igusa**, *Higher complex torsion and the framing principle*, Mem. Amer. Math. Soc. 835, Amer. Math. Soc., Providence, RI (2005) MR Zbl

[18] **K Igusa**, *Axioms for higher torsion invariants of smooth bundles*, J. Topol. 1 (2008) 159–186 MR Zbl

[19] **M A Kervaire**, **J W Milnor**, *Groups of homotopy spheres, I*, Ann. of Math. 77 (1963) 504–537 MR Zbl

[20] **M Krannich**, *A homological approach to pseudoisotopy theory, I*, Invent. Math. 227 (2022) 1093–1167 MR Zbl

[21] **M Krannich**, **O Randal-Williams**, *Diffeomorphisms of discs and the second Weiss derivative of BTop*(-), preprint (2021) arXiv 2109.03500

[22] **A Kupers**, **O Randal-Williams**, *The cohomology of Torelli groups is algebraic*, Forum Math. Sigma 8 (2020) art. id. e64 MR Zbl

[23] **N Perlmutter**, *Homological stability for diffeomorphism groups of high-dimensional handlebodies*, Algebr. Geom. Topol. 18 (2018) 2769–2820 MR Zbl

[24] **G Raptis**, **W Steimle**, *On the map of Bökstedt–Madsen from the cobordism category to A–theory*, Algebr. Geom. Topol. 14 (2014) 299–347 MR Zbl

[25] **G Raptis**, **W Steimle**, *Parametrized cobordism categories and the Dwyer–Weiss–Williams index theorem*, J. Topol. 10 (2017) 700–719 MR Zbl

[26] **L C Siebenmann**, *Classification of sliced families of manifold structures*, from "Foundational essays on topological manifolds, smoothings, and triangulations", Ann. of Math. Studies 88, Princeton Univ. Press (1977) Essay V, 215–290 MR Zbl

[27] **F Waldhausen**, **B r Jahren**, **J Rognes**, *Spaces of PL manifolds and categories of simple maps*, Ann. of Math. Studies 186, Princeton Univ. Press (2013) MR Zbl

Fachbereich Mathematik und Informatik, Westfälische Wilhelms-Universität Münster
Münster, Germany

johannes.ebert@uni-muenster.de

Received: 19 August 2021 Revised: 14 December 2021

Algebraic & Geometric Topology **23**:5 (2023) 2347–2368
DOI: 10.2140/agt.2023.23.2347
Published: 25 July 2023

Intrinsic symmetry groups of links

CHARLES LIVINGSTON

The set of isotopy classes of ordered n–component links in S^3 is acted on by the symmetric group $\mathbb{S}_n$ via permutation of the components. The subgroup $\mathbb{S}(L) \subset \mathbb{S}_n$ is defined to be the set of elements in the symmetric group that preserve the ordered isotopy type of L as an unoriented link. The study of these groups was initiated in 1969, but the question of whether or not every subgroup of $\mathbb{S}_n$ arises as such an intrinsic symmetry group of some link has remained open. We provide counterexamples; in particular, if $n \geq 6$, then there does not exist an n–component link L for which $\mathbb{S}(L)$ is the alternating group $\mathbb{A}_n$.

57K10

1 Introduction

The *oriented diffeomorphism group* of an ordered link $L = \{L_1, \dots, L_n\} \subset S^3$ consists of all orientation-preserving diffeomorphisms of S^3 that preserve the link setwise. We denote this group by $\mathcal{D}(L)$. The action of $\mathcal{D}(L)$ on the components of L defines a homomorphism from $\mathcal{D}(L)$ to the symmetric group $\mathbb{S}_n$; its image is denoted by $\mathbb{S}(L)$. A basic question asks whether every subgroup $H \subset \mathbb{S}_n$ arises as $\mathbb{S}(L)$ for some n–component link. We provide obstructions. Our examples of groups that do not arise are the alternating groups $\mathbb{A}_n$ for $n \geq 6$.

Theorem *If $n \geq 6$, then there does not exist an ordered n–component link L that satisfies $\mathbb{S}(L) = \mathbb{A}_n$.*

The study of symmetries of links is usually placed in the context of an extension of the symmetric group, called the *Whitten group*,

$$\Gamma_n = \mathbb{Z}_2 \oplus ((\mathbb{Z}_2)^n \rtimes \mathbb{S}_n).$$

In the semidirect product, $\mathbb{S}_n$ acts on $(\mathbb{Z}_2)^n$ by permuting the coordinates. If we let $\mathcal{D}^*(L)$ denote the set of diffeomorphisms of an n–component link L, including those that reverse the orientation of S^3, then there is a natural map of $\mathcal{D}^*(L)$ to Γ_n. The first

$\mathbb{Z}_2$ factor keeps track of the orientation of S^3 and the remaining $\mathbb{Z}_2$ factors track the orientations of the components of L. The image of this map is denoted by $\Sigma(L)$. The question of which subgroups of Γ_n arise as $\Sigma(L)$ for some n–component link L was considered by Fox and Whitten in the mid-1960s, first appearing in print in 1969 [17].

There is a quotient map $\Phi\colon \Gamma_n \to \mathbb{S}_n$ which carries $\Sigma(L) \cap \big(0 \oplus ((\mathbb{Z}_2)^n \rtimes \mathbb{S}_n)\big)$ to $\mathbb{S}(L)$. Thus, we have the following corollary:

Corollary *If $n \geq 6$ and $H \subset \Gamma_n$ has the property that $\Phi\big(H \cap \big(0 \oplus ((\mathbb{Z}_2)^n \rtimes \mathbb{S}_n)\big)\big) = \mathbb{A}_n$, then there is no link L with $\Sigma(L) = H$. In particular, if $n \geq 6$, then the subgroup $0 \oplus (0 \oplus \mathbb{A}_n) \subset \Gamma_n$ is not of the form $\Sigma(L)$ for any n–component link.*

Summary of proof The basic idea of our approach is as follows. For a given link L there is a Jaco–Shalen–Johannson (JSJ) decomposition of the complement of L into hyperbolic and Seifert fibered components $\{C_i\}$. This decomposition is unique up to isotopy. We first observe that, if $\mathbb{S}(L)$ does not contain an index two subgroup, then one of the C_i (say C_1) is invariant under the action of $\mathcal{D}(L)$ up to isotopy.

If C_1 is hyperbolic, we can replace the action of $\mathcal{D}(L)$ restricted to C_1 with a finite group of isometries of C_1. We then use a reembedding of C_1 into S^3 (as first described by Budney in [3]) to extend that action to S^3. It follows from results such as Boileau, Leeb and Porti [2] that the action on S^3 is conjugate to a linear action. We then find that $\mathbb{S}(L)$ is a quotient of a finite subgroup of SO(4). Finally, a group-theoretic analysis reduces the problem to the simpler one of considering quotients of finite subgroups of SO(3), which are enumerated.

In contrast to the hyperbolic case, if C_1 is Seifert fibered, then the diffeomorphism group of C_1 itself is large, sufficiently so that we can construct enough symmetries of L to show that $\mathbb{S}(L) = \mathbb{S}_n$.

Outline Section 2 describes the general theory of intrinsic symmetry groups of oriented links, as first considered by Fox and Whitten [17]. Sections 3 and 4 describe the classical case of knots, $n = 1$, and results for the case of $n = 2$. Section 5 presents prime, nonsplit links, with full symmetry group for all n.

In Section 6 we describe JSJ decompositions, the associated tree diagrams, and a proof that, in the case of $\mathbb{S}(L) = \mathbb{A}_n$, some component of the decomposition is fixed (up to isotopy) by the action of the diffeomorphism group. Section 7 explains how that distinguished component can be reembedded into S^3 as the complement of a link. The reembedding is used in Section 8 to show that if the fixed component is hyperbolic, then $\mathbb{S}(L)$ is a subgroup of a quotient of a finite subgroup of SO(4). Finally, in Section 9 we present the Seifert fibered case. In the concluding Section 10, we present a few questions and include an example of a four-component link L with $\mathbb{S}(L) = \mathbb{A}_4$.

Notational comment We are calling the groups studied here the *intrinsic symmetry groups* of links. The *symmetry group* of a link consists of the group of diffeomorphisms of S^3 that leave the link invariant, modulo isotopy. Even for knots, these symmetry groups include, for instance, all dihedral groups.

Acknowledgments I have benefited from comments from Ryan Budney, whose work provides a backdrop for our approach. Nathan Dunfield provided an example of a four-component link L with $\mathbb{S}(L) = \mathbb{A}_4$. I was also helped by discussions with Jim Davis, Allan Edmonds, Charlie Frohman, Michael Larsen, Swatee Naik and Dylan Thurston. Suggestions from the referee led to several improvements in the exposition.

This work was supported by a grant from the National Science Foundation, NSF-DMS-1505586.

2 The general setting of oriented links

We now describe the general theory of *intrinsic symmetry groups* of links. This theory was initially developed by Fox and was first presented by Whitten in [17]. To be precise, we will momentarily consider links in three-manifolds that are diffeomorphic to S^3, rather than work specifically with S^3. In this setting we have the following definition: an *n–component link* is an ordered $(n+1)$–tuple of oriented manifolds, $L = (S, L_1, L_2, \dots, L_n)$, where S is diffeomorphic to S^3 and the L_i are disjoint submanifolds of S, each diffeomorphic to S^1. The set of n–component links will be denoted by $\mathcal{L}_n$.

Given a second link $L' = (S', L'_1, L'_2, \dots, L'_n)$, an *orientation-preserving diffeomorphism* from L to L' is an orientation-preserving diffeomorphism $F\colon S \to S'$ such that $F(L_i) = L'_i$ as oriented manifolds for all i.

For any oriented manifold M, $-M$ denotes its orientation reverse. Let $\mathbb{Z}_2$ be the cyclic group of order two written multiplicatively: $\mathbb{Z}_2 = \{1, -1\}$. If $\epsilon = -1 \in \mathbb{Z}_2$, we will let $\epsilon M = -M$, and if $\epsilon = 1 \in \mathbb{Z}_2$, we will let $\epsilon M = M$. The group $\mathbb{Z}_2 \oplus (\mathbb{Z}_2)^n$ acts on $\mathcal{L}_n$ by changing the orientations of the factors. The symmetric group $\mathbb{S}_n$ acts on $\mathcal{L}_n$ by permuting the component knots. These actions do not commute, but together define an action on the set of knots by the *Whitten group*

$$\Gamma_n = \mathbb{Z}_2 \oplus ((\mathbb{Z}_2)^n \rtimes \mathbb{S}_n).$$

In this semidirect product, $\mathbb{S}_n$ acts on the n–fold product by permuting the coordinates. To be precise, given an element $s = (\eta, (\epsilon_1, \dots, \epsilon_n), \rho) \in \Gamma_n$ and an n–component link L, we let

$$sL = (\eta S, \epsilon_1 L_{\rho(1)}, \cdots, \epsilon_n L_{\rho(n)}).$$

Notice that these group actions are defined to be on the left. Thus, elements in $\mathbb{S}_n$ are multiplied right to left.

Definition 2.1 For a link $L \in \mathcal{L}_n$, the *intrinsic symmetry group* of L is the subgroup $\Sigma(L) = \{s \in \Gamma_n \mid sL \cong L\} \subset \Gamma_n$. Note that "$\cong$" indicates the existence of an orientation- and order-preserving diffeomorphism.

There are two fundamental questions regarding such link symmetries:

Problem 1 Given an n–component link L, determine $\Sigma(L)$.

Problem 2 For each subgroup $H \subset \Gamma_n$, does there exist an n–component link L such that $\Sigma(L) = H$?

The first can be effectively answered for low crossing number links with programs such as SnapPy [6]. The second is the focus of this paper; we present the first examples of groups that cannot arise as the symmetry group of a link.

2.1 Restricting to the oriented category and basic observations

There is a canonical index two subgroup $\overline{\Gamma}_n \subset \Gamma_n$ consisting of elements of the form

$$(1, (\epsilon_1, \dots, \epsilon_n), \rho).$$

This subgroup maps onto $\mathbb{S}_n$. We leave it to the reader to verify the following, which implies that any constraint on what groups occur as $\mathbb{S}(L)$ places a constraint on what groups can arise as $\Sigma(L)$:

Theorem 2.2 *The image of $\Sigma(L) \cap \overline{\Gamma}_n$ in $\mathbb{S}_n$ is precisely $\mathbb{S}(L)$.*

After the initial sections of this paper, we will be restricting our work to orientation-preserving diffeomorphisms of S^3 and will work with unoriented links. We will use the following conventions, which were summarized in the introduction:

(1) Links will all be of the form $L = (S^3, L_1, L_2, \dots, L_n)$, where S^3 has some fixed orientation and the L_i are disjoint unoriented submanifolds, each diffeomorphic to S^1.

(2) We will consider diffeomorphisms of the link that are orientation-preserving on S^3 and that possibly permute the set of L_i.

(3) The set of such diffeomorphisms will be denoted by $\mathcal{D}(L)$.

(4) Given $F \in \mathcal{D}(L)$, we have
$$(S^3, F(L_1), F(L_2), \dots, F(L_n)) = (S^3, L_{\rho(1)}, L_{\rho(2)} \dots, L_{\rho(n)})$$
for some $\rho \in \mathbb{S}_n$. This defines a homomorphism $\Phi\colon \mathcal{D}(L) \to \mathbb{S}_n$.

(5) The image Φ in $\mathbb{S}_n$ is denoted by $\mathbb{S}(L)$.

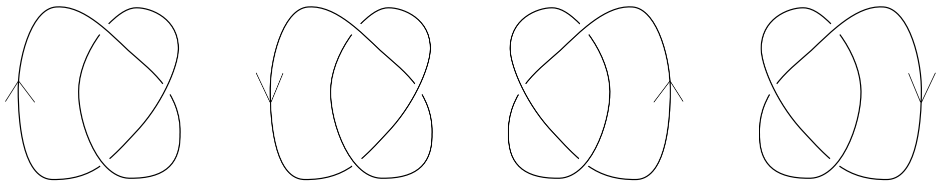
3 Examples: knots

Before restricting to the orientation-preserving diffeomorphism group, in this section and the next we will summarize what is known in general for links of one and of two components. Then, in Section 5, we show that for all n there is a prime, nonsplittable n–component link L with $\Sigma(L) = \Gamma_n$.

Let $n = 1$. The symmetric group $\mathbb{S}_1$ is trivial and thus the first Whitten group is $\Gamma_1 \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$. The knots $(1,-1)K$, $(-1,1)K$ and $(-1,-1)K$ have been called the *reverse*, K^r, the *mirror image*, $m(K)$, and the *reversed mirror image*, $m(K)^r$, respectively. (Older references have called the reverse of K the *inverse*. The name "reverse" is used to distinguish it from the concordance inverse, which is represented by the reversed mirror image.) Figure 1 illustrates the possibilities. A detailed account of the key results in the study of knot symmetries is contained in [8]. Here is a brief summary.

The group Γ_1 has five subgroups: the entire group, the trivial subgroup and the three subgroups containing exactly one of the nontrivial elements of Γ_1. Each is realized as $\Sigma(K)$ for some knot K.

- The unknot and the figure eight knot, 4_1, have full symmetry group. They are called *fully amphicheiral*.
- The trefoil knot is reversible. Dehn showed that it does not equal its mirror image, a fact that can now be proved using such invariants as the signature or the Jones polynomial. Thus, 3_1 is *reversible*.



K $\quad(1,-1)K = K^r$ $\quad(-1,1)K = m(K)$ $\quad(-1,1)K = m(K)^r$

Figure 1: Symmetries of knots.

- Trotter [14] proved the existence of nonreversible knots. His examples in [14] have nonzero signature and thus have trivial symmetry group. We say that such knots are *chiral*. Hartley [7] proved that 9_{32} is nonreversible and, since it has nonzero signature, it too is chiral.
- Kawauchi [9] proved that $K = 8_{17}$ is nonreversible. It is easily seen that $K = m(K)^r$, and thus 8_{17} is *negative amphicheiral*.
- The simplest example of a low crossing number knot that is nonreversible and for which $K = m(K)$ is $12a_{147}$, which was detected by the program SnapPy. (Presumably the general techniques developed by Hartley in [7] would also show that this knot is not reversible.) More complicated examples of such *positive amphicheiral* knots were first discovered by Trotter.

4 Two-component links

Here we summarize the results of [1; 5] concerning two-component links. We have that $\Gamma_2 = \mathbb{Z}_2 \oplus ((\mathbb{Z}_2)^2 \rtimes \mathbb{S}_2)$ is of order 16. In [1; 5], the authors describe the 27 conjugacy classes of subgroups of Γ_2. They then show that tables of prime, nonsplittable links provide examples of links realizing 21 of these subgroups. One of the missing subgroups is Γ_2 itself. This is clearly the symmetry group of the unlink; in a note on MathOverflow [4], Budney showed that Γ_2 is the symmetry group of a nonsplittable Brunnian link. We will expand on that example in the next section.

To conclude this section, we list the subgroups that are currently not known to be the symmetry groups of two-component links, where τ denotes the transposition in $\mathbb{S}_2$:

- $\langle (1, (-1, 1)\tau) \rangle \cong \mathbb{Z}_4$.
- $\langle (1, (-1, 1)), (-1, (1, 1)) \rangle \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$.
- $\langle (1, (1, -1)), (-1, (-1, 1)) \rangle \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

- $\langle(-1,(-1,1)),(1,(-1,1)\tau)\rangle \cong D_4$, the dihedral group with four elements.
- $\langle(1,(1,-1)),(1,(-1,1)),(-1,(1,1))\rangle \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

5 Fully amphicheiral links for all *n*

In Figure 2, we illustrate a knot K in a solid torus D. Two parallel strands of K are tied in a knot J, where J is chosen to be fully amphicheiral; the figure eight knot would be sufficient. As oriented pairs, we have $(D,K) \cong (-D,K) \cong (D,-K) \cong (-D,-K)$.

Budney's example [4] of a two-component link L with full symmetry group $\Sigma(L) = \Gamma_2$ is formed from the Hopf link by replacing neighborhoods of each component with copies of (D,K). An example of a three-component link with full symmetry group is built in the same way, starting with the Borromean link. Notice that, in both these examples, the links are Brunnian. Problem (5) in Section 10 asks: Does there exist a Brunnian link with four or more components with full symmetry group?

We conclude this section with an elementary observation:

Theorem 5.1 *For every n, there exists a prime, nonsplittable link L for which $\Sigma(L) \cong \Gamma_n$.*

Proof To form an n–component link with full symmetry group, proceed as follows: Starting with any nontrivial fully amphicheiral knot J', form a link by replacing J' with n parallel copies of J'; formally, form the $(0,n)$–companion of J'. (Again, the simplest example would be to let J' be the figure eight knot.) Next, replace a neighborhood of each component of that link with a copy of (D,K) as illustrated in Figure 2, built

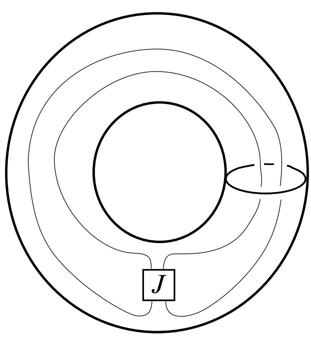


Figure 2: Companion.

using the fully amphicheiral knot J. Innermost circle arguments, dating to the work of Schubert [13], can be used to show that this link L is prime and nonsplittable. The fact that we used parallel copies of J' implies that the components can be freely permuted. By replacing the components with (D, K), we have ensured that the components can be independently reversed. The fact that J and J' are fully amphicheiral ensures that there is an orientation-reversing diffeomorphism of S^3 that preserves the link. □

6 Torus decompositions and tree diagrams

A principal tool in understanding knot and link complements is the Jaco–Shalen–Johannson torus decomposition, which we refer to as the JSJ decomposition. An excellent resource is [3], which contains details for the results we summarize here.

Let X be the complement of a nonsplittable link L in S^3. The JSJ decomposition of X is given by a finite family of disjoint incompressible embedded tori, $\{T_i\}$, with the property that each component of the complement of $\bigcup T_i$ has either a complete hyperbolic structure or is Seifert fibered. There is the additional condition that no T_i is boundary parallel and that no two of the T_i are parallel. Up to isotopy, there is a unique minimal set $\{T_i\}$ with these properties; this set provides the JSJ decomposition. No two T_i in the decomposition are isotopic.

We can associate a finite tree $\mathrm{Tr}(L)$ to this decomposition, as follows: Let the components of $X \setminus \bigcup T_i$ be denoted by $\{C_i\}$. The vertices of the $\mathrm{Tr}(L)$ correspond to the C_j. Two vertices are joined by an edge if the closures of the corresponding C_i intersect; there is one edge for each T_i. When possible, we will use the names C_i and T_i to denote the vertices and edges. We will say that a component C_i *contains a component* $L_j \in L$ if L_j is in the closure of C_i.

6.1 The subtrees $\mathrm{Tr}_L(K)$ and $\widehat{\mathrm{Tr}}(L)$

Let K be a component of L. Its orbit under the action of $\mathcal{D}(L)$ is a sublink of L, $\{K_1, \dots, K_l\}$ for some $l \geq 1$ with $K_1 = K$. Each K_i is contained in a vertex of $\mathrm{Tr}(L)$. The set of such vertices is denoted by $\{D_1, \dots, D_k\}$. Since the action of $\mathcal{D}(L)$ on the set of K_i is transitive, each D_j contains the same number of components of L. In particular, k divides l. Later we will expand on this observation.

The vertices $\{D_1, \dots, D_k\}$ in $\mathrm{Tr}(L)$ span a unique minimal subtree, which we denote by $\mathrm{Tr}_L(K)$. In the case that the action of $\mathcal{D}(L)$ is transitive on L, the orbit of K is all of L, and we write $\widehat{\mathrm{Tr}}(L) = \mathrm{Tr}_L(K)$. (Notice that $\widehat{\mathrm{Tr}}(L)$ need not equal $T(L)$; for instance, vertices of $T(L)$ of valence one that do not contain components of L are not included in $\widehat{\mathrm{Tr}}(L)$.)

Theorem 6.1 *If $\mathcal{D}(L)$ acts transitively on L, then the tree $\widehat{\mathrm{Tr}}(L)$ either contains exactly one vertex, or its valence one vertices are precisely the set $\{D_1, \dots, D_k\}$.*

Proof It is an elementary observation that, in the subtree of a tree spanned by the set of vertices $\{D_j\}$, the only vertices of valence one correspond to elements in the set $\{D_j\}$, and that, if there is more than one D_j, then at least one of them is a vertex of valence one. We need to see that each D_j has valence one.

Suppose that the vertex D_1 is of valence one in $\widehat{\mathrm{Tr}}(L)$ and that it contains L_1. Let D_2 be another vertex and suppose it contains L_2. There is an element $F \in \mathcal{D}(L)$ such that $F(L_1) = L_2$. The map F is isotopic relative to L to a diffeomorphism F' that preserves the JSJ decomposition. This F' induces an automorphism of $\mathrm{Tr}(L)$ that leaves $\widehat{\mathrm{Tr}}(L)$ invariant. Thus, there is an automorphism of $\widehat{\mathrm{Tr}}(L)$ that carries D_1 to D_2. It follows that D_2 is of valence one in $\widehat{\mathrm{Tr}}(L)$. □

6.2 The group $\mathcal{D}^*(L)$

Fix a JSJ decomposition of $S^3 \setminus L$.

Definition 6.2 We let $\mathcal{D}^*(L) \subset \mathcal{D}(L)$ be the subgroup consisting of elements that leave the JSJ decomposition invariant.

Theorem 6.3 *The image of $\mathcal{D}^*(L)$ in $\mathbb{S}_n$ equals $\mathbb{S}(L)$.*

Proof Given an element in $\mathbb{S}(L)$, there is a diffeomorphism $F \in \mathcal{D}(L)$ that maps to it. We have that F is isotopic relative to L to an element $F' \in \mathcal{D}^*(L)$. The map F' induces the same permutation of the components of L as does F. □

Theorem 6.4 *In the case that $\mathcal{D}^*(L)$ acts transitively on the components of L, the action of $\mathcal{D}^*(L)$ on $\widehat{\mathrm{Tr}}(L)$ factors through an action of $\mathbb{S}(L)$ on $\widehat{\mathrm{Tr}}(L)$.*

Proof An automorphism of a tree is completely determined by its action on the valence one vertices of the tree. We leave this elementary observation to the reader. □

6.3 The structure of $\widehat{\mathrm{Tr}}(L)$ when $\mathbb{S}(L) = \mathbb{A}_n$

In Figure 3 we provide an example of a labeled tree to serve as a model for the discussion that follows.

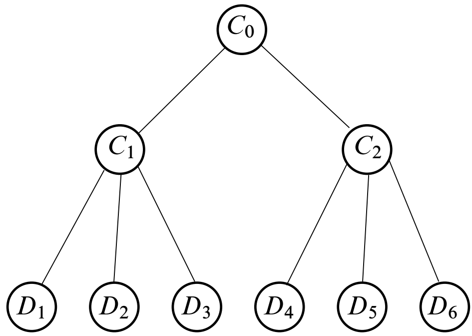


Figure 3: Tree diagram for a sublink K of L on which $\mathcal{D}(L)$ acts transitively.

Lemma 6.5 *If $\mathbb{S}(L) = \mathbb{A}_n$ and $\widehat{\mathrm{Tr}}(L)$ contains more than one vertex, then each D_i contains exactly one L_1 and the number of vertices in the set $\{D_i\}$ is n.*

Proof Suppose that D_1 contains L_1 and L_2 and that D_2 contains L_3 and L_4. Then the permutation $(123) \in \mathbb{A}_n$ does not induce an action on $\widehat{\mathrm{Tr}}(L)$. $\square$

Theorem 6.6 *If $\mathbb{S}(L) = \mathbb{A}_n$ with $n \geq 3$, then $\widehat{\mathrm{Tr}}(L)$ is a rooted tree with either exactly one vertex, C, or with n vertices of valence one. In the second case, there is a unique vertex with valence greater than two; the tree $\widehat{\mathrm{Tr}}(L)$ is built from that high valence vertex C by attaching n linear branches, all of the same length. The vertex C is invariant under the action of $\mathbb{A}_n$ on $\widehat{\mathrm{Tr}}(L)$.*

Proof Figure 4 is a schematic of a tree. We are asserting that $\widehat{\mathrm{Tr}}(L)$ is of this form.

We have seen that each D_i contains precisely one L_i and these are the valence one vertices of $\widehat{\mathrm{Tr}}(L)$. A tree with more than two valence one vertices always contains

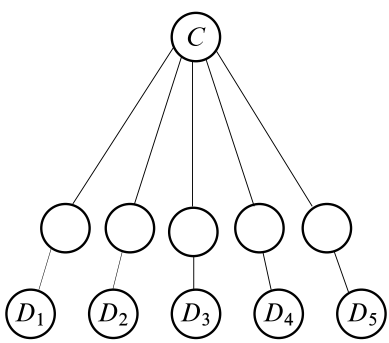


Figure 4: Possible tree diagram $\widehat{\mathrm{Tr}}(L)$ for a five-component link L on which $\mathbb{S}(L) = \mathbb{A}_5$.

some vertex with valence greater than 2. It remains to show that there is a unique such vertex of valence greater than two. (For an example of the sort of tree we need to rule out, build a tree from two copies of the graph illustrated in Figure 4 by joining the roots with single edge.)

An elementary exercise shows that, for any tree on which $\mathbb{A}_n$ acts with this action transitive on the vertices of valence one, there is an invariant vertex or edge: proceed by induction, removing all valence one vertices and their adjacent edges from the tree.

We next observe that, in the case that the symmetry group is $\mathbb{A}_n$, there must be an invariant vertex. The action of the symmetry group of the tree is transitive on its valence one vertices, so, if there is an invariant edge, some elements must reverse that edge. It follows that the subgroup of the symmetry group that does not reverse the edge is index two. But $\mathbb{A}_n$ does not contain an index two subgroup for $n \geq 3$. $\square$

6.4 The structure of the core C in the case that $\mathbb{S}(L) = \mathbb{A}_n$

Suppose that $\mathbb{S}(L) = \mathbb{A}_n$. Then, by Theorem 6.6, there is a core C in the JSJ decomposition of L. This core is acted on by $\mathcal{D}^*(L)$. The boundary of C is the union of two sets of tori, $\{T_1, \dots, T_n\} \cup \{S_1, \dots, S_m\}$. Each T_i bounds a submanifold $W_i \subset S^3$ that contains the link component L_i and does not contain C. A schematic appears in Figure 5. In this diagram we have included extra edges showing $\widehat{\mathrm{Tr}}(L)$ might be a proper subtree of $\mathrm{Tr}(L)$ and that C might have more than n boundary components.

Let $\mathcal{D}(C)$ be the diffeomorphism group of the core C. It contains a subgroup $\mathcal{D}(C, T)$ that leaves invariant the set of T_i. This group maps to $\mathbb{S}_n$ via its action on $\{T_i\}$.

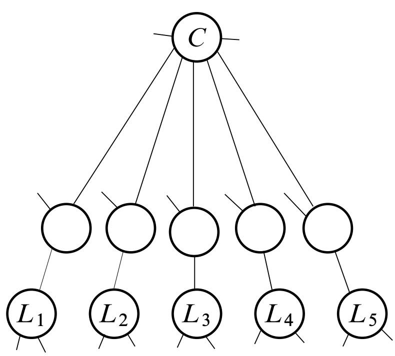


Figure 5: Possible tree diagram $\widehat{\mathrm{Tr}}(L)$ for a five-component link L on which $\mathcal{D}^*(L)$ acts transitively.

Theorem 6.7 *In the case that $\mathbb{S}(L) = \mathbb{A}_n$ with $n \geq 5$, with core C, the group $\mathcal{D}(C,T)$ acts on $\{T_i\}$ as either $\mathbb{A}_n$ or $\mathbb{S}_n$.*

Proof It is clear that the action contains $\mathbb{A}_n$. The only subgroups of $\mathbb{S}_n$ that contain $\mathbb{A}_n$ are $\mathbb{A}_n$ and $\mathbb{S}_n$. □

Notice that it might happen that there are elements of $\mathcal{D}(C,T)$ that do not map to elements of $\mathbb{A}_n$; it is possible that not every action on C extends to S^3.

7 Reembeddings

Reembeddings appear in two different ways in our proof. In the case of C hyperbolic, we embed C in S^3 as a link complement. In the Seifert fibered case, we embed C into a closed Seifert fibered space as the complement of a set of regular fibers. In this section, we describe the embedding into S^3.

In the previous section, some of the (torus) boundary components of the core C were denoted by T_i. We will now see that by using reembeddings we can view these T_i, along with the other boundary components S_i of C, as peripheral tori for a link in S^3. This is presented in [3], where Budney gave a reembedding theorem for submanifolds of S^3. Here we present a slightly enhanced version of that result, keeping track of boundary curves. First we set up some notation.

Let $X \subset S^3$ be a compact, connected submanifold with one of its boundary components a torus T. The complement of T consists of two spaces, Y_1 and Y_2. We have $H_1(Y_1) \cong \mathbb{Z} \cong H_1(Y_2)$. We assume $X \subset Y_1$. When needed, we will write these as $Y_1(X,T)$ and $Y_2(X,T)$.

We have that $\ker(H_1(T) \to H_1(Y_1)) \cong \mathbb{Z}$. The generator can be represented by a simple closed curve we denote by l. Similarly, a representative of $\ker(H_1(T) \to H_1(Y_2)) \cong \mathbb{Z}$ is denoted by m. There is no natural orientation for these choices. However, we can assume that they are oriented so that the intersection number of m and l is 1 with respect to the orientation of T viewed as the boundary of Y_1. We can also assume that m and l intersect transversely in exactly one point. With this setup, we have the following:

Theorem 7.1 *There exists an orientation-preserving embedding $F: X \to S^3$ such that $F(T)$ is the boundary of a tubular neighborhood of a knot in S^3 having meridian $F(m)$ and longitude $F(l)$.*

Proof An embedded torus in S^3 bounds (on one side or the other) a solid torus, which we denote by W. If $Y_2 = W$, then m is the meridian of W and F can be taken to be the identity.

If $Y_1 = W$, then form the boundary union $Z = Y_1 \cup W'$, where W' is a solid torus, attached so that its meridian is identified with m and its longitude is identified with l. Then Z is the union of two solid tori and the choice of identification ensures that $H_1(Z) = 0$. Thus, $Z \cong S^3$. □

Corollary 7.2 *Suppose that $X \subset S^3$ is a compact manifold with boundary a union of tori $\{T_1, \dots, T_k\}$. There exists a link $L = \{L_1, \dots, L_k\}$ and an orientation-preserving homeomorphism $F\colon X \to S^3 \setminus \nu(L)$, where $\nu(L)$ is an open tubular neighborhood. Furthermore, it can be assumed that F preserves meridians and longitudes.*

Corollary 7.3 *With $X \subset S^3$ and L as in Corollary 7.2, suppose that a diffeomorphism $g\colon S^3 \to S^3$ satisfies $g(X) = X$. Then the diffeomorphism of $F(X)$ given as the composition $F \circ g \circ F^{-1}$ extends to a diffeomorphism of (S^3, L).*

Note Not every diffeomorphism of X determines a diffeomorphism of L. It is essential here that the diffeomorphism of X extends to S^3.

7.1 Summary theorem

Theorem 7.4 *Suppose that $\mathbb{S}(L) = \mathbb{A}_n$. Then there is a link $(L'_1, \dots, L'_n, J_1, \dots J_m)$ with complement diffeomorphic to C and that is either hyperbolic or Seifert fibered. The mapping class group of this link has a subgroup that preserves $(L'_1, \dots L'_n)$. The image of this subgroup in $\mathbb{S}_n$ is either $\mathbb{A}_n$ or $\mathbb{S}_n$.*

Proof To prove this using the previous results, we need to show that a JSJ decomposition exists — that is, that L is nonsplittable. If L does split, it splits as a union of nonsplit sublinks, say $D_1, \dots, D_k$, where each D_i is contained in a ball that does not intersect the other D_j. The transitivity of the $\mathbb{A}_n$–actions implies that the D_i are identical links. Thus, we can write $D_i = \{D_i^1, \dots, D_i^m\}$ for some m that is independent of i.

If $m = 1$, then L is consists of n copies of a knot J, each copy in a separate ball. In this case, the symmetry group would be $\mathbb{S}_n$. If $m = n$, then we are in the nonsplit case, as desired.

Finally, if $1 < m < n$, then any element of $\mathcal{M}$ that carries D_1^1 to D_2^1 must carry D_1^2 to some D_2^i. But not every element of $\mathbb{A}_n$ behaves in this way. □

To complete the proof that $\mathbb{A}_n$ for $n \geq 6$ is not the intrinsic symmetry group of any link, we consider the hyperbolic and Seifert fibered cases separately.

8 The case of C hyperbolic

We use the notion of *core* as in the previous section.

Theorem 8.1 *If $\mathbb{A}_n \subset \mathbb{S}(L)$ and the core C is hyperbolic, then some finite subgroup of* SO(4) *contains a finite subgroup having $\mathbb{A}_n$ as a quotient.*

Proof For each element $\phi \in \mathcal{D}(C, \partial C)$ that extends to S^3, let ϕ' denote an isometry that is isotopic to ϕ relative to the boundary. Note that the actions of ϕ and ϕ' on the finite set of components $\{\partial C\}$ are the same. The set of ϕ' generates a subgroup of $\mathrm{Isom}(C)$. This is necessarily a finite group, H. The group H contains the subgroup $H' \subset H$ that leaves invariant the set $\{L'_1, \dots, L'_n\}$. The image of H' in $\mathbb{S}_n$ contains $\mathbb{A}_n$. By restricting to a further subgroup H'', we can assume the image is precisely $\mathbb{A}_n$.

By results such as [12; 2], any finite subgroup of $\mathrm{Diff}(S^3)$, such as H'', is isomorphic to a subgroup of SO(4). □

Corollary 8.2 *If $\mathbb{A}_n \subset \mathbb{S}(L)$ then $n \leq 5$.*

Proof This follows from the results of the next subsection. □

8.1 The only subgroup of SO(4) that maps onto a noncyclic simple group is isomorphic to $\mathbb{A}_5$

We prove somewhat more than this.

Theorem 8.3 *If A is a nonabelian simple group and a subgroup $H \subset$ SO(4) surjects onto A, then $A \cong \mathbb{A}_5$.*

Denote the surjection from H to A by $\phi\colon H \to A$. We begin by recalling the structure of SO(4).

The set of unit quaternions is homeomorphic to S^3 and as a Lie group is isomorphic to SU(2). Quotienting by ± 1 yields a two-fold cover $\mathrm{SU}(2) \to \mathrm{SO}(3)$.

Let x and y be unit quaternions and view elements $v \in \mathbb{R}^4$ as quaternions. Then x and y define a homomorphism $\psi_{x,y}\colon \mathrm{SU}(2) \times \mathrm{SU}(2) \to \mathrm{SO}(4)$ by $\psi_{x,y}(v) = xvy^{-1}$. This yields a two-fold covering of $\mathrm{SU}(2) \times \mathrm{SU}(2) \to \mathrm{SO}(4)$. Hence,

$$\mathrm{SO}(4) \cong (\mathrm{SU}(2) \times \mathrm{SU}(2))/\langle(-1,-1)\rangle.$$

There is a two-fold covering space $q\colon (\mathrm{SU}(2)\times\mathrm{SU}(2))/\langle(-1,-1)\rangle \to \mathrm{SO}(3)\times\mathrm{SO}(3)$. We thus have the diagram

$$\begin{tikzcd}
(\mathrm{SU}(2)\times\mathrm{SU}(2))/\langle(-1,-1)\rangle \arrow[r, "\cong"] \arrow[d, "\text{two-fold cover } q"'] & \mathrm{SO}(4) \\
\mathrm{SO}(3)\times\mathrm{SO}(3) &
\end{tikzcd}$$

We will write elements of $\mathrm{SO}(4)$ and of $\mathrm{SO}(3)\times\mathrm{SO}(3)$ as equivalence classes of pairs of unit quaternions.

Lemma 8.4 *The map ϕ induces a surjection $\phi'\colon q(H)\subset \mathrm{SO}(3)\times\mathrm{SO}(3)\to A$.*

Proof If the map $q\colon H\to q(H)$ is an isomorphism, then this is trivially true. It is possible that $q\colon H\to q(H)$ is two-to-one, which can occur if and only if the central element $(1,-1)\in H$. In this case, $q(H)\cong H/\langle(1,-1)\rangle$. Since A is nonabelian and simple, the image of $(1,-1)$ in A is trivial. □

Lemma 8.5 *Let $G\subset \mathrm{SO}(3)\times\mathrm{SO}(3)$. Let G_1 and G_2 be the images of the projections of G onto the first and second factors of the product. If $\phi'\colon G\to A$ where A is nonabelian and simple, then a subgroup of G_1 or G_2 maps onto A. In particular, A is a quotient of a finite subgroup of* $\mathrm{SO}(3)$.

Proof Let $F=G\cap(\mathrm{SO}(3)\times\{1\})$. We have that F is a normal subgroup of G and, thus, $\phi'(F)=A$ or $\phi'(F)=\{1\}$. In the first case, we are done, so assume that $\phi'(F)=\{1\}$.

We now define a surjective homomorphism $\phi''\colon G_2\to A$. Given $y\in G_2$, there exists an element $x\in G_1$ such that $(x,y)\in G$. Set $\phi''(y)=\phi'((x,y))$. To see that this is well defined, notice that, if $(x_1,y)\in G$ and $(x_2,y)\in G$, then $x_1x_2^{-1}\in F$. Thus, $\phi'((x_1,y))=\phi'((x_2,y))$. It is easily checked that ϕ' is surjective and is a homomorphism. □

Lemma 8.6 *The group $\mathbb{A}_5$ is the only finite noncyclic simple group contained in* $\mathrm{SO}(3)$.

Proof The finite subgroups of $\mathrm{SO}(3)$ are classified. Here is the list of possibilities:

- Cyclic groups $A_n\cong\mathbb{Z}_n$.
- Dihedral groups D_n.
- Tetrahedral group $E_6\cong\mathbb{A}_4$.

- Octahedral group $E_7 \cong \mathbb{S}_4$.
- Icosahedral group $E_8 \cong \mathbb{A}_5$.

The subgroups of the dihedral group are either dihedral, and thus not simple, or cyclic. The smallest nonabelian simple group is $\mathbb{A}_5$. □

9 The case of C Seifert fibered

We begin with a basic example.

Example 9.1 Consider the $(n+2)$–component link L formed as follows. Let T be a standardly embedded torus in S^3 and form the (np, nq)–torus link on T with $q > p > 1$ relatively prime. Add to this the cores of two solid tori bounded by T. There is a Seifert fibration of S^3 with the torus link represented by regular fibers and the two cores being neighborhoods of singular fibers of type p/q and q/p.

We leave it to the reader to confirm that, for this link, $\Sigma(L) \cong \mathbb{Z}_2 \oplus \mathbb{S}_n$. It should be clear how the components of the (np, nq)–torus link can be freely permuted. The $\mathbb{Z}_2$ arises from a diffeomorphism that reverses all the components.

Two exercises arise here. The first is to show that every symmetry fixes the two core circles. The second is to show that the complement of this link is homeomorphic to the complement of $n+2$ fibers of the Hopf fibration of S^3.

More examples can be built from this one. Let $J \subset S^1 \times B^2$ be a knot for which $\partial(S^1 \times B^2)$ is incompressible in the complement of J. A new link can be formed by replacing neighborhoods of the components of L with copies of $S^1 \times B^2$. Then the symmetry group of this new link will be isomorphic to either $\mathbb{Z}_2 \oplus \mathbb{S}_{n-2}$ or $\mathbb{S}_{n-2}$, depending on the symmetry type of J.

9.1 C is the complement of regular fibers in a closed Seifert manifold

Example 9.2 Figure 6 provides a schematic of one possible case in which the core C is Seifert fibered. Some of the labels in the diagram will be explained later. A link L can be formed by filling each T_i with pairs $(S^1 \times B^2, J_i)$ and the S_i are filled with either solid tori or nontrivial knot complements. There are constraints required for this to produce a link in S^3 and we do not assert that in all cases in which C is Seifert fibered it will be of this form. We illustrate it to provide a good model to have in mind

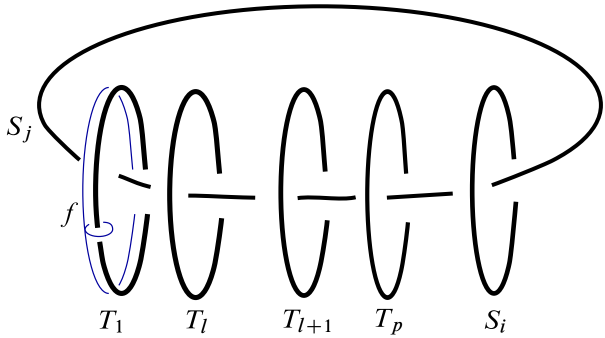


Figure 6: Possible Seifert fibered core C.

as we develop the notation and arguments that follow. Another good model is provided by Example 9.1.

Notice the complement of this link is homeomorphic to the complement of the link formed by giving the parallel strands a full twist. In this case, all the components, including the horizontal one, are fibers of the Hopf fibration of S^3. More generally we have the following:

Theorem 9.3 *The core C is diffeomorphic to the complement of a set of regular fibers in a closed Seifert manifold.*

Proof Build a manifold M by attaching solid tori to the boundary components of C so that each longitude is identified with the fiber of the fibration of C. Then the Seifert fibration of C extends to M and the cores of the solid tori are regular fibers. □

We now fix the choice of that M and its Seifert fibration.

9.2 Notation and a basis for $H_1(T_i)$

For each T_i, there is a basis of $H_1(T_i)$ represented by a pair of curves, $\{f_i, g_i\}$; since T_i bounds the solid torus neighborhood of a regular fiber, we let f_i denote the fiber and let g_i denote the meridian of the solid torus.

Each torus T_i bounds a submanifold of S^3 that contains the component L_i; denote it by W_i. All the pairs (W_i, L_i) are diffeomorphic, so we choose one and denote it by (W, K) with boundary T. We have that T contains a canonical longitude that is null-homologous in W, which we denote by λ; choose a second curve intersecting it once and denote it by μ.

We now see that (S^3, L) is built from C by attaching copies of W to the T_i using attaching maps we denote by G_i. (Other manifolds have to be attached along the other boundary components of C, which we have denoted by S_i.) Denote the images of $\{\lambda, \mu\}$ under G_i by $\{\lambda_i, \mu_i\}$.

Theorem 9.4 *The intersection number of λ_i with f_i is nonzero.*

Proof Our proof depends on the uniqueness of the fibrations of Seifert fibered manifolds, up to isotopy. This does not hold for all Seifert manifolds (eg $S^1 \times B^2$), but Waldhausen [15; 16] proved that, if the Seifert fibered manifold M is sufficiently large, that is, if it contains an incompressible surface that is not boundary parallel, then the fibration is unique. (See also [10].) In the case that the three-manifold has four or more boundary components, it is clearly sufficiently large. The preimage of a circle in the base space that bounds two of the boundary components is an incompressible torus and is not boundary parallel.

We now claim that the λ_i are not fibers of the fibration. Consider $i \neq j$ and the pair λ_i and λ_j. Any element of $\mathcal{D}(L)$ that maps L_i to L_j carries λ_i to $\pm\lambda_j$. Self-homeomorphisms of Seifert fibered spaces with more than three boundary components preserve fibers up to isotopy, so, if λ_i is a fiber, then λ_j is also a fiber.

Suppose that are λ_i and λ_j are fibers. Then there is a vertical annulus A in C joining λ_i to λ_j. There are also surfaces B_i and B_j in W_i and W_j with boundaries λ_i and λ_j. The union of A with $B_i \cup B_j$ is a closed surface in S^3. There is also a curve on T_i meeting this surface in exactly one point. This is impossible in S^3. □

9.3 Maps between the T_i

Without loss of generality, we will focus on T_1 and T_2. We denote a chosen element in $\mathcal{D}(L)$ that carries L_1 to L_2 by F. Note that we can assume $F(f_1) = f_2$, $F(\lambda_1) = \lambda_2$ and $F(\mu_1) = \mu_2$. However, maps of C do not necessarily preserve the g_i. We can assume that $F(g_1) = g_2 + wf_2$ for some w.

For both values of i we have constants such that

$$\lambda_i = \alpha_i f_i + \beta_i g_i, \quad \mu_i = \delta_i f_i + \gamma_i g_i.$$

Applying F to the set with $i = 1$ and renaming variables, we have

$$\lambda_1 = \alpha f_1 + \beta g_1, \quad \mu_1 = \delta f_1 + \gamma g_1$$

and

$$\lambda_2 = (\alpha f_2 + \beta g_2) + \beta w f_2, \quad \mu_2 = (\delta f_2 + \gamma g_2) + \gamma w f_2.$$

9.4 Constructing the transposition

Theorem 9.5 *There is a diffeomorphism G of C that interchanges T_1 and T_2 and is the identity on all other boundary components of C. The map G can be chosen so that it preserves the f_1 and satisfies $G(g_1) = g_2 + wf_2$ and $G(g_2) = g_1 - wf_2$.*

Proof Using the fact that T_i are boundaries of regular fibers, there is a diffeomorphism G of C that interchanges T_1 and T_2 that also preserves the pairs $\{f_i, g_i\}$. This map can be assumed to be the identity on the other components.

There is a vertical annulus in C joining f_1 and f_2. We can perform a w–fold twist along this annulus. This is the identity map on all boundary components other than T_1 and T_2. On T_1 and T_2, it preserves the f_1 and f_2, it maps g_1 to $g_1 - wf_1$ and it maps g_2 to $g_2 + wf_2$. □

9.5 Main theorem in the Seifert fibered case

Theorem 9.6 *If $\mathbb{A}_n \subset \mathbb{S}(L)$ and the associated core C is Seifert fibered, then there is an element $H \in \mathcal{D}(L)$ which transposes L_1 and L_2. Equivalently, $\mathbb{S}(L) = \mathbb{S}_n$.*

Proof The map G given in Theorem 9.5 satisfies

$$G(\lambda_1) = \alpha f_2 + \beta(g_2 + wf_2) \quad \text{and} \quad G(m_1) = \delta f_2 + \gamma(g_2 + wf_2).$$

It also satisfies

$$G(\lambda_2) = \alpha f_1 + \beta(g_1 - wf_1) + \beta w f_1 \quad \text{and} \quad G(\mu_1) = \delta f_1 + \gamma(g_1 - wf_1) + \gamma w f_1.$$

Simplifying shows that this interchanges the attaching maps of W to T_1 and T_2, and thus extends as desired. □

10 Questions

(1) For the four-component link illustrated in Figure 7, the group of symmetries that preserve string orientations is isomorphic to $\mathbb{A}_4$. This example was found by Nathan Dunfield using the program SnapPy [6], where it is listed as $L12a2007$. We have illustrated the link so that each component is in a regular neighborhood of a face of the standard projection of the tetrahedron to $S^2 = \mathbb{R}^2 \cup \infty$. Recall that the orientation-preserving symmetry group of the tetrahedron is isomorphic to $\mathbb{A}_4$.

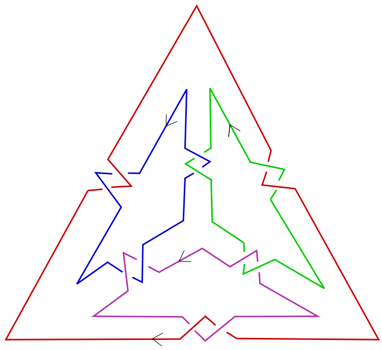


Figure 7: The link $L12a2007$.

Notice that rotation about the vertical axis interchanges two components, so the unoriented symmetry group is $\mathbb{S}(L) = \mathbb{S}_4$. However, if we build a new link by forming the connected sum of each component of $L12a2007$ with the same nonreversible knot J, then any symmetry of this new link, L', would have to preserve the orientations of the components. Thus, $\mathbb{S}(L') = \mathbb{A}_4$.

Similar constructions will likely produce links with symmetry groups that are isomorphic to polyhedral groups. For instance, using the dodecahedron would yield a 12–component link with symmetry group A_5. It is not clear how to reduce the number of components without changing the symmetry group, and we are left with the following question:

Does there exist a five-component link L with $\mathbb{S}(L) = \mathbb{A}_5$?

(2) The Fox–Whitten group Γ_n maps onto $\mathbb{S}_n$, and thus the obstructions we have developed here provide obstructions to groups $G \subset \Gamma_n$ from being oriented intrinsic symmetry groups of links. Can the techniques used here provide finer obstructions in the oriented case?

(3) As a particular example of (2), can any of the unknown cases for two-component links described in Section 4 be eliminated as possible intrinsic symmetry groups?

(4) If a subgroup $H \subset \mathbb{S}_n$ or $H \subset \Gamma_n$ is the intrinsic symmetry group for a link, is it the intrinsic symmetry group of a nonsplit link or of an irreducible link?

(5) A natural class of links consists of Brunnian links; these are nonsplittable but become the unlink upon removing any one of the components. The links produced in Theorem 5.1 having symmetry group $\mathbb{S}_n$ are not Brunnian. The examples of two-component and three-component links with $\mathbb{S}(L) = \mathbb{S}_n$ that precede the proof of that

theorem are Brunnian. Hence, we ask: For all $n \geq 3$, does there exist a Brunnian link L with $\mathbb{S}(L) = \mathbb{S}_n$?

(6) Another class to consider is alternating links, and presumably there are strong constraints on $\mathbb{S}(L)$ for these.

(7) Let M be a compact three-manifold with n torus boundary components, $\partial_i(M)$. Choose a basis of $H_1(\partial_i(M))$ for each i. One can form a Whitten-like group $\Omega_n = \mathbb{Z}_2 \oplus (G^n \rtimes \mathbb{S}_n)$, where G is the automorphism group of $\mathbb{Z} \oplus \mathbb{Z}$. Each manifold M gives rise to a subgroup of Ω_n. What subgroups arise in this way? This is particularly interesting in the case that the interior of M has a complete hyperbolic structure.

(8) The previous question can be modified. Given a subgroup $H \subset \mathbb{S}_n$, is there a complete hyperbolic three-manifold with n cusps such that the H represents the permutations of the cusps that are realized by isometries of M? In relation to this, Paoluzzi and Porti [11] proved that every finite group is the isometry group of the complement of a hyperbolic link in S^3. Notice that their isometries need not extend to S^3. Applying their construction to a subgroup of $\mathbb{S}_n$ does not produce an n–component link.

References

[1] **M Berglund**, **J Cantarella**, **M P Casey**, **E Dannenberg**, **W George**, **A Johnson**, **A Kelley**, **A LaPointe**, **M Mastin**, **J Parsley**, **J Rooney**, **R Whitaker**, *Intrinsic symmetry groups of links with 8 and fewer crossings*, Symmetry 4 (2012) 143–207 MR Zbl

[2] **M Boileau**, **B Leeb**, **J Porti**, *Geometrization of 3–dimensional orbifolds*, Ann. of Math. 162 (2005) 195–290 MR Zbl

[3] **R Budney**, *JSJ-decompositions of knot and link complements in S^3*, Enseign. Math. 52 (2006) 319–359 MR Zbl

[4] **R Budney**, *Is there a 2 component link with full symmetry*?, MathOverflow post (2012) Available at `http://mathoverflow.net/questions/91493`

[5] **J Cantarella**, **J Cornish**, **M Mastin**, **J Parsley**, *The 27 possible intrinsic symmetry groups of two-component links*, Symmetry 4 (2012) 129–142 MR Zbl

[6] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy: a computer program for studying the geometry and topology of 3–manifolds* (2016) Available at `http://snappy.computop.org/`

[7] **R Hartley**, *Identifying noninvertible knots*, Topology 22 (1983) 137–145 MR Zbl

[8] **J A Hillman**, *Symmetries of knots and links, and invariants of abelian coverings, I*, Kobe J. Math. 3 (1986) 7–27 MR Zbl

[9] **A Kawauchi**, *The invertibility problem on amphicheiral excellent knots*, Proc. Japan Acad. Ser. A Math. Sci. 55 (1979) 399–402 MR Zbl

[10] **P Orlik**, *Seifert manifolds*, Lecture Notes in Math. 291, Springer (1972) MR Zbl

[11] **L Paoluzzi**, **J Porti**, *Hyperbolic isometries versus symmetries of links*, Topology Appl. 156 (2009) 1140–1147 MR Zbl

[12] **J Porti**, *Geometrization of three manifolds and Perelman's proof*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. 102 (2008) 101–125 MR Zbl

[13] **H Schubert**, *Knoten und Vollringe*, Acta Math. 90 (1953) 131–286 MR Zbl

[14] **H F Trotter**, *Non-invertible knots exist*, Topology 2 (1963) 275–280 MR

[15] **F Waldhausen**, *Eine Klasse von 3–dimensionalen Mannigfaltigkeiten, I*, Invent. Math. 3 (1967) 308–333 MR Zbl

[16] **F Waldhausen**, *Eine Klasse von 3–dimensionalen Mannigfaltigkeiten, II*, Invent. Math. 4 (1967) 87–117 MR Zbl

[17] **W C Whitten, Jr**, *Symmetries of links*, Trans. Amer. Math. Soc. 135 (1969) 213–222 MR Zbl

Department of Mathematics, Indiana University
Bloomington, IN, United States

livingst@indiana.edu

Received: 1 November 2021 Revised: 13 January 2022

Algebraic & Geometric Topology **23**:5 (2023) 2369–2388
DOI: 10.2140/agt.2023.23.2369
Published: 25 July 2023

Loop homotopy of 6–manifolds over 4–manifolds

RUIZHI HUANG

Let M be the 6–manifold M arising as the total space of the sphere bundle of a rank 3 vector bundle over a simply connected closed 4–manifold. We show that, after looping, M is homotopy equivalent to a product of loops on spheres in general. This particularly implies a cohomological rigidity property of M after looping. Furthermore, passing to rational homotopy we show that such an M is Koszul.

55P15, 55P35, 57R19; 55P10, 55P40, 55P62

1 Introduction

Classification of manifolds is a fundamental problem in geometry and topology. Numerous investigations have been made around this problem in both the smooth and topological categories. For instance, in the general case, Wall [33; 35] studied $(n-1)$–connected $2n$–manifolds and $(n-1)$–connected $(2n+1)$–manifolds. For concrete cases with specified dimension, Barden [2] classified simply connected 5–manifolds, and Wall [34], Jupp [23] and Zhubr [37; 38] classified simply connected 6–manifolds. More recently, Kreck and Su [25] classified certain nonsimply connected 5–manifolds, while Crowley and Nordström [15] and Kreck [24] studied the classification of various kinds of 7–manifolds.

In the literature mentioned, the homotopy classification of M was usually carried out as a byproduct of a system of invariants. However, it is almost impossible to extract nontrivial homotopy information of M directly from the classification. On the other hand, unstable homotopy theory is a powerful tool for studying the homotopy properties of manifolds preserved by suspending or looping. From the suspension viewpoint, So and Theriault [31] determined the homotopy type of the suspension of connected 4–manifolds, while Huang [19] studied the suspension of simply connected 6–manifolds. From the loop viewpoint, Beben and Theriault [6] studied the loop decompositions of $(n-1)$–connected $2n$–manifolds, while Beben and Wu [8] and Huang and Theriault [20] studied the loop decompositions of the $(n-1)$–connected $(2n+1)$–manifolds. The homotopy groups of these manifolds were also investigated by Samik Basu and Somnath Basu [3; 4] from different point of view. Moreover, a theoretical method of loop decomposition was developed by Beben and Theriault [7], which is quite useful for studying the homotopy of manifolds.

We study the loop homotopy of certain simply connected 6–manifolds constructed from 4–manifolds. Let N be a simply connected closed 4–manifold with $H^2(N;\mathbb{Z}) \cong \mathbb{Z}^{\oplus d}$ for $d \geq 1$. A rank 3 vector bundle ξ over N is classified by a map $f\colon N \to B\mathrm{SO}(3)$, where $B\mathrm{SO}(3)$ is the classifying space of the special orthogonal group $\mathrm{SO}(3)$. The sphere bundle of ξ

$$S^2 \xrightarrow{i} M \xrightarrow{p} N \tag{1}$$

defines the closed 6–manifold M. Since the integral cohomologies of N and S^2 are free and concentrated in even degree, the Serre spectral sequence of (1) collapses, and $H^*(M;\mathbb{Z}) \cong H^*(N;\mathbb{Z}) \otimes H^*(S^2;\mathbb{Z})$. Our main result is the following theorem, which will be proved in Section 4.

Theorem 1.1 *Let N be a simply connected closed 4–manifold with $H^2(N;\mathbb{Z}) \cong \mathbb{Z}^{\oplus d}$ for $d \geq 1$. Let M be the total manifold of the sphere bundle of a rank 3 vector bundle over N. Then:*

- *If $d = 1$,*
$$\Omega M \simeq S^1 \times \Omega S^2 \times \Omega S^5.$$
- *If $d \geq 2$,*
$$\Omega M \simeq S^1 \times \Omega S^2 \times \Omega(S^2 \times S^3) \times \Omega\big(J \vee (J \wedge \Omega(S^2 \times S^3))\big),$$
where $J = \bigvee_{i=1}^{d-2}(S^2 \vee S^3)$.

From Theorem 1.1 and its proof, it can be easily seen that the decompositions in Theorem 1.1 are compatible with the S^2–bundle (1) after looping. In particular, this means that though the fibre bundle (1) does not split in general, its loop does. Moreover, as discussed in [6, page 217], the term $J \vee (J \wedge \Omega(S^2 \times S^3))$ in the second decomposition of Theorem 1.1 is a bouquet of spheres. Hence by the Hilton–Milnor theorem, we see that ΩM is homotopy equivalent to a product of loops on spheres with S^1. Additionally, since the decompositions of Theorem 1.1 only depend on the value of d, which is determined by and determines $H^2(M;\mathbb{Z})$, we have a rigidity property for M after looping.

Corollary 1.2 *Let M and M' be two 6–manifolds satisfying the conditions of Theorem 1.1. Then $\Omega M \simeq \Omega M'$ if and only if $H^2(M;\mathbb{Z}) \cong H^2(M';\mathbb{Z})$.* □

Theorem 1.1 can be improved if we pass from integral homotopy to rational homotopy. Indeed, by Theorem 1.1 it is straightforward to compute the homotopy groups of M in terms of those of spheres. However, there is an additional Lie algebra structure on the homotopy groups of any *CW* complex X. In rational homotopy theory, the graded Lie algebra $\pi_*(\Omega X)\otimes\mathbb{Q}$ is called the *homotopy Lie algebra* of X, and X is called *coformal* if the rational homotopy type of X is completely determined by its homotopy Lie algebra. If X is further *formal*, that is the homotopy type of X is determined by the graded commutative algebra $H^*(X;\mathbb{Q})$, then X is *Koszul* in the sense of Berglund [9, Definition 1.1]. In the latter case, $H^*(X;\mathbb{Q})$ is a *Koszul algebra* and $\pi_*(\Omega X)\otimes\mathbb{Q}$ is a *Koszul Lie algebra* [9]. The following theorem concerns these additional structures on M of the type in Theorem 1.1.

Theorem 1.3 *Let N be a simply connected closed 4–manifold with $H^2(N;\mathbb{Z})\cong\mathbb{Z}^{\oplus d}$. Let M be the total manifold of the sphere bundle of a rank 3 vector bundle over N. Then*:

- *If $d=1$, M is not coformal.*
- *If $d\geq 2$, M is Koszul, and there is an isomorphism of graded Lie algebras*

$$\pi_*(\Omega M)\otimes\mathbb{Q}\cong H^*(M;\mathbb{Q})^{!\mathscr{L}ie},$$

where $(-)^{!\mathscr{L}ie}$ is the Koszul dual Lie functor defined in [9, Section 2].

We turn to the remaining case, when $d=0$, that is, $N\cong S^4$. Note, we still have the 6–manifold M as constructed in (1). Though the homotopy classification of such manifolds was almost determined by Yamaguchi [36], this case is surprisingly much harder than the general one. We will explain this point after the statement of our result in this case. Let $\eta_2\colon S^3\to S^2$ be the Hopf map. For any integer n, let $S^m\{n\}$ be the homotopy fibre of the degree n map on S^m.

Theorem 1.4 *Let M be the total space of the sphere bundle of a rank 3 vector bundle over S^4. Then M has a cell structure of the form*

$$M \simeq S^2 \cup_{k\eta_2} e^4 \cup e^6,$$

where $k \in \mathbb{Z}$. Let $k = p_1^{r_1} \cdots p_\ell^{r_\ell}$ be the prime decomposition of k. Further:

- *If k is odd,*
$$\Omega M \simeq S^1 \times \prod_{j=1}^{\ell} S^3\{p_j^{r_j}\} \times \Omega S^7.$$
- *If $k = 2^r$ with $r \geq 3$,*
$$\Omega M \simeq S^1 \times S^3\{2^r\} \times \Omega S^7.$$

Note that we still have cohomological rigidity in this case, since the homotopy type of ΩM only depends on k, which is determined by the square of a generator in $H^2(M;\mathbb{Z})$. But it is less interesting since the cohomological rigidity of M without looping holds except for the case when k is even and M is Spin [36]. Further note that Theorem 1.4 is only a partial result. The difficulty in this case is due to the fact that the proof of Theorem 1.4 heavily relies on a result of Huang and Theriault [20] on the loop decomposition of 2–connected 7–manifolds. As discussed in [20, Section 6], the case when $k = 2^r m$ with m odd and greater than 1 is much more difficult. Also, since it is known that $S^3\{2\}$ is not an H–space (see Cohen [11]), we cannot expect a decomposition of the form $\Omega M \simeq S^1 \times S^3\{2\} \times \Omega S^7$ for the case when $k = 2$. In contrast, the rational homotopy of M in this case is simple. As shown in Lemma 5.2, M is rationally homotopy equivalent to $\mathbb{C}P^3$ or $S^2 \times S^4$. Moreover, it is well known that $\mathbb{C}P^3$ is not coformal (see Neisendorfer and Miller [27, Example 4.7]), while $S^2 \times S^4$ is Koszul; see Berglund [9, Examples 5.1 and 5.4].

Before we close the introduction, let us make two remarks. Firstly, our results provide further evidence on the Moore conjecture. Recall that the Moore conjecture states that a simply connected finite *CW* complex Z is rationally elliptic if and only if it has a finite homotopy exponent at all primes, or equivalently, Z is rationally hyperbolic if and only if it has unbound homotopy exponent at some prime. For M in our context, it is elliptic if and only if $d \leq 2$, and in any of these cases M has a finite homotopy exponent at all primes by Cohen, Moore and Neisendorfer [12; 13] and James [21]. When $d \geq 3$, M is hyperbolic such that ΩM has $\Omega(S^2 \vee S^3)$ as product summand, hence it has no bound on its homotopy exponent for any prime p; see Neisendorfer and Selick [28] or Boyde [10] for instance. Secondly, Amorós and Biswas [1] characterized simply connected rationally elliptic compact Kähler threefolds in terms of Hodge diamonds, and in particular, their second Betti numbers satisfy $b_2 \leq 3$. For M in our context, this is equivalent to $d \leq 2$, and our decompositions provide further information on the homotopy of M. For instance, the homotopy groups of M can be computed in terms of those of spheres.

The paper is organized as follows. In Section 2 we classify rank 3 bundles over the 4–manifold N. In Section 3, we prove Lemma 3.1, which implies that under Lemma 2.1 one component of the classifying map f of the bundle ξ over N is trivial in a special case. This is crucial for proving Theorem 1.1. In Section 4, we prove Theorem 1.1 by dividing it into two cases. Section 5 is devoted to the remaining case when $d = 0$ and we prove Theorem 1.4 there. We discuss the rational homotopy of 6–manifolds and prove Theorem 1.3 in Section 6.

Acknowledgements Ruizhi Huang was supported by the National Natural Science Foundation of China (grants 11801544 and 11688101), and the Chen Jingrun Future Star Program of the AMSS. He would like to thank Professor Stephen Theriault for the international online lecture series *Loop space decomposition*, which stimulated his research interest in the homotopy of 6–manifolds. He also want to thank Professor Yang Su for helpful discussions on obstructions to trivializing vector bundles.

2 Rank 3 bundles over 4–manifolds

In this section, we discuss necessary knowledge of rank 3 vector bundles over simply connected 4–manifolds, which will be used in the subsequent sections. There are various ways to study the classification of vector bundles. Here, we adopt an approach from a homotopy theoretical point of view for later use.

Let N be a simply connected 4–manifold such that $H^2(N;\mathbb{Z}) \cong \mathbb{Z}^{\oplus d}$ with $d \geq 0$. A rank 3 vector bundle ξ over N is classified by a map $f\colon N \to B\mathrm{SO}(3)$. The sphere bundle of ξ

$$S^2 \xrightarrow{i} M \xrightarrow{p} N$$

defines the closed 6–manifold M. For N, there is the homotopy cofiber sequence

$$S^3 \xrightarrow{\phi} \bigvee_{i=1}^{d} S^2 \xrightarrow{\rho} N \xrightarrow{q} S^4 \xrightarrow{\Sigma\phi} \bigvee_{i=1}^{d} S^3, \tag{2}$$

where ϕ is the attaching map of the top cell of N, ρ is the injection of the 2–skeleton, and q is the pinch map onto the top cell. Let $s\colon S^1 \cong \mathrm{SO}(2) \to \mathrm{SO}(3)$ be the canonical inclusion of Lie groups.

Lemma 2.1 *There is a surjection*

$$\Phi\colon [S^4, B\mathrm{SO}(3)] \times [N, BS^1] \to [N, B\mathrm{SO}(3)]$$

of pointed sets that restricts to q^* on $[S^4, B\mathrm{SO}(3)]$, and to $(Bs)_*$ on $[N, BS^1]$.

Proof By (2), there is the exact sequence of pointed sets

$$0 = \left[\bigvee_{i=1}^{d} S^3, B\mathrm{SO}(3)\right] \to [S^4, B\mathrm{SO}(3)] \xrightarrow{q^*} [N, B\mathrm{SO}(3)] \xrightarrow{\rho^*} \left[\bigvee_{i=1}^{d} S^2, B\mathrm{SO}(3)\right] \to [S^3, B\mathrm{SO}(3)] = 0,$$

in a strong sense: there is an action of $[S^4, B\mathrm{SO}(3)]$ on $[N, B\mathrm{SO}(3)]$ through q^* such that the sets $\rho^{*-1}(x)$, for $x \in \left[\bigvee_{i=1}^{d} S^2, B\mathrm{SO}(3)\right]$, are precisely the orbits. It is known that $\left[\bigvee_{i=1}^{d} S^2, B\mathrm{SO}(3)\right] \cong \bigoplus_d \mathbb{Z}/2\mathbb{Z}$ and $[S^4, B\mathrm{SO}(3)] \cong \mathbb{Z}$. Moreover, there is the commutative diagram

$$\begin{tikzcd}
{[N, BS^1]} \arrow[r, "\rho^*", "\cong"'] \arrow[d, "(Bs)_*"] & \left[\bigvee_{i=1}^{d} S^2, BS^1\right] \arrow[r, "\cong"] \arrow[d, "(Bs)_*"] & \bigoplus_d \mathbb{Z} \arrow[d, "\bigoplus_d \rho_2"] \\
{[N, B\mathrm{SO}(3)]} \arrow[r, "\rho^*"] & \left[\bigvee_{i=1}^{d} S^2, B\mathrm{SO}(3)\right] \arrow[r, "\cong"] & \bigoplus_d \mathbb{Z}/2\mathbb{Z}
\end{tikzcd}$$

where ρ^* is an isomorphism onto $\left[\bigvee_{i=1}^{d} S^2, BS^1\right] \cong \bigoplus_d \mathbb{Z}$ and ρ_2 is the mod 2 reduction, hence $(Bs)_*$ is surjective onto $\left[\bigvee_{i=1}^{d} S^2, B\mathrm{SO}(3)\right]$. Now for any $f \in [N, B\mathrm{SO}(3)]$ we have $\rho^*(f) = (Bs)_*(x)$ for some $x \in \left[\bigvee_{i=1}^{d} S^2, BS^1\right]$. Write $\alpha = (\rho^{*-1})(x)$. Then $Bs_*(\alpha)$ and f belong to the same orbit of the action, for they have same image in $\left[\bigvee_{i=1}^{d} S^2, B\mathrm{SO}(3)\right]$ through ρ^*. Hence, there exists an $f' \in [S^4, B\mathrm{SO}(3)]$ such that $q^*(f') \cdot (Bs_*(\alpha)) = f$. □

From Lemma 2.1 and its proof, for the classifying map $f\colon N \to B\mathrm{SO}(3)$, we have associated a pair of maps

(3) $$(f', \alpha) \in [S^4, B\mathrm{SO}(3)] \times [N, BS^1],$$

such that $q^*(f')\cdot(Bs_*(\alpha)) = f$ and $\omega_2(\xi) \equiv \alpha \mod 2$. We also notice that if $\rho^*(f) \neq 0$, or equivalently ξ is non-Spin, the element α can be always chosen to be primitive, that is, α is not divisible by any integer k with $k \neq \pm 1$. This is important for our later use.

Let $\pi\colon W \to N$ be a map from a closed manifold W. The pullback of the bundle ξ along π has an associated sphere bundle

$$S^2 \xrightarrow{\iota} Z \xrightarrow{\mathfrak{p}} W,$$

which defines the closed manifold Z. The following lemma is critical for proving Proposition 4.1.

Lemma 2.2 *Suppose for W there is a homotopy cofibration*

$$W_{m-1} \xrightarrow{\varrho} W \xrightarrow{\mathfrak{q}} S^m,$$

such that $\pi \circ \varrho$ factors as

$$W_{m-1} \xrightarrow{\pi_\prime} \bigvee_{i=1}^{d} S^2 \xrightarrow{\rho} N$$

for some $\pi_\prime$, where W_{m-1} is the $(m-1)$–skeleton of W. Then if $f' \circ q \circ \pi$ and $\alpha \circ \pi$ are both nullhomotopic, the bundle $\pi^*(\xi)$ is trivial, and in particular

$$Z \cong S^2 \times W.$$

Proof By the assumption, there is a diagram of homotopy cofibrations

$$\begin{tikzcd} W_{m-1} \arrow[r, "\varrho"] \arrow[d, "\pi_\prime"] & W \arrow[r, "\mathfrak{q}"] \arrow[d, "\pi"] & S^m \arrow[d, "\pi'"] \\ \bigvee_d S^2 \arrow[r, "\rho"] & N \arrow[r, "q"] & S^4 \end{tikzcd}$$

which defines the map π'. It follows that there is a morphism of exact sequences of pointed sets

$$\begin{tikzcd} {[S^4, B\mathrm{SO}(3)]} \arrow[r, "q^*"] \arrow[d, "\pi'^*"] & {[N, B\mathrm{SO}(3)]} \arrow[r, "\rho^*"] \arrow[d, "\pi^*"] & {\left[\bigvee_{i=1}^d S^2, B\mathrm{SO}(3)\right]} \arrow[d, "\pi_\prime^*"] \\ {[S^m, B\mathrm{SO}(3)]} \arrow[r, "\mathfrak{q}^*"] & {[W, B\mathrm{SO}(3)]} \arrow[r, "\varrho^*"] & {[W_{m-1}, B\mathrm{SO}(3)]} \end{tikzcd}$$

such that the action of $[S^4, B\mathrm{SO}(3)]$ on $[N, B\mathrm{SO}(3)]$ is compatible with the action of $[S^m, B\mathrm{SO}(3)]$ on $[W, B\mathrm{SO}(3)]$ through π'^*. Hence, by (3), the classifying map $f\circ\pi$ of $\pi^*(\xi)$ satisfies

$$\begin{aligned} f\circ\pi &= \pi^*\big(q^*(f')\cdot(Bs_*(\alpha))\big) = \mathfrak{q}^*(\pi'^*(f'))\cdot\pi^*(Bs_*(\alpha)) \\ &= \pi^*(q^*(f'))\cdot\pi^*((Bs_*(\alpha))) = (f'\circ q\circ\pi)\cdot Bs_*(\alpha\circ\pi), \end{aligned}$$

which is nullhomotopic by the assumption. □

Lemma 2.1 also gives a byproduct on the classification of rank 3 vector bundles over N via characteristic classes, which could also be proved by other methods, like the classical obstruction theory.

Proposition 2.3 *A rank 3 vector bundle ξ over N is completely determined by its second Stiefel–Whitney class $\omega_2(\xi)$ and its first Pontryagin class $p_1(\xi)$.*

Proof Given two rank 3 vector bundles ξ_1 and ξ_2 over N, suppose that $\omega_2(\xi_1)=\omega_2(\xi_2)$ and $p_1(\xi_1)=p_1(\xi_2)$. We want to show that $\xi_1\cong\xi_2$, or equivalently, $f_1\simeq f_2$, where $f_1, f_2\colon N\to B\mathrm{SO}(3)$ are the classifying maps of ξ_1 and ξ_2, respectively. By Lemma 2.1 and (3), $f_1=q^*(f_1')\cdot(Bs_*(\alpha))$ for a pair of maps $(f_1',\alpha)\in[S^4, B\mathrm{SO}(3)]\times[N, BS^1]$ such that $\omega_2(\xi_1)\equiv\alpha \mod 2$. Since $\omega_2(\xi_1)=\omega_2(\xi_2)$, there exists $f_2'\in[S^4, B\mathrm{SO}(3)]$ such that $f_2=q^*(f_2')\cdot(Bs_*(\alpha))$. It follows that to show $f_1\simeq f_2$, it suffices to show $f_1'\simeq f_2'$. Indeed, for either ξ_i the expression of f_i can be explicitly described as

$$f_i\colon N\xrightarrow{\mu'} N\vee S^4\xrightarrow{\alpha\vee f_i'} BS^1\vee B\mathrm{SO}(3)\xrightarrow{Bs\vee\mathrm{id}} B\mathrm{SO}(3)\vee B\mathrm{SO}(3)\xrightarrow{\nabla} B\mathrm{SO}(3),$$

where μ' is the coaction map and ∇ is the folding map. In particular, it is easy to see that

$$p_1(\xi_i)=q^*(p_1(f_i'))+\alpha^2, \tag{4}$$

where we denote by $p_1(f_i')$ the first Pontryagin class of the bundle over S^4 determined by f_i'. Since $p_1(\xi_1)=p_1(\xi_2)$, (4) implies that $q^*(p_1(f_1'))=q^*(p_2(f_i'))$. Moreover, it is clear that $q^*\colon H^4(S^4;\mathbb{Z})\to H^4(N;\mathbb{Z})$ is an isomorphism. Hence $p_1(f_1')=p_1(f_2')$. Now since $[S^4, B\mathrm{SO}(3)]\simeq\mathbb{Z}$, and the morphism $\frac{1}{4}p_1\colon[S^4, B\mathrm{SO}(3)]\to H^4(S^4;\mathbb{Z})$ sending each map to one fourth of the first Pontryagin class of the associated bundle is an isomorphism [18], we see that $f_1'\simeq f_2'$. Then $f_1\simeq f_2$ and the proposition follows. □

3 The induced map between top cells

Let N be a simply connected closed 4–manifold such that $H^2(N;\mathbb{Z}) \cong \mathbb{Z}^{\oplus d}$ with $d \geq 1$. Consider the circle bundle

$$S^1 \xrightarrow{j} Y \xrightarrow{\pi} N$$

classified by a primitive element $\beta \in H^2(N;\mathbb{Z})$, which defines the simply connected 5–manifold Y. By [16, Lemma 1], Y has cell structure of the form

$$Y \simeq \bigvee_{d-1} (S^2 \vee S^3) \cup e^5.$$

Then, by the cellular approximation theorem, there is the diagram of homotopy cofibration

$$\begin{tikzcd} \bigvee_{d-1}(S^2 \vee S^3) \arrow[r, "\varrho"] \arrow[d] & Y \arrow[r, "\mathfrak{q}"] \arrow[d, "\pi"] & S^5 \arrow[d, "\pi'"] \\ \bigvee_d S^2 \arrow[r, "\rho"] & N \arrow[r, "q"] & S^4 \end{tikzcd} \tag{5}$$

where the bottom cofibration is part of (2), ϱ is the inclusion of the 3–skeleton of Y followed by the quotient $\mathfrak{q}$, and π' is induced from π. In this section, we prove the following key lemma for understanding rank 3–bundles over Y in a special case. Let $[N]$ be the fundamental class of N. Let $\langle x \cup y, [N]\rangle \in \mathbb{Z}$ be the canonical pairing for any cohomology classes x, $y \in H^2(N;\mathbb{Z})$.

Lemma 3.1 *The induced map π' in (5) is nullhomotopic when $\langle \beta^2, [N]\rangle$ is odd.*

Proof The primitive element β is represented by a map $\beta\colon N \to \mathbb{C}P^\infty \simeq K(\mathbb{Z},2)$. By the cellular approximation theorem, β factors through $\mathbb{C}P^2$,

$$\beta\colon N \xrightarrow{\tilde{\beta}} \mathbb{C}P^2 \xrightarrow{x} \mathbb{C}P^\infty,$$

which defines the map $\tilde{\beta}$, and x represents a generator $x \in H^2(\mathbb{C}P^2;\mathbb{Z})$. The factorization gives a diagram of circle bundles

$$\begin{tikzcd} S^1 \arrow[r, "j"] \arrow[d, equal] & Y \arrow[r, "\pi"] \arrow[d, "\hat{\beta}"] & N \arrow[d, "\tilde{\beta}"] \\ S^1 \arrow[r] & S^5 \arrow[r, "\pi_0"] & \mathbb{C}P^2 \end{tikzcd} \tag{6}$$

where the bundle in the second row is classified by x, and $\widehat{\beta}$ is the induced map. By the cellular approximation theorem, there is a homotopy commutative diagram

(7)
$$\begin{tikzcd}
& Y \arrow[rr, "\pi"] \arrow[dl, "q"'] \arrow[dd, "\widehat{\beta}"] & & N \arrow[dl, "q"'] \arrow[dd, "\widetilde{\beta}"] \\
S^5 \arrow[rr, "\pi'"] \arrow[dd, "\widehat{\beta}'"] & & S^4 \arrow[dd, "\widetilde{\beta}'"] & \\
& S^5 \arrow[rr, "\pi_0"] \arrow[dl, equal] & & \mathbb{C}P^2 \arrow[dl, "q_0"] \\
S^5 \arrow[rr, "\pi'_0"] & & S^4 &
\end{tikzcd}$$

where the rear and top faces are the right squares in (6) and (5), respectively, q_0 is the quotient map onto the top cell of $\mathbb{C}P^2$, π'_0 is defined to be $q_0 \circ \pi_0$, and $\widehat{\beta}'$ and $\widetilde{\beta}'$ are the induced maps. By the homotopy commutativity of the right face of (7), the assumption that $\langle \beta^2, [N] \rangle$ is odd is equivalent to $\widetilde{\beta}'$ having odd degree. Further, since the homotopy cofibre of π_0 is $\mathbb{C}P^3$, for which the Steenrod operation $\mathrm{Sq}^2\colon H^4(\mathbb{C}P^3; \mathbb{Z}/2\mathbb{Z}) \to H^6(\mathbb{C}P^3; \mathbb{Z}/2\mathbb{Z})$ is trivial, we obtain that $\pi'_0 = q_0 \circ \pi_0$ is nullhomotopic. Now consider the front face of (7). Combining the above arguments and the fact that $\pi_5(S^4) \cong \mathbb{Z}/2\mathbb{Z}\{\eta_4\}$ [32], we see that $\pi' \simeq \widetilde{\beta}' \circ \pi' \simeq \pi'_0 \circ \widehat{\beta}'$ is nullhomotopic. □

4 Proof of Theorem 1.1

Let N be a simply connected 4–manifold such that $H^2(N; \mathbb{Z}) \cong \mathbb{Z}^{\oplus d}$ with $d \geq 1$. A rank 3 vector bundle ξ over N is classified by a map $f\colon N \to B\mathrm{SO}(3)$ with the associated sphere bundle

$$S^2 \xrightarrow{i} M \xrightarrow{p} N,$$

which defines the closed 6–manifold M. Recall, by Lemma 2.1 and (3), the classifying map $f\colon N \to B\mathrm{SO}(3)$ for the bundle ξ is determined by

$$(f', \alpha) \in [S^4, B\mathrm{SO}(3)] \times [N, BS^1]$$

such that $f = q^*(f') \cdot (Bs)_*(\alpha)$ and $\omega_2(\xi) \equiv \alpha \bmod 2$, where q and s are defined before Lemma 2.1. Moreover, by the discussion after Lemma 2.1, when ξ is non-Spin we suppose that α is primitive.

For the loop homotopy of M, we may study S^1–bundles over M pulled back from those over the 4–manifold N. Consider the circle bundle

(8)
$$S^1 \xrightarrow{j} Y \xrightarrow{\pi} N$$

classified by a primitive element $\beta \in H^2(N;\mathbb{Z})$, which defines the simply connected 5–manifold Y. Based on the previous remark on the choice of α, we make the following convention on the choice of β:

- $\beta = \alpha$ if ξ is non-Spin, or
- β can be any primitive element if ξ is Spin.

The remainder of this section is devoted to the proof of Theorem 1.1 by dividing it into two cases according to the parity of $\langle \beta^2, [N] \rangle$. In Section 4.1, we prove Theorem 1.1 using Lemma 3.1 under the assumption that $\langle \beta^2, [N] \rangle$ is odd. This is the case when the circle bundle (8) plays an essential role. However, when $\langle \beta^2, [N] \rangle$ is even, we have to apply a different method to prove Theorem 1.1. This is done in Section 4.2.

4.1 Case I: $\langle \beta^2, [N] \rangle$ is odd

In this case, by the choice of the circle bundle (8), consider the pullback of fibre bundles

$$\begin{tikzcd} & S^2 \arrow[r, equal] \arrow[d, "\iota"] & S^2 \arrow[d, "i"] \\ S^1 \arrow[r, "J"] \arrow[d, equal] & X \arrow[r, "\psi"] \arrow[d, "\mathfrak{p}"] & M \arrow[d, "p"] \\ S^1 \arrow[r, "j"] & Y \arrow[r, "\pi"] & N \end{tikzcd} \tag{9}$$

which defines the closed 7–manifold X with bundle projections ψ and $\mathfrak{p}$ onto M and Y, respectively. We show that the induced bundle over Y in (9) is trivial in this case.

Proposition 4.1 *If $\langle \beta^2, [N] \rangle$ is odd, then the bundle $\pi^*(\xi)$ defined in (9) is trivial, and, in particular,*

$$X \cong S^2 \times Y.$$

Proof By Lemma 3.1, π' is nullhomotopic. This implies that $f' \circ q \circ \pi \simeq f' \circ \pi' \circ \mathfrak{q}$ is nullhomotopic by the homotopy commutativity of the right square in (5).

If ξ is non-Spin, then $\beta = \alpha$. We obtain the homotopy fibration $Y \xrightarrow{\pi} N \xrightarrow{\alpha} BS^1$, which implies that $\alpha \circ \pi$ is nullhomotopic, hence so is $(Bs_*)(\alpha \circ \pi)$. Then by Lemma 2.2 the classifying map $f \circ \pi$ of the bundle $\pi^*(\xi)$ is nullhomotopic, and the proposition follows in this case.

If ξ is Spin, by Lemma 2.1 the classifying map $f\colon N \to B\mathrm{SO}(3)$ of ξ is in the image of q^*, that is, there exists a map $f'\colon S^4 \to B\mathrm{SO}(3)$ such that $f' \circ q \simeq f$, and then the bundle ξ is the pullback of the bundle ξ' over S^4 classified by f'. Hence $f\circ\pi\simeq f'\circ q\circ\pi$ is nullhomotopic by the previous argument, and then the bundle $\pi^*(\xi)$ is trivial. In particular, $X\cong S^2\times Y$ and the proposition follows in this case. □

Proof of Theorem 1.1 in Case I As in the beginning of this subsection, consider the circle bundle $S^1\xrightarrow{j}Y\xrightarrow{\pi}N$ classified by the primitive element $\alpha\in H^2(N;\mathbb{Z})$. Then by Proposition 4.1, the total space X of the sphere bundle of $\pi^*(\xi)$ satisfies $X\cong S^2\times Y$. Hence, by (9),

$$\Omega M\simeq S^1\times\Omega X\simeq S^1\times\Omega S^2\times\Omega Y. \tag{10}$$

If $d=1$, then Y has to be S^5, and hence $\Omega M\simeq S^1\times\Omega S^2\times\Omega S^5$. If $d\geq 2$, by [7, Example 4.4] or [3] there is a homotopy equivalence

$$\Omega Y\simeq\Omega(S^2\times S^3)\times\Omega\big(J\vee(J\wedge\Omega(S^2\times S^3))\big) \tag{11}$$

with $J=\bigvee_{i=1}^{d-2}(S^2\vee S^3)$. Combining (10) with (11), we obtain the loop decomposition of M in the theorem. □

4.2 Case II: $\langle\beta^2,[N]\rangle$ is even

In this case, the induced bundle $\pi^*(\xi)$ defined in (9) may not be trivial, and we need to apply a different method to prove Theorem 1.1. Indeed, in this case we can work with the sphere bundle $S^2\xrightarrow{i}M\xrightarrow{p}N$ directly, and show that it splits after looping.

Proposition 4.2 *If $\langle\beta^2,[N]\rangle$ is even, the sphere bundle $S^2\xrightarrow{i}M\xrightarrow{p}N$ of ξ defined in (1) is homotopically trivial after looping, and in particular*

$$\Omega M\simeq\Omega S^2\times\Omega N.$$

Proof By Poincaré duality there exists a class $\alpha\in H^2(N;\mathbb{Z})$ such that $\langle\alpha\cup\beta,[N]\rangle=1$. Since by assumption $\langle\beta^2,[N]\rangle$ is even, $\alpha\neq\beta$. Hence by [6, proof of proposition 3.2 and Lemma 3.3] there is a Poincaré duality space Q such that $H^*(Q;\mathbb{Z})\cong H^*(S^2\times S^2;\mathbb{Z})$ as graded rings, $\Omega Q\simeq\Omega S^2\times\Omega S^2$, and there is a map

$$h\colon N\to Q$$

such that Ωh has a right homotopy inverse and $h^*(x)=\alpha$ with $x\in H^2(Q;\mathbb{Z})$ a generator. Let us fix a homotopy equivalence $e\colon\Omega S^2\times\Omega S^2\to\Omega Q$ defined in [6, Lemma 2.3] with its inverse denoted by e^{-1}.

Recall, ξ is determined by a pair of maps $(f', \alpha) \in [S^4, B\mathrm{SO}(3)] \times [N, BS^1]$. By Lemma 2.1, define a rank 3 vector bundle ζ over Q by $(f', x) \in [S^4, B\mathrm{SO}(3)] \times [Q, BS^1]$. It follows that $\xi = h^*(\zeta)$ and there is a pullback of sphere bundles

(12)
$$\begin{tikzcd} S^2 \arrow[r, "i"] \arrow[d, equal] & M \arrow[r, "p"] \arrow[d, "\tilde{h}"] & N \arrow[d, "h"] \\ S^2 \arrow[r, "\tilde{i}"] & \tilde{Q} \arrow[r, "\tilde{p}"] & Q \end{tikzcd}$$

where the second row is the sphere bundle of ζ and $\tilde{h}$ is the induced map. Since $H^*(Q;\mathbb{Z})$ and $H^*(S^2;\mathbb{Z})$ are concentrated in even degrees, the Serre spectral sequence for the fibration $S^2 \to \tilde{Q} \to Q$ collapses for degree reasons, and then

$$H^*(\tilde{Q};\mathbb{Z}) \cong H^*(S^2;\mathbb{Z}) \otimes H^*(Q;\mathbb{Z}).$$

Apply the loop functor to (12). It is clear that there is a map $i_1 \times i_2 \colon S^1 \times S^1 \to \Omega\tilde{Q}$ such that the composition

$$S^1 \times S^1 \xrightarrow{i_1 \times i_2} \Omega\tilde{Q} \xrightarrow{\Omega\tilde{p}} \Omega Q \xrightarrow{e^{-1}} \Omega S^2 \times \Omega S^2$$

is homotopic to $E \times E$ with $E \colon S^1 \to \Omega S^2$ the suspension map. By the universal property of $\Omega\Sigma$, there is a unique extension $I \colon \Omega S^2 \times \Omega S^2 \to \Omega\tilde{Q}$ of $i_1 \times i_2$ up to homotopy such that

$$\Omega S^2 \times \Omega S^2 \xrightarrow{I} \Omega\tilde{Q} \xrightarrow{\Omega\tilde{p}} \Omega Q \xrightarrow{e^{-1}} \Omega S^2 \times \Omega S^2$$

is homotopic to the identity. Therefore, the sphere bundle of ζ splits after looping to give

$$\Omega\tilde{Q} \simeq \Omega S^2 \times \Omega Q \simeq \Omega S^2 \times \Omega S^2 \times \Omega S^2.$$

In particular, $\Omega\tilde{i}$ has a left homotopy inverse $\tilde{r}$, which implies that $\tilde{r} \circ \Omega\tilde{h}$ is a left homotopy inverse of Ωi. Then the sphere bundle in the top row of (12) splits after looping, and in particular $\Omega M \simeq \Omega S^2 \times \Omega N$. □

Proof of Theorem 1.1 in Case II Since $\langle \beta^2, [N] \rangle$ is even and β is primitive, we have $d \geq 2$. By Proposition 4.2, $\Omega M \simeq \Omega S^2 \times \Omega N$. Further, by [6, Theorem 1.3] there is a homotopy equivalence

$$\Omega N \simeq S^1 \times \Omega(S^2 \times S^3) \times \Omega\big(J \vee (J \wedge \Omega(S^2 \times S^3))\big)$$

with $J = \bigvee_{i=1}^{d-2} (S^2 \vee S^3)$. Then in this case the theorem follows by combining the above decompositions. □

5 The case when $d = 0$

In this section, we study the case when $d = 0$ and prove Theorem 1.4 as an immediate corollary of Propositions 5.3 and 5.4. Indeed, we work in a slightly more general context, that is, we study the loop decomposition of the closed 6–manifold M with cell structure of the form

$$M \simeq S^2 \cup e^4 \cup e^6. \tag{13}$$

Notice that M in Theorem 1.4, as the total space of an S^2–bundle over S^4, is an example of (13). Yamaguchi [36] almost determined the homotopy classification of M in (13) with correction by [5; 29], and summarized the criterion for whether M has the same homotopy type as an S^2–bundle over S^4 in [36, Remark 4.8] based on [30].

By (13) there are generators $x \in H^2(M;\mathbb{Z})$ and $y \in H^4(M;\mathbb{Z})$ such that

$$x^2 = ky \tag{14}$$

for some $k \in \mathbb{Z}$. Consider the S^1–bundle

$$S^1 \xrightarrow{j} X \to M \tag{15}$$

classified by $x \in H^2(M;\mathbb{Z}) \cong [M, BS^1]$, which defines the closed 7–manifold X. Let $P^n(k)$ be the Moore space such that $\widetilde{H}^*(P^n(k);\mathbb{Z}) \cong \mathbb{Z}/k\mathbb{Z}$ if $* = n$ and 0 otherwise [26].

Lemma 5.1 *If $k \neq 0$, there is a homotopy equivalence*

$$X \simeq P^4(k) \cup e^7.$$

Proof The lemma can be proved directly by analyzing the Serre spectral sequence of the fibration $X \to M \xrightarrow{x} BS^1$ induced from (15). Here we provide an alternative proof using results in geometric topology. By [22, Theorem 1.3], X is homotopy equivalent to the total space of an S^3–bundle over S^4. Then by the homotopy classification of S^3–bundles over S^4 [14; 30], X is homotopy equivalent to $P^4(k') \cup e^7$ for some $k' \in \mathbb{Z}$. Notice that $\pi_3(X) \cong \pi_3(M) \cong \pi_3(S^2 \cup_{k\eta_2} e^4) \cong \mathbb{Z}/k$, where $\eta_2 \in \pi_3(S^2)$ is the Hopf element. Then $k = k'$ because $\pi_3(P^4(k') \cup e^7) \cong \mathbb{Z}/k'$, and the lemma follows. □

Lemma 5.1 has an immediate consequence on the rational homotopy of M.

Lemma 5.2 *Let M be a closed 6–manifold with cell structure of the form (13). Then if $k \neq 0$ there is a rational homotopy equivalence $M \simeq_{\mathbb{Q}} \mathbb{C}P^3$, and if $k = 0$ then $M \simeq_{\mathbb{Q}} S^2 \times S^4$.*

Proof Let $x^2 = ky$ for some $k \in \mathbb{Q}$, where $x, y \in H^*(M;\mathbb{Q})$ are two generators with $\deg(x) = 2$. By Poincaré duality, it is easy to see that the cohomology algebra $H^*(M;\mathbb{Q})$ is determined by k, and is isomorphic to $H^*(\mathbb{C}P^3;\mathbb{Q})$ if $k \neq 0$ and $H^*(S^2 \times S^4;\mathbb{Q})$ if $k = 0$. Since every simply connected 6–manifold is formal [27, Proposition 4.6], the rational homotopy type of M is determined by its rational cohomology algebra $H^*(M;\mathbb{Q})$. Hence $M \simeq_{\mathbb{Q}} \mathbb{C}P^3$ or $M \simeq_Q S^2 \times S^4$. $\square$

5.1 The subcase when k is odd

When k is odd, the loop decomposition of the Poincaré complex $P^4(k) \cup e^7$ was determined by Huang and Theriault [20]. For any prime p, let $S^m\{p^r\}$ be the homotopy fibre of the degree p^r map on S^m. Let $k = p_1^{r_1} \cdots p_\ell^{r_\ell}$ be the prime decomposition of k. By [20, Theorem 1.1], when k is odd there is a homotopy equivalence

$$\Omega(P^4(k) \cup e^7) \simeq \prod_{j=1}^{\ell} S^3\{p_j^{r_j}\} \times \Omega S^7. \tag{16}$$

Proposition 5.3 *Let M be a closed 6–manifold with cell structure of the form $S^2 \cup_{k\eta_2} e^4 \cup e^6$. If k is odd, then M has the same homotopy type as an S^2–bundle over S^4, and there is a homotopy equivalence*

$$\Omega M \simeq S^1 \times \prod_{j=1}^{\ell} S^3\{p_j^{r_j}\} \times \Omega S^7. \tag{17}$$

Proof The homotopy equivalence (17) follows immediately from Lemma 5.1, (15) and (16). For the first statement, recall that there is the fibre bundle [18, Section 1.1]

$$S^2 \to \mathbb{C}P^3 \to S^4,$$

classified by a generator of $\pi_4(BSO(3)) \cong \mathbb{Z}$. Pulling back this bundle along a self-map of S^4 of degree k, we obtain the 6–manifold M' in the following diagram of S^2–bundles:

$$\begin{tikzcd}
S^2 \arrow[r] \arrow[d, equal] & M' \arrow[r] \arrow[d] & S^4 \arrow[d, "k"] \\
S^2 \arrow[r] & \mathbb{C}P^3 \arrow[r] & S^4
\end{tikzcd}$$

It is easy to see that $x'^2 = ky'$, where $x' \in H^2(M';\mathbb{Z})$ and $y' \in H^4(M';\mathbb{Z})$ are two generators. By [36, Corollary 4.6], when k is odd the homotopy type of M is uniquely determined by k, and hence $M \simeq M'$. $\square$

5.2 The subcase when k is even

In [20, Section 6], Huang and Theriault showed that for $P^4(2^r) \cup e^7$ with $r \geq 3$, there is an homotopy equivalence

$$\Omega(P^4(2^r) \cup e^7) \simeq S^3\{2^r\} \times \Omega S^7, \tag{18}$$

provided there is a map $P^4(2^r) \cup e^7 \to S^4$ inducing a surjection in mod-2 homology.

Proposition 5.4 *Let M be a closed 6–manifold with cell structure of the form $S^2 \cup_{2^r \eta_2} e^4 \cup e^6$. If $r \geq 3$, then there is a homotopy equivalence*

$$\Omega M \simeq S^1 \times S^3\{2^r\} \times \Omega S^7.$$

Proof Recall by Lemma 5.1 and its proof that $X \simeq P^4(2^r) \cup e^7$ and X is homotopy equivalent to the total space of an S^3–bundle over S^4

$$S^3 \to X \xrightarrow{q} S^4.$$

It is clear that $q_*\colon H_4(X;\mathbb{Z}/2\mathbb{Z}) \to H_4(S^4;\mathbb{Z}/2\mathbb{Z})$ is surjective. Hence, by (18), $\Omega X \simeq S^3\{2^r\} \times \Omega S^7$. The lemma then follows from (15) immediately. $\square$

6 Coformality of 6–manifolds

In this section, we study the rational homotopy theory of 6–manifolds as an application of our decompositions in Theorem 1.1. We briefly recall some necessary terminology used in this section, and for a detailed account of rational homotopy theory one can refer to the standard literature [17].

Recall, a *CW* complex X is rationally *formal* if its rational homotopy type is determined by the graded commutative algebra $H^*(X;\mathbb{Q})$, and is rationally *coformal* if its rational homotopy type is determined by the graded Lie algebra $\pi_*(\Omega X) \otimes \mathbb{Q}$, which is called the *homotopy Lie algebra* of X and is denoted by L_X. Suppose $(\Lambda V_X, d)$ is a *Sullivan model* of X. The differential d equals $\sum_{i\geq 0} d_i$ with $d_i\colon V_X \to \Lambda^{i+1} V_X$, and $(\Lambda V_X, d)$ is *minimal* if the linear part d_0 equals 0. In the latter case, V_X is dual to $\pi_*(\Omega X) \otimes \mathbb{Q}$.

Moreover, X is coformal if and only if it has a *purely quadratic* Sullivan model $C^*(L_X, 0) = (\Lambda(sL_X)^\#, d_1)$, where $C^*(-)$ is the *commutative cochain algebra functor*, s is the suspension, and # is the dual operation.

Proposition 6.1 *Let M be a 6–manifold as in Theorem 1.1 such that $d \geq 2$. Then M is coformal.*

Proof Consider the S^2–bundle

$$S^2 \xrightarrow{i} M \xrightarrow{p} N \tag{19}$$

in (9). By [27, Proposition 4.4] N is coformal since $d \geq 2$, and hence has a minimal Sullivan model of the form $C^*(L_N, 0) = (\Lambda(sL_N)^\#, d_1)$ as the associated commutative cochain algebra of $(L_N, 0)$ [17, Example 7 in Chapter 24(f)]. Let

$$\hat{p}\colon C^*(L_N, 0) \to (C^*(L_N) \otimes \Lambda(a, b), d)$$

be a relative minimal Sullivan model of p whose quotient $(\Lambda(a, b), \bar{d})$ is a minimal Sullivan model of S^2 with $db = a^2$ and $\deg(a) = 2$. It follows that there is the short exact sequence of the linear part of the model of (19),

$$0 \to ((sL_N)^\#, 0) \to ((sL_N)^\# \oplus \mathbb{Q}(a, b), d_0) \to (\mathbb{Q}(a, b), 0) \to 0, \tag{20}$$

such that $H^*((sL_N)^\# \oplus \mathbb{Q}(a, b), d_0)$ is dual to $\pi_*(M) \otimes \mathbb{Q}$. However, since the homotopy groups of (19) split by Theorem 1.1 and its proof, we see from (20) that the linear part d_0 equals 0 for $(sL_N)^\# \oplus \mathbb{Q}(a, b)$, and hence $(C^*(L_N) \otimes \Lambda(a, b), d)$ is a minimal model of M.

To show M is coformal, it suffices to show that the differential d is quadratic on $\mathbb{Q}(a, b)$ in $(C^*(L_N) \otimes \Lambda(a, b), d)$. Since N is simply connected, $(sL_N)^\#$ concentrates in degrees larger than or equal to 2. So, by the minimality of $(C^*(L_N) \otimes \Lambda(a, b), d)$ and degree reasons,

$$da = 0 \quad \text{and} \quad db = a^2 + ay + \sum_i z_i w_i$$

for some degree 2 elements $y, z_i, w_i \in (sL_N)^\#$. So $d = d_1$ in $(C^*(L_N) \otimes \Lambda(a, b), d)$. This shows that M is coformal. □

Proof of Theorem 1.3 First, it is well known that $\mathbb{C}P^i$ is not coformal for $i \geq 2$ by [27, Example 4.7]. If $d = 1$, then M is determined by a fibre bundle $S^2 \to M \to \mathbb{C}P^2$. It has a model of the form

$$(\Lambda(c, x), dx = c^3) \to (\Lambda(c, x, a, b), \tilde{d}) \to (\Lambda(a, b), db = a^2),$$

where $\deg(c) = \deg(a) = 2$. By degree reasons $\tilde{d}(a) = 0$, and $\tilde{d}(b) = a^2 + kc^2$ for some $k \in \mathbb{Q}$, which implies that $(\Lambda(c, x, a, b), \tilde{d})$ is minimal. However, $\tilde{d}$ is not quadratic as $\tilde{d}(x) = c^3$. Hence M is not coformal.

When $d \geq 2$, by Proposition 6.1 M is coformal. Moreover, Neisendorfer and Miller [27, Proposition 4.6] showed that every simply connected 6–manifold is formal. Hence, by [9, Theorem 1.2], M is Koszul. By [9, Theorem 1.3], there is an isomorphism of graded Lie algebras

$$\pi_*(\Omega M) \otimes \mathbb{Q} \cong H^*(M;\mathbb{Q})^{!\mathscr{L}ie},$$

where $(-)^{!\mathscr{L}ie}$ is the Koszul dual Lie functor. □

References

[1] **J Amorós**, **I Biswas**, *Compact Kähler manifolds with elliptic homotopy type*, Adv. Math. 224 (2010) 1167–1182 MR Zbl

[2] **D Barden**, *Simply connected five-manifolds*, Ann. of Math. (2) 82 (1965) 365–385 MR Zbl

[3] **S Basu**, *The homotopy type of the loops on $(n-1)$–connected $(2n+1)$–manifolds*, from "Algebraic topology and related topics" (M Singh, Y Song, J Wu, editors), Springer (2019) 1–25 MR Zbl

[4] **S Basu**, **S Basu**, *Homotopy groups of highly connected manifolds*, Adv. Math. 337 (2018) 363–416 MR Zbl

[5] **H-J Baues**, *The degree of maps between certain 6–manifolds*, Compositio Math. 110 (1998) 51–64 MR Zbl

[6] **P Beben**, **S Theriault**, *The loop space homotopy type of simply-connected four-manifolds and their generalizations*, Adv. Math. 262 (2014) 213–238 MR Zbl

[7] **P Beben**, **S Theriault**, *Homotopy groups of highly connected Poincaré duality complexes*, Doc. Math. 27 (2022) 183–211 MR Zbl

[8] **P Beben**, **J Wu**, *The homotopy type of a Poincaré duality complex after looping*, Proc. Edinb. Math. Soc. 58 (2015) 581–616 MR Zbl

[9] **A Berglund**, *Koszul spaces*, Trans. Amer. Math. Soc. 366 (2014) 4551–4569 MR Zbl

[10] **G Boyde**, *p–hyperbolicity of homotopy groups via K–theory*, Math. Z. 301 (2022) 977–1009 MR Zbl

[11] **F R Cohen**, *Applications of loop spaces to some problems in topology*, from "Advances in homotopy theory" (S M Salamon, B Steer, W A Sutherland, editors), London Math. Soc. Lecture Note Ser. 139, Cambridge Univ. Press (1989) 11–20 MR Zbl

[12] **F R Cohen**, **J C Moore**, **J A Neisendorfer**, *The double suspension and exponents of the homotopy groups of spheres*, Ann. of Math. (2) 110 (1979) 549–565 MR Zbl

[13] **F R Cohen**, **J C Moore**, **J A Neisendorfer**, *Torsion in homotopy groups*, Ann. of Math. (2) 109 (1979) 121–168 MR Zbl

[14] **D Crowley**, **C M Escher**, *A classification of S^3–bundles over S^4*, Differential Geom. Appl. 18 (2003) 363–380 MR Zbl

[15] **D Crowley**, **J Nordström**, *The classification of 2–connected 7–manifolds*, Proc. Lond. Math. Soc. 119 (2019) 1–54 MR Zbl

[16] **H Duan**, **C Liang**, *Circle bundles over 4–manifolds*, Arch. Math. (Basel) 85 (2005) 278–282 MR Zbl

[17] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Math. 205, Springer (2001) MR Zbl

[18] **F Hirzebruch**, **T Berger**, **R Jung**, *Manifolds and modular forms*, Aspects of Mathematics 20, Vieweg, Brunswick (1992) MR Zbl

[19] **R Huang**, *Suspension homotopy of 6–manifolds*, Algebr. Geom. Topol. 23 (2023) 439–460 MR Zbl

[20] **R Huang**, **S Theriault**, *Loop space decompositions of $(2n-2)$–connected $(4n-1)$–dimensional Poincaré duality complexes*, Res. Math. Sci. 9 (2022) art. id. 53 MR Zbl

[21] **I M James**, *On the suspension sequence*, Ann. of Math. (2) 65 (1957) 74–107 MR Zbl

[22] **Y Jiang**, *Regular circle actions on 2–connected 7–manifolds*, J. Lond. Math. Soc. 90 (2014) 373–387 MR Zbl

[23] **P E Jupp**, *Classification of certain 6–manifolds*, Proc. Cambridge Philos. Soc. 73 (1973) 293–300 MR Zbl

[24] **M Kreck**, *On the classification of 1–connected 7–manifolds with torsion free second homology*, J. Topol. 11 (2018) 720–738 MR Zbl

[25] **M Kreck**, **Y Su**, *On 5–manifolds with free fundamental group and simple boundary links in S^5*, Geom. Topol. 21 (2017) 2989–3008 MR Zbl

[26] **J Neisendorfer**, *Algebraic methods in unstable homotopy theory*, New Mathematical Monographs 12, Cambridge Univ. Press (2010) MR Zbl

[27] **J Neisendorfer**, **T Miller**, *Formal and coformal spaces*, Illinois J. Math. 22 (1978) 565–580 MR Zbl

[28] **J A Neisendorfer**, **P S Selick**, *Some examples of spaces with or without exponents*, from "Current trends in algebraic topology, I" (R M Kane, S O Kochman, P S Selick, V P Snaith, editors), CMS Conf. Proc. 2, Amer. Math. Soc., Providence, RI (1982) 343–357 MR Zbl

[29] **S Oka**, review of [36], Mathematical Reviews (1982) Available at http://msp.org/idx/mr/0672527

[30] **S Sasao**, *On homotopy type of certain complexes*, Topology 3 (1965) 97–102 MR Zbl

[31] **T So**, **S Theriault**, *The suspension of a* 4*–manifold and its applications*, preprint (2019) arXiv 1909.11129

[32] **H Toda**, *Composition methods in homotopy groups of spheres*, Annals of Mathematics Studies 49, Princeton Univ. Press (1962) MR Zbl

[33] **C T C Wall**, *Classification of* $(n-1)$*–connected* $2n$*–manifolds*, Ann. of Math. (2) 75 (1962) 163–189 MR Zbl

[34] **C T C Wall**, *Classification problems in differential topology, V: On certain* 6*–manifolds*, Invent. Math. 1 (1966) 355–374 MR Zbl

[35] **C T C Wall**, *Classification problems in differential topology, VI: Classification of* $(s-1)$*–connected* $(2s+1)$*–manifolds*, Topology 6 (1967) 273–296 MR Zbl

[36] **K Yamaguchi**, *On the homotopy type of CW complexes with the form* $S^2 \cup e^4 \cup e^6$, Kodai Math. J. 5 (1982) 303–312 MR Zbl

[37] **A V Zhubr**, *Classification of simply connected six-dimensional spinor manifolds*, Izv. Akad. Nauk SSSR Ser. Mat. 39 (1975) 839–859 MR Zbl In Russian; translated in Math. USSR-Izv. 9 (1975) 793–812

[38] **A V Zhubr**, *Closed simply connected six-dimensional manifolds: proofs of classification theorems*, Algebra i Analiz 12 (2000) 126–230 MR Zbl In Russian; translated in St. Petersburg Math. J. 12 (2001) 605–680

Institute of Mathematics, Chinese Academy of Sciences
Beijing, China

huangrz@amss.ac.cn

https://sites.google.com/site/hrzsea/

Received: 9 November 2021 Revised: 2 December 2021

Algebraic & Geometric Topology **23**:5 (2023) 2389–2414
DOI: 10.2140/agt.2023.23.2389
Published: 25 July 2023

Infinite families of higher torsion in the homotopy groups of Moore spaces

STEVEN AMELOTTE
FREDERICK R COHEN
YUXIN LUO

We give a refinement of the stable Snaith splitting of the double loop space of a Moore space and use it to construct infinite periodic families of elements of order p^{r+1} in the homotopy groups of mod p^r Moore spaces. For odd primes p, our splitting implies that the homotopy groups of the mod p^{r+1} Moore spectrum are summands of the unstable homotopy groups of each mod p^r Moore space.

55P35, 55P42, 55Q51, 55Q52

1 Introduction

The purpose of this note is to combine three standard results in homotopy theory:

(1) the construction of elements of order p^{r+1} in the homotopy groups of the mod p^r Moore space $P^n(p^r)$, as described in [8];

(2) the stable splitting of $\Omega^2 P^n(p^r)$ first proved by Snaith [21]; and

(3) the introduction of v_1–periodic self-maps by Adams in his work on the image of the J–homomorphism.

In their fundamental work on the homotopy theory of Moore spaces, Cohen, Moore and Neisendorfer [7; 8; 17; 18] proved that $P^n(p^r) = S^{n-1} \cup_{p^r} e^n$ has homotopy exponent exactly p^{r+1} when p is a prime number greater than 3. We will refer to elements of this maximal possible order in $\pi_*(P^n(p^r))$ as *higher torsion* elements. The main results of this paper give additional infinite families of higher torsion elements in the homotopy groups of odd primary Moore spaces which are different from those constructed via Samelson products in [8]. The main technical ingredient is a slightly finer stable decomposition of $\Omega^2 P^{2n+1}(p^r)$ which essentially follows from a combination of results (1) and (2) above. The reason that this question arose is because of computations of Roman Mikhailov and Jie Wu, who asked about "functorial elements" of order p^{r+1} in the homotopy groups of mod p^r Moore spaces not given by those in [8].

By considering the integral homology of the double loop space of a Moore space, it is clear that certain spherical homology classes force the classical Snaith splitting of $\Omega^2 P^{2n+1}(p^r)$ to stably decompose further than previously described. This new stable splitting allows for the construction of new higher torsion elements which are detected by K–theory but not detected in the ordinary homology of any iterated loop space of a Moore space, unlike the elements of order p^{r+1} in [8] which have nontrivial Hurewicz images in the homology of $\Omega^2 P^{2n+1}(p^r)$ (see Lemma 4.1).

The main results are described next. Recall that the Snaith splitting gives a functorial stable homotopy equivalence

$$\Sigma^\infty \Omega^2 \Sigma^2 X \simeq \Sigma^\infty \bigvee_{j=1}^{\infty} D_j(\Omega^2 \Sigma^2 X)$$

for any path-connected CW–complex X, where the stable summands are given by suspension spectra of the extended powers $D_j(\Omega^2\Sigma^2 X) = \mathcal{C}_2(j)_+ \wedge_{\Sigma_j} X^{\wedge j}$, and $\mathcal{C}_2(j)$ denotes the space of j little 2–cubes disjointly embedded in $\mathbb{R}^2$. In the case that X is an odd-dimensional sphere S^{2n-1}, the stable summands $D_j(\Omega^2 S^{2n+1})$ of $\Omega^2 S^{2n+1}$ have been well studied; they are p–locally contractible unless $j \equiv 0$ or $1 \bmod p$, in which case they can be identified with suitably suspended Brown–Gitler spectra. In particular, after localizing at a prime p, they are stably indecomposable. Below we consider the case of an odd-dimensional Moore space and the stable summands $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ which map naturally onto these Brown–Gitler spectra by the map $\Omega^2\Sigma^2 q$ where $q\colon P^{2n-1}(p^r) \to S^{2n-1}$ is the pinch map.

Theorem 1.1 *Suppose p is prime and $n > 1$.*

(a) *If $p \geq 3$ and $r \geq 1$, then $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ is stably homotopy equivalent to*

$$P^{2np^k-2}(p^{r+1}) \vee X_{p^k}$$

for some finite CW–complex X_{p^k} for all $k \geq 1$.

(b) *If $p = 2$ and $r > 1$, then $D_2(\Omega^2 P^{2n+1}(2^r))$ is homotopy equivalent to*

$$P^{4n-2}(2^{r+1}) \vee X_2$$

for some 4–cell complex $X_2 = P^{4n-3}(2^r) \cup CP^{4n-2}(2)$.

(c) *If $p = 2$ and $r = 1$, then $D_2(\Omega^2 P^{2n+1}(2))$ is a stably indecomposable 6–cell complex.*

The reason the stable splitting of $\Omega^2 P^{2n+1}(p^r)$ described by Theorem 1.1 has implications for the unstable homotopy groups of $P^{2n+1}(p^r)$ is that maps

$$P^{2np^k-2}(p^{r+1}) \to \Omega^2 P^{2n+1}(p^r)$$

admitting stable retractions exist unstably when p is odd (and in only a few cases when $p=2$; see Section 6).

Theorem 1.2 *Let p be an odd prime, $r \geq 1$ and $n > 1$. Then for every $k \geq 1$ there exist homotopy commutative diagrams*

$$\begin{tikzcd} P^{2np^k-2}(p^{r+1}) \arrow[r] \arrow[dr, "E^\infty"'] & \Omega^2 P^{2n+1}(p^r) \arrow[d] \\ & QP^{2np^k-2}(p^{r+1}) \end{tikzcd} \qquad \begin{tikzcd} P^{(4n-2)p^k-2}(p^{r+1}) \arrow[r] \arrow[dr, "E^\infty"'] & \Omega^2 P^{2n}(p^r) \arrow[d] \\ & QP^{(4n-2)p^k-2}(p^{r+1}) \end{tikzcd}$$

where E^∞ is the stabilization map (*ie*, *unit of the adjunction* $\Sigma^\infty \dashv \Omega^\infty$).

The loop space decompositions of odd primary Moore spaces given in [8; 9] imply that the stable homotopy groups $\pi_*^s(P^n(p^r))$ are in a certain sense retracts of the unstable homotopy groups $\pi_*(P^n(p^r))$. Different loop space decompositions were used by Chen and Wu [4] to obtain the same result for 2–primary Moore spaces, and other examples of spaces whose stable and unstable homotopy groups share this property are given by Beben and Wu [3]. As a consequence of Theorem 1.2, the stable homotopy groups of the mod p^{r+1} Moore spectrum retract off the unstable homotopy groups of each mod p^r Moore space in a similar sense when p is an odd prime (see Corollary 5.1).

This observation clearly suggests that $\pi_*(P^n(p^r))$ contains many $\mathbb{Z}/p^{r+1}$ summands when p is odd. To generate explicit examples, in Section 6 we use desuspensions of Adams maps in conjunction with Theorem 1.2 to construct infinite v_1–periodic families of higher torsion elements and obtain the following. Let $q = 2(p-1)$.

Theorem 1.3 *Let p be an odd prime, $r \geq 1$ and $n > 1$. Then for all sufficiently large k,*

$$\pi_{2np^k-1+tqp^r}(P^{2n+1}(p^r)) \quad \text{and} \quad \pi_{(4n-2)p^k-1+tqp^r}(P^{2n}(p^r))$$

contain $\mathbb{Z}/p^{r+1}$ summands for every $t \geq 0$.

Remark 1.4 A lower bound on k in Theorem 1.3 is required to ensure the existence of unstable Adams maps $v_1 \colon P^{\ell+qp^r}(p^{r+1}) \to P^\ell(p^{r+1})$ which induce isomorphisms in K–theory. See Section 6 for a more precise statement. In particular, in the most interesting case when $r=1$, we only require $k \geq 1$.

For $p=2$, unstable maps analogous to those in Theorem 1.2 rarely exist for reasons related to the divisibility of the Whitehead square. When $r>1$, the inclusion of the stable summand $P^{4n-2}(2^{r+1})$ of $\Omega^2 P^{2n+1}(2^r)$ given by Theorem 1.1(b) exists unstably if and only if $n=2$ or 4. As examples in these cases, we describe unstable v_1–periodic families of higher torsion elements in $\pi_*(P^5(2^r))$ and $\pi_*(P^9(2^r))$ for some small values of r (Theorem 6.6).

Remark 1.5 When p is odd, the $r>1$ case of Theorem 1.1(a) follows quickly from an unstable product decomposition of $\Omega^2 P^{2n+1}(p^r)$ proved by Neisendorfer. More precisely, [19, Theorem 1] shows that $\Omega S^{2np^k-1}\{p^{r+1}\}$ is a retract of $\Omega^2 P^{2n+1}(p^r)$ for all $k\geq 1$, and it is readily checked that $P^{2np^k-2}(p^{r+1})$ is a stable retract of $\Omega S^{2np^k-1}\{p^{r+1}\}$; cf [2, Proposition 4.1]. The $r=1$ case of Theorem 1.1(a) gives some evidence for conjectures surrounding the unstable homotopy type of $\Omega^2 P^{2n+1}(p)$ considered by Cohen, Moore and Neisendorfer [7; 19], Gray [14] and Theriault [22].

In Section 2 we describe the homology of the double loop space of $P^{2n+1}(p^r)$ and establish some basic properties of its Snaith summands. In Section 3 we review Cohen, Moore and Neisendorfer's construction of higher torsion elements and, in particular, their work on the homotopy and homology Bockstein spectral sequences for the single loop space of a Moore space. In Section 4 we compute higher Bocksteins in the homology of the double loop space of a Moore space and prove the splittings of Theorem 1.1. In Section 5 we derive Theorem 1.2 from Theorem 1.1 and discuss implications for the unstable homotopy groups of odd primary Moore spaces. Finally, in Section 6 we construct the infinite families of higher torsion elements discussed above.

Acknowledgements This material is based upon work partly supported by the National Science Foundation under grant DMS-1929284 while Amelotte was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI.

2 Homology of the stable summands $D_j(\Omega^2 P^{2n+1}(p^r))$

The homology of $\Omega^n\Sigma^n X$ taken with field coefficients as a filtered algebra was worked out in [6]. A short summary of that information elucidates the homology of the stable Snaith summands, usually denoted by $D_{n,j}X$. In the applications below where $n=2$, the j^{th} stable summand $D_{2,j}X$ of $\Omega^2\Sigma^2 X$ will be denoted by $D_j(\Omega^2\Sigma^2 X)$. All homology groups have $\mathbb{Z}/p$ coefficients unless indicated otherwise.

Start with the connected graded vector space

$$V = \overline{H}_*(X)$$

given by the reduced mod p homology of a path-connected space X. Next, consider the reduced homology of the suspension ΣX, denoted by σV. Form the free graded Lie algebra generated by σV, denoted by

$$L[\sigma V].$$

In addition, consider the free graded restricted Lie algebra

$$L^p[\sigma V].$$

This restricted Lie algebra is isomorphic to the module of primitive elements in the tensor algebra generated by σV, so the tensor algebra is a primitively generated Hopf algebra.

A basis for $L^p[\sigma V]$ is given by the union of

(1) a basis $\mathcal{B} = \{b_\alpha \mid \alpha \in I\}$ for $L[\sigma(V)]_{\text{odd}}$, the elements of odd degree in $L[\sigma(V)]$;
(2) a basis $\mathcal{C} = \{c_\gamma \mid \gamma \in J\}$ for $L[\sigma(V)]_{\text{even}}$, the elements of even degree in $L[\sigma(V)]$; and
(3) a basis for the $(p^k)^{\text{th}}$ powers of $L[\sigma(V)]_{\text{even}}$, say $\mathcal{PC} = \{{c_\gamma}^{p^k} \mid \gamma \in J,\ k \geq 1\}$.

It follows from the Bott–Samelson theorem that a basis for the module of primitives in the mod p homology of $\Omega\Sigma^2 X$ is given by

$$\mathcal{B} \cup \mathcal{C} \cup \mathcal{PC}.$$

The mod p homology of $\Omega^2\Sigma^2 X$ can now be described using the preparations of the previous paragraph. First, for each $(p^k)^{\text{th}}$ power $x = {c_\gamma}^{p^k} \in \mathcal{PC}$, let $\sigma^{-1}x$ denote the formal desuspension, lowering the degree of x by one. Let $\beta\sigma^{-1}x$ denote the formal first Bockstein of $\sigma^{-1}{c_\gamma}^{p^k}$, with degree $|x|-2$. Let Ψ denote the set of elements given by

$$\Psi = \{\sigma^{-1}x, \beta\sigma^{-1}x \mid x \in \mathcal{PC}\}.$$

Theorem 2.1 [6] *If X is a path-connected CW–complex and p is an odd prime, then the mod p homology $H_*(\Omega^2\Sigma^2 X)$ is isomorphic as an algebra to the free graded commutative algebra generated by*

$$\sigma^{-1}\mathcal{B} \cup \sigma^{-1}\mathcal{C} \cup \Psi.$$

Remark 2.2 The result in case $p = 2$ is different and mildly simpler (see [6] and the remarks preceding Lemma 2.5 below).

The homology of iterated loop-suspensions are free E_n–algebras naturally equipped with more algebraic structure than we will need here, but we briefly mention two homology operations which will often be used below to label elements of $H_*(\Omega^2\Sigma^2 X)$ when $X = P^{2n-1}(p^r)$; namely, the Dyer–Lashof operation

$$Q_1\colon H_n(\Omega^2\Sigma^2 X) \to H_{np+p-1}(\Omega^2\Sigma^2 X)$$

and the Browder bracket

$$\lambda\colon H_n(\Omega^2\Sigma^2 X) \otimes H_m(\Omega^2\Sigma^2 X) \to H_{n+m+1}(\Omega^2\Sigma^2 X).$$

The images of Q_1 and λ contain the transgressions of $(p^k)^{\text{th}}$ powers and iterated commutators, respectively, of primitive elements in the tensor Hopf algebra $H_*(\Omega\Sigma^2 X)$. See [6] for a description of the generators in $\sigma^{-1}\mathcal{B} \cup \sigma^{-1}\mathcal{C} \cup \Psi$ in terms of these operations.

We recall that the set of indecomposables $\sigma^{-1}\mathcal{B} \cup \sigma^{-1}\mathcal{C} \cup \Psi$ is also graded by weights as in [6] in such a way that the reduced mod p homology of $D_j(\Omega^2\Sigma^2 X)$ is spanned by the monomials of weight j in the free graded commutative algebra generated by $\sigma^{-1}\mathcal{B} \cup \sigma^{-1}\mathcal{C} \cup \Psi$.

Explicitly, for $X = P^{2n-1}(p^r)$, let $\{u, v\}$ be a basis for the graded vector space $V = \overline{H}_*(P^{2n-1}(p^r))$ with $|u| = 2n-2$, $|v| = 2n-1$ and $\beta^{(r)} v = u$. Then, if p is odd, Theorem 2.1 implies that $H_*(\Omega^2 P^{2n+1}(p^r))$ is a free graded commutative algebra on generators

$$u,\ v,\ \lambda(u,u),\ \lambda(u,v),\ \lambda(u,\lambda(u,v)),\ \lambda(v,\lambda(u,v)), \ldots \in \sigma^{-1}\mathcal{B} \cup \sigma^{-1}\mathcal{C},$$

$$\left.\begin{array}{l} Q_1 v,\ \beta^{(1)} Q_1 v,\ Q_1^2 v,\ \beta^{(1)} Q_1^2 v, \ldots \\ Q_1\lambda(u,u),\ \beta^{(1)} Q_1\lambda(u,u),\ Q_1^2\lambda(u,u),\ \beta^{(1)} Q_1^2\lambda(u,u), \ldots \\ Q_1\lambda(u,\lambda(u,v)),\ \beta^{(1)} Q_1\lambda(u,\lambda(u,v)), \ldots \\ \vdots \end{array}\right\} \in \Psi,$$

with weights defined by

$$\begin{aligned} \mathrm{wt}(u) &= \mathrm{wt}(v) = 1, \\ \mathrm{wt}(Q_1^k x) &= \mathrm{wt}(\beta^{(1)} Q_1^k x) = p^k \mathrm{wt}(x), \\ \mathrm{wt}(\lambda(x,y)) &= \mathrm{wt}(x) + \mathrm{wt}(y), \end{aligned}$$

and extended to all monomials by $\mathrm{wt}(xy) = \mathrm{wt}(x) + \mathrm{wt}(y)$.

We will often use the same notation, u and v, to denote generators of the bottom two mod p homology groups of $P^{2n+1}(p^r)$, $\Omega P^{2n+1}(p^r)$ and $\Omega^2 P^{2n+1}(p^r)$, indicating degrees with subscripts when necessary.

To prove the splittings in Theorem 1.1, we will need to know the top two homology groups of the Snaith summands $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ explicitly. Although the list of generators above is specific to the case of p odd, the following lemma holds for all primes p.

Lemma 2.3 *For each $k \geq 0$, $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ is a $(2np^k-2p^k-1)$–connected $(2np^k-1)$–dimensional space with*

(a) $H_{2np^k-1}(D_{p^k}(\Omega^2 P^{2n+1}(p^r))) = \mathrm{span}\{Q_1^k v\}$,

(b) $H_{2np^k-2}(D_{p^k}(\Omega^2 P^{2n+1}(p^r))) = \mathrm{span}\{\beta^{(1)} Q_1^k v, \mathrm{ad}_\lambda^{p^k-1}(v)(u)\}$,

where $\mathrm{ad}_\lambda^{p^k-1}(v)(u)$ denotes the (p^k-1)–fold iterated Browder bracket

$$\lambda(v, \lambda(v, \dots, \lambda(v, u), \dots)).$$

Proof Since the reduced mod p homology of $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ consists of the elements of homogeneous weight p^k in $H_*(\Omega^2 P^{2n+1}(p^r))$, the connectivity and dimension of $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ follow from the fact that the weight p^k monomials of lowest and highest homological degree are u^{p^k} and $Q_1^k v$, respectively, with $|u^{p^k}| = 2np^k - 2p^k$ and $|Q_1^k v| = 2np^k - 1$.

Observe that any nonzero iterated Browder bracket with arguments in $\{u, v\}$ must involve $u \in H_{2n-2}(\Omega^2 P^{2n+1}(p^r))$ since $\lambda(v, v) = 0$ (being the transgression of the graded commutator of an even degree element with itself in the tensor algebra $H_*(\Omega P^{2n+1}(p^r))$). Parts (a) and (b) now follow easily by inspection of monomials of weight p^k in homological degrees $2np^k - 1$ and $2np^k - 2$. □

Remark 2.4 For $k = 0$, the span in Lemma 2.3(b) is 1–dimensional since $\beta^{(1)} v$ and $\mathrm{ad}_\lambda^0(v)(u)$ coincide if $r = 1$, and $\beta^{(1)} v = 0$ if $r > 1$. Of course, in this case $D_{p^k}(\Omega^2 P^{2n+1}(p^r)) = D_1(\Omega^2 P^{2n+1}(p^r))$ is simply $P^{2n-1}(p^r)$. For all $k \geq 1$, $\dim H_{2np^k-2}(D_{p^k}(\Omega^2 P^{2n+1}(p^r))) = 2$.

In the $p = 2$ case we will need to know the homology of $D_2(\Omega^2 P^{2n+1}(2^r))$ as a module over the Steenrod algebra. The mod 2 homology generators of weight 2 differ somewhat from those appearing in the list above for odd primes. First, since $H_*(\Omega^2 P^{2n+1}(2^r))$ is a polynomial algebra, we have the quadratic generator v^2 in addition to u^2 and uv. Second, the Browder bracket $\lambda(u,u)$ is trivial since this class represents the transgression of the commutator $[u,u]=u^2+u^2=0$ in the tensor algebra $H_*(\Omega P^{2n+1}(2^r))=T(u,v)$ over $\mathbb{Z}/2$. On the other hand, since $u^2\in H_{4n-2}(\Omega P^{2n+1}(2^r))$ is primitive, this class transgresses to a generator Q_1u in $H_{4n-3}(\Omega^2 P^{2n+1}(2^r))$, unlike in the odd primary case.

It follows that $D_2(\Omega^2 P^{2n+1}(2^r))$ is a 6–cell complex with

$$\begin{aligned}
H_{4n-1}(D_2(\Omega^2 P^{2n+1}(2^r))) &= \mathrm{span}\{Q_1v\},\\
H_{4n-2}(D_2(\Omega^2 P^{2n+1}(2^r))) &= \mathrm{span}\{v^2,\ \lambda(u,v)\},\\
H_{4n-3}(D_2(\Omega^2 P^{2n+1}(2^r))) &= \mathrm{span}\{uv,\ Q_1u\},\\
H_{4n-4}(D_2(\Omega^2 P^{2n+1}(2^r))) &= \mathrm{span}\{u^2\}.
\end{aligned}$$

Lemma 2.5 *The action of the Steenrod algebra on $H_*(D_2(\Omega^2 P^{2n+1}(2^r)))$ is determined by*

(a) $\mathrm{Sq}^1_* Q_1 v = \begin{cases} v^2+\lambda(u,v) & \text{if } r=1,\\ v^2 & \text{if } r>1,\end{cases}$

(b) $\mathrm{Sq}^2_* Q_1 v = \begin{cases} Q_1u & \text{if } r=1,\\ 0 & \text{if } r>1,\end{cases}$

(c) $\mathrm{Sq}^2_* v^2 = \mathrm{Sq}^1_* uv = \begin{cases} u^2 & \text{if } r=1,\\ 0 & \text{if } r>1.\end{cases}$

Proof By [6, III.3.10], $\beta^{(1)}Q_1x = x^2+\lambda(x,\beta^{(1)}x)$ for $x\in H_*(\Omega^2\Sigma^2X;\mathbb{Z}/2)$ with $|x|$ odd. Part (a) follows since $\beta^{(1)}v=u$ if $r=1$, and $\beta^{(1)}v=0$ if $r>1$.

Part (b) follows from the Nishida relation $\mathrm{Sq}^2_*Q_1 = Q_1\mathrm{Sq}^1_*$.

For part (c), the Cartan formula and the fact that $\mathrm{Sq}^1_* = \beta^{(1)}$ is a derivation on the Pontryagin ring $H_*(\Omega^2 P^{2n+1}(2^r))$ imply that $\mathrm{Sq}^2_* v^2$ and $\mathrm{Sq}^1_* uv$ are as claimed.

Since the Browder bracket satisfies the Cartan formula [6, III.1.2(7)]

$$\mathrm{Sq}^n_*\lambda(x,y) = \sum_{i+j=n} \lambda(\mathrm{Sq}^i_* x, \mathrm{Sq}^j_* y),$$

we have $\mathrm{Sq}^2_*\lambda(u,v) = \mathrm{Sq}^1_*\lambda(u,v) = 0$. The relations above therefore determine all nontrivial Steenrod operations in $H_*(D_2(\Omega^2 P^{2n+1}(2^r)))$. □

3 Review of the work of Cohen, Moore and Neisendorfer

To prepare for the proofs of Theorems 1.1 and 1.2 in the next sections, we briefly review some of the work of Cohen, Moore and Neisendorfer [8; 17] on torsion in the homotopy groups of Moore spaces.

Recall that the mod p^r homotopy groups of a space X are defined by

$$\pi_n(X;\mathbb{Z}/p^r)=[P^n(p^r),X].$$

Provided $p^r>2$, there are splittings [20, Proposition 6.2.2]

$$P^n(p^r)\wedge P^m(p^r)\simeq P^{n+m}(p^r)\vee P^{n+m-1}(p^r)$$

for $n,m\geq 2$ which allow for the definition of mod p^r Samelson products

$$\pi_n(\Omega X;\mathbb{Z}/p^r)\otimes\pi_m(\Omega X;\mathbb{Z}/p^r)\to\pi_{n+m}(\Omega X;\mathbb{Z}/p^r).$$

Together with the Bockstein differential, this gives $\pi_*(\Omega X;\mathbb{Z}/p^r)$ the structure of a differential graded Lie algebra when $p\geq 5$ and $r\geq 1$, and a differential graded quasi-Lie algebra when $p=3$ and $r\geq 2$; see [17; 20]. The mod p Hurewicz map

$$h\colon \pi_*(\Omega X;\mathbb{Z}/p)\to H_*(\Omega X;\mathbb{Z}/p)$$

intertwines mod p Samelson products with commutators in the Pontryagin ring and commutes with Bockstein differentials, thereby inducing a morphism of spectral sequences from the mod p homotopy Bockstein spectral sequence $(E^s_\pi(\Omega X),\beta^{(s)})$ to the mod p homology Bockstein spectral sequence $(E^s_H(\Omega X),\beta^{(s)})$.

Consider $\pi_*(\Omega P^{2n+1}(p^r);\mathbb{Z}/p)$. In degrees $2n$ and $2n-1$, denote the mod p reduction of the adjoint of the identity map and its r^{th} Bockstein by

$$\nu\colon P^{2n}(p)\to\Omega P^{2n+1}(p^r)\quad\text{and}\quad\beta^{(r)}\nu=\mu\colon P^{2n-1}(p)\to\Omega P^{2n+1}(p^r),$$

respectively. Then the Hurewicz images $h(\nu)=v$ and $h(\mu)=u$ generate

$$H_{2n}(\Omega P^{2n+1}(p^r);\mathbb{Z}/p)\quad\text{and}\quad H_{2n-1}(\Omega P^{2n+1}(p^r);\mathbb{Z}/p),$$

respectively, and by the Bott–Samelson theorem,

$$H_*(\Omega P^{2n+1}(p^r);\mathbb{Z}/p)\cong T(u,v)\cong UL(u,v), \tag{1}$$

where $L(u,v)$ is the free differential graded Lie algebra on two generators u and v with differential $\beta^{(r)}v=u$.

In any graded (quasi-)Lie algebra L (more generally, any graded module with an antisymmetric bracket operation), let $\mathrm{ad}(x)(y)=[x,y]$ for $x,y\in L$. Define $\mathrm{ad}^0(x)(y)=y$ and inductively define $\mathrm{ad}^k(x)(y)=\mathrm{ad}(x)(\mathrm{ad}^{k-1}(x)(y))$ for $k\geq 1$. To detect higher torsion in $\pi_*(P^{2n+1}(p^r))$, Cohen, Moore and Neisendorfer [8] consider the mod p Samelson products

$$\tau_k(\nu)=\mathrm{ad}^{p^k-1}(\nu)(\mu),$$
$$\sigma_k(\nu)=\frac{1}{2p}\sum_{j=1}^{p^k-1}\binom{p^k}{j}[\mathrm{ad}^{j-1}(\nu)(\mu),\mathrm{ad}^{p^k-j-1}(\nu)(\mu)]$$

in $\pi_*(\Omega P^{2n+1}(p^r);\mathbb{Z}/p)$ and their mod p Hurewicz images $\tau_k(v)$ and $\sigma_k(v)$ defined similarly in terms of graded commutators. Since the tensor algebra (1) is acyclic with respect to the differential $\beta^{(r)}$, the homology Bockstein spectral sequence collapses at the $(r+1)^{\text{st}}$ page and no higher differentials in the homotopy Bockstein spectral sequence can be detected by the Hurewicz map:

$$E_H^1(\Omega P^{2n+1}(p^r))=\cdots=E_H^r(\Omega P^{2n+1}(p^r))=T(u,v),$$
$$\bar{E}_H^{r+1}(\Omega P^{2n+1}(p^r))=0.$$

In particular, $\tau_k(v)\in H_{2np^k-1}(\Omega P^{2n+1}(p^r))$ is killed by the differential

$$\beta^{(r)}v^{p^k}=\tau_k(v)$$

for all $k\geq 0$. To tease out higher torsion, Cohen, Moore and Neisendorfer instead compute the homology Bockstein spectral sequence of the loops on the fibre $F^{2n+1}(p^r)$ of the pinch map $q\colon P^{2n+1}(p^r)\to S^{2n+1}$, where lifts

$$\tau_k'(\nu),\sigma_k'(\nu)\in\pi_*(\Omega F^{2n+1}(p^r);\mathbb{Z}/p)$$

of $\tau_k(\nu)$ and $\sigma_k(\nu)$, and their Hurewicz images $\tau_k'(v)$ and $\sigma_k'(v)$, are shown to survive to the $(r+1)^{\text{st}}$ page, at least when p is odd.

Theorem 3.1 [8, Theorem 10.3] *Let p be an odd prime and $r\geq 1$. Then there is an isomorphism of differential graded Hopf algebras*

$$E_H^{r+1}(\Omega F^{2n+1}(p^r))\cong\Lambda(\tau_0'(v),\tau_1'(v),\tau_2'(v),\dots)\otimes\mathbb{Z}/p[\sigma_1'(v),\sigma_2'(v),\dots],$$

where $|\tau_k'(v)|=2np^k-1$, $|\sigma_k'(v)|=2np^k-2$ and

$$\beta^{(r+1)}\tau_k'(v)=\ell\sigma_k'(v),\quad \ell\neq 0,$$

for $k\geq 1$.

Since the homology classes $\tau'_k(v) = h(\tau'_k(\nu))$ support nontrivial Bocksteins $\beta^{(r+1)}$ for $k \geq 1$, the same is true of the mod p homotopy classes

$$\tau'_k(\nu) \in \pi_{2np^k-1}(\Omega F^{2n+1}(p^r); \mathbb{Z}/p).$$

It follows that there exist maps

$$\delta'_k \colon P^{2np^k-1}(p^{r+1}) \to \Omega F^{2n+1}(p^r) \tag{2}$$

for each $k \geq 1$ which satisfy $(\delta'_k)_*(v_{2np^k-1}) = \tau'_k(v)$ in mod p homology and induce split monomorphisms in integral homology.

To exhibit nontrivial classes in the image of $\beta^{(r+1)}$ in $E_\pi^{r+1}(\Omega P^{2n+1}(p^r))$, Cohen, Moore and Neisendorfer show that the composition of $\beta^{(r+1)}\tau'_k(\nu)$ with

$$\Omega F^{2n+1}(p^r) \to \Omega P^{2n+1}(p^r)$$

does not represent zero in $E_\pi^{r+1}(\Omega P^{2n+1}(p^r))$, thereby proving the following.

Theorem 3.2 [8; 17] *Let p be an odd prime and $r \geq 1$. Then $\pi_{2np^k-1}(P^{2n+1}(p^r))$ contains a $\mathbb{Z}/p^{r+1}$ summand for every $k \geq 1$.*

Remark 3.3 It follows from Theorem 3.2 and the loop space decomposition for even-dimensional odd primary Moore spaces [8, Theorem 1.1],

$$\Omega P^{2n+2}(p^r) \simeq S^{2n+1}\{p^r\} \times \Omega\Big(\bigvee_{j=0}^{\infty} P^{4n+2nj+3}(p^r)\Big), \tag{3}$$

that $\pi_*(P^n(p^r))$ contains $\mathbb{Z}/p^{r+1}$ summands for all $n \geq 3$ when p is odd.

4 Splittings of $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$

In this section we prove Theorem 1.1 in a series of lemmas and discuss the stable homotopy type of $\Omega^2 P^{2n+1}(p^r)$. We assume throughout that $n > 1$.

The higher torsion discussed in the previous section is not reflected in the homology of the single loop space of a Moore space since $H_*(\Omega P^n(p^r))$ is acyclic with respect to $\beta^{(r)}$. The next lemma shows that it becomes visible in homology after looping twice.

Lemma 4.1 *Let p be an odd prime and $r \geq 1$. Then in the mod p homology Bockstein spectral sequence of $\Omega^2 P^{2n+1}(p^r)$,*

$$\beta^{(r+1)}\mathrm{ad}_\lambda^{p^k-1}(v)(u) \neq 0$$

in $E_H^{r+1}(\Omega^2 P^{2n+1}(p^r))$ *for each* $k \geq 1$*. Moreover, there exist maps*

$$\delta_k \colon P^{2np^k-2}(p^{r+1}) \to \Omega^2 P^{2n+1}(p^r)$$

for each $k \geq 1$ *which satisfy* $(\delta_k)_*(v_{2np^k-2}) = \mathrm{ad}_\lambda^{p^k-1}(v)(u)$ *in mod* p *homology and induce split monomorphisms in integral homology.*

Proof Consider the composite

$$P^{2np^k-1}(p^{r+1}) \xrightarrow{\delta'_k} \Omega F^{2n+1}(p^r) \to \Omega P^{2n+1}(p^r),$$

where δ'_k is the map from (2) and the second map is the fibre inclusion of the looped pinch map $\Omega q \colon \Omega P^{2n+1}(p^r) \to \Omega S^{2n+1}$. In mod p homology, $(\delta'_k)_*(v_{2np^k-1}) = \tau'_k(v)$ and $(\delta'_k)_*(u_{2np^k-2}) = \ell\sigma'_k(v)$ for $\ell \neq 0$ by Theorem 3.1 and naturality of the Bockstein. Since these classes map to $\tau_k(v), \ell\sigma_k(v) \in H_*(\Omega P^{2n+1}(p^r))$, the composite above induces a monomorphism in mod p homology (with Bocksteins acting trivially on the image).

Define $\delta_k \colon P^{2np^k-2}(p^{r+1}) \to \Omega^2 P^{2n+1}(p^r)$ to be the adjoint of the composite above. Let $\tau_k^\lambda(v)$ denote the iterated Browder bracket

$$\tau_k^\lambda(v) = \mathrm{ad}_\lambda^{p^k-1}(v)(u) \in H_{2np^k-2}(\Omega^2 P^{2n+1}(p^r)),$$

which is the transgression of the iterated commutator

$$\tau_k(v) = \mathrm{ad}^{p^k-1}(v)(u) \in H_{2np^k-1}(\Omega P^{2n+1}(p^r)).$$

It follows that $(\delta_k)_*(v_{2np^k-2}) = \tau_k^\lambda(v)$. Similarly, $(\delta_k)_*(u_{2np^k-3}) = \ell\sigma_k^\lambda(v) \neq 0$, where $\sigma_k^\lambda(v)$ denotes the transgression of $\sigma_k(v) \in H_{2np^k-2}(\Omega P^{2n+1}(p^r))$.

We now have a map $\delta_k \colon P^{2np^k-2}(p^{r+1}) \to \Omega^2 P^{2n+1}(p^r)$ inducing

$$\begin{tikzcd} v_{2np^k-2} \arrow[r, mapsto] \arrow[d, bend right, "\beta^{(r+1)}"'] & \tau_k^\lambda(v) \\ u_{2np^k-3} \arrow[r, mapsto] & \ell\sigma_k^\lambda(v) \end{tikzcd}$$

in homology and it remains to show that $\beta^{(r+1)}\tau_k^\lambda(v) \neq 0$. It suffices by naturality of $\beta^{(r+1)}$ to show that $\sigma_k^\lambda(v)$ does not represent zero in

$$E_H^{r+1}(\Omega^2 P^{2n+1}(p^r)) = H_*(E_H^r(\Omega^2 P^{2n+1}(p^r)), \beta^{(r)}).$$

First note that $\tau_k^\lambda(v)$ and $\sigma_k^\lambda(v)$ are $\beta^{(s)}$–cycles for $s \leq r$ since v_{2np^k-2} and u_{2np^k-3} are. To see that they are not $\beta^{(s)}$–boundaries for any $s \leq r$, consider the Snaith splitting of $\Omega^2 P^{2n+1}(p^r)$. Since all Bocksteins must respect the induced splitting in homology and $\tau_k^\lambda(v)$ and $\sigma_k^\lambda(v)$ lie in the homology of the stable summand $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$, it follows from Lemma 2.3 that for degree reasons the only class x which could potentially satisfy $\beta^{(s)}x = \sigma_k^\lambda(v)$ is a linear combination of $\beta^{(1)}Q_1^k v$ and $\tau_k^\lambda(v)$. But $\beta^{(s)}\beta^{(1)}Q_1^k v = \beta^{(s)}\tau_k^\lambda(v) = 0$ for all $s \le r$. Therefore $\tau_k^\lambda(v)$ and $\sigma_k^\lambda(v)$ represent nontrivial classes in $E_H^{r+1}(\Omega^2 P^{2n+1}(p^r))$, where the differential $\beta^{(r+1)}\tau_k^\lambda(v) = \ell\sigma_k^\lambda(v)$ is forced. □

As a partial 2–primary analogue of Lemma 4.1, we show that the class $\mathrm{ad}_\lambda^{2^k-1}(v)(u)$ supports a higher Bockstein when $k=1$.

Lemma 4.2 *Let $p=2$ and $r \ge 1$. Then in the mod 2 homology Bockstein spectral sequence of $\Omega^2 P^{2n+1}(2^r)$,*

$$\beta^{(r+1)}\lambda(u,v) = Q_1 u$$

in $E_H^{r+1}(\Omega^2 P^{2n+1}(2^r))$.

Proof We give a direct chain level calculation similar to the proof of [6, III.3.10]. Consider the Σ_2–invariant map $\theta\colon \mathcal{C}_2(2)\times X\times X \to X$ given by the action of the little 2–cubes operad on $X = \Omega^2 P^{2n+1}(2^r)$. Let e_k and α be as defined in [16, Section 6] and let a and b be chains representing v and u, respectively, with $d(a) = 2^r b$. Then

$$\begin{aligned} d((\alpha+1)e_1\otimes a\otimes b) &= (\alpha^2-1)e_0\otimes a\otimes b - 2^r(\alpha+1)e_1\otimes b\otimes b \\ &= -2^{r+1}e_1\otimes b\otimes b. \end{aligned}$$

Since θ_* commutes with d, it follows from the definitions of λ and Q_1 (cf [6; 16]) that $\beta^{(s)}\lambda(u,v) = 0$ for $s \le r$ and $\beta^{(r+1)}\lambda(u,v) = Q_1 u$. □

Let $i\colon S^{n-1} \to P^n(p^r)$ denote the inclusion of the bottom cell and $\eta\colon S^n \to S^{n-1}$ the Hopf map.

Lemma 4.3 *Let $n \ge 4$ and $r \ge 1$. Then:*

(a) $\pi_{n-1}(P^n(p^r)) = \mathbb{Z}/p^r\langle i\rangle$.

(b) $\pi_n(P^n(p^r)) = \begin{cases} \mathbb{Z}/2\langle i\eta\rangle & \text{if } p=2, \\ 0 & \text{if } p \text{ is odd.} \end{cases}$

Proof Both parts follow immediately from the sequence

$$\pi_j(S^{n-1}) \to \pi_j(S^{n-1}) \to \pi_j(P^n(p^r)) \to \pi_j(S^n) \to \pi_j(S^n)$$

induced by the cofibration defining $P^n(p^r)$, which is exact for $j = n-1, n$ by the Blakers–Massey theorem. Note that the degree p^r map on S^{n-1} induces multiplication by p^r on $\pi_n(S^{n-1}) = \mathbb{Z}/2\langle\eta\rangle$ since η is a suspension for $n \geq 4$ (whereas $S^2 \xrightarrow{2} S^2$ induces multiplication by 4 on $\pi_3(S^2)$, implying $\pi_3(P^3(2)) = \mathbb{Z}/4$, eg). $\square$

We are now ready to prove the splittings of Theorem 1.1, parts (a) and (b) of which are restated below as Lemmas 4.4 and 4.5, respectively.

Lemma 4.4 *If p is an odd prime and $r \geq 1$, then $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ is stably homotopy equivalent to*

$$P^{2np^k-2}(p^{r+1}) \vee X_{p^k}$$

for some finite CW–complex X_{p^k} for all $k \geq 1$.

Proof Suppose p is an odd prime, $r \geq 1$ and let $k \geq 1$. By Lemma 4.1, the map

$$\delta_k\colon P^{2np^k-2}(p^{r+1}) \to \Omega^2 P^{2n+1}(p^r)$$

induces a monomorphism in mod p homology with

$$(\delta_k)_*(v_{2np^k-2}) = \mathrm{ad}_\lambda^{p^k-1}(v)(u), \quad (\delta_k)_*(u_{2np^k-3}) = \beta^{(r+1)}\mathrm{ad}_\lambda^{p^k-1}(v)(u).$$

Since these elements have weight p^k in $H_*(\Omega^2 P^{2n+1}(p^r))$, by stabilizing δ_k and composing with the Snaith splitting, we obtain a stable map

$$P^{2np^k-2}(p^{r+1}) \to D_{p^k}(\Omega^2 P^{2n+1}(p^r))$$

with the same image in homology.

It therefore suffices to produce a map $f_k\colon D_{p^k}(\Omega^2 P^{2n+1}(p^r)) \to P^{2np^k-2}(p^{r+1})$ with

$$(f_k)_*(\mathrm{ad}_\lambda^{p^k-1}(v)(u)) = v_{2np^k-2}. \tag{4}$$

By collapsing the $(2np^k-4)$–skeleton of $D_{p^k}(\Omega^2 P^{2n+1}(p^r))$ to a point, we are left with a complex with cells only in dimensions $2np^k-3$, $2np^k-2$ and $2np^k-1$ (by Lemma 2.3) of the form

$$\left(\bigvee_{i=1}^{d} S^{2np^k-3}\right) \cup e^{2np^k-2} \cup e^{2np^k-2} \cup e^{2np^k-1}, \tag{5}$$

where $d = \dim H_{2np^k-3}(D_{p^k}(\Omega^2 P^{2n+1}(p^r)))$ and the top three cells carry the homology classes $Q_1^k v$, $\beta^{(1)} Q_1^k v$ and $\mathrm{ad}_\lambda^{p^k-1}(v)(u)$. Since

$$\mathrm{ad}_\lambda^{p^k-1}(v)(u) \in H_{2np^k-2}(D_{p^k}(\Omega^2 P^{2n+1}(p^r)))$$

supports a nontrivial $(r+1)^{\text{st}}$ Bockstein by Lemma 4.1, we may assume (altering by a self-homotopy equivalence if necessary) that the inclusion of one of the bottom cells in (5) has Hurewicz image $\beta^{(r+1)}\mathrm{ad}_\lambda^{p^k-1}(v)(u)$. Then by further collapsing a wedge $\bigvee_{i=1}^{d-1} S^{2np^k-3}$ of the other bottom cells to a point, we obtain a 4–cell complex C with mod p homology Bockstein spectral sequence given by

$$\begin{array}{lcc}
2np^k-1 & Q_1^k v & \\
 & \big\downarrow{\scriptstyle\beta^{(1)}} & \\
2np^k-2 & \beta^{(1)}Q_1^k v & \mathrm{ad}_\lambda^{p^k-1}(v)(u) \\
 & & \big\downarrow{\scriptstyle\beta^{(r+1)}} \\
2np^k-3 & & \beta^{(r+1)}\mathrm{ad}_\lambda^{p^k-1}(v)(u)
\end{array}$$

and a map $D_{p^k}(\Omega^2 P^{2n+1}(p^r)) \to C$ inducing an epimorphism in homology.

It follows from the description of $H_*(C)$ above that

$$C \simeq P^{2np^k-2}(p^{r+1}) \cup_\alpha e^{2np^k-2} \cup_\gamma e^{2np^k-1}$$

for some attaching maps α and γ. Since $\beta^{(1)}Q_1^k v \in H_{2np^k-2}(C)$ does not support any differential in the Bockstein spectral sequence and every nonzero element $\alpha \in \pi_{2np^k-3}(P^{2np^k-2}(p^{r+1})) = \mathbb{Z}/p^{r+1}$ is detected by a Bockstein, we conclude that α is trivial. Next we consider

$$\begin{aligned}\gamma \in \pi_{2np^k-2}&(P^{2np^k-2}(p^{r+1}) \vee S^{2np^k-2}) \\ &= \pi_{2np^k-2}(P^{2np^k-2}(p^{r+1})) \oplus \pi_{2np^k-2}(S^{2np^k-2}).\end{aligned}$$

By Lemma 4.3, $\pi_{2np^k-2}(P^{2np^k-2}(p^{r+1})) = 0$ since p is odd, and since the top homology class $Q_1^k v \in H_{2np^k-1}(C)$ supports a nontrivial first Bockstein differential, it follows that the second component of γ is of degree $\pm p$. Therefore,

$$C \simeq P^{2np^k-2}(p^{r+1}) \vee P^{2np^k-1}(p).$$

Finally, using this splitting we define the map f_k by the composite

$$f_k\colon D_{p^k}(\Omega^2 P^{2n+1}(p^r)) \to C \simeq P^{2np^k-2}(p^{r+1}) \vee P^{2np^k-1}(p) \xrightarrow{\pi_1} P^{2np^k-2}(p^{r+1}),$$

where the first map is the quotient map described in the previous paragraph and π_1 is the projection onto the first wedge summand. By construction, f_k satisfies (4) so the assertion follows. □

Lemma 4.5 *If $r > 1$, then there is a homotopy equivalence*

$$D_2(\Omega^2 P^{2n+1}(2^r)) \simeq P^{4n-2}(2^{r+1}) \vee X_2$$

for some 4–cell complex $X_2 = P^{4n-3}(2^r) \cup CP^{4n-2}(2)$.

Proof Let $r>1$ and note that the mod 2 homology generators $u, v \in H_*(\Omega^2 P^{2n+1}(2^r))$ in respective degrees $2n-2$, $2n-1$ give a basis for the homology of the first stable summand $D_1(\Omega^2 P^{2n+1}(2^r)) = P^{2n-1}(2^r)$ of $\Omega^2 P^{2n+1}(2^r)$. Next, a basis for the quadratic part of $H_*(\Omega^2 P^{2n+1}(2^r))$ is given by the classes

$$\begin{tikzcd}
4n-1 & Q_1 v \arrow[d, bend right, "\beta^{(1)}"'] & \\
4n-2 & v^2 & \lambda(v,u) \arrow[d, bend left, "\beta^{(r+1)}"] \\
4n-3 & uv \arrow[d, bend right, "\beta^{(r)}"'] & Q_1 u \\
4n-4 & u^2 &
\end{tikzcd}$$

with Bockstein differentials acting as indicated by Lemmas 2.5 and 4.2 and the fact that $\beta^{(r)} v = u$. It follows that $D_2(\Omega^2 P^{2n+1}(2^r))$ has the homotopy type of a 6–cell complex

$$D_2(\Omega^2 P^{2n+1}(2^r)) \simeq P^{4n-3}(2^r) \cup_\alpha e^{4n-3} \cup_\gamma e^{4n-2} \cup_\delta e^{4n-2} \cup_\epsilon e^{4n-1}$$

with homology as above, where the bottom Moore space carries the homology classes uv and u^2. As in the proof of Lemma 4.4, the attaching map α is null homotopic since every nonzero element of $\pi_{4n-4}(P^{4n-3}(2^r)) = \mathbb{Z}/2^r$ is detected by a Bockstein and $Q_1 u$ supports no differential in the homology Bockstein spectral sequence. The next attaching maps γ and δ may therefore be regarded as elements of

$$\pi_{4n-3}(P^{4n-3}(2^r) \vee S^{4n-3}) = \mathbb{Z}/2 \oplus \mathbb{Z},$$

where the first summand is generated by $i\eta$ by Lemma 4.3. Naturality and the morphism of cofibrations

$$\begin{tikzcd}
S^{4n-3} \arrow[r, "\eta"] \arrow[d, equal] & S^{4n-4} \arrow[r] \arrow[d, "i"] & \Sigma^{4n-6}\mathbb{C}P^2 \arrow[d] \\
S^{4n-3} \arrow[r, "i\eta"] & P^{4n-3}(2^r) \arrow[r] & C_{i\eta}
\end{tikzcd}$$

imply that $i\eta$ is detected by Sq^2_* since η is. Since Sq^2_* acts trivially on

$$H_*(D_2(\Omega^2 P^{2n+1}(2^r)))$$

when $r>1$ by Lemma 2.5, the first components of γ and δ must therefore be trivial. Without loss of generality, we may assume the second components of γ and δ are trivial and degree $\pm 2^{r+1}$, respectively, since we have a basis $\{v^2, \lambda(v,u)\}$ of

$H_{4n-2}(D_2(\Omega^2 P^{2n+1}(2^r)))$ where v^2 is a permanent cycle and $\lambda(v,u)$ supports a nontrivial $\beta^{(r+1)}$.

We now have a homotopy equivalence

$$D_2(\Omega^2 P^{2n+1}(2^r)) \simeq (P^{4n-3}(2^r) \vee P^{4n-2}(2^{r+1}) \vee S^{4n-2}) \cup_\epsilon e^{4n-1},$$

where $v^2 \in H_{4n-2}(D_2(\Omega^2 P^{2n+1}(2^r)))$ corresponds to the fundamental class of the $(4n-2)$–sphere on the right. Denote the components of the attaching map ϵ by

$$\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3) \in \pi_{4n-2}(P^{4n-3}(2^r)) \oplus \pi_{4n-2}(P^{4n-2}(2^{r+1})) \oplus \pi_{4n-2}(S^{4n-2}).$$

It suffices to show that ϵ_2 and ϵ_3 are trivial and degree ± 2, respectively. Clearly the first Bockstein $\beta^{(1)} Q_1 v = v^2$ on the top class of $D_2(\Omega^2 P^{2n+1}(2^r))$ implies ϵ_3 is of degree ± 2. To see that ϵ_2 is trivial, collapse the bottom Moore space $P^{4n-3}(2^r)$ to a point and repeat the argument above analyzing the attaching map γ. □

To conclude the proof of Theorem 1.1, it remains to show that $D_2(\Omega^2 P^{2n+1}(2))$ is a stably indecomposable 6–cell complex. This follows immediately from homological considerations: by Lemmas 2.5 and 4.2, $H_*(D_2(\Omega^2 P^{2n+1}(2)))$ clearly does not admit any nontrivial decomposition respecting Steenrod and higher Bockstein operations.

5 Proof of Theorem 1.2

In this section we derive Theorem 1.2 from Theorem 1.1 and discuss some implications for the unstable homotopy groups of odd primary Moore spaces.

Proof of Theorem 1.2 Suppose p is an odd prime and $r \geq 1$. Then for each $k \geq 1$, the map $\delta_k \colon P^{2np^k-2}(p^{r+1}) \to \Omega^2 P^{2n+1}(p^r)$ from Lemma 4.1 admits a stable retraction by Lemma 4.4. Taking adjoints in the resulting homotopy commutative diagram

$$\begin{tikzcd}
\Sigma^\infty P^{2np^k-2}(p^{r+1}) \arrow[r, "\Sigma^\infty \delta_k"] \arrow[dr, equal] & \Sigma^\infty \Omega^2 P^{2n+1}(p^r) \arrow[d] \\
& \Sigma^\infty P^{2np^k-2}(p^{r+1})
\end{tikzcd}$$

yields the desired factorization of the unstable map

$$E^\infty \colon P^{2np^k-2}(p^{r+1}) \to QP^{2np^k-2}(p^{r+1})$$

through $\Omega^2 P^{2n+1}(p^r)$.

In order to similarly factor the stabilization map of a mod p^{r+1} Moore space through $\Omega^2 P^{2n}(p^r)$, we reduce to the odd-dimensional case using the fact that $\Omega^2 P^{4n-1}(p^r)$ is an unstable retract of $\Omega^2 P^{2n}(p^r)$ by the loop space decomposition (3). Explicitly, there is a map

$$P^{(4n-2)p^k-2}(p^{r+1}) \xrightarrow{\delta_k} \Omega^2 P^{4n-1}(p^r) \to \Omega^2 P^{2n}(p^r)$$

admitting a stable retraction, so the argument above implies that the stabilization map of $P^{(4n-2)p^k-2}(p^{r+1})$ factors through $\Omega^2 P^{2n}(p^r)$. □

Note that if p is prime and $\pi_j(P^{2np^k-2}(p^{r+1}))$ is in the stable range such that the map $E^\infty\colon P^{2np^k-2}(p^{r+1}) \to QP^{2np^k-2}(p^{r+1})$ is an isomorphism on $\pi_j(\cdot)$, then Theorem 1.2 implies that $\pi_j(P^{2np^k-2}(p^{r+1}))$ is a summand of $\pi_{j+2}(P^{2n+1}(p^r))$. Since for any given $j \in \mathbb{Z}$ we have $\pi_j^s(P^{2np^k-2}(p^{r+1})) = \pi_j(P^{2np^k-2}(p^{r+1}))$ for k sufficiently large, it follows that every stable homotopy group of a mod p^{r+1} Moore space is a summand of $\pi_*(P^{2n+1}(p^r))$.

Rephrasing a little, we have the following consequence of Theorem 1.2. Let $\mathbb{S}/p^r$ denote the mod p^r Moore spectrum; that is, the cofibre of $\mathbb{S} \xrightarrow{p^r} \mathbb{S}$ where $\mathbb{S}$ is the sphere spectrum.

Corollary 5.1 *Let p be an odd prime and $r \geq 1$. Then for each $j \in \mathbb{Z}$, $\pi_j(\mathbb{S}/p^{r+1})$ is a summand of $\pi_{2np^k+j-1}(P^{2n+1}(p^r))$ for every sufficiently large k.*

Proof For each $j \in \mathbb{Z}$,

$$\pi_j(\mathbb{S}/p^{r+1}) = \pi^s_{j+2np^k-3}(P^{2np^k-2}(p^{r+1})) = \pi_{j+2np^k-3}(P^{2np^k-2}(p^{r+1}))$$

for all sufficiently large k. Therefore the first commutative diagram in Theorem 1.2 implies that $\pi_j(\mathbb{S}/p^{r+1})$ retracts off

$$\pi_{j+2np^k-3}(\Omega^2 P^{2n+1}(p^r)) = \pi_{j+2np^k-1}(P^{2n+1}(p^r)).$$ □

Remark 5.2 The second commutative diagram in Theorem 1.2 implies that similar results hold for the unstable homotopy groups of even-dimensional odd primary Moore spaces.

6 v_1–periodic families

In this section we construct new infinite families of higher torsion elements in the unstable homotopy groups of Moore spaces using Theorem 1.2 and periodic self-maps

$$v_1\colon P^{n+q_r}(p^r) \to P^n(p^r),$$

as introduced by Adams [1] in his study of the image of the J–homomorphism. Here,

$$q_r = \begin{cases} qp^{r-1} & \text{if } p \text{ is odd,} \\ \max(8, 2^{r-1}) & \text{if } p = 2, \end{cases}$$

where $q = 2(p-1)$ and v_1 induces an isomorphism in K–theory. Such maps exist unstably provided $n \geq 2r + 3$ by [12] and desuspend further to $P^3(p)$ in case p is odd and $r = 1$ by [10].

Restricting each iterate $v_1^t = v_1 \circ \Sigma^{q_r} v_1 \circ \cdots \circ \Sigma^{tq_r} v_1$ of v_1 to the bottom cell gives an infinite family of maps

$$S^{n+tq_r-1} \to P^{n+tq_r}(p^r) \xrightarrow{v_1^t} P^n(p^r)$$

which generate $\mathbb{Z}/p^r$ summands in $\pi_{n+tq_r-1}(P^n(p^r))$ for $t \geq 0$, and composing with the pinch map $q\colon P^n(p^r) \to S^n$ gives rise to the first studied infinite families in the stable homotopy groups of spheres. For example, if p is odd, these composites form the α–family and generate the p–component of the image of J in $\pi_{tq_r-1}(\mathbb{S})$; see [11, Proposition 1.1]. To generate $\mathbb{Z}/p^{r+1}$ summands in $\pi_*(P^n(p^r))$ when p is odd, we apply the same procedure to mod p^{r+1} Moore spaces and compose into $P^n(p^r)$ along the maps in the diagrams of Theorem 1.2.

When $p = 2$, analogous unstable maps $\delta_1\colon P^{4n-2}(2^{r+1}) \to \Omega^2 P^{2n+1}(2^r)$ realizing the stable splittings of Theorem 1.1 only exist when $n = 2$ or 4, as we show below. In the $n = 2$ case, no Adams self-map of the mod 2^{r+1} Moore spectrum desuspends far enough to precompose δ_1 with. Instead, we show that an infinite family of elements of order 8 in the homotopy groups of spheres constructed in [15] factors through $P^6(8)$ and injects along $\delta_1\colon P^6(8) \to \Omega^2 P^5(4)$; see Theorem 6.6 below.

6.1 The odd primary case

Let $n > 1$. As usual, for an odd prime p we let $q = 2(p-1)$. The following is a more precise statement of Theorem 1.3.

Theorem 6.1 *Let p be an odd prime and $r \geq 1$.*

(a) *If $k \geq \log_p((r+4)/n)$, then $\pi_{2np^k-1+tqp^r}(P^{2n+1}(p^r))$ contains a $\mathbb{Z}/p^{r+1}$ summand for every $t \geq 0$.*

(b) *If $k \geq \log_p((r+3)/(2n-1))$, then $\pi_{(4n-2)p^k-1+tqp^r}(P^{2n}(p^r))$ contains a $\mathbb{Z}/p^{r+1}$ summand for every $t \geq 0$.*

Proof By Theorem 1.2, $E^\infty\colon P^{2np^k-2}(p^{r+1}) \to QP^{2np^k-2}(p^{r+1})$ factors as a composite

$$P^{2np^k-2}(p^{r+1}) \xrightarrow{\delta_k} \Omega^2 P^{2n+1}(p^r) \to QP^{2np^k-2}(p^{r+1}).$$

Note, the restriction of δ_k to the bottom cell defines an element of $\pi_{2np^k-1}(P^{2n+1}(p^r))$ of order p^{r+1}. The bound on k ensures that $2np^k - 2 \geq 2(r+1)+4$, which implies that an unstable representative $v_1\colon P^{2np^k-2+qp^r}(p^{r+1}) \to P^{2np^k-2}(p^{r+1})$ of the Adams map exists by [12, Proposition 2.11]. That the restriction of any iterate v_1^t to the bottom cell has order p^{r+1} follows from the fact that v_1^t induces an isomorphism in K–theory. Therefore the composite

$$S^{2np^k-3+tqp^r} \to P^{2np^k-2+tqp^r}(p^{r+1}) \xrightarrow{v_1^t} P^{2np^k-2}(p^{r+1}) \xrightarrow{\delta_k} \Omega^2 P^{2n+1}(p^r)$$

also has order p^{r+1} for all $t \geq 0$ since composing further into $QP^{2np^k-2}(p^{r+1})$ gives the adjoint of the restriction of $\Sigma^\infty v_1^t$ to the bottom cell. Part (b) is proved similarly. □

Remark 6.2 The proof above shows that each δ_k generates an infinite v_1–periodic family in $\pi_*(P^{2n+1}(p^r); \mathbb{Z}/p^{r+1})$ giving rise to an infinite family of higher torsion elements in $\pi_*(P^{2n+1}(p^r))$. We point out that in the loop space decomposition [9]

$$\Omega P^{2n+1}(p^r) \simeq T^{2n+1}\{p^r\} \times \Omega\Big(\bigvee_\alpha P^{n_\alpha}(p^r)\Big),$$

each of these elements lands in the homotopy of the bottom indecomposable factor $T^{2n+1}\{p^r\}$, and many more infinite families than are indicated here can be obtained by applying the Hilton–Milnor theorem to the second factor and iterating our construction above.

6.2 The 2–primary case

We consider next the problem of desuspending the inclusion of the stable summand $P^{4n-2}(2^{r+1})$ of $\Omega^2 P^{2n+1}(2^r)$ given by Theorem 1.1(b) and mimicking the construction above of unstable v_1–periodic families of higher odd primary torsion elements.

Note that a homotopy commutative diagram

$$\begin{tikzcd} P^{4n-2}(2^{r+1}) \arrow[r] \arrow[dr, "E^\infty"'] & \Omega^2 P^{2n+1}(2^r) \arrow[d] \\ & QP^{4n-2}(2^{r+1}) \end{tikzcd} \tag{6}$$

cannot exist unless $r > 1$ since $D_2(\Omega^2 P^{2n+1}(2))$ is stably indecomposable by Theorem 1.1(c). Furthermore, such a factorization implies $Q_1 u \in H_{4n-3}(\Omega^2 P^{2n+1}(2^r))$ is spherical since only this class lies in the image of the $(r+1)^{\text{st}}$ Bockstein in degree $4n-3$. For $r = 1$, it follows from the proposition below that this class is spherical only in Kervaire invariant dimensions.

Proposition 6.3 [23, Proposition 2.21] *The class $u^2 \in H_{4n-2}(\Omega P^{2n+1}(2))$ is spherical if and only if the Whitehead square $w_{2n-1} \in \pi_{4n-3}(S^{2n-1})$ is divisible by 2.*

For $r > 1$, the same argument leads to the following.

Proposition 6.4 *Let $r > 1$. The following conditions are equivalent*:

(a) $Q_1 u \in H_{4n-3}(\Omega^2 P^{2n+1}(2^r))$ *is spherical*;

(b) $u^2 \in H_{4n-2}(\Omega P^{2n+1}(2^r))$ *is spherical*;

(c) $n = 1, 2$ *or* 4.

Proof If a map $f : S^{4n-3} \to \Omega^2 P^{2n+1}(2^r)$ has mod 2 reduced Hurewicz image $Q_1 u$, then the adjoint of f factors as

$$f' : S^{4n-2} \xrightarrow{\Sigma f} \Sigma\Omega^2 P^{2n+1}(2^r) \xrightarrow{\sigma} \Omega P^{2n+1}(2^r),$$

where σ induces the homology suspension

$$\sigma_* : H_*(\Omega^2 P^{2n+1}(2^r)) \to H_{*+1}(\Omega P^{2n+1}(2^r)).$$

Thus $\sigma_*(Q_1 u) = u^2$ is the Hurewicz image of f'.

Conversely, given $g' : S^{4n-2} \to \Omega P^{2n+1}(2^r)$ with $g'_*(\iota_{4n-2}) = u^2$, the adjoint of g' factors as

$$g : S^{4n-3} \xrightarrow{E} \Omega S^{4n-2} \xrightarrow{\Omega g'} \Omega^2 P^{2n+1}(2^r).$$

Consider the morphism of path-loop fibrations induced by g'. Since u^2 transgresses to $Q_1 u$ in the Serre spectral sequence associated to the path-loop fibration over $\Omega P^{2n+1}(2^r)$, it follows by naturality that $g_*(\iota_{4n-3}) = Q_1 u$. Therefore conditions (a) and (b) are equivalent.

If $n = 1, 2$ or 4, then the adjoint of the Hopf invariant one map $S^{4n-1} \to S^{2n}$ has Hurewicz image $\iota_{2n-1}^2 \in H_{4n-2}(\Omega S^{2n})$, so the composite

$$S^{4n-2} \to \Omega S^{2n} \xrightarrow{\Omega i} \Omega P^{2n+1}(2^r)$$

has Hurewicz image u^2.

Conversely, if $u^2 \in H_{4n-2}(\Omega P^{2n+1}(2^r))$ is spherical, then the proof given in [23] of Proposition 6.3 above shows that $i \circ w_{2n-1}$ is null homotopic in the diagram

$$\begin{tikzcd} \Omega P^{2n}(2^r) \arrow[r] & E \arrow[r, "f"] & S^{2n-1} \arrow[r, "i"] & P^{2n}(2^r) \\ & & S^{4n-3} \arrow[ul, dashed, "\ell"] \arrow[u, "w_{2n-1}"'] & \end{tikzcd}$$

where the top row is a fibration sequence. By [5, Lemma 21.1],

$$[u, v] \in H_{4n-3}(\Omega P^{2n}(2^r))$$

is spherical and so is its image in $H_{4n-3}(E)$. As in [23], it follows that the $(4n-3)$–skeleton of E is homotopy equivalent to $S^{2n-1} \vee S^{4n-3}$ and $f|_{S^{4n-3}}$ is null homotopic. Therefore a lift ℓ may be chosen to factor through $f|_{S^{2n-1}}$, which is of degree 2^r. Since the degree 2 map induces multiplication by 2 on $\pi_{4n-3}(S^{2n-1})$ by Barratt's distributivity formula [5, Proposition 4.3], the Whitehead square w_{2n-1} is divisible by 2^r, which implies $n \in \{1, 2, 4\}$ since $r > 1$. $\square$

By Proposition 6.4, diagrams of the form (6) inducing the stable splittings of Theorem 1.1(b) (where $n > 1$ is assumed) cannot exist if $n \neq 2$ or 4. We verify that such diagrams do exist in these two exceptional dimensions.

Theorem 6.5 *Let $r > 1$. Then there exist homotopy commutative diagrams*

$$\begin{tikzcd} P^6(2^{r+1}) \arrow[r] \arrow[dr, "E^\infty"'] & \Omega^2 P^5(2^r) \arrow[d] \\ & QP^6(2^{r+1}) \end{tikzcd} \qquad \begin{tikzcd} P^{14}(2^{r+1}) \arrow[r] \arrow[dr, "E^\infty"'] & \Omega^2 P^9(2^r) \arrow[d] \\ & QP^{14}(2^{r+1}) \end{tikzcd}$$

Proof Let $n = 2$ or 4 and let $f\colon S^{4n-3} \to \Omega^2 P^{2n+1}(2^r)$ be a map with mod 2 reduced Hurewicz image $Q_1 u$. Then by Lemma 4.2, the integral Hurewicz image of f is a generator of $H_{4n-3}(\Omega^2 P^{2n+1}(2^r); \mathbb{Z}) \cong \mathbb{Z}/2^{r+1}$, so f has order at least 2^{r+1}. That f has order at most 2^{r+1} follows from [5, Proposition 13.3], so f extends to a map $\bar{f}\colon P^{4n-2}(2^{r+1}) \to \Omega^2 P^{2n+1}(2^r)$ with $\beta^{(r+1)} \bar{f}_*(v_{4n-2}) = Q_1 u$. Since $r > 1$, the Snaith splitting and Theorem 1.1(b) give a composite

$$\Omega^2 P^{2n+1}(2^r) \to QD_2(\Omega^2 P^{2n+1}(2^r)) \to QP^{4n-2}(2^{r+1}),$$

which is an epimorphism on $H_{4n-2}(\cdot)$ and $H_{4n-3}(\cdot)$. It follows that the composition of $\bar{f}$ with the composite above is $(4n-2)$–connected and hence homotopic to the stabilization map $E^\infty\colon P^{4n-2}(2^{r+1}) \to QP^{4n-2}(2^{r+1})$ up to a self-equivalence. $\square$

Theorem 6.6 *Let $r = 2$ or 3. Then*

(a) $\pi_{3+8t}(P^5(4))$ *contains a* $\mathbb{Z}/8$ *summand for every* $t \geq 1$;

(b) $\pi_{7+8t}(P^9(2^r))$ *contains a* $\mathbb{Z}/2^{r+1}$ *summand for every* $t \geq 1$.

Proof We begin with part (b). Consider the second diagram in Theorem 6.5 and let $r = 2$ or 3. Then for stability reasons, an unstable Adams map

$$v_1 \colon P^{n+8}(2^{r+1}) \to P^n(2^{r+1})$$

exists for $n = 14$. As in the odd primary case, restricting any iterate

$$v_1^{t-1} \colon P^{6+8t}(2^{r+1}) \to P^{14}(2^{r+1})$$

to the bottom cell yields a homotopy class of order 2^{r+1} and stable order 2^{r+1}. The resulting composition with the map $P^{14}(2^{r+1}) \to \Omega^2 P^9(2^r)$ therefore generates a $\mathbb{Z}/2^{r+1}$ summand in $\pi_{5+8t}(\Omega^2 P^9(2^r)) = \pi_{7+8t}(P^9(2^r))$ by Theorem 6.5.

For part (a), we use the fact that an unstable Adams map $v_1 \colon P^{n+8}(8) \to P^n(8)$ exists for $n \geq 9$ with the property that the composite

$$N_t \colon S^{8t} \xrightarrow{i} P^{1+8t}(8) \xrightarrow{v_1^{t-1}} P^9(8) \xrightarrow{\nu^\sharp} S^5$$

has order 8 in $\pi_{8t}(S^5)$ for all $t \geq 1$ by [15, Theorem E]. Here $\nu^\sharp$ denotes an extension of $\nu \colon S^8 \to S^5$. Suspending once, an extension $\nu^\sharp \colon P^{10}(8) \to S^6$ of $\nu \colon S^9 \to S^6$ can be chosen to factor through the pinch map $q \colon P^6(8) \to S^6$ (we postpone a proof of this claim to Lemma 6.7 below). Combining this with the first diagram in Theorem 6.5, we obtain a homotopy commutative diagram

$$\begin{tikzcd}
P^{2+8t}(8) \arrow[r, "v_1^{t-1}"] & P^{10}(8) \arrow[r] & P^6(8) \arrow[r] \arrow[d, "q"] \arrow[dr, "E^\infty"'] & \Omega^2 P^5(4) \arrow[d] \\
S^{1+8t} \arrow[u, "i"] \arrow[rr, "\Sigma N_t"] & & S^6 \arrow[dr, "E^\infty"'] & QP^6(8) \arrow[d, "Qq"] \\
& & & QS^6
\end{tikzcd}$$

The composite $E^\infty \circ \Sigma N_t$ is adjoint to $\Sigma^\infty N_t \in \pi_{8t-5}(\mathbb{S})$ and therefore has order 8 since the proof of [15, Theorem E] shows N_t has real e–invariant $b/8$ where b is odd. Hence the composite $S^{1+8t} \to \Omega^2 P^5(4)$ has order at least 8. Since $8i = 0$, the theorem follows. $\square$

It remains to prove the following factorization of $\nu \in \pi_9(S^6)$ used in the proof above.

Lemma 6.7 *There is a homotopy commutative diagram*

$$\begin{tikzcd} P^{10}(8) \arrow[r] & P^6(8) \arrow[d, "q"] \\ S^9 \arrow[u, "i"] \arrow[r, "\nu"] & S^6 \end{tikzcd}$$

Proof Since $8 \circ \nu = \nu \circ 8 = 0$ in $\pi_9(S^6)$, ν lifts to the fibre $S^6\{8\}$ of the degree 8 map, and since $\mathrm{sk}_9(S^6\{8\}) \simeq P^6(8)$, it follows that ν factors as $S^9 \xrightarrow{\ell} P^6(8) \xrightarrow{q} S^6$. It suffices to show that ℓ has order 8. The fibre F of the pinch map q has the homotopy type of a CW–complex $S^5 \cup e^{10} \cup e^{15} \cup \cdots$ where the first attaching map is $8w_5 = 0$ by [13, Corollary 5.8]. It follows that $\pi_9(F) = \mathbb{Z}/2$. The short exact sequence

$$0 = \pi_{10}(S^6) \to \pi_9(F) \to \pi_9(P^6(8)) \xrightarrow{q_*} \pi_9(S^6) \xrightarrow{0} \pi_8(F)$$

therefore implies $\pi_9(P^6(8)) = \mathbb{Z}/2 \oplus \mathbb{Z}/8$. In particular, ℓ has order 8. □

References

[1] **J F Adams**, *On the groups $J(X)$, IV*, Topology 5 (1966) 21–71 MR Zbl

[2] **S Amelotte**, *The fibre of the degree 3 map, Anick spaces and the double suspension*, Proc. Edinb. Math. Soc. 63 (2020) 830–843 MR Zbl

[3] **P Beben**, **J Wu**, *Modular representations and the homotopy of low rank p–local CW–complexes*, Math. Z. 273 (2013) 735–751 MR Zbl

[4] **W Chen**, **J Wu**, *Decomposition of loop spaces and periodic problem on π_**, Algebr. Geom. Topol. 13 (2013) 3245–3260 MR Zbl

[5] **F R Cohen**, *A course in some aspects of classical homotopy theory*, from "Algebraic topology" (H R Miller, D C Ravenel, editors), Lecture Notes in Math. 1286, Springer (1987) 1–92 MR Zbl

[6] **F R Cohen**, **T J Lada**, **J P May**, *The homology of iterated loop spaces*, Lecture Notes in Math. 533, Springer (1976) MR Zbl

[7] **F R Cohen**, **J C Moore**, **J A Neisendorfer**, *The double suspension and exponents of the homotopy groups of spheres*, Ann. of Math. 110 (1979) 549–565 MR Zbl

[8] **F R Cohen**, **J C Moore**, **J A Neisendorfer**, *Torsion in homotopy groups*, Ann. of Math. 109 (1979) 121–168 MR Zbl

[9] **F R Cohen**, **J C Moore**, **J A Neisendorfer**, *Exponents in homotopy theory*, from "Algebraic topology and algebraic K–theory" (W Browder, editor), Ann. of Math. Stud. 113, Princeton Univ. Press (1987) 3–34 MR Zbl

[10] **F R Cohen**, **J A Neisendorfer**, *Note on desuspending the Adams map*, Math. Proc. Cambridge Philos. Soc. 99 (1986) 59–64 MR Zbl

[11] **M C Crabb**, **K Knapp**, *Adams periodicity in stable homotopy*, Topology 24 (1985) 475–486 MR Zbl

[12] **D M Davis**, **M Mahowald**, *Some remarks on v_1–periodic homotopy groups*, from "Adams memorial symposium on algebraic topology, II" (N Ray, G Walker, editors), London Math. Soc. Lecture Note Ser. 176, Cambridge Univ. Press (1992) 55–72 MR Zbl

[13] **B Gray**, *On the homotopy groups of mapping cones*, Proc. London Math. Soc. 26 (1973) 497–520 MR Zbl

[14] **B Gray**, *EHP spectra and periodicity, I: Geometric constructions*, Trans. Amer. Math. Soc. 340 (1993) 595–616 MR Zbl

[15] **J R Klein**, **W Richter**, *Poincaré duality and periodicity*, Algebr. Geom. Topol. 11 (2011) 1961–1985 MR Zbl

[16] **J P May**, *A general algebraic approach to Steenrod operations*, from "The Steenrod algebra and its applications" (F P Peterson, editor), Lecture Notes in Math. 168, Springer (1970) 153–231 MR Zbl

[17] **J A Neisendorfer**, *3–Primary exponents*, Math. Proc. Cambridge Philos. Soc. 90 (1981) 63–83 MR Zbl

[18] **J A Neisendorfer**, *The exponent of a Moore space*, from "Algebraic topology and algebraic K–theory" (W Browder, editor), Ann. of Math. Stud. 113, Princeton Univ. Press (1987) 35–71 MR Zbl

[19] **J Neisendorfer**, *Product decompositions of the double loops on odd primary Moore spaces*, Topology 38 (1999) 1293–1311 MR Zbl

[20] **J Neisendorfer**, *Algebraic methods in unstable homotopy theory*, New Mathematical Monographs 12, Cambridge Univ. Press (2010) MR Zbl

[21] **V P Snaith**, *A stable decomposition of $\Omega^n S^n X$*, J. London Math. Soc. 7 (1974) 577–583 MR Zbl

[22] **S D Theriault**, *Anick's spaces and the double loops of odd primary Moore spaces*, Trans. Amer. Math. Soc. 353 (2001) 1551–1566 MR Zbl

[23] **J Wu**, *Homotopy theory of the suspensions of the projective plane*, Mem. Amer. Math. Soc. 769, Amer. Math. Soc., Providence, RI (2003) MR Zbl

Department of Mathematics, Western University
London, ON, Canada

Department of Mathematics, University of Rochester
Rochester, NY, United States

Department of Mathematics, University of Rochester
Rochester, NY, United States

samelot@uwo.ca, yuxinluo@rochester.edu

Received: 24 November 2021 Revised: 26 February 2022

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 23 Issue 5 (pages 1935–2414) 2023