AGT

Algebraic & Geometric Topology

Volume 17 (2017)

Issue 4 (pages 1917–2564)

msp

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

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

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

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebraic & Geometric Topology 17 (2017) 1917–1951

The mapping cone formula in Heegaard Floer homology and Dehn surgery on knots in S^3

FYODOR GAINULLIN

We write down an explicit formula for the $+$ version of the Heegaard Floer homology (as an absolutely graded vector space over an arbitrary field) of the results of Dehn surgery on a knot K in S^3 in terms of homological data derived from $\mathrm{CFK}^\infty(K)$. This allows us to prove some results about Dehn surgery on knots in S^3. In particular, we show that for a fixed manifold there are only finitely many alternating knots that can produce it by surgery. This is an improvement on a recent result by Lackenby and Purcell. We also derive a lower bound on the genus of knots depending on the manifold they give by surgery. Some new restrictions on Seifert fibred surgery are also presented.

57M27, 57M25

1 Introduction

Dehn surgery is a fundamental technique in 3–manifold topology. Indeed, we can construct any 3–manifold[1] beginning with any other 3–manifold and performing Dehn surgery enough times. However, it is a highly nontrivial and widely open problem to understand what manifolds can be obtained by doing Dehn surgery once (even starting from the "simplest" 3–manifold, namely S^3) and what knots yield a fixed manifold by surgery.

Heegaard Floer theory is a relatively recent collection of powerful tools in low-dimensional topology. It has many aspects and provides invariants in many different contexts. In this paper, we are only concerned with the 3–manifold and knot invariants (defined in Ozsváth and Szabó [18; 17] and Rasmussen [24]). The collections of 3–manifold invariants and knot invariants are connected via the surgery formula that expresses the Heegaard Floer homology of a 3–manifold obtained by surgery on a given knot in terms of the Heegaard Floer homology data of the knot (see Ozsváth and Szabó [23]). This makes Heegaard Floer homology an especially suitable tool for investigating questions about Dehn surgery.

[1] In this paper, whenever we say "3–manifold" we mean "closed connected orientable 3–manifold".

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.1917

A natural question about Dehn surgery is whether there are manifolds that can be obtained by surgery on infinitely many distinct knots in S^3. The answer is "yes"; see Osoinach [13] or Teragaito [27]. There is still hope, however, that perhaps this does not happen for some nice classes of knots.

One interesting and well-studied class of knots is that of alternating knots. At first sight, their diagrammatic definition seems to have little to do with the geometric-topological properties of these knots. However, this is not so; see, for example, Lackenby and Purcell [7]. In particular, they prove the following:

Theorem 1 [7, Theorem 1.3] *For any closed* 3*–manifold* M *with sufficiently large Gromov norm, there are at most finitely many prime alternating knots* K *and fractions* p/q *such that* M *is obtained by* p/q *surgery along* K*.*

In fact, the statement about fractions p/q can be deduced, for example, from Ni and Wu [11, Theorem 1.2]. We will also show in this paper that given any manifold Y there is a universal bound on q for such fractions, which also implies that they are finite in number. Using techniques that are very different from those used in [7] we are able to establish the following improvement of this theorem.

Theorem 2 *Let* Y *be a* 3*–manifold. There are at most finitely many alternating knots* $K \subset S^3$ *such that* $Y = S^3_{p/q}(K)$*.*

Heegaard Floer homology is also very useful in bounding genera of various surfaces. In particular, knot Floer homology determines the genus of a knot; see Ozsváth and Szabó [16]. Combining this with information about surgery often allows one to put restrictions on genera of knots admitting certain surgeries. For example, if surgery on a knot K produces an L–space Y (a generalisation of lens spaces; see below for the definition), then $2g(K)-1 \le |H_1(Y)|$, where by $g(K)$ we mean the genus of K (see eg [23, Corollary 1.4]).

We derive a bound which is in some respects "opposite" to the bound for L–spaces. It is a lower bound which can be nontrivial only for non-L–spaces. For the statement of the theorem below and the rest of the paper note that we work over an arbitrary field $\mathbb{F}$. Heegaard Floer homology is then an $\mathbb{F}[U]$–module and we denote the action of U simply by multiplication. For a rational homology sphere Y, $\mathrm{HF}_{\mathrm{red}}(Y)$ denotes its reduced Floer homology.

Theorem 3 *For any knot* $K \subset S^3$ *and any* $p/q \in \mathbb{Q}$ *we have*

$$U^{g(K)+\lceil g_4(K)/2\rceil} \cdot \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = 0.$$

We remark that if K is an L–space knot, then $U^{\lceil g_4(K)/2\rceil}\cdot \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = 0$. Moreover, for any $N > 0$ and $p > 0$ there is a 3–manifold Y which can be obtained by a surgery on a knot in S^3 such that $U^N \cdot \mathrm{HF}_{\mathrm{red}}(Y) \neq 0$ and $|H_1(Y)| = p$.

Here $g_4(K)$ is the slice genus of K. We obviously have $\lceil \frac{1}{2} g_4(K)\rceil \leq \lceil \frac{1}{2} g(K)\rceil$, so the theorem does give a lower bound for $g(K)$.

A different lower bound for the knot genus producing non-L–spaces has been found by Jabuka [3, Theorem 1.3], but unlike our bound, it also depends on the denominator of the slope. Note also that there exists a manifold for which the genus of knots producing it is not bounded above [27].

More recently, Jabuka [4, Corollary 1.5] has produced a new lower bound on the genus that does not involve the denominator of the slope. He also obtained the ranks of $\widehat{\mathrm{HF}}$ for the result of surgery on a knot in S^3. His genus bound appears to be quite different from ours.

Using the genus bound of Theorem 3 and some other considerations we are able to prove results about Seifert fibred surgery on knots in S^3. Wu (improving on the results of Ozsváth and Szabó [20]) has proven the following (the definitions of Seifert orientation and torsion coefficients will be provided later):

Theorem 4 [29, Theorems 1.2 and 1.3] *Let $K \subset S^3$ be a knot. Suppose there is a rational number $p/q > 0$ such that $Y = S^3_{p/q}(K)$ is Seifert fibred.*

If Y is a positively oriented Seifert fibred space, then all the torsion coefficients $t_i(K)$ are nonnegative and $\widehat{\mathrm{HFK}}(K, g(K))$ is supported in even degrees. In particular, $\deg \Delta_K = g(K)$.

If Y is a negatively oriented Seifert fibred space and $0 < p/q < 3$, then for all $i > 0$ the torsion coefficients $t_i(K)$ are nonpositive. If Y is a negatively oriented Seifert fibred space, $g(K) > 1$ and $2g(K) - 1 > p/q$, then $\widehat{\mathrm{HFK}}(K, g(K))$ is supported in odd degrees. In particular, $\deg \Delta_K = g(K)$.

We are able to prove the following.

Theorem 5 *Let $K \subset S^3$ be a knot. Suppose there is a rational number $p/q > 0$ such that $Y = S^3_{p/q}(K)$ is a negatively oriented Seifert fibred space. Then*

- $U^{g(K)} \cdot \mathrm{HF}_{\mathrm{red}}(Y) = 0$;
- *if $0 < p/q \leq 3$, then all the torsion coefficients $t_i(K)$ are nonpositive (including $t_0(K)$) and* $\deg \Delta_K = g(K)$;

- *more generally, if* $i \geq \lfloor \frac{1}{2}(\lceil p/q \rceil - \sqrt{\lceil p/q \rceil}) \rfloor$*, then* t_i *is nonpositive;*
- *if* $g(K) > \lfloor \frac{1}{2}(\lceil p/q \rceil - \sqrt{\lceil p/q \rceil}) \rfloor$*, then* $\deg \Delta_K = g(K)$*;*
- *if* $U^{\lfloor |H_1(Y)|/2 \rfloor} \cdot \mathrm{HF}_{\mathrm{red}}(Y) \neq 0$ *then* $\deg \Delta_K = g(K)$*.*

In all statements where $\deg \Delta_K = g(K)$ *we have that* $\widehat{\mathrm{HFK}}(K, g(K))$ *is supported in odd degrees.*

After the proof of Theorem 3 in Section 5, we describe negatively oriented Seifert fibred spaces Y for which the power of U needed to annihilate $\mathrm{HF}_{\mathrm{red}}(Y)$ gets arbitrarily large compared to the order of the first homology group.

Theorem 5 combined with the result of Wu has the following straightforward corollary.

Corollary 6 *Suppose* $Y = S^3_{p/q}(K)$ *is a Seifert fibred rational homology sphere. If* $|H_1(Y)| \leq 3$*, then all the torsion coefficients of* K *have the same sign and* $\deg \Delta_K = g(K)$*.*

To prove Theorems 2 and 3 we need to study the mapping cone formula, which connects the Heegaard Floer data of the knot with the Heegaard Floer homology of the manifolds obtained by surgery on it. Given a knot K in S^3 there is a doubly filtered complex $C = \mathrm{CFK}^\infty(K)$ associated to it. The doubly filtered homotopy type of this complex is a knot invariant, from which all the flavours of knot Floer homology are derived.

In fact, the mapping cone formula states that given C and a certain chain homotopy equivalence which identifies $C\{i \geq 0\}$ with $C\{j \geq 0\}$ we can determine $\mathrm{HF}^+(S^3_{p/q}(K))$ completely for any rational p/q.

In Section 3 we explicitly describe $\mathrm{HF}^+(S^3_{p/q}(K))$ as an absolutely graded vector space in terms of homological data from $\mathrm{CFK}^\infty(K)$, with no reference to the chain homotopy equivalence mentioned above. For a large part this has already been done (see Ni and Zhang [12], Ni and Wu [11] and Ozsváth and Szabó [23]), but the results are scattered across multiple papers, sometimes not in explicit form, and we consider it useful to have them collected in one place. While all the results of this section concerning positive surgeries have been shown before, as far as we are aware, the results for negative and zero surgeries (contained in Sections 3.2 and 3.3, respectively) are new.

This allows us to derive some other applications as well, a few of which we mention here.

Theorem 7 *Suppose* K *is a nontrivial knot and* $Y = S^3_{p/q}(K)$*. Then*

$$|q| \leq |H_1(Y)| + \dim \mathrm{HF}_{\mathrm{red}}(Y).$$

The existence of such a bound for the denominator seems to be known to some experts in Heegaard Floer homology (it could be deduced from [23, Proposition 9.6]) but we have not seen it explicitly stated. We will use this fact in the proof of Theorem 2.

A bound on the number of slopes that produce a given manifold has also been obtained by Lackenby [6, Theorem 2.9] in quite a general but somewhat different setting (in particular, due to homological conditions, it does not deal with surgeries in S^3).

Theorem 8 *Let K be an L–space knot and $p/q \leq 1$ a rational number. Then $S^3_{p/q}(K)$ and p/q determine the Alexander polynomial of K.*

We remark that the Alexander polynomial determines the knot Floer homology of an L–space knot; see Ozsváth and Szabó [21]. Conversely, the Alexander polynomial of an L–space knot determines the HF^+ of all surgeries on it. In particular, if for two L–space knots K and K' we have that $\mathrm{HF}^+(S^3_{p/q}(K)) \cong \mathrm{HF}^+(S^3_{p/q}(K'))$ for some $p/q \leq 1$, then we also have $\mathrm{HF}^+(S^3_{p'/q'}(K)) \cong \mathrm{HF}^+(S^3_{p'/q'}(K'))$ for all other p'/q'. This theorem also implies that if two torus knots $T_{r,s}$, $T_{r',s'}$ with $r, s, r', s' > 0$ have the same surgery with the same slope ≤ 1, then they are the same. Presumably, this can also be obtained by more elementary methods. Note, however, that there do exist positive integral slopes for which there are two distinct torus knots with the same surgeries at these slopes; see Ni and Zhang [12, Example 1.1].

Teragaito [27] constructs a small Seifert fibred space Y and a sequence of knots $K_n \subset S^3$ such that $K_n(-4) = Y$.[2] Moreover, the genus of the knots K_n is unbounded. Incidentally, this shows that we cannot hope for an upper bound on knot genus for knots giving some arbitrary manifold by surgery. In Section 7 we show that Y can only be obtained by (-4)–surgery and we find the Alexander polynomial of the knots K_n. It is, in fact, possible to find the Heegaard Floer homology of all manifolds obtained by surgery on each K_n.

The organisation of this paper is as follows. In Section 2 we review some definitions, notation and the mapping cone formula. In Section 3 we derive the expression for the Heegaard Floer homology of surgeries on a knot. In Section 4 we prove Theorem 2, in Section 5 we prove Theorem 3 and in Section 6 we prove Theorem 5. Finally, in Section 7 we present some other applications of the mapping cone formula to questions in Dehn surgery.

Acknowledgements I would like to thank Tye Lidman, Jake Rasmussen, András Juhász, Marc Lackenby and Duncan McCoy for their very valuable suggestions and comments on the earlier drafts of this paper (Tye Lidman, in particular, suggested that

[2] In fact, Teragaito constructs $-Y$, his knots are the mirror images of K_n and the slope he uses is 4. It is more convenient for us to work with this orientation.

Theorem 2 could be proven using techniques of this paper). I am also very grateful to Marco Marengon, Tom Hockenhull and V S Pyasetkii for many important comments on the structure and presentation of this paper. This paper greatly benefited from a visit to the University of Texas at Austin, and many interesting and enlightening conversations that I had there. For this opportunity, I am very thankful to the Doris Chen Award, and the help of my supervisor Dorothy Buck. I am particularly grateful to Dorothy for her continued encouragement and support over the course of my PhD studies. Finally, I would like to thank the referee for helpful comments on the drafts of this paper.

2 The mapping cone formula

In this section, we set up notation and review the rational surgery formula of Ozsváth and Szabó [23]. We largely follow the exposition and notation of Ni and Wu in [11].

Given a knot K in S^3 we can associate to it a doubly filtered complex $C = \mathrm{CFK}^\infty(K)$. We denote generators of this complex by $[\boldsymbol{x}, i, j]$, where this generator has filtration $(i, j) \in \mathbb{Z} \times \mathbb{Z}$. By [24, Lemma 4.5] the complex C is homotopy equivalent (as a filtered complex) to a complex for which all filtration-preserving differentials are trivial. In other words, at each filtration level we replace the group, viewed as a chain complex with the filtration preserving differential, by its homology. From now on we work with this *reduced* complex.

The complex C is invariant under the shift by the vector $(-1, -1)$. There is an action of a formal variable U on C which is simply the translation by the vector $(-1, -1)$. In other words, the group at the filtration level (i, j) is the same as the one at the filtration level $(i-1, j-1)$ and U is the identity map from the first one to the second. Of course, U is a chain map. In C the map U is invertible (but note that it will not be in various subcomplexes and quotients), so C is an $\mathbb{F}[U, U^{-1}]$–module.

This means that as an $\mathbb{F}[U, U^{-1}]$–module C is generated by the elements with the first filtration level $i = 0$. In the reduced complex the group at filtration level $(0, j)$ is denoted $\widehat{\mathrm{HFK}}(K, j)$ and is known as the knot Floer homology of K (at Alexander grading j).

The complex C is absolutely $\mathbb{Z}$–graded. In fact, the complex C is the complex used to compute the (∞–flavour of the) Heegaard Floer homology of S^3, and the knot provides an additional filtration for it. By grading the Heegaard Floer homology of S^3 we obtain the grading on C. The map U decreases this grading by 2.

Using the filtration on C we can define the following quotients of it:

$$A_k^+(K) = C\{i \geq 0 \text{ or } j \geq k\}, \quad k \in \mathbb{Z},$$
$$B^+ = C\{i \geq 0\} \cong \mathrm{CF}^+(S^3).$$

We also define two chain maps $v_k, h_k\colon A_k^+(K) \to B^+$. The first one is just the projection (ie it sends to zero all generators with $i < 0$ and acts as the identity map for everything else). The second one is the composition of three maps: firstly we project to $C\{j \geq k\}$, then we multiply by U^k (this shifts everything by the vector $(-k, -k)$) and finally we apply a chain homotopy equivalence that identifies $C\{j \geq 0\}$ with $C\{i \geq 0\}$. Such a chain homotopy equivalence exists because the two complexes both represent $\mathrm{CF}^+(S^3)$ and by general theory [17] there is a chain homotopy equivalence between them, induced by the moves between the Heegaard diagrams. We usually do not know the explicit form of this chain homotopy equivalence.

Genus detection, alluded to before, has the following form:

Theorem 9 (Ozsváth and Szabó [16, Theorem 1.2]) *Let $K \subset S^3$ be a knot.*

Then $g(K) = \max\{j \in \mathbb{Z} \mid \widehat{\mathrm{HFK}}(K, j) \neq 0\}$.

From this (together with symmetries of C) we can see that the maps v_k (resp. h_k) are isomorphisms if $k \geq g$ (resp. $k \leq -g$).

We define chain complexes

$$\mathcal{A}_{i,p/q}^+(K) = \bigoplus_{n \in \mathbb{Z}} (n, A_{\lfloor (i+pn)/q \rfloor}^+(K)), \quad \mathcal{B}^+ = \bigoplus_{n \in \mathbb{Z}} (n, B^+).$$

The first entry in the brackets here is simply a label used to distinguish different copies of the same group. There is a chain map $D_{i,p/q}^+$ from $\mathcal{A}_{i,p/q}^+(K)$ to $\mathcal{B}^+$ defined by taking sums of all maps v_k, h_k with appropriate domains and requiring that the map v_k goes to the group with the same label n and h_k increases the label by 1. Explicitly, $D_{i,p/q}^+(\{(k, a_k)\}_{k \in \mathbb{Z}}) = \{(k, b_k)\}_{k \in \mathbb{Z}}$, where

$$b_k = v_{\lfloor (i+pk)/q \rfloor}^+(a_k) + h_{\lfloor (i+p(k-1))/q \rfloor}^+(a_{k-1}).$$

Each of $A_k^+(K)$ and B^+ inherits a relative $\mathbb{Z}$–grading from the one on C. Let $\mathbb{X}_{i,p/q}^+$ denote the mapping cone of $D_{i,p/q}^+$. We fix a relative $\mathbb{Z}$–grading on the whole of it by requiring that the maps v_k, h_k (and so $D_{i,p/q}^+$) decrease it by 1. The following is proven in [23].

Theorem 10 (Ozsváth and Szabó [23, Theorem 1.1]) *There is a relatively graded isomorphism of $\mathbb{F}[U]$–modules*

$$H_*(\mathbb{X}_{i,p/q}^+) \cong \mathrm{HF}^+(S_{p/q}^3(K), i).$$

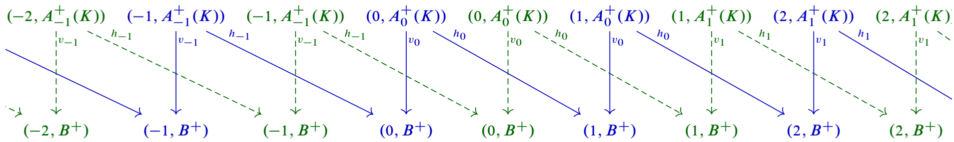


Figure 1: Schematic representation of the portion of the complex whose mapping cone gives the Heegaard Floer homology of the surgery on a knot. This case illustrates $\frac{2}{3}$–surgery; the blue (solid) and green (dashed) subcomplexes represent two different Spin^c–structures on the resulting space.

The index i in $\mathrm{HF}^+(S^3_{p/q}(K), i)$ stands for the Spin^c–structure. The numbering of Spin^c–structures we refer to is defined in [23], but we do not need precise details of how to obtain this numbering for our purposes.

We can also determine the absolute grading on the mapping cone. The group $\mathcal{B}^+$ is independent of the knot. Now if we insist that the absolute grading on the mapping cone for the unknot should coincide with the grading of HF^+ of the surgery on it (ie $d(L(p,q),i)$), this fixes the grading on $\mathcal{B}^+$. We then use this grading to fix the grading on $\mathbb{X}^+_{i,p/q}$ for arbitrary knots; this grading then is the correct grading, ie it coincides with the one HF^+ should have.

The mental picture we have of the mapping cone theorem is illustrated in Figure 1. We have two rows of groups. The bottom row is just the row of identical groups B^+. The upper row consists of the various "hook" groups $A_k^+(K)$. Specifically, if the surgery slope is p/q, in the upper row we repeat each group q times. We then have vertical arrows pointing down for the maps v_k, and the arrows for the maps h_k are slanted. More precisely, they go p groups to the right (if p is negative, this means $-p$ to the left). This creates $|p|$ subcomplexes, connected by a zig-zag set of arrows. Each such zig-zag subcomplex corresponds to a Spin^c–structure on the manifold that is the result of the surgery. To obtain the Heegaard Floer homology of this manifold we need to take the mapping cone of this chain map.

For our purposes, it suffices to pass to the homology of the mapping cone under consideration. Let

$$\boldsymbol{A}_k^+(K) = H_*(A_k^+(K)), \qquad \boldsymbol{B}^+ = H_*(B^+),$$
$$\mathbb{A}^+_{i,p/q}(K) = H_*(\mathcal{A}^+_{i,p/q}(K)), \quad \mathbb{B}^+ = H_*(\mathcal{B}^+)$$

and let $\boldsymbol{v}_k, \boldsymbol{h}_k, \boldsymbol{D}^+_{i,p/q}$ denote the maps induced by $v_k, h_k, D^+_{i,p/q}$, respectively, in homology.

When we talk about $\mathbb{A}^+_{i,p/q}(K)$ as an absolutely graded group, we mean the grading that it inherits from the absolute grading of the mapping cone that we described above.

Since $B^+ \cong \mathrm{CF}^+(S^3)$, we have $\boldsymbol{B}^+ \cong \mathcal{T}_d^+$, where $\mathcal{T}_d^+ \cong \mathbb{F}[U^{-1}] = \mathbb{F}[U, U^{-1}]/U\mathbb{F}[U]$, d signifies the grading of 1 and multiplication by U decreases the grading by 2. We sometimes call this module the *tower*. When we are not interested in the absolute grading we omit the subscript.

Recall that the short exact sequence

$$0 \longrightarrow \mathcal{B}^+ \xrightarrow{i} \mathbb{X}^+_{i,p/q} \xrightarrow{j} \mathcal{A}^+_{i,p/q}(K) \longrightarrow 0$$

induces the exact triangle

$$\begin{tikzcd} \mathbb{A}^+_{i,p/q}(K) \arrow[rr, "\boldsymbol{D}^+_{i,p/q}"] & & \mathbb{B}^+ \arrow[d, "i_*"] \\ & & H_*(\mathbb{X}^+_{i,p/q}) \cong \mathrm{HF}^+(S^3_{p/q}(K), i) \arrow[ull, "j_*"] \end{tikzcd} \tag{1}$$

All maps in these sequences are U–equivariant. This triangle is the main tool in the calculations of the next section. In particular, if the surgery slope is positive, then the map $\boldsymbol{D}^+_{i,p/q}$ will be surjective, so the triangle above implies that $\mathrm{HF}^+(S^3_{p/q}(K), i) \cong \ker \boldsymbol{D}^+_{i,p/q}$.

3 Calculations

In this section we want to use the mapping cone formula to calculate the Heegaard Floer homology for the results of surgery on a knot in S^3. Given a rational homology sphere Y and a Spin^c–structure $\mathfrak{s}$, we have $\mathrm{HF}^+(Y, \mathfrak{s}) = \mathcal{T}_d^+ \oplus \mathrm{HF}_{\mathrm{red}}(Y, \mathfrak{s})$, where $d = d(Y, \mathfrak{s})$ is called the *correction term* and $\mathrm{HF}_{\mathrm{red}}(Y, \mathfrak{s})$ is a finite-dimensional $\mathbb{F}[U]$–module annihilated by a big enough power of U, called the reduced Floer homology of Y in Spin^c–structure $\mathfrak{s}$. The sum of these groups over all Spin^c–structures is called the reduced Floer homology of Y and is denoted by $\mathrm{HF}_{\mathrm{red}}(Y)$.

We state a weaker version of [22, Theorem 2.3].

Theorem 11 (Ozsváth and Szabó) *There is an integer N such that for all $m \geq N$ and all $i \in \mathbb{Z}/m\mathbb{Z}$ there is an isomorphism of relatively graded $\mathbb{F}[U]$–modules*

$$A_k^+(K) \cong \mathrm{HF}^+(K_m, i),$$

where $k \equiv i \pmod m$ and $|k| \leq \frac{1}{2}m$.

In particular, each $A_k^+(K)$ is an HF^+ of a rational homology sphere in a certain Spin^c–structure, hence by the previous paragraph we can decompose it as $A_k^+(K) \cong A_k^T(K) \oplus A_k^{\mathrm{red}}(K)$, where $A_k^{\mathrm{red}}(K)$ is a finite-dimensional vector space in the kernel of some power of U and $A_k^T(K) \cong \mathcal{T}^+$.

We will need to talk about the Euler characteristic of the groups $A_k^{\mathrm{red}}(K)$, so we need to fix an absolute $\mathbb{Z}/2\mathbb{Z}$–grading for them. We do so by requiring that for the purposes of this grading each group $A_k^T(K)$ lies entirely in grading 0 and then using the relative $\mathbb{Z}/2\mathbb{Z}$–grading (induced by the parity of the relative $\mathbb{Z}$–grading) on $A_k^+(K)$.

A rational homology sphere Y is called an *L–space* if $\mathrm{HF}_{\mathrm{red}}(Y,\mathfrak{s}) = 0$ for all Spin^c–structures $\mathfrak{s}$. A knot $K \subset S^3$ is called an *L–space knot* if some positive surgery on it is an L–space. In fact, it is known that a p/q surgery on an L–space knot is an L–space if and only if $p/q \geq 2g(K)-1$ (here $g(K)$ is, as usual, the genus of K). In particular, all large surgeries on L–space knots are L–spaces, hence for any L–space knot K we have $A_k^{\mathrm{red}}(K) = 0$ for all k.

In the same way we can decompose the complexes of the exact triangle (1):

$$\mathbb{A}^T_{i,p/q}(K) = \bigoplus_{n\in\mathbb{Z}} (n, A^T_{\lfloor (i+pn)/q \rfloor}(K)), \quad \mathbb{A}^{\mathrm{red}}_{i,p/q}(K) = \bigoplus_{n\in\mathbb{Z}} (n, A^{\mathrm{red}}_{\lfloor (i+pn)/q \rfloor}(K)).$$

We can also decompose the map $\boldsymbol{D}^+_{i,p/q} = \boldsymbol{D}^T_{i,p/q} \oplus \boldsymbol{D}^{\mathrm{red}}_{i,p/q}$, where the first map is the restriction of $\boldsymbol{D}^+_{i,p/q}$ to $\mathbb{A}^T_{i,p/q}(K)$ and the second one is the restriction to $\mathbb{A}^{\mathrm{red}}_{i,p/q}(K)$. Note that $\boldsymbol{D}^+_{i,p/q} = \boldsymbol{D}^T_{i,p/q}$ for L–space knots.

Now the restrictions of $\boldsymbol{v}_k$ and $\boldsymbol{h}_k$ to $A_k^T(K)$ are multiplications by some powers of U, which we denote by V_k and H_k, respectively. (This is because at large gradings these maps are isomorphisms.) The following are some useful properties of V_k and H_k (see [11]):

- $V_k \geq V_{k+1}$ for any $k \in \mathbb{Z}$;
- $H_k \leq H_{k+1}$ for any $k \in \mathbb{Z}$;
- $V_k = H_{-k}$ for any $k \in \mathbb{Z}$;
- $V_k \to +\infty$ as $k \to -\infty$;
- $H_k \to +\infty$ as $k \to +\infty$;
- $V_k = 0$ for $k \geq g(K)$;
- $H_k = 0$ for $k \leq -g(K)$.

In other words, the V_k form a nonincreasing unbounded sequence of nonnegative numbers which become zero at $g(K)$, and $H_k = V_{-k}$.

We now separate into three different cases. Firstly, we cover the case of positive surgery slope. Secondly, we treat negative surgeries. The third case is the zero surgery.

3.1 Positive surgeries

The next lemma is used to establish that $\boldsymbol{D}^T_{i,p/q}$ is surjective when $p/q > 0$. We state it in a slightly more general form because in this form it also applies to other manifolds.

Lemma 12 *Let $X = Y = \bigoplus_{i\in\mathbb{Z}}(i,\mathcal{T}^+)$, $X' = \bigoplus_{i\neq 0}(i,\mathcal{T}^+)$ and let the maps*

$$\alpha_i\colon (i,\mathcal{T}^+) \to (i,\mathcal{T}^+), \quad \beta_i\colon (i,\mathcal{T}^+) \to (i+1,\mathcal{T}^+)$$

be multiplications by U^{a_i} and U^{b_i}, respectively. Suppose further that

- *there is a number N such that $a_i = 0$ for $i \geq N$ and $b_i = 0$ for $i \leq -N$, and*
- *$a_i \to +\infty$ as $i \to -\infty$ and $b_i \to +\infty$ as $i \to +\infty$.*

Define D to be the sum of the maps α_i and β_i. Then the restriction of D to X' is surjective.

The setting here is very similar to the one described by Figure 1, only we choose one of the zig-zag complexes and all the groups in both the top and the bottom row are the towers; see Figure 2.

Proof This is essentially what Ni and Wu prove in [11, Lemma 2.8]. We will show that, for any $n \geq 0$ and $i \leq 0$, (i, U^{-n}) is in the image of the restriction of D to X'. The conclusion will then follow by symmetry and linearity.

We clearly have $(i,U^{-n}) = \beta_{i-1}(i-1, U^{-n-b_{i-1}})$. Define $\xi = \{(i,\xi_i)\}_{i\in\mathbb{Z}} \in X'$ recursively by

$$\xi_s = \begin{cases} 0 & \text{if } s \geq i, \\ U^{-n-b_{i-1}} & \text{if } s = i-1, \\ (-1)^{s-i+1}U^{a_{s+1}-b_s}\cdot\xi_{s+1} & \text{otherwise.} \end{cases}$$

In a way, after we set that $\xi_s = 0$ for $s \geq i$, this is the only possible definition (up to the kernel of D). This is because the arrow "slanted to the right" has to be used to cancel the rightmost element in the lower row, hence we know what element in its codomain we have to choose so that it indeed cancels. This tells us what the image of the "vertical" arrow is and hence what the next "slanted" arrow has to cancel, etc.

Since we have $a_{s+1} - b_s \to +\infty$ as $s \to -\infty$, ξ only has a finite number of nonzero coordinates and hence is a well-defined element of X'. It is also easy to see that $D(\xi) = (i, U^{-n})$. □

$$\begin{tikzcd}[column sep=small]
(-2,\mathcal{T}^+) \arrow[dr, "\times U^{b_{-2}}"] & & (-1,\mathcal{T}^+) \arrow[dl, "\times U^{a_{-1}}"'] \arrow[dr, "\times U^{b_{-1}}"] & & (0,\mathcal{T}^+) \arrow[dl, "\times U^{a_0}"'] \arrow[dr, "\times U^{b_0}"] & & (1,\mathcal{T}^+) \arrow[dl, "\times U^{a_1}"'] \arrow[dr, "\times U^{b_1}"] & & (2,\mathcal{T}^+) \arrow[dl, "\times U^{a_2}"'] \\
& (-1,\mathcal{T}^+) & & (0,\mathcal{T}^+) & & (1,\mathcal{T}^+) & & (2,\mathcal{T}^+) &
\end{tikzcd}$$

Figure 2: Maps and groups of Lemmas 12 and 13

Let $\tau_d(N)$ be a submodule of $\mathcal{T}_d^+$ generated by $\{U^{-n}\}_{0\le n\le N-1}$. As before, we omit the subscript in the absence of the absolute grading.

The setting of the next lemma is less general; indeed, we use more information about the numbers V_k and H_k.

Lemma 13 *To the assumptions of Lemma 12 add the following:*

- *(a_i) is a nonincreasing sequence;*
- *(b_i) is a nondecreasing sequence;*
- *$a_i \le b_i$ for $i \ge 0$;*
- *$a_i \ge b_i$ for $i < 0$.*

Put absolute gradings on X and Y by the rule that the maps α_i and β_i decrease it by 1, the multiplication by U decreases it by 2 and $1 \in (0,\mathcal{T}^+) \subset Y$ has grading $d-1$, where d is some rational number.

Then if $a_0 \ge b_{-1}$ we have

$$\ker D \cong \mathcal{T}^+_{d-2a_0} \oplus \bigoplus_{n\ge 1} \tau_{d_n^-}(b_{-n}) \oplus \bigoplus_{n\ge 1} \tau_{d_n^+}(a_n).$$

Otherwise

$$\ker D \cong \mathcal{T}^+_{d-2b_{-1}} \oplus \bigoplus_{n\ge 2} \tau_{d_n^-}(b_{-n}) \oplus \bigoplus_{n\ge 0} \tau_{d_n^+}(a_n).$$

The isomorphisms are as absolutely graded $\mathbb{F}[U]$–modules. The numbers $d_n^\pm$ are defined by

$$\begin{aligned}
d_0^\pm &= d - 2\max\{a_0, b_{-1}\},\\
d_{n+1}^- &= d_n^- + 2(a_{-n} - b_{-(n+1)}),\\
d_{n+1}^+ &= d_n^+ + 2(b_n - a_{n+1}).
\end{aligned}$$

Proof The two cases are completely analogous, so we will assume $a_0 \ge b_{-1}$. First, following [11, proof of Proposition 1.6] we define $\rho^T\colon \mathcal{T}^+_{d-2a_0} \to \ker D$ as follows. If we write $\rho^T(\eta) = \{(s,\xi_s)\}_{s\in\mathbb{Z}}$, we set $\xi_0 = \eta$ and determine the other components by

$$\xi_s = \begin{cases} -U^{b_{s-1}-a_s}\xi_{s-1} & \text{if } s > 0,\\ -U^{a_{s+1}-b_s}\xi_{s+1} & \text{if } s < 0.\end{cases}$$

In effect, we want to simply send the tower to the tower in the 0-component of the upper group. But it is not in the kernel of D, so we need to correct for that. In fact, we also want the map to be an $\mathbb{F}[U]$–module homomorphism, which is the reason for considering the cases $a_0 \geq b_{-1}$ and $a_0 < b_{-1}$ separately.

Notice that we always multiply by a nonnegative power of U: if $s > 0$, then $b_{s-1} \geq a_{s-1} \geq a_s$; if $s = -1$, this is the assumption $a_0 \geq b_{-1}$; if $s < -1$, then $a_{s+1} \geq b_{s+1} \geq b_s$. Thus the map is indeed an $\mathbb{F}[U]$–module homomorphism.

As before, $\xi_s = 0$ if $|s|$ is very big, so the map is well-defined. The map ρ^T is one-to-one because its 0-component is (ie $\xi_0 = \eta$). It is also graded correctly (ie the map ρ^T sends homogeneous elements of absolute grading d to homogeneous elements of grading d) because $(0, U^{-a_0}) \in X$ is sent to $(0, 1) \in Y$ by α_0, which has grading $d-1$. Thus $(0, 1) \in X$ has grading $d - 2a_0$, since to descend from $(0, U^{-a_0}) \in X$ to $(0, 1) \in X$ we need to multiply by U^{a_0} and multiplication by U has grading -2.

We have identified the tower in the kernel. Now we need to deal with the rest of it. Below we prove that the rest of the kernel consist of the kernels of the maps $\alpha_i + \beta_i$ for each i, except the one at which the tower is situated (ie $i = 0$). It is easy to see that the kernel of $\alpha_i + \beta_i$ is isomorphic to $\tau(\min(a_i, b_i))$.

If $\nu = \{(s, \nu_s)\}_{s\in\mathbb{Z}} \in \ker D$, by subtracting elements in the image of ρ^T we may assume that $\nu \in X'$, ie $\nu_0 = 0$. Without loss of generality, there exists $s < 0$ such that $\nu_s \neq 0$. To finish the proof we need to show that $U^{b_s} \cdot \nu_s = 0$ (recall that in this range $b_s \leq a_s$). Suppose this is not so and $0 \neq U^{b_s} \cdot \nu_s$. Since ν is in the kernel, it has to be cancelled by something. It follows that we must have $\beta_s(\nu_s) + \alpha_{s+1}(\nu_{s+1}) = 0$. Thus $0 \neq U^{b_s} \cdot \nu_s = -U^{a_{s+1}} \nu_{s+1}$ implies that $0 \neq U^{b_{s+1}} \nu_{s+1}$, as $a_{s+1} \geq b_{s+1}$ if $s < -1$. By proceeding in this way it follows that $\nu_0 \neq 0$, ie $\nu \notin X'$, a contradiction. □

The two lemmas above can be readily translated into results about surgery. The d–invariant formula (2) from the corollary below is [11, Proposition 1.6].

Corollary 14 *If $p/q > 0$ the map $\boldsymbol{D}^T_{i,p/q}$ is surjective. It follows that so is $\boldsymbol{D}^+_{i,p/q}$, and we conclude that $\mathrm{HF}^+(S^3_{p/q}(K), i) \cong \ker \boldsymbol{D}^+_{i,p/q}$.*

If $\lfloor i/q \rfloor \leq -\lfloor (i-p)/q \rfloor$, then

$$\ker \boldsymbol{D}^T_{i,p/q} \cong \mathcal{T}^+_d \oplus \bigoplus_{n\geq 1} \tau_{d_n^-}(H_{\lfloor (i-np)/q \rfloor}) \oplus \bigoplus_{n\geq 1} \tau_{d_n^+}(V_{\lfloor (i+np)/q \rfloor}).$$

Otherwise

$$\ker \boldsymbol{D}^T_{i,p/q} \cong \mathcal{T}^+_d \oplus \bigoplus_{n\geq 2} \tau_{d_n^-}(H_{\lfloor (i-np)/q \rfloor}) \oplus \bigoplus_{n\geq 0} \tau_{d_n^+}(V_{\lfloor (i+np)/q \rfloor}).$$

Here

$$d = d(S^3_{p/q}(K), i) = d(L(p,q), i) - 2\max\{V_{\lfloor i/q \rfloor}, H_{\lfloor (i-p)/q \rfloor}\}, \tag{2}$$

and

$$d_n^- = d + 2\sum_{k=0}^{n-1}(V_{\lfloor (i-kp)/q \rfloor} - H_{\lfloor (i-(k+1)p)/q \rfloor}),$$

$$d_n^+ = d + 2\sum_{k=0}^{n-1}(H_{\lfloor (i+kp)/q \rfloor} - V_{\lfloor (i+(k+1)p)/q \rfloor}).$$

Proof This is a straightforward application of Theorem 10 and Lemmas 12 and 13 after renumbering of the groups and maps; objects numbered with $\lfloor (i+np)/q \rfloor$ correspond to the ones numbered with n in Lemmas 12 and 13.

To fix the grading, note that the grading of $\mathbb{B}^+$ does not depend on the knot, but only on the surgery slope. Thus to grade it we can take the unknot U. For the unknot we have $V_i = 0$ for $i \geq 0$ and $V_i = i$ for $i < 0$. Hence $0 = V_{\lfloor i/q \rfloor} \geq H_{\lfloor (i-p)/q \rfloor} = 0$, and by the same argument as we used for an arbitrary knot, the grading of 1 in $(0, \boldsymbol{A}^+_{\lfloor i/q \rfloor}(U))$ is the d–invariant of the surgery, which we know to be $d(L(p,q), i)$ in this case. Since $V_{\lfloor i/q \rfloor} = 0$, we find that the grading of 1 in $(0, \boldsymbol{B}^+)$ is $d(L(p,q), i) - 1$. This allows us to fix the d–invariants for all other knots.

We can fix $d_n^{\pm}$ by the fact that the maps $\boldsymbol{v}_k$ and $\boldsymbol{h}_k$ reduce it by 1 and the multiplication by U reduces it by 2. □

As we noted before, for L–space knots, $\boldsymbol{D}^+_{i,p/q} = \boldsymbol{D}^T_{i,p/q}$. Let K be a knot and $\Delta_K(T) = a_0 + \sum_i a_i(T^i + T^{-i})$ be its symmetrised Alexander polynomial, with normalisation convention $\Delta_K(1) = 1$. Define its *torsion coefficients* $t_i(K)$ for $i \geq 0$ by

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

Clearly, if we know all the torsion coefficients, we know the Alexander polynomial. For L–space knots, $V_k = t_k$ for $k \geq 0$ (this follows, for example, from [23, Theorem 1.2]), so Corollary 14 determines the Heegaard Floer homology of positive surgeries on an L–space knot in terms of its Alexander polynomial.

The next proposition expresses the Heegaard Floer homology of positive surgeries for arbitrary knots in terms of data from CFK^∞. This proposition is essentially [12, Proposition 3.5].

Proposition 15 *As absolutely graded vector spaces,*

$$\ker \boldsymbol{D}^+_{i,p/q} \cong \ker \boldsymbol{D}^T_{i,p/q} \oplus \mathbb{A}^{\mathrm{red}}_{i,p/q}(K).$$

Moreover, $\ker \boldsymbol{D}^T_{i,p/q}$ *is actually a submodule of* $\ker \boldsymbol{D}^+_{i,p/q}$.

Proof This is a straightforward exercise in linear algebra.

Given vector spaces U, V, W and linear maps $\rho_U\colon U \to W$ and $\rho_V\colon V \to W$ such that ρ_U is surjective, $\ker(\rho_U \oplus \rho_V) \cong \ker \rho_U \oplus V$.

There exists a map $\rho_U^*\colon W \to V$ such that $\rho_U \circ \rho_U^* = \mathrm{id}_W$. In the graded situation we can make ρ_U^* send homogeneous elements to homogeneous elements. Then we can define $T\colon \ker \rho_U \oplus V \to \ker(\rho_U \oplus \rho_V)$ by $T(x \oplus y) = (x - \rho_U^* \circ \rho_V(y)) \oplus y$. Since in our case $\rho_U \oplus \rho_V$ is graded, T is an isomorphism of graded vector spaces. □

Let

$$\mathbb{A}^s(K) = \bigoplus_{k\in\mathbb{Z}} A^{\mathrm{red}}_k(K).$$

This is a finite-dimensional vector space, as each $A^{\mathrm{red}}_k(K)$ is and $A^{\mathrm{red}}_k(K) = 0$ for $|k| \geq g(K)$. We define $\delta(K) = \dim \mathbb{A}^s(K)$. Note that $\delta(K) = 0$ if and only if K is an L–space knot. The following proposition generalises [11, Proposition 5.3]:

Proposition 16 [12, Corollary 3.6] *Let $K \subset S^3$ be a knot and $p/q > 0$. Then*

$$\text{(3)}\quad \dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = q\delta(K) + qV_0 + 2q\sum_{i=1}^{g-1} V_i - \sum_{i=0}^{p-1} \max(V_{\lfloor i/q\rfloor}, H_{\lfloor (i-p)/q\rfloor}).$$

Proof Since

$$\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = \sum_{i=0}^{p-1} \dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K), i),$$

combining Proposition 15 and Corollary 14 we see that

$$\begin{aligned}
&\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) \\
&\quad = \sum_{i\in\mathbb{Z}} \dim A^{\mathrm{red}}_{\lfloor i/q\rfloor}(K) + \sum_{i\geq 0} V_{\lfloor i/q\rfloor} + \sum_{i\geq 1} H_{\lfloor -i/q\rfloor} - \sum_{i=0}^{p-1} \max(V_{\lfloor i/q\rfloor}, H_{\lfloor (i-p)/q\rfloor}) \\
&\quad = q\sum_{k\in\mathbb{Z}} \dim A^{\mathrm{red}}_k(K) + q\sum_{i=0}^{g-1} V_i + q\sum_{i=-(g-1)}^{-1} H_i - \sum_{i=0}^{p-1} \max(V_{\lfloor i/q\rfloor}, H_{\lfloor (i-p)/q\rfloor}) \\
&\quad = q\delta(K) + qV_0 + 2q\sum_{i=1}^{g-1} V_i - \sum_{i=0}^{p-1} \max(V_{\lfloor i/q\rfloor}, H_{\lfloor (i-p)/q\rfloor}).
\end{aligned}$$

□

Now we are ready to prove Theorem 7.

Theorem 7 *Suppose K is a nontrivial knot and $Y = S^3_{p/q}(K)$. Then*

$$|q| \le |H_1(Y)| + \dim \mathrm{HF}_{\mathrm{red}}(Y).$$

Proof This is an easy consequence of Ni and Zhang's formula of Proposition 16 (by taking the mirror image we may assume $p/q > 0$). We have

$$\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + \sum_{i=0}^{p-1} \max(V_{\lfloor i/q \rfloor}, H_{\lfloor (i-p)/q \rfloor}) = q\delta(K) + qV_0 + 2q\sum_{i=1}^{g-1} V_i \ge q(\delta(K) + V_0).$$

Recall that $\delta(K) = \dim \mathbb{A}^s(K)$, so it is nonnegative and $\delta(K) = 0$ if and only if K is an L–space knot, in which case $V_k = 0$ if and only if $k \ge g(K)$, so for nontrivial L–space knots $V_0 \ne 0$. If $V_0 = 0$ then all V's (and H's) are zero and as $\delta(K) \ne 0$ by the previous sentence, we clearly get $q \le \dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$.

So suppose $V_0 \ne 0$. Then

$$\begin{aligned}\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + pV_0 &\ge \dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + \sum_{i=0}^{p-1} \max(V_{\lfloor i/q \rfloor}, H_{\lfloor (i-p)/q \rfloor}) \\ &\ge q(\delta(K) + V_0).\end{aligned}$$

Finally, we have

$$\begin{aligned} q &\le \frac{\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + pV_0}{\delta(K) + V_0} \\ &= \frac{\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))}{\delta(K) + V_0} + \frac{pV_0}{\delta(K) + V_0} \\ &\le \dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + p. \end{aligned}$$

□

3.2 Negative surgeries

In the case when $p/q < 0$ the map $\boldsymbol{D}^+_{i,p/q}$ is no longer surjective. However, we can show that the cokernel consists of exactly the tower part and the kernel is the reduced Floer homology $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K), i)$. We start with a general lemma, which is similar to Lemmas 12 and 13. The main difference is in that the β_i maps go to the groups labelled with a smaller index.

Lemma 17 *Let $X = Y = \bigoplus_{i\in\mathbb{Z}}(i, \mathcal{T}^+)$ and let the maps*

$$\alpha_i\colon (i,\mathcal{T}^+) \to (i,\mathcal{T}^+), \quad \beta_i\colon (i,\mathcal{T}^+) \to (i-1,\mathcal{T}^+)$$

be multiplications by U^{a_i} and U^{b_i}, respectively. Suppose further that a_i and b_i have the following properties:

- *There is a number N such that $a_i = 0$ for $i \geq N$ and $b_i = 0$ for $i \leq -N$;*
- *$a_i \to +\infty$ as $i \to -\infty$ and $b_i \to +\infty$ as $i \to +\infty$;*
- *$a_i \geq b_i$ for $i < 0$ and $a_i \leq b_i$ for $i \geq 0$.*

Then no element of $(-1, \mathcal{T}^+) \subset Y$ is in the image of D and $(-1, \mathcal{T}^+) \subset Y$ generates the cokernel of D. The kernel of D has the form

$$\ker D \cong \bigoplus_{i\in\mathbb{Z}} \tau(\min(a_i, b_i)).$$

Proof As all of the maps α_i, β_i are surjective, it is easy to see that the cokernel of D is generated by the (equivalence classes of) elements in any one of $(i, \mathcal{T}^+) \subset Y$. Suppose $\eta = \{(s, \eta_s)\}_{s\in\mathbb{Z}} = D(\xi)$ with $\eta_s = 0$ for $s \neq -1$. Let $\xi = \{(s, \xi_s)\}_{s\in\mathbb{Z}}$.

Without loss of generality (by symmetry) we may assume that $\alpha_{-1}(\xi_{-1}) \neq 0$. Since $a_{-1} \geq b_{-1}$ it follows that $\beta_{-1}(\xi_{-1}) \neq 0$. Since $\eta_{-2} = 0 = \beta_{-1}(\xi_{-1}) + \alpha_{-2}(\xi_{-2})$, we have $\alpha_{-2}(\xi_{-2}) \neq 0$, and hence $\xi_{-2} \neq 0$. Continuing in the same way we conclude that ξ is not supported on a finite set and hence no such ξ can exist.

Similarly to the proof of Lemma 13, we want to show that the kernel of D separates into the kernels of maps $\alpha_i + \beta_i$. This will finish the proof.

Now let $\xi = \{(s, \xi_s)\}_{s\in\mathbb{Z}} \in \ker D$. As before, without loss of generality we assume there is $n < 0$ such that $\beta_n(\xi_n) \neq 0$. Then $\alpha_{n-1}(\xi_{n-1}) \neq 0$, so $\beta_{n-1}(\xi_{n-1}) \neq 0$. Proceeding inductively we again reach a contradiction to ξ being finitely supported. □

The previous lemma describes the action of $\boldsymbol{D}^T_{i,p/q}$ when $p/q < 0$. We make this explicit in the next lemma.

Lemma 18 *Let $p < 0$, $q > 0$. Then*

$$\operatorname{coker} \boldsymbol{D}^T_{i,p/q} \cong \mathcal{T}^+_d,$$

where $d = d(L(p,q), i)$, and

$$\ker \boldsymbol{D}^T_{i,p/q} \cong \bigoplus_{n\geq 1} \tau_{d_n^-}(H_{\lfloor (i-np)/q\rfloor}) \bigoplus_{n\geq 0} \tau_{d_n^+}(V_{\lfloor (i+np)/q\rfloor}).$$

Here

$$d_0^+ = d + 1 - 2H_{\lfloor i/q \rfloor},$$

$$d_n^- = d_0^+ + 2\sum_{k=0}^{n-1}(V_{\lfloor (i-kp)/q \rfloor} - H_{\lfloor (i-(k+1)p)/q \rfloor}),$$

$$d_n^+ = d_0^+ + 2\sum_{k=0}^{n-1}(H_{\lfloor (i+kp)/q \rfloor} - V_{\lfloor (i+(k+1)p)/q \rfloor}).$$

Proof This is a straightforward application of Lemma 17. Objects that are labelled with $\lfloor (i+np)/q \rfloor$ in the mapping cone correspond to the ones labelled with $-n$ in Lemma 17. In particular, take $a_n = V_{\lfloor (i-np)/q \rfloor}$ and $b_n = H_{\lfloor (i-np)/q \rfloor}$. The grading comes from the fact that this works in the same way for the unknot (the towers in the cokernel coincide for all knots). Just as in Corollary 14, we get the values of $d_n^{\pm}$ by the fact that the maps $\boldsymbol{v}_k$, $\boldsymbol{h}_k$ have grading -1 and the multiplication by U has grading -2. □

Just as Corollary 14 is sufficient for positive surgeries on L–space knots, so is Lemma 18 for negative surgeries on L–space knots. We observe that in this case the Alexander polynomial also determines the Heegaard Floer homology of the surgeries. Lemma 18 also implies that negative p/q surgeries on L–space knots have the same d–invariants as the lens spaces $L(p,q)$, so do not depend on the particular L–space knot. The next proposition extends our analysis to arbitrary knots.

Proposition 19 *Let $p < 0$, $q > 0$. As absolutely graded $\mathbb{F}[U]$–modules, we have*

$$\operatorname{coker} \boldsymbol{D}^+_{i,p/q} \cong \mathcal{T}^+_d.$$

As absolutely graded vector spaces, we have

$$\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K), i) \cong \ker \boldsymbol{D}^+_{i,p/q} \cong \ker \boldsymbol{D}^T_{i,p/q} \oplus \mathcal{A},$$

where $\mathbb{A}^{\mathrm{red}}_{i,p/q}(K) \cong \mathcal{A} \oplus \tau_\delta(N_{i,p/q})$, $\delta = d(L(p,q),i)+1$, and $N_{i,p/q}$ is characterised by

$$d = d(S^3_{p/q}(K), i) = d(L(p,q), i) + 2N_{i,p/q}.$$

In fact, $N_{i,p/q} = \max\{\overline{V}_{\lfloor i/q \rfloor}, \overline{H}_{\lfloor (i+p)/q \rfloor}\}$, where $\overline{V}_k$, $\overline{H}_k$ are for the mirror image of K the same as V_k, H_k are for K.

Proof Recall that no element in $(-1, \boldsymbol{B}^+)$ is in the image of the map $\boldsymbol{D}^T_{i,p/q}$. Since $\mathbb{A}^{\mathrm{red}}_{i,p/q}(K)$ lies in the kernel of the multiplication by a big enough power of U, so does its image under $\boldsymbol{D}^+_{i,p/q}$. Hence $\boldsymbol{D}^+_{i,p/q}$ only "chops off" a finite piece of the tower.

More precisely, let N be the largest integer such that $U^{-N+1} \in (-1, \boldsymbol{B}^+)$ appears as a term of some element η in the image of $\boldsymbol{D}^+_{i,p/q}$.

We claim that then U^{-N+k} is also in the image for all $k \geq 1$. This is easily seen by an inductive argument: 1 is in the image, as $1 = U^{N-1}\eta$; U^{-1} is, because 1 is and $U^{N-2}\eta$ is. Proceeding in the same way, we establish the claim.

Thus the cokernel of $\boldsymbol{D}^+_{i,p/q}$ is generated by U^{-N-k} for $k \geq 0$, none of which are in its image. Thus the map i_* from the exact triangle (1) injects $\langle\{U^{-N-k}\}_{k\geq 0}\rangle_{\mathbb{F}}$ into $\mathrm{HF}^+(S^3_{p/q}(K), i)$. Since $U^{-N+1} \in (-1, \boldsymbol{B}^+)$ is in the image of $\boldsymbol{D}^+_{i,p/q}$, it is in the kernel of i_* and we have $U \cdot i_*(U^{-N}) = 0$. Hence the image of i_* is exactly the tower $\mathcal{T}^+_d$ with $d = d(S^3_{p/q}(K), i)$. By Lemma 18, $1 \in (-1, \boldsymbol{B}^+)$ has grading $d(L(p,q), i)$, so $d(S^3_{p/q}(K), i) = d(L(p,q), i) + 2N$.

By the first isomorphism theorem and exactness of (1), we have

$$\ker \boldsymbol{D}^+_{i,p/q} = \operatorname{im} j_* \cong \mathrm{HF}^+(S^3_{p/q}(K), i)/\ker j_* = \mathrm{HF}^+(S^3_{p/q}(K), i)/\operatorname{im} i_*.$$

Since $\operatorname{im} i_*$ is the tower, we have

$$\ker \boldsymbol{D}^+_{i,p/q} \cong \mathrm{HF}^+(S^3_{p/q}(K), i)/\operatorname{im} i_* \cong \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K), i).$$

The rest is just linear algebra again. We can split $\mathbb{A}^{\mathrm{red}}_{i,p/q}(K)$ into the part that goes isomorphically to the base of the tower, which is not in the image of $\boldsymbol{D}^T_{i,p/q}$ (ie $(-1, \boldsymbol{B}^+) \cap \operatorname{im} \boldsymbol{D}^+_{i,p/q}$) and the part that goes into the image of $\boldsymbol{D}^T_{i,p/q}$. We then proceed as in the proof of Proposition 15.

The fact that $N_{i,p/q} = \max\{\bar{V}_{\lfloor i/q \rfloor}, \bar{H}_{\lfloor (i-p)/q \rfloor}\}$ follows from taking the mirror image of K and comparing with the formula already obtained for the correction terms from Corollary 14. We have

$$\begin{aligned} 2N_{i,p/q} &= d(S^3_{p/q}(K), i) - d(L(p,q), i) \\ &= -d(S^3_{-p/q}(m(K)), i) + d(L(-p,q), i) \\ &= \max\{\bar{V}_{\lfloor i/q \rfloor}, \bar{H}_{\lfloor (i+p)/q \rfloor}\}, \end{aligned}$$

where $m(K)$ is the mirror image of K. □

We can also express the total rank of $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K), i)$ as follows.

Proposition 20 *We have*

$$\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = q\delta(K) + qV_0 + 2q\sum_{i=1}^{g-1} V_i - \sum_{i=0}^{p-1} N_{i,p/q}.$$

Proof The proof is virtually the same as for Proposition 16. □

3.3 Zero surgeries

We now treat the case of zero surgeries. For the case of L–space knots the formula for the Heegaard Floer homology of the zero surgery was derived in [14, Theorem 7.2]. The main tool we use is due to Ozsváth and Szabó:

Theorem 21 [19, Theorem 9.19] *There is a U–equivariant exact triangle*

$$\begin{tikzcd} \mathrm{HF}^+(S^3) \arrow[r, "F^+_{;i}"] & \displaystyle\bigoplus_{j\equiv i \pmod m} \mathrm{HF}^+(S^3_0(K), j) \arrow[d, "F^+_{0;i}"] \\ & \mathrm{HF}^+(S^3_m(K), i) \arrow[ul, "F^+_{m;i}"] \end{tikzcd} \tag{4}$$

Moreover, the map $F^+_{m;i}$ is equal to the one induced by the surgery cobordism.

Given i, we can make m in (4) so big that

$$\bigoplus_{j\equiv i \pmod m} \mathrm{HF}^+(S^3_0(K), j) = \mathrm{HF}^+(S^3_0(K), i).$$

From now on we assume that m is at least that large.

The group $A_0^+(K) \cong A_0^T(K) \oplus A_0^{\mathrm{red}}(K)$ is relatively $\mathbb{Z}$–graded. If we fix an absolute $\mathbb{Q}$–grading for any element of $A_0^+(K)$, the relative grading will fix the absolute grading for all the elements. In particular, it will absolutely grade $A_0^{\mathrm{red}}(K)$.

In the statement of the next proposition (but not necessarily in the proof), we use the grading of $A_0^{\mathrm{red}}(K)$ induced by grading the tower $A_0^T(K)$ in such a way that the grading of 1 is $\frac{1}{2} - 2V_0$.

Proposition 22 *Let $k \neq 0$. Then, as $\mathbb{Z}/2\mathbb{Z}$–graded vector spaces, we have*

$$\mathrm{HF}^+(S^3_0(K), k) \cong \tau(V_{|k|}) \oplus A_k^{\mathrm{red}}(K). \tag{5}$$

As absolutely $\mathbb{Q}$–graded vector spaces, we have

$$\mathrm{HF}^+(S^3_0(K), 0) \cong \mathcal{T}^+_{-1/2+2\bar{V}_0} \oplus \mathcal{T}^+_{1/2-2V_0} \oplus \mathcal{A}. \tag{6}$$

Here $\mathcal{A} \oplus \tau_{1/2}(\bar{V}_0) \cong A_0^{\mathrm{red}}(K)$ as absolutely graded vector spaces, where the absolute grading of $A_0^{\mathrm{red}}(K)$ is as described above.

Proof The first part is immediate from [14, proof of Theorem 7.2]. Note that $\mathrm{HF}^+(S^3_m(K), k) \cong \mathcal{T} \oplus A_k^{\mathrm{red}}(K)$ (recall that we are assuming that m is large). In the proof of [14, Theorem 7.2] Ozsváth and Szabó show that the restriction of $F^+_{m;i}$ to the tower part is surjective and its kernel is $\mathbb{F}[U^{-1}]/U^{-V_{|k|}}$. So we are done by the same elementary linear algebra as in the proof of Proposition 15.

For the second part, note that we can assign absolute gradings, as we are dealing with a torsion Spin^c–structure. As shown in [19, Theorem 10.4], $\mathrm{HF}^\infty(S^3_0(K),0)$ is a direct sum of two copies of $\mathbb{Z}[U,U^{-1}]$ that lie in different relative $\mathbb{Z}/2\mathbb{Z}$–gradings. This is equivalent to saying that the difference of the absolute gradings between the elements from the different summands is always odd. As in the case of rational homology spheres, the exact sequence

$$\cdots \to \mathrm{HF}^-(Y,\mathfrak{s}) \to \mathrm{HF}^\infty(Y,\mathfrak{s}) \to \mathrm{HF}^+(Y,\mathfrak{s}) \to \cdots$$

establishes that

$$\mathrm{HF}^+(S^3_0(K),0) \cong \mathcal{T}_{d_1} \oplus \mathcal{T}_{d_2} \oplus \mathcal{A},$$

where $\mathcal{A} = \mathrm{HF}_{\mathrm{red}}(S^3_0(K),0)$ is a finitely generated $\mathbb{F}[U]$–module in the kernel of some large enough power of U.

In fact, combining [14, Proposition 4.12] with the d–invariant formula of Ni and Wu stated in Corollary 14, we obtain $d_1 = -\frac{1}{2} + 2\overline{V}_0$ and $d_2 = \frac{1}{2} - 2V_0$.

The last step in the proof is determining $\mathcal{A}$. The maps $F^+_{;0}$ and $F^+_{0;0}$ from the exact triangle (4) have gradings $-\frac{1}{2}$ and $\frac{1}{4}(m-3)$, respectively, by [14, Lemma 7.11]. The map $F^+_{m;0}$ is not graded but is a sum of graded maps, and the set of grading shifts of these maps is $\{\frac{1}{4}(1-m(2k-1)^2)\}_{k\in\mathbb{Z}}$.

Since $\mathrm{HF}^+(S^3) \cong \mathcal{T}^+_0$ and the grading of the map $F^+_{;0}$ is $-\frac{1}{2}$, $\mathcal{T}^+_{1/2-2V_0}$ is not in the image of $F^+_{;0}$, hence the map $F^+_{0;0}$ is an isomorphism between $\mathcal{T}^+_{1/2-2V_0}$ and the tower part of $\mathrm{HF}^+(S^3_m(K),0)$, which is equal to $\mathcal{T}^+_{(m-1)/4-2V_0}$ by Proposition 15. Hence the restriction of the map $F^+_{m;0}$ to the tower part of $\mathrm{HF}^+(S^3_m(K),0)$ is zero. As in the proof of Proposition 19, the restriction of $F^+_{m;0}$ to $\mathrm{HF}_{\mathrm{red}}(S^3_m(K),0)$ maps a subgroup of the form $\tau(N)$ isomorphically to the base of the tower $\mathrm{HF}^+(S^3) \cong \mathcal{T}^+_0$. By the grading considerations again we see that $N = \overline{V}_0$.

Recall from Proposition 15 that $\mathrm{HF}^+(S^3_m(K),0) \cong \mathcal{T}^+_{(m-1)/4-2V_0} \oplus A^{\mathrm{red}}_0(K)$ (the grading here is such that the relative grading is as it should be). Let the maximal grading of a nontrivial element in $A^{\mathrm{red}}_0(K)$ be $\frac{1}{4}(m-1) - 2V_0 + C$.

Consider one homogeneous summand of $F^+_{m;0}$ with grading $\frac{1}{4}(1-m(2k-1)^2)$. It maps the element of $A^{\mathrm{red}}_0(K)$ of maximal grading to an element with grading

$$\tfrac{1}{4}(m-1) - 2V_0 + C + \tfrac{1}{4}(1-m(2k-1)^2) = \tfrac{1}{4}\big(m(1-(2k-1)^2) - 8V_0 + 4C\big).$$

If $k \neq 0, 1$ we have $1-(2k-1)^2 < 0$ and so by making m sufficiently large we can make sure that $\frac{1}{4}(m(1-(2k-1)^2) - 8V_0 + 4C) < 0$, and, as all nontrivial elements in the image have grading ≥ 0, this means that all components with $k \neq 0, 1$ are zero.

Thus we can assume that the map $F^+_{m;0}$ has grading $\frac{1}{4}(1-m)$. As discussed above the map $F^+_{m;0}$ maps a subgroup of $A^{\mathrm{red}}_0(K)$ of the form $\tau(\bar{V}_0)$ isomorphically to such a subgroup at the lower end of the tower $\mathrm{HF}^+(S^3) \cong \mathcal{T}^+_0$. Therefore 1 in $\tau(\bar{V}_0)$ must have grading $\frac{1}{4}(m-1)$.

The rest of $A^{\mathrm{red}}_0(K)$ will be in the kernel of $F^+_{m;0}$ and thus in the image of $\mathcal{A}$ by $F^+_{0;0}$. Now noting that the grading of the map $F^+_{0;0}$ is $\frac{1}{4}(m-3)$ finishes the proof. □

Torsion coefficients of the Alexander polynomial of a knot describe the Euler characteristics of the groups $A^{\mathrm{red}}_k(K)$, which we can see for example by combining Theorems 10.14 and 10.17 of [19] (though a more direct proof is also possible). This has also been shown in [12, Lemma 3.2].

Lemma 23 *For $k \geq 0$ we have*

$$t_k(K) = V_k + \chi(A^{\mathrm{red}}_k(K)). \tag{7}$$

Recall that the absolute $\mathbb{Z}/2\mathbb{Z}$ grading used to calculate the Euler characteristics here is fixed by the requirement that the tower $A^T_k(K)$ lies entirely in grading 0.

4 Proof of Theorem 2

In this section we prove:

Theorem 2 *Let Y be a 3–manifold. There are at most finitely many alternating knots $K \subset S^3$ such that $Y = S^3_{p/q}(K)$.*

The strategy of our proof is as follows. We first want to restrict the possible Alexander polynomials of knots that yield a given 3–manifold Y by surgery. We then want to show that, out of this restricted set, only finitely many can be Alexander polynomials of alternating knots. This will finish the proof, due to the next proposition, which was proved by Moore and Starkston. We provide the proof for the reader's convenience (and since it is nice and short).

Proposition 24 [9, Proposition 5.1] *There is only a finite number of alternating knots with a given Alexander polynomial.*

Proof By the Bankwitz theorem [1, Theorem 5.5] the determinant $\det(K)$ of an alternating knot K is greater than or equal to the minimal crossing number of K. Thus there are only finitely many alternating knots with a given determinant. The classical result [26, page 213] (or definition) $\det(K) = |\Delta_K(-1)|$ finishes the proof. $\square$

For a knot $K \subset S^3$, let $m(K)$ be its mirror image. Clearly, K is alternating if and only if $m(K)$ is. Since $S^3_{p/q}(K) = -S^3_{-p/q}(m(K))$ we can assume that the surgery slope is positive (if nonzero).

For Y a rational homology sphere and $q > 0$ a natural number, define

$$M(Y,q) = \frac{1}{2}\Bigg(\sum_{0\le i\le p-1} d(L(p,q),i) - \sum_{\mathfrak{s}\in \mathrm{Spin}^c(Y)} d(Y,\mathfrak{s})\Bigg),$$

where $p = |H_1(Y)|$.

Theorem 7 shows that for any rational homology sphere Y there is some number $n(Y)$ such that $Y \neq S^3_{p/q}(K)$ for any K and $|q| > n(Y)$.

If Y is obtained by $p/q > 0$ surgery on K, then by (2) the numbers V_k for K satisfy

$$M(Y,q) = \sum_{i=0}^{p-1} \max\{V_{\lfloor i/q\rfloor}, V_{-\lfloor (i-p)/q\rfloor}\}.$$

Combining this with Proposition 16 we get

$$\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + M(S^3_{p/q}(K),q) = q\Bigg(\delta(K) + V_0 + 2\sum_{i\ge 1} V_i\Bigg).$$

This formula implies the inequality

$$\frac{\dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) + M(S^3_{p/q}(K),q)}{q} \ge \sum_{k\ge 0}(V_k + \dim A_k^+(K)).$$

Now let

$$c(Y) = \max_{1\le q\le n(Y)} \left\{\frac{\dim \mathrm{HF}_{\mathrm{red}}(Y) + M(Y,q)}{q}\right\}.$$

The inequality above implies that if a rational homology sphere Y is obtained by surgery on a knot K with associated sequence $\{V_k\}_{k\ge 0}$, then

$$c(Y) \ge \sum_{k\ge 0}(V_k + \dim A_k^+(K)). \tag{8}$$

Lemma 25 *Suppose Y is a rational homology sphere obtained by a $p/q>0$ surgery on a knot $K\subset S^3$. Then*

$$\sum_{i\geq 0}|t_i(K)|\leq c(Y). \tag{9}$$

Proof It follows from Lemma 23 that for each $k\geq 0$ we have

$$|t_k(K)|=|V_k+\chi(A_k^+(K))|\leq V_k+|\chi(A_k^+(K))|\leq V_k+\dim A_k^+(K).$$

Combining with (8) yields the result. □

Let S_Y be some set of knots in S^3 that give a rational homology sphere Y by surgery (not necessarily all such knots and not necessarily alternating). Denote by $g(S_Y)$ and $\Delta(S_Y)$ the sets of genera and of Alexander polynomials, respectively, of knots in S_Y.

Lemma 26 *If $g(S_Y)$ is finite, then so is $\Delta(S_Y)$.*

Proof We clearly have $t_i(K)=0$ for all $K\in S_Y$ and all $i\geq\max(g(S_Y))$. By Lemma 25, $\sum_{i\geq 0}|t_i(K)|$ is bounded above, so we clearly have finitely many sequences $\{t_i(K)\}$ for $K\in S_Y$. Now observe that the torsion coefficients determine the Alexander polynomial, so this results in at most finitely many possible Alexander polynomials. □

A theorem of Murasugi is crucial for our proof:

Theorem 27 [10, Theorem 1.1] *Let $K\subset S^3$ be an alternating knot and*

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

be its Alexander polynomial. Then $a_i\neq 0$ for $0\leq i\leq g(K)$.

The next lemma is the last step before we can prove Theorem 2.

Lemma 28 *Let $K\subset S^3$ be an alternating knot that gives a rational homology sphere Y by surgery. Then*

$$g(K)\leq 3c(Y).$$

Proof Suppose $g(K)\geq 3c(Y)+1$. Note that $a_g=t_{g-1}(K)\neq 0$. We claim that there are three consecutive indices i, $i+1$ and $i+2\leq g$ with $t_i(K)=t_{i+1}(K)=t_{i+2}(K)=0$. It then follows that $a_{i+1}=0$, which is a contradiction to Theorem 27.

To prove the claim, suppose there is no such consecutive triple of zero torsion coefficients. Then

$$\sum_{i\geq 0}|t_i(K)| = \sum_{k\geq 0}(|t_{3k}(K)|+|t_{3k+1}(K)|+|t_{3k+2}(K)|) \geq \left\lfloor \tfrac{1}{3}(g-1)\right\rfloor + 1 \geq c(Y)+1,$$

which contradicts Lemma 25.

We have thus established that $g \leq 3c(Y)$. □

Proof of Theorem 2 Suppose Y is a rational homology sphere. Then by Lemma 28 there is a genus bound for alternating knots that give Y by surgery, so by Lemma 26 the set of Alexander polynomials of such alternating knots is finite.

If Y is obtained by 0–surgery on K, then Propositions 10.14 and 10.17 of [19] show that the Alexander polynomial of K can be deduced directly from the Heegaard Floer homology of Y.

Proposition 24 now finishes the proof. □

5 The genus bound

We now turn to the proof of Theorem 3, which we restate here.

Theorem 3 *For any knot $K \subset S^3$ and any $p/q \in \mathbb{Q}$ we have*

$$U^{g(K)+\lceil g_4(K)/2\rceil} \cdot \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = 0.$$

Lemma 29 *Let K be a knot in S^3 with genus g. Then for any $k \in \mathbb{Z}$ we have*

$$U^g \cdot A_k^{\mathrm{red}}(K) = 0.$$

Proof By the conjugation symmetry we may assume that $k \geq 0$. Let $C = \mathrm{CFK}^\infty(K)$, $\Delta_k = C\{i < 0 \text{ and } j \geq k\}$. This is a subquotient of C (ie a subcomplex of a quotient). Note that $U^g \cdot \Delta_k = 0$, as this is the maximal possible "height" of this complex. We illustrate the complexes Δ_k, $A_k^+(K)$, B^+ in Figure 3.

We have an exact sequence

$$0 \to \Delta_k \to A_k^+(K) \to B^+ \to 0$$

which leads to an exact U–equivariant triangle

(10)
$$\begin{tikzcd} H_*(\Delta_k) \arrow[r, "i_*"] & A_k^+(K) \arrow[d, "v_k"] \\ & B^+ \arrow[ul] \end{tikzcd}$$

Since $\boldsymbol{v}_k$ is surjective, we in fact have a short exact sequence

$$0 \to H_*(\Delta_k) \to \boldsymbol{A}_k^+(K) \to \boldsymbol{B}^+ \to 0,$$

so $H_*(\Delta_k) \cong \ker \boldsymbol{v}_k$ and hence $U^g \cdot \ker \boldsymbol{v}_k = 0$.

Recall that

$$\boldsymbol{A}_k^+(K) = \boldsymbol{A}_k^T(K) \oplus \boldsymbol{A}_k^{\mathrm{red}}(K),$$

and similarly we can decompose the map $\boldsymbol{v}_k = \boldsymbol{v}_k^T \oplus \boldsymbol{v}_k^{\mathrm{red}}$ into components. The map $\boldsymbol{v}_k^T$ is surjective. We claim that $\boldsymbol{A}_k^{\mathrm{red}}(K) \cong \ker \boldsymbol{v}_k / \ker \boldsymbol{v}_k^T$. From this the conclusion of the lemma follows immediately.

To prove the claim we construct an isomorphism from $\ker \boldsymbol{v}_k / \ker \boldsymbol{v}_k^T$ to $\boldsymbol{A}_k^{\mathrm{red}}(K)$. Let $x \in \ker \boldsymbol{v}_k \setminus \ker \boldsymbol{v}_k^T$. Then send an equivalence class of x to $\boldsymbol{A}_k^{\mathrm{red}}(K)$ by projection. This map is well defined, because two different elements with the same projection are in $\ker \boldsymbol{v}_k^T$. Clearly this is also a surjective $\mathbb{F}[U]$–module homomorphism. □

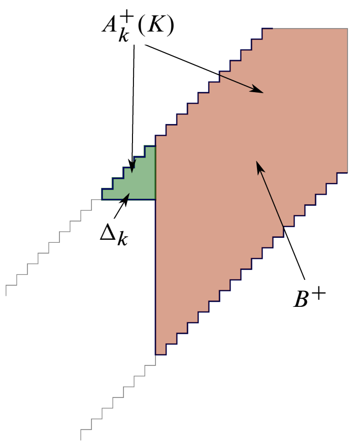


Figure 3: Complexes Δ_k, $A_k^+(K)$ and B^+ inside CFK^∞

The previous lemma clearly implies:

Corollary 30 *The following relation holds:*

$$U^g \cdot \mathbb{A}_{i,p/q}^{\mathrm{red}}(K) = 0.$$

Proof of Theorem 3 Note that by [25, Theorem 2.3] we have $V_0 \leq \left\lceil \frac{1}{2} g_4(K) \right\rceil$.

If the slope is negative the reduced part is exactly equal to the kernel of $\boldsymbol{D}_{i,p/q}^+$. So suppose $x \in \ker \boldsymbol{D}_{i,p/q}^+$. By Corollary 30, $U^g \cdot x \in \ker \boldsymbol{D}_{i,p/q}^T$. But by Lemma 18 the kernel of $\boldsymbol{D}^T_{i,p/q}$ consist of the summands of the type $\tau(N)$ with $N \le V_0 \le \lceil \frac{1}{2} g_4(K) \rceil$, so

$$U^{\lceil g_4(K)/2 \rceil} \cdot \ker \boldsymbol{D}^T_{i,p/q} = 0.$$

Now suppose the slope is positive. If we assume that $x \in \ker \boldsymbol{D}^+_{i,p/q}$ then we still have $U^g \cdot x \in \ker \boldsymbol{D}^T_{i,p/q}$, and $U^{\lceil g_4(K)/2 \rceil} \cdot \ker \boldsymbol{D}^T_{i,p/q}$ is contained in the tower part by Corollary 14.

Similarly, the case of zero surgery follows immediately from Proposition 22. This finishes the proof. □

Since by Corollary 14 and Lemma 18 the reduced Floer homology of surgeries on L–space knots consists only of a direct sum of $\mathbb{F}[U]$–modules of the form $\tau(V_k)$, we see that if K is an L–space knot, then $U^{\lceil g_4(K)/2 \rceil} \cdot \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = 0$.

In order to construct examples for which this genus bound gets arbitrarily large, note that every negative surgery on a knot contains a summand of the form $\tau(V_0)$. So if V_0 is large, the genus bound will also be large, independent of the absolute value of the negative slope we use. In particular, we can choose any order of the first homology we like.

For L–space knots, $V_0 = t_0$ can be read from the Alexander polynomial; in particular, this is true for torus knots $T_{p,q}$ with $p, q > 0$.

Suppose we have an L–space knot K with Alexander polynomial

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

Then the coefficients alternate between 1 and -1, with the first nontrivial coefficient being 1 [21, Theorem 1.2]. So we clearly have

$$t_0 \ge \#\{a_i = 1, i > 0\} \ge \tfrac{1}{4}(\#\{a_i \ne 0\} - 1) \ge \tfrac{1}{4}(\Delta_K(-1) - 1).$$

Consider the torus knots $T_{p,2}$ for p positive odd. They have Alexander polynomials of the form

$$\frac{(T^{2p} - 1)(T - 1)}{(T^p - 1)(T^2 - 1)} = T^{p-1} - T^{p-2} + \cdots + 1,$$

which evaluates to p at -1.

Moreover, these examples are actually negatively oriented (see next section) small Seifert fibred spaces, which is interesting in light of the next section.

We note that a result similar to Theorem 3 can be obtained for a knot in any L-space rational homology sphere, the bound being in terms of the width of the knot Floer homology rather than the genus.

6 Seifert fibred surgery

The aim of this section is to prove:

Theorem 5 *Let $K \subset S^3$ be a knot. Suppose there is a rational number $p/q > 0$ such that $Y = S^3_{p/q}(K)$ is a negatively oriented Seifert fibred space. Then*

- $U^{g(K)} \cdot \mathrm{HF}_{\mathrm{red}}(Y) = 0$*;*
- *if $0 < p/q \le 3$, then all the torsion coefficients $t_i(K)$ are nonpositive (including $t_0(K)$) and* $\deg \Delta_K = g(K)$*;*
- *more generally, if $i \ge \lfloor \frac{1}{2}(\lceil p/q \rceil - \sqrt{\lceil p/q \rceil}) \rfloor$, then t_i is nonpositive;*
- *if $g(K) > \lfloor \frac{1}{2}(\lceil p/q \rceil - \sqrt{\lceil p/q \rceil}) \rfloor$, then* $\deg \Delta_K = g(K)$*;*
- *if $U^{\lfloor |H_1(Y)|/2 \rfloor} \cdot \mathrm{HF}_{\mathrm{red}}(Y) \neq 0$, then* $\deg \Delta_K = g(K)$.

In all statements where $\deg \Delta_K = g(K)$ *we have that $\widehat{\mathrm{HFK}}(K, g(K))$ is supported in odd degrees.*

Proof First we need to define the Seifert orientation for Seifert fibred spaces. Following [20] we say that Y has *positive Seifert orientation* if $-Y$ bounds $W(\Gamma)$, where Γ is a weighted tree which has either negative definite or negative semidefinite intersection form. For the construction of the 4–manifold $W(\Gamma)$ from the weighted tree Γ, see [15]. We say that Y has *negative Seifert orientation* if $-Y$ has positive Seifert orientation.

Using [15, Corollary 1.4] (together with the inversion of the absolute $\mathbb{Z}/2\mathbb{Z}$–grading on the reduced homology upon reversing the orientation) we can see that if Y has a negative Seifert orientation, then its reduced Floer homology is concentrated in the odd $\mathbb{Z}/2\mathbb{Z}$–grading and that it bounds a negative definite 4–manifold with torsion-free first homology group.

Lemma 31 *Let $K \subset S^3$ be a knot. Suppose there is a rational number $p/q > 0$ such that $Y = S^3_{p/q}(K)$ is a negatively oriented Seifert fibred space. Then $A^{\mathrm{red}}_k(K)$ is supported in odd $\mathbb{Z}/2\mathbb{Z}$–grading for every k.*

Proof of Lemma 31 As an absolutely $\mathbb{Z}/2\mathbb{Z}$–graded group, each $A^{\mathrm{red}}_k(K)$ is a subgroup of $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ by Proposition 15. Since $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ is supported in odd grading, so must be each $A^{\mathrm{red}}_k(K)$. □

Denote by $\widetilde{g}$ the minimal index i for which $V_i = 0$. As above, denote by a_i the coefficient of the Alexander polynomial of K corresponding to T^i. If $\widetilde{g} < g(K)$, then by Lemma 23 we have

$$a_{g(K)} = t_{g(K)-1} = \chi(A^{\mathrm{red}}_{g(K)-1}), \tag{11}$$

so, in particular, $a_{g(K)} \neq 0$ if all $A_k^{\mathrm{red}}(K)$ are supported in the same $\mathbb{Z}/2\mathbb{Z}$–grading, since in this case

$$A^{\mathrm{red}}_{g(K)-1} \cong \mathrm{HF}^+(S^3_0(K), g-1) \cong \widehat{\mathrm{HFK}}(K, g(K)) \neq 0.$$

It follows that in this case $\deg(\Delta_K) = g(K)$ and $\widehat{\mathrm{HFK}}(K, g(K))$ is supported in odd degrees.

Moreover, if $\widetilde{g} = 0$, then $V_k = 0$ for all $k \geq 0$, so that

$$t_k = \chi(A_k^{\mathrm{red}}(K)) \leq 0.$$

We now need to establish conditions which ensure that $\widetilde{g} = 0$ or $\widetilde{g} < g(K)$.

McCoy [8, Lemma 2.3] slightly modified the proof of [2, Theorem 1.1] by Greene to show that if $S^3_{p/q}(K)$ bounds a negative-definite 4–manifold with torsion-free first homology, then

$$2\widetilde{g} \leq n - \sqrt{n},$$

where $n = \lceil p/q \rceil$.

It follows that if $p/q \leq 3$ then $\widetilde{g} = 0$.

More generally, if $i \geq \lfloor (n - \sqrt{n})/2 \rfloor$, where $n = \lceil p/q \rceil$, then $i \geq \widetilde{g}$ and hence $V_i = 0$. It follows that $t_i = \chi(A_i^{\mathrm{red}}(K)) \leq 0$. If $g(K) > \left\lfloor \frac{1}{2}(\lceil p/q \rceil - \sqrt{\lceil p/q \rceil}) \right\rfloor$, then $g(K) > \widetilde{g}$ as well.

For the improvement of the genus bound, note that all the summands of $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ coming from the V_i (ie of the form $\tau(V_i)$) are situated in the even grading and therefore must vanish. Now the proof of Theorem 3 shows that $U^{g(K)} \cdot \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = 0$.

Now if

$$U^{\lfloor |H_1(Y)|/2 \rfloor} \cdot \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) \neq 0,$$

then $\left\lfloor \frac{1}{2}|H_1(Y)| \right\rfloor \leq g(K) - 1$, so $\frac{1}{2}(|H_1(Y)| + 1) \leq g(K)$. On the other hand,

$$\widetilde{g} \leq \tfrac{1}{2}(\lceil p/q \rceil - \sqrt{\lceil p/q \rceil}) < \tfrac{1}{2}(p/q + 1) \leq \tfrac{1}{2}(p+1) = \tfrac{1}{2}(|H_1(Y)| + 1) \leq g(K).$$

It follows from (11) that $\deg(\Delta_K) = g(K)$. □

We end this section with the following question:

Question 32 Does there exist a knot $K \subset S^3$ with $\deg(\Delta_K) \neq g(K)$ and with a Seifert fibred surgery?

7 Some other applications of the mapping cone formula

In this section, we demonstrate some other applications of the results obtained in Section 3.

Theorem 8 *Let K be an L–space knot and $p/q \leq 1$ a rational number. Then $S^3_{p/q}(K)$ and p/q determine the Alexander polynomial of K.*

Proof If the slope is zero this is immediate from Proposition 22. If the slope is negative this also easily follows from Lemma 18; by looking at $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ we can work out a sequence of numbers that represents all the torsion coefficients with some repetitions (they are orders of cyclic $\mathbb{F}[U]$–modules). But we know the number of repetitions because we know the slope. From this we deduce all the torsion coefficients (in the correct order, as they form a monotone sequence), and hence the Alexander polynomial.

If the slope is in the interval $(0, 1]$ the reasoning is the same; Corollary 14 allows us to work out the torsion coefficients, since we know how many times each occurs. The only torsion coefficient we might not be able to work out from the module structure of $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ is t_0 if the slope is 1. But in this case, it can be worked out from the d–invariant formula of Ni and Wu from Corollary 14. □

Sometimes we can work out a lot about the Heegaard Floer homology associated to a knot from a surgery on it even if it is not an L-space knot.

Proposition 33 *The small Seifert fibred space $Y = S^2((2,1),(6,-1),(7,-2))$ can only be obtained by (-4)–surgery. All knots producing it are non-L–space knots.*

Proof We find the HF^+ of this space using the computer program `HFNem2` by Çağrı Karakurt [5]. There are four Spin^c–structures $\{\mathfrak{s}_i\}_{i=0}^3$, and HF^+ in them have the form

$$\begin{aligned}
\mathrm{HF}^+(Y,\mathfrak{s}_0) &\cong \mathcal{T}_{-3/4},\\
\mathrm{HF}^+(Y,\mathfrak{s}_1) &\cong \mathcal{T}_0 \oplus \tau_0(1),\\
\mathrm{HF}^+(Y,\mathfrak{s}_2) &\cong \mathcal{T}_{1/4},\\
\mathrm{HF}^+(Y,\mathfrak{s}_3) &\cong \mathcal{T}_0 \oplus \tau_0(1).
\end{aligned}$$

Using Theorem 7 we can restrict the possible slopes to $\{\pm 4, \pm\frac{4}{3}, \pm\frac{4}{5}\}$. Calculating the correction terms of $L(4,1) = L(4,-3) = L(4,5)$ and $L(4,-1) = L(4,3) = L(4,-5)$ we notice that only $L(4,-1)$ has correction terms such that the difference of each of them with some correction term of Y is an integer. This means that the slope has to be in $\{-4, \frac{4}{3}, -\frac{4}{5}\}$.

We also notice that the d–invariants of Y coincide exactly with the d–invariants of the lens space $L(4,-1)$. By the d–invariant formula (2) we conclude that $V_0 = 0$. A similar argument using the d–invariant formula for negative surgeries in Proposition 19 establishes that $\overline{V}_0 = 0$.

Now using the total dimension formulas of Propositions 16 and 20 we conclude

$$2 = \dim \mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K)) = q\delta(K),$$

which is impossible for $q = 3$ or $q = -5$.

Comparing the labelling of Spin^c–structures we see that the order in which we listed $\mathrm{HF}^+(Y, \mathfrak{s}_i)$ above corresponds to $i = 0, 1, 2$ and 3.

If Y could be obtained by (-4)–surgery on an L–space knot, then the fact that $V_0 = 0$ would imply that its genus is zero, ie it is the unknot. However, Y is not a lens space. □

It seems worth noticing that in fact there are infinitely many knots K_n that produce Y from the proposition above; see [27]. In fact, $K_0 = \overline{9_{42}}$. The spaces resulting from p/q–surgeries on these knots have rather similar Floer homologies; in particular, all the correction terms are the same (and coincide with the correction terms of the lens space $L(p,q)$) and the total rank of reduced Floer homology is $2q$.

Moreover, we can work out the Heegaard Floer homology of all surgeries on these knots and their Alexander polynomials. Teragaito [27, Remark 6.1] mentions that K_n has genus $2n+2$. In [17, Corollary 4.5] it is shown that

$$\widehat{\mathrm{HFK}}(K, g(K)) \cong \mathrm{HF}^+(S^3_0(K), g-1),$$

so it is nontrivial by Theorem 9, and thus by Proposition 22 and the fact that $V_0 = 0$ for the present examples, we get that $A^{\mathrm{red}}_{\pm(g(K)-1)}$ have to be nontrivial. By description of the Heegaard Floer homology of Y in the proof of Proposition 33 we conclude that $A^{\mathrm{red}}_{2n+1}(K_n) = A^{\mathrm{red}}_{-(2n+1)}(K_n) = \tau(1)$ and $A^{\mathrm{red}}_k(K_n) = 0$ for any $k \neq \pm 2n+1$. Using Proposition 19 we can also fix the gradings, and then using results from Section 3 deduce the Heegaard Floer homology of all surgeries on these knots.

Proposition 34 *The Alexander polynomial of K_0 is $-1+2(T+T^{-1})-(T^2+T^{-2})$. For $n \neq 0$ the Alexander polynomial is given by*

$$\Delta_{K_n}(T) = 1-(T^{2n}+T^{-2n})+2(T^{2n+1}+T^{-(2n+1)})-(T^{2n+2}+T^{-(2n+2)}).$$

Proof From the discussion above, $V_0 = 0$ and the only nontrivial $A^{\mathrm{red}}_k(K)$ are $A^{\mathrm{red}}_{2n+1}(K_n) = A^{\mathrm{red}}_{-(2n+1)}(K_n) = \tau(1)$. Moreover, since the reduced parts of the Heegaard Floer homology of (-4)–surgery are in absolute $\mathbb{Z}/2\mathbb{Z}$–grading 0, it means that $A^{\mathrm{red}}_{\pm(2n+1)}$ are in grading 1. (We can see from the description of the absolute grading on the mapping cone and Lemma 18 that for negative surgeries the $\mathbb{Z}/2\mathbb{Z}$–grading of $\mathbb{A}^+_{i,p/q}(K)$ switches from what we have defined it to be in the mapping cone.) Now Lemma 23 implies that $t_{2n+1}=-1$ and $t_i=0$ for all other $i\geq 0$. □

By a straightforward argument involving $\mathbb{Z}/2\mathbb{Z}$–grading considerations and dimension count it is not difficult to establish that in fact for $n>0$ we have

$$\widehat{\mathrm{HFK}}(K_n,2n+2)\cong\widehat{\mathrm{HFK}}(K_n,2n)\cong\mathbb{F}\quad\text{and}\quad\widehat{\mathrm{HFK}}(K_n,2n+1)\cong\mathbb{F}^2.$$

7.1 Property S

Heegaard Floer homology has been very successful in restricting cosmetic surgeries on knots in S^3 (see [11; 23; 28]). In this subsection, we define a class of knots that do not admit purely cosmetic surgeries.

Definition 35 Let $r_1,r_2\in\mathbb{Q}$ and let $K\subset S^3$ be a knot. The surgeries on K with slopes r_1 and r_2 are called cosmetic if $S^3_{r_1}(K)$ is homeomorphic to $S^3_{r_2}(K)$. They are called purely cosmetic if $S^3_{r_1}(K)\cong S^3_{r_2}(K)$, by which we mean that there exists an orientation-preserving homeomorphism between them.

We now begin defining the property that will imply the nonexistence of purely cosmetic surgeries.

Definition 36 We say that a rational homology sphere Y has property S if $\mathrm{HF}_{\mathrm{red}}(Y)$ is all concentrated in the same absolute $\mathbb{Z}/2\mathbb{Z}$–grading.

Definition 37 We say that a knot $K\subset S^3$ has property S if $S^3_{p/q}(K)$ has property S for some $p/q\neq 0$.

Proposition 38 *A knot K has property S if and only if $S^3_{p/q}(K)$ has property S for any $p/q\geq 2g(K)-1$.*

Proof Suppose $S^3_{p/q}(K)$ has property S. By taking the mirror of the knot, we may assume $p/q>0$.

Then by looking at Corollary 14 and Proposition 15 we see that for all k all elements of $A^{\mathrm{red}}_k(K)$ are in the same $\mathbb{Z}/2\mathbb{Z}$–grading. This is enough for all elements of $\mathrm{HF}^+(S^3_{p/q}(K))$ for $p/q\geq 2g(K)-1$ to be concentrated in the same $\mathbb{Z}/2\mathbb{Z}$–grading. □

Corollary 39 *A nontrivial knot with property S admits no purely cosmetic surgeries.*

Proof The proof is completely analogous to the proof of [11, Corollary 3.12]. In fact, Ni and Wu show that if Y can be obtained by a purely cosmetic surgery, then the Euler characteristic of $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ has to be 0. They also show that V_0 and $\bar{V}_0$ have to be zero for a knot that admits cosmetic surgeries. This implies that $V_i = H_i = 0$ for $i \geq 0$, so we do not have any $\tau(V_i)$ groups in the reduced Floer homology. A knot with property S has all the $A^+_k(K)$ groups concentrated in the same $\mathbb{Z}/2\mathbb{Z}$–grading, and in the case at hand these are the groups that constitute the reduced Floer homology. Therefore, in this case, the Euler characteristic of $\mathrm{HF}_{\mathrm{red}}(S^3_{p/q}(K))$ is equal to (plus or minus) its rank, so it is an L–space. However, if an L–space knot has $V_0 = 0$, then it is trivial. □

Ni and Wu [11, Corollary 3.12] show that Seifert fibred spaces cannot be obtained by purely cosmetic surgeries. We can extend this result as follows.

Corollary 40 *There are no purely cosmetic surgeries on knots with nonzero Seifert fibred surgeries.*

Proof By [15] Seifert fibred rational homology spheres have property S. □

We remark that there are knots which do not have this property, for example 9_{44}. Indeed, $+1$ and -1–surgeries on this knot have the same HF^+, but are not homeomorphic [23, Section 9].

References

[1] **R H Crowell**, *Nonalternating links*, Illinois J. Math. 3 (1959) 101–120 MR

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

[3] **S Jabuka**, *Heegaard Floer genus bounds for Dehn surgeries on knots*, J. Topol. 7 (2014) 523–542 MR

[4] **S Jabuka**, *Heegaard Floer groups of Dehn surgeries*, J. Lond. Math. Soc. 92 (2015) 499–519 MR

[5] **Ç Karakurt**, `HFNem2` (2013) Magma code Available at `https://tinyurl.com/HFNem2`

[6] **M Lackenby**, *Dehn surgery on knots in 3–manifolds*, J. Amer. Math. Soc. 10 (1997) 835–864 MR

[7] **M Lackenby**, **J S Purcell**, *Cusp volumes of alternating knots*, Geom. Topol. 20 (2016) 2053–2078 MR

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

[9] **A H Moore**, **L Starkston**, *Genus-two mutant knots with the same dimension in knot Floer and Khovanov homologies*, Algebr. Geom. Topol. 15 (2015) 43–63 MR

[10] **K Murasugi**, *On the Alexander polynomial of the alternating knot*, Osaka Math. J. 10 (1958) 181–189 MR

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

[12] **Y Ni**, **X Zhang**, *Characterizing slopes for torus knots*, Algebr. Geom. Topol. 14 (2014) 1249–1274 MR

[13] **J K Osoinach, Jr**, *Manifolds obtained by surgery on an infinite number of knots in S^3*, Topology 45 (2006) 725–733 MR

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

[15] **P Ozsváth**, **Z Szabó**, *On the Floer homology of plumbed three-manifolds*, Geom. Topol. 7 (2003) 185–224 MR

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

[17] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and knot invariants*, Adv. Math. 186 (2004) 58–116 MR

[18] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. 159 (2004) 1027–1158 MR

[19] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and three-manifold invariants: properties and applications*, Ann. of Math. 159 (2004) 1159–1245 MR

[20] **P Ozsváth**, **Z Szabó**, *On Heegaard Floer homology and Seifert fibered surgeries*, from "Proceedings of the Casson Fest" (C Gordon, Y Rieck, editors), Geom. Topol. Monogr. 7, Geom. Topol. Publ., Coventry (2004) 181–203 MR

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

[22] **P S Ozsváth**, **Z Szabó**, *Knot Floer homology and integer surgeries*, Algebr. Geom. Topol. 8 (2008) 101–153 MR

[23] **P S Ozsváth**, **Z Szabó**, *Knot Floer homology and rational surgeries*, Algebr. Geom. Topol. 11 (2011) 1–68 MR

[24] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard (2003) MR Available at `http://search.proquest.com/docview/305332635`

[25] **J Rasmussen**, *Lens space surgeries and a conjecture of Goda and Teragaito*, Geom. Topol. 8 (2004) 1013–1031 MR

[26] **D Rolfsen**, *Knots and links*, corrected reprint, Math. Lect. Ser. 7, Publish or Perish, Houston (1990) MR

[27] **M Teragaito**, *A Seifert fibered manifold with infinitely many knot-surgery descriptions*, Int. Math. Res. Not. 2007 (2007) art. id rnm 028, 16 pages MR

[28] **J Wang**, *Cosmetic surgeries on genus one knots*, Algebr. Geom. Topol. 6 (2006) 1491–1517 MR

[29] **Z Wu**, *On mapping cones of Seifert fibered surgeries*, J. Topol. 5 (2012) 366–376 MR

Department of Mathematics, Imperial College London
South Kensington Campus, London, SW7 2AZ, United Kingdom

fyodor.gainullin@gmail.com

Received: 7 July 2015 Revised: 26 November 2016

Algebraic & Geometric Topology 17 (2017) 1953–2011

msp

The C_2–spectrum $\mathrm{Tmf}_1(3)$ and its invertible modules

MICHAEL A HILL
LENNART MEIER

We explore the C_2–equivariant spectra $\mathrm{Tmf}_1(3)$ and $\mathrm{TMF}_1(3)$. In particular, we compute their C_2–equivariant Picard groups and the C_2–equivariant Anderson dual of $\mathrm{Tmf}_1(3)$. This implies corresponding results for the fixed-point spectra $\mathrm{TMF}_0(3)$ and $\mathrm{Tmf}_0(3)$. Furthermore, we prove a real Landweber exact functor theorem.

55N34, 55P42

1 Introduction

The spectrum TMF of topological modular forms comes in many variants. While TMF itself arises from the moduli stack of elliptic curves $\mathcal{M}_{\mathrm{ell}}$, there is also a spectrum Tmf associated with the compactification $\overline{\mathcal{M}}_{\mathrm{ell}}$. Finally, tmf is defined as the connective cover of Tmf. It has been the spectrum tmf and its cohomology that have been so far most relevant to applications (see eg Behrens, Hill, Hopkins and Mahowald [9] and Behrens and Pemmaraju [12] for applications to generalized Toda–Smith complexes and Ando, Hopkins and Rezk [2], Mahowald and Hopkins [45] and Hill [26] for applications to string bordism).

It is often simpler to work with topological modular forms with level structures. Among the many possibilities, the most relevant for us will be $\mathrm{TMF}_1(n)$ and $\mathrm{TMF}_0(n)$ corresponding to the moduli stacks $\mathcal{M}_1(n)$ and $\mathcal{M}_0(n)$. The former stack classifies elliptic curves with a chosen point of exact order n and the latter elliptic curves with a chosen subgroup of order n. Note that for $n \geq 2$, the spectrum $\mathrm{TMF}_1(n)$ is Landweber exact, while $\mathrm{TMF}_0(n)$ is not in general, as will be explained in Section 4.1.

Besides providing simpler variants of TMF, there are several reasons to care about TMF with level structures. First, we mention the $Q(l)$–spectra defined by Behrens (see Behrens [7] and Behrens and Ormsby [10]), which are built from TMF with level structures and provide approximations of the $K(2)$–local sphere. Second, as shown in Behrens, Ormsby, Stapleton and Stojanoska [11], there is an injective map

$$\pi_* \mathrm{TMF} \wedge \mathrm{TMF} \to \prod_{i \in \mathbb{Z}, j \geq 0} \pi_* \mathrm{TMF}_0(3^j) \times \pi_* \mathrm{TMF}_0(5^j),$$

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.1953

important in the study of cooperations of TMF and tmf. As a last point, we mention that Lurie defines in [40] a sheaf of E_∞–ring spectra on the n–torsion of the universal elliptic curve over $\mathcal{M}_{\mathrm{ell}}$ whose global sections provide the value of C_n–equivariant TMF at a point; if we invert n, this can be analyzed in terms of the $\mathrm{TMF}_1(k)$ for $k \mid n$.

In each of these cases, it would be interesting to have compactified and connective variants. As a first step, Hill and Lawson overcame in [29] certain technical obstacles to define E_∞–ring spectra $\mathrm{Tmf}_1(n)$ and $\mathrm{Tmf}_0(n)$ corresponding to the compactified moduli stacks $\overline{\mathcal{M}}_1(n)$ and $\overline{\mathcal{M}}_0(n)$. One can then define $\mathrm{tmf}_1(n)$ and $\mathrm{tmf}_0(n)$ as the connective covers of these spectra and they form good connective models for $\mathrm{TMF}_1(n)$ and $\mathrm{TMF}_0(n)$ if n is small. The aim of this article is to explore these spectra in the case $n = 3$ with methods from real homotopy theory.

Real homotopy theory is the study of genuine equivariant C_2–spectra, also sometimes known as real spectra. The theory has its origins in Atiyah's article [4] on real K-theory and came to new prominence through the work of Hu and Kriz [32] and the work of Hill, Hopkins and Ravenel [28] on the Kervaire invariant one problem.

The spectra $\mathrm{TMF}_1(3)$ and $\mathrm{Tmf}_1(3)$ inherit C_2–actions from an algebrogeometrically defined C_2–action on $\overline{\mathcal{M}}_1(3)$. We will view them as cofree C_2–spectra (as explained in Sections 2.2 and 4.1) so that

$$\mathrm{TMF}_1(3)^{C_2} \simeq \mathrm{TMF}_1(3)^{hC_2} \simeq \mathrm{TMF}_0(3)$$

and

$$\mathrm{Tmf}_1(3)^{C_2} \simeq \mathrm{Tmf}_1(3)^{hC_2} \simeq \mathrm{Tmf}_0(3).$$

We define the C_2–spectrum $\mathrm{tmf}_1(3)$ as the C_2–equivariant connective cover of $\mathrm{Tmf}_1(3)$.

Mahowald and Rezk [46] have already computed the homotopy groups of $\mathrm{TMF}_0(3)$ and a similar computation actually produces the $\mathrm{RO}(C_2)$–graded C_2–equivariant homotopy groups of $\mathrm{tmf}_1(3)$ and hence $\mathrm{TMF}_1(3)$. Using this computation, we show that $\mathrm{tmf}_1(3)$ has a real orientation and is more precisely a form of $BP\mathbb{R}\langle 2\rangle$. This implies in particular that there exists a form of $BP\mathbb{R}\langle 2\rangle$ that is a strictly commutative C_2–spectrum, while it was not known before that there is a form of $BP\mathbb{R}\langle 2\rangle$ with any kind of ring structure.

Moreover, we show that $\mathrm{TMF}_1(3)$ is real Landweber exact in the sense that there is an isomorphism

$$M\mathbb{R}_\star(X) \otimes_{MU_{2*}} \mathrm{TMF}_1(3)_{2*} \to \mathrm{TMF}_1(3)_\star X,$$

natural in a C_2–spectrum X. Here $M\mathbb{R}_\star(X)$ denotes the $\mathrm{RO}(C_2)$–graded C_2–equivariant homology groups of X with respect to the real bordism spectrum $M\mathbb{R}$ and similarly for $\mathrm{TMF}_1(3)_\star X$.

As $\overline{\mathcal{M}}_1(3)$ is proper over $\operatorname{Spec}\mathbb{Z}\big[\tfrac{1}{3}\big]$, one expects a manifestation of Serre duality in $\mathrm{Tmf}_1(3)$. A suitable duality to look for in the topological setting is *Anderson duality*, an integral version of Brown–Comenetz duality. For example, Stojanoska computed in [60] that $\mathrm{Tmf}\big[\tfrac{1}{2}\big]$ is Anderson self-dual in the sense that $I_{\mathbb{Z}[\frac{1}{2}]}\mathrm{Tmf}\big[\tfrac{1}{2}\big] \simeq \Sigma^{21}\mathrm{Tmf}\big[\tfrac{1}{2}\big]$. We want to compute the C_2–equivariant Anderson dual $I_{\mathbb{Z}[\frac{1}{3}]}\mathrm{Tmf}_1(3)$ of $\mathrm{Tmf}_1(3)$. While it is an easy calculation that nonequivariantly $I_{\mathbb{Z}[\frac{1}{3}]}\mathrm{Tmf}_1(3) \simeq \Sigma^9\mathrm{Tmf}_1(3)$, this equivalence does *not* hold C_2–equivariantly. Rather, we get the following:

Theorem *There is a C_2–equivariant equivalence*

$$I_{\mathbb{Z}[\frac{1}{3}]}\mathrm{Tmf}_1(3) \simeq \Sigma^{5+2\rho}\mathrm{Tmf}_1(3),$$

where ρ denotes the regular representation of C_2. It follows that

$$I_{\mathbb{Z}[\frac{1}{3}]}\mathrm{Tmf}_0(3) \simeq (\Sigma^{5+2\rho}\mathrm{Tmf}_1(3))^{hC_2}.$$

Thus, the self-duality of $\mathrm{Tmf}_0(3)$ is not fully apparent in the integer-graded homotopy groups

$$\pi_*\mathrm{Tmf}_0(3) \cong \pi_*^{C_2}\mathrm{Tmf}_1(3),$$

but only in the $\mathrm{RO}(C_2)$–graded homotopy groups $\pi_\star^{C_2}\mathrm{Tmf}_1(3)$. Likewise, the resulting universal coefficient sequence uses $\mathrm{RO}(C_2)$–graded homotopy groups. Indeed, for C_2–spectra X the theorem implies a short exact sequence

$$0\to\mathrm{Ext}^1\big(R^{C_2}_{(a-6)+(b-2)\rho}(X),\mathbb{Z}\big[\tfrac{1}{3}\big]\big)\to R_{C_2}^{a+b\rho}(X)\to\mathrm{Hom}\big(R^{C_2}_{(a-5)+(b-2)\rho}(X),\mathbb{Z}\big[\tfrac{1}{3}\big]\big)\to 0$$

with $R=\mathrm{Tmf}_1(3)$ and Hom and Ext computed over $\mathbb{Z}\big[\tfrac{1}{3}\big]$. We prove the theorem by an application of the slice spectral sequence. There has been similar work by Ricka [56] on Anderson duality of integral versions of Morava K-theory; our results have been obtained independently.

Next we turn to the topic of Picard groups. Given an E_∞–ring spectrum R, its Picard group $\mathrm{Pic}(R)$ is defined as the group of invertible R–module spectra up to weak equivalence. From the perspective of Bunke and Nikolaus [14], these are the global twists of the associated cohomology theory and define a natural grading of R–homology groups. The Picard group was first introduced into stable homotopy theory by Hopkins; recent work of Mathew and Stojanoska [51] then significantly extended our toolbox for its computation. They show that all invertible TMF–modules are suspensions of TMF so that $\mathrm{Pic}(\mathrm{TMF}) \cong \mathbb{Z}/576$. In contrast, they show that Pic(Tmf) contains exotic elements that are not suspensions of Tmf and compute $\mathrm{Pic}(\mathrm{Tmf}) \cong \mathbb{Z}\oplus\mathbb{Z}/24$.

We will use their methods to understand $\mathrm{Pic}(\mathrm{TMF}_0(3))$ and $\mathrm{Pic}(\mathrm{Tmf}_0(3))$, but add a dash of equivariant homotopy theory. The maps

$$\mathrm{Tmf}_0(3) \to \mathrm{Tmf}_1(3) \quad \text{and} \quad \mathrm{TMF}_0(3) \to \mathrm{TMF}_1(3)$$

are faithful C_2–Galois extensions in the sense of Rognes; see Mathew and Meier [50, Theorem 7.12]. As explained in Section 6.1, Galois descent then shows that

$$\mathrm{Pic}(\mathrm{Tmf}_0(3)) \cong \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)) \quad \text{and} \quad \mathrm{Pic}(\mathrm{TMF}_0(3)) \cong \mathrm{Pic}_{C_2}(\mathrm{TMF}_1(3)),$$

where $\mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3))$ denotes the group of invertible C_2–module spectra over $\mathrm{Tmf}_1(3)$ and similarly for $\mathrm{Pic}_{C_2}(\mathrm{TMF}_1(3))$. First we prove:

Theorem *Every invertible* $\mathrm{TMF}_0(3)$*–module is an (integral) suspension of* $\mathrm{TMF}_0(3)$*. Thus,*

$$\mathrm{Pic}_{C_2}(\mathrm{TMF}_1(3)) \cong \mathrm{Pic}(\mathrm{TMF}_0(3)) \cong \mathbb{Z}/48.$$

The analogous theorem for $\mathrm{Tmf}_0(3)$ is not true, but we have the following equivariant refinement:

Theorem *Every invertible* C_2*–equivariant* $\mathrm{Tmf}_1(3)$*–module is an equivariant suspension* $\Sigma^V \mathrm{Tmf}_1(3)$*, for an element* $V \in \mathrm{RO}(C_2)$*. The corresponding homomorphism*

$$\mathrm{RO}(C_2) \to \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)), \quad V \mapsto \Sigma^V \mathrm{Tmf}_1(3)$$

is thus surjective and has kernel generated by $8 - 8\sigma$*, for* σ *the sign representation. Therefore,*

$$\mathrm{Pic}(\mathrm{Tmf}_0(3)) \cong \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)) \cong \mathbb{Z} \oplus \mathbb{Z}/8.$$

We remark that invertible modules over TMF with level structure occur in the study of equivariant TMF, for example those defined by representation spheres. We hope that our results on Picard groups may have relevance there.

We give a short overview of the structure of this article. Section 2 discusses preliminaries from equivariant homotopy theory. In particular, it is about the passage from spectra with G–action to genuine G–spectra and to their connective covers and how a commutative multiplication under this passage is preserved; furthermore, we discuss the $\mathrm{RO}(G)$–graded homotopy fixed point spectral sequence and the slice spectral sequence. Section 3 is about real orientability and the real Landweber exact functor theorem; it concludes with the definition and basic properties of forms of $BP\mathbb{R}\langle n\rangle$ and $E\mathbb{R}(n)$. Section 4 introduces the main characters $\mathrm{Tmf}_0(3)$ and $\mathrm{Tmf}_1(3)$ and their variants, discusses their relationship and computes the $\mathrm{RO}(C_2)$–graded homotopy groups of $\mathrm{tmf}_1(3)$; here we also present our applications to forms of $BP\mathbb{R}\langle 2\rangle$ and $E\mathbb{R}(2)$. Section 5 computes the slices of $\mathrm{Tmf}_1(3)$ and applies this to compute its equivariant Anderson dual. Section 6 is about computations of Picard groups, especially those of $\mathrm{TMF}_0(3)$ and $\mathrm{Tmf}_0(3)$. As a step, we prove a generalization of a result of Baker and Richter [5] to give a conceptual computation of $\mathrm{Pic}(\mathrm{TMF}_1(3))$. Note that Sections 5 and 6 are independent and also independent of Section 3.

Conventions For a scheme X with an action by a group scheme G, we denote by X/G the *stack* quotient. Furthermore, for a (pre)sheaf $\mathcal{F}$ of spectra, $\pi_*\mathcal{F}$ will always denote the *sheafified* homotopy groups, ie the sheafification of $U \mapsto \pi_*(\mathcal{F}(U))$.

Acknowledgements We thank John Greenlees, Akhil Mathew and the referee for their comments on earlier versions of this article. We also thank the Hausdorff Institute for its hospitality during the time when the first version of this article was completed. Hill was supported by NSF DMS-1307896 and the Sloan Foundation. Meier was supported by DFG SPP 1786.

2 G–spectra and equivariant homotopy

After giving some basics on (genuine) G–spectra, we will treat in detail how to go from a spectrum with a G–action to a genuine G–spectrum, why this move preserves commutative multiplications and why the same is true for the passage to connective covers. After this, we will discuss the $\mathrm{RO}(G)$–graded homotopy fixed point spectral sequence and the slice filtration.

2.1 Conventions on equivariant spectra

We work in the category of genuine G–spectra for a finite group G, and our particular model will be orthogonal G–spectra. These were introduced by Mandell and May [47], though we draw heavily from [28] and also recommend [58] for a slightly different point of view on the same subject matter. In particular, a G–spectrum will always mean an orthogonal G–spectrum indexed on a complete G–universe, and morphisms are equivariant maps.

For each $H \subset G$ and for each G–spectrum X, we have stable homotopy groups

$$\pi_n^H(X) = \mathrm{colim}_V [S^{V+\mathbb{R}^n}, X(V)]^H,$$

where the colimit is taken over the finite dimensional representations of G (or more simply, over the cofinal subsystem of sums of the regular representation), and for any representation V, the space S^V is the 1–point compactification. Recall finally that a map is a weak equivalence if it induces an isomorphism on these equivariant stable homotopy groups for all $H \subset G$. These are the weak equivalences in the standard model structures on Sp^G which give the genuine equivariant stable homotopy category; this extends the ordinary equivariant Spanier–Whitehead category described by Adams.

Since we are considering the genuine model structure, the homotopy objects are naturally Mackey-functor valued: for any two G–spectra X and Y, the assignment

$$T \mapsto [T_+ \wedge X, Y]^G$$

extends to an additive functor from the Burnside category of finite G–sets to abelian groups. In general, we will denote the obvious Mackey functor extension of classical objects like homotopy groups with an underline. In particular, we can rephrase the above condition on weak equivalences as simply that a map $f\colon X \to Y$ is a weak equivalence if it induces an isomorphism of homotopy Mackey functors $\underline{\pi}_* X \to \underline{\pi}_* Y$.

2.1.1 RO(G)–grading and distinguished representations Since we are working genuine equivariantly, the representation spheres S^V are elements of the Picard group of the homotopy category $\mathrm{Ho}(\mathrm{Sp}^G)$. In particular, all of our $\mathbb{Z}$–graded homotopy groups extend to $\mathrm{RO}(G)$–graded homotopy groups, and similarly for Mackey functors, via the assignment

$$T \mapsto [T_+ \wedge S^V \wedge X, Y]^G.$$

We will use this combined structure extensively, and when $X = S^0$, we will simply denote these groups by $\underline{\pi}_V(Y)$. Note that to be precise, we have to choose (once and for all) for every element of $\mathrm{RO}(G)$ an actual invertible G–spectrum and not just a class in the Picard group and by abuse of notation we will denote it also by S^V for $V \in \mathrm{RO}(G)$. Every such choice results in $\underline{\pi}_\star$ being a lax symmetric monoidal functor by [38, Appendix A].

We single out several representations.

Notation 2.1 (1) Let ρ denote the regular representation of G.

(2) Let $\bar{\rho}$ denote the quotient of ρ by the trivial summand.

(3) Let σ denote the nontrivial 1–dimensional real representation of C_2.

There are several distinguished homotopy classes of maps between representation spheres we shall need. If V is a representation of G with no fixed points, then let

$$a_V\colon S^0 \to S^V$$

denote the inclusion of the fixed points into the V sphere. This map is not null, and no iterate of it is null. However, its restriction to any subgroup H such that $V^H \neq \{0\}$ is null-homotopic. This shows the following standard fact in equivariant stable homotopy theory.

Lemma 2.2 *Given a G–spectrum X, the geometric fixed points of X can be computed as the G–fixed points of*

$$X[a_{\bar{\rho}}^{-1}] = \mathrm{hocolim}(X \xrightarrow{a_{\bar{\rho}}} \Sigma^{\bar{\rho}} X \xrightarrow{a_{\bar{\rho}}} \cdots).$$

Proof The homotopy colimit

$$S^{\infty\bar{\rho}} = \mathrm{hocolim}(S^0 \xrightarrow{a_{\bar{\rho}}} S^{\bar{\rho}} \xrightarrow{a_{\bar{\rho}}} \cdots)$$

is a model for the space $\widetilde{E}\mathcal{P}$, where $\mathcal{P}$ is the family of proper subgroups of G. The geometric fixed points are computed by smashing X with $\widetilde{E}\mathcal{P}$ and taking fixed points, from which the result follows. □

2.1.2 G–equivariant homology theories

Definition 2.3 Let G be a finite group. An *(ungraded) G–equivariant homology theory* is an exact functor $h_0\colon \mathrm{Ho}(\mathrm{Sp}^G) \to \mathrm{Ab}$ to the category of abelian groups (or any other abelian category) that sends (possibly infinite) coproducts to direct sums.

To such an ungraded homology theory we can associate an $\mathrm{RO}(G)$–graded version as follows: For a given element $V \in \mathrm{RO}(G)$, we consider the chosen invertible G–spectrum S^V and define $h_V(X)$ as $h_0(S^{-V} \wedge X)$. The resulting functor is also called an $\mathrm{RO}(G)$–*graded homology theory*. We will write h_V for $h_V(S^0)$.

For a G–spectrum E, we can define a G–equivariant homology theory by

$$X \mapsto E_0(X) = \pi_0^G(E \wedge X),$$

and we clearly get a natural isomorphism $E_\star(X) \cong \pi_\star^G(E \wedge X)$ of the $\mathrm{RO}(G)$–graded theories as well.

2.2 Passage from naive to genuine

The spectra which arise from algebraic geometry machines are almost never given to us as orthogonal G–spectra for some group G. Instead, they will be commutative ring spectra together with an action of G. There is a natural, homotopically meaningful way to prolong this to a genuine G–spectrum in a way which respects the multiplicative structure: passage to the cofree spectrum. It is easiest to explain this in two steps: extending a naive G–spectrum to a genuine one and then controlling the multiplicative structure.

2.2.1 Additive structure Denote by Sp_u^G the category of orthogonal spectra with G–action. We consider an equivariant map $X \to Y$ to be an equivalence if it is a stable equivalence of the underlying nonequivariant orthogonal spectra. Since the inclusion of trivial representations of G into a complete universe induces an equivalence

$$I\colon \mathrm{Sp}_u^G \to \mathrm{Sp}^G$$

of categories [47, Theorem V.1.5], we may consider any spectrum with a G–action as an orthogonal G–spectrum indexed on a complete universe. The functor I is not homotopical, however.

In contrast, the functor

$$\mathrm{Sp}_u^G \to \mathrm{Sp}^G, \quad X \mapsto IF(EG_+, X)$$

preserves weak equivalences and defines therefore a derived functor for $I\colon \mathrm{Sp}_u^G \to \mathrm{Sp}^G$. Here $F(-,-)$ is understood to be the derived function spectrum so that it includes a fibrant replacement of X. We call G–spectra *cofree* if they are up to weak equivalence in the image of $IF(EG_+,-)$.

In particular, using the cofree functor $IF(EG_+,-)$, we may view any spectrum with a G–action as a genuine G–spectrum. We will use this to view $\mathrm{TMF}_1(3)$ and $\mathrm{Tmf}_1(3)$ as C_2–spectra.

2.2.2 Multiplicative concerns The homotopical behavior of the cofree functor on commutative ring spectra is most easily understood via an operadic approach using instead E_∞–ring spectra. Let $\mathcal{O}$ be an E_∞ operad (for example, the linear isometries operad). As the model category of orthogonal spectra fulfills the monoid axiom by [48, Theorem 12.1], [59, Theorem 4] implies that the category of $\mathcal{O}$–algebras in orthogonal spectra with G–action has a projective model structure. Thus, if R is an $\mathcal{O}$–algebra with G–action, there exists an $\mathcal{O}$–algebra with G–action that is fibrant as a spectrum and weakly equivalent to R.

The equivalence of categories I above is strong symmetric monoidal, so in particular, it takes $\mathcal{O}$–algebras to $\mathcal{O}$–algebras in orthogonal G–spectra indexed on a complete universe. Here, the group G acts trivially on the operad, so this is the prototypical example of a naive N_∞ operad in the sense of [13]. Applying $IF(EG_+,-)$ takes R to $IF(EG_+,R)$, which is an algebra over $F(EG_+,\mathcal{O})$ if R is fibrant. However, this operad is a G-E_∞ operad [13, Theorem 6.25]. In particular, since the category of algebras over a G-E_∞ operad is Dwyer–Kan equivalent to the category of equivariant commutative ring spectra by [13, Theorem A.6], we conclude the following.

Theorem 2.4 *If R is a commutative ring spectrum with a G–action via commutative ring maps, then $IF(EG_+,R)$ is an equivariant commutative ring spectrum. More precisely, one can functorially associate to R an equivariant commutative ring spectrum R' such that R' and $IF(EG_+,R)$ are equivalent as E_∞–algebras.*

In particular, this will immediately imply that $\mathrm{TMF}_1(3)$ and $\mathrm{Tmf}_1(3)$ can actually be viewed as C_2–equivariant commutative ring spectra. Deducing a similar result for the equivariant connective cover of $\mathrm{Tmf}_1(3)$ will require a simple result undoubtedly known to the experts. The proof is also standard; we include it for completeness. Before proceeding, recall the following result about the connectivity of symmetric powers.

Lemma 2.5 *If X is a $(k-1)$–connected equivariant spectrum with $k \geq 0$, then for all $n \geq 1$,*

$$\mathrm{Sym}^n(X) = X^{\wedge n}/\Sigma_n$$

is also $(k-1)$–connected.

Proof This follows from the weak equivalence

$$E_G\Sigma_{n+} \wedge_{\Sigma_n} X^{\wedge n} \to \mathrm{Sym}^n(X).$$

As in [28, (B.171) and (B.177)], we can reduce the statement of the lemma by this equivalence to the following statement: the G–spectrum $\bigwedge_{G/H} S^l$ is $(k-1)$–connected for all $H \subset G$ and all $l \geq k$. This is clear, as $\bigwedge_{G/H} S^l \cong S^{\mathrm{ind}_H^G l}$ and ind_H^G contains an l–dimensional trivial summand. □

Remark 2.6 If X is $(k-1)$–connected for $k > 1$, then we do not always get a bump in the connectivity of the symmetric powers as happens classically. For n sufficiently large, the n^{th} symmetric power is more highly connected than X, but for low values of n, they are often equally connected. The reason for this is the norm: if $[G:H]=m$, then there is a canonical homotopy class of maps

$$N_H^G i_H^* X \to \mathrm{Sym}^m(X)$$

coming from any inclusion of $G \times \Sigma_m/\Gamma$ into $E_G\Sigma_m$, where Γ is the graph of the homomorphism $G \to \Sigma_m$ defining G/H as a G–set; see [28; 13]. In particular, if ρ_H^G is the representation $\mathrm{Ind}_H^G \mathbb{R}$ and if $\bar{\rho}_H^G$ is the quotient of ρ_H^G by the trivial summand, then for any class $x \in \pi_k^G(X)$, we have an element

$$a_{\bar{\rho}_H^G}^k N_H^G i_H^*(x) \in \pi_k^G(\mathrm{Sym}^m(X)).$$

Checking on the case of spheres shows that these maps are generically nontrivial. This is the only complicating factor in the proof of the following theorem, since it means that the k^{th} homotopy Mackey functor of the free commutative ring spectrum on something $(k-1)$–connected is strictly larger than

$$\underline{\pi}_k S^0 \oplus \underline{\pi}_k(X).$$

Theorem 2.7 *If R is a G–equivariant commutative ring spectrum, then there is a commutative ring spectrum structure on the (-1)–connected cover r of R such that the canonical map $r \to R$ is a map of commutative ring spectra.*

Proof We will inductively build a series of (-1)–connected commutative ring spectra r^k over R for $-1 \le k$ such that the induced map on homotopy groups $\underline{\pi}_j(r^k) \to \underline{\pi}_k(R)$ is an isomorphism for $0 \le j \le k$ (this condition is vacuous when $k = -1$). Let r^{-1} denote the zero sphere, which maps to R via the unit.

Assume that we have built $r^{k-1} \to R$ as above. We can assume that r^{k-1} is cofibrant in the positive model structure on equivariant commutative rings as in [28, Proposition B.130]. To build r^k, we first choose a surjective map

$$\bigoplus_{i \in I_k} \underline{A}_{G/H_i} \to \underline{\pi}_k(R),$$

where $\underline{A}_{G/H_i}$ is the Mackey functor $\underline{\pi}_k(G/H_{i+} \wedge S^k)$. Any such surjective map can be realized topologically as a map

$$\bigvee_{i \in I_k} G/H_{i+} \wedge S^k \xrightarrow{j_k} R,$$

and this induces a map of commutative ring spectra

$$e_k = \mathbb{P}\big(\bigvee_{i \in I_k} G/H_{i+} \wedge S^k\big) \to R,$$

where $\mathbb{P}$ denotes the free commutative ring spectrum functor. Smashing this with the map $r^{k-1} \to R$ gives a map

$$e_k \wedge r^{k-1} \xrightarrow{J_k} R.$$

This is the correct derived smash product by [28, Proposition 2.30]. The map $S^0 \to e_k$ induces an isomorphism in homotopy groups through dimension $(k-1)$ by Lemma 2.5, and the Künneth spectral sequence of Lewis and Mandell [38] implies that the map J_k induces an isomorphism in homotopy in dimensions between 0 and $(k-1)$ and a surjection in dimension k.

At this point, the argument is classical. Let F_k denote the fiber of $e_k \wedge r^{k-1} \to R$, and let f_k denote the (-1)–connected cover of F_k. Since the map $e_k \wedge r^{k-1} \to R$ was a map of commutative ring spectra, the composite

$$\mathrm{Sym}^n(f_k) \to \mathrm{Sym}^n(F_k) \to e_k \wedge r^{k-1} \to R$$

is null for all $n > 0$. In particular, if we let r^k denote the pushout in commutative ring spectra

$$\begin{tikzcd} \mathbb{P}(f_k) \arrow[r] \arrow[d] & e_k \wedge r^{k-1} \arrow[d] \\ S^0 \arrow[r] & r^k \end{tikzcd}$$

then we have an extension of J_k over r^k. Note that r^k is actually equivalent to the *derived* smash product $S^0 \wedge_{\mathbb{P}(f_k)} (e_k \wedge r^{k-1})$ because S^0 is a cofibrant $\mathbb{P}(f_k)$–module with respect to a monoidal model structure [28, Proposition B.137].

We have a cofiber sequence $\overline{\mathbb{P}}(f_k) \to \mathbb{P}(f_k) \to S^0$. Because r^k is a retract of $e_k \wedge r^{k-1}$, this induces short exact sequences

$$0 \to \underline{\pi}_i(\overline{\mathbb{P}}(f_k) \wedge_{\mathbb{P}(f_k)} (e_k \wedge r^{k-1})) \to \underline{\pi}_i(e_k \wedge r^{k-1}) \to \underline{\pi}_i(r^k) \to 0$$

for every $i \in \mathbb{Z}$. Since $\overline{\mathbb{P}}(f_k) \cong \bigvee_{n \geq 1} \mathrm{Sym}^n(f_k)$, Lemma 2.5 implies that $\overline{\mathbb{P}}(f_k)$ is $(k-1)$–connected. By the Künneth spectral sequence, $\overline{\mathbb{P}}(f_k) \wedge_{\mathbb{P}(f_k)} (e_k \wedge r^{k-1})$ is thus also $(k-1)$–connected. Therefore, $\underline{\pi}_i(e_k \wedge r^{k-1}) \to \underline{\pi}_i r^k$ is an isomorphism for $i \leq k-1$ and hence so is $\underline{\pi}_i r^k \to \underline{\pi}_i R$.

For the analysis of $\underline{\pi}_k$, consider the diagram:

$$\begin{tikzcd} \underline{\pi}_k(\overline{\mathbb{P}}(f_k)) \arrow[r] \arrow[d] & \underline{\pi}_k(f_k) \arrow[d] & \\ \underline{\pi}_k(\overline{\mathbb{P}}(f_k) \wedge_{\mathbb{P}(f_k)} (e_k \wedge r^{k-1})) \arrow[r, "\iota_*"] & \underline{\pi}_k(e_k \wedge r^{k-1}) \arrow[r, "(J_k)_*"] & \underline{\pi}_k(R) \end{tikzcd}$$

We know that $\underline{\pi}_k(f_k)$ surjects onto the kernel of $(J_k)_*$. As f_k is a summand of $\overline{\mathbb{P}}(f_k)$, ι_* also must surject onto the kernel of $(J_k)_*$. Thus, $\underline{\pi}_k r^k \cong \mathrm{coker}(\iota_*)$ maps injectively into $\underline{\pi}_k R$ and also surjectively because already $(J_k)_*$ is surjective. Now define r as the colimit of the r^k. Clearly, r is connective and the maps $r^k \to R$ extend to a map $r \to R$ that induces an isomorphism in π_i for $i \geq 0$. □

2.3 The RO(G)–graded homotopy fixed point spectral sequence

If V is a virtual representation of G, then by tracing through the adjunctions, we see that

$$\pi_V^G F(EG_+, X) \cong [S^V \wedge EG_+, X]^G \cong \pi_0^G F(EG_+, S^{-V} \wedge X).$$

For a G–spectrum E denote by $\mathrm{HF}(E)$ the homotopy fixed point spectral sequence for E (as constructed eg in [18, Section 6]). Let $V_1, \dots, V_n$ be representatives of the isomorphism classes of nontrivial irreducible real representations of G. We define the $\mathrm{RO}(G)$–graded homotopy fixed point spectral sequence for E as

$$\mathrm{HF}^{\mathrm{RO}(G)}(E) = \bigoplus_{a_i \in \mathbb{Z}} \mathrm{HF}(a_1 V_1 + \cdots + a_n V_n, E),$$

where

$$\mathrm{HF}(a_1 V_1 + \cdots + a_n V_n, E) = \mathrm{HF}(S^{-a_1 V_1} \wedge \cdots \wedge S^{-a_n V_n} \wedge E).$$

We can use the twisting isomorphisms of the symmetric monoidal structure on G–spectra to define isomorphisms

$$S^{-a_1 V_1} \wedge \cdots \wedge S^{-a_n V_n} \wedge S^{-b_1 V_1} \wedge \cdots \wedge S^{-b_n V_n} \cong S^{-(a_1+b_1)V_1} \wedge \cdots \wedge S^{-(a_n+b_n)V_n}.$$

As in [18], a multiplication $E \wedge E \to E$ then defines multiplicative pairings

$$\begin{aligned}\mathrm{HF}(a_1 V_1 + \cdots + a_n V_n, E) &\otimes \mathrm{HF}(b_1 V_1 + \cdots + b_n V_n, E) \\ &\to \mathrm{HF}((a_1 + b_1) V_1 + \cdots + (a_n + b_n) V_n, E).\end{aligned}$$

As explained in [38, Appendix A], we can choose the isomorphisms above so that this actually defines an associative and commutative multiplication on $\mathrm{HF}^{\mathrm{RO}(G)}(E)$. We summarize in the following proposition.

Proposition 2.8 *If E is a G–spectrum with a multiplication up to homotopy, then there is a multiplicative $\mathrm{RO}(G)$–graded spectral sequence*

$$E_2^{s,V} = H^s(G; \pi_0(S^{-V} \wedge E)) \implies \pi^G_{V-s} F(EG_+, E).$$

In particular, the Leibniz rule states that for elements $x \in E_r^{s,V}$ and $y \in E_r^{t,W}$ with $V = a_0 + a_1 V_1 + \cdots + a_n V_n$, we have $d_r(xy) = d_r(x)y + (-1)^{a_0} x d_r(y)$.

Note that while the $\mathrm{RO}(G)$–graded homotopy fixed point spectral sequence decomposes additively in infinitely many summands, we package them into one spectral sequence for the sake of a more efficient multiplicative presentation. In our later computations, our generating permanent cycles will sit in nonintegral degrees.

2.4 The C_2–equivariant slice filtration

The C_2–equivariant slice filtration was introduced by Dugger in his study of Atiyah's real K–theory. This was generalized by Hopkins, Ravenel and Hill [28] to arbitrary finite groups in the solution to the Kervaire invariant one problem. We will recall some of the basic properties here. A more detailed treatment can be found in [28] or [27].

Proposition 2.9 [28, Proposition 4.20 and Lemma 4.23] *For any C_2–equivariant spectrum E, the odd slices are determined by the formula*

$$P^{2n-1}_{2n-1}(E) = \Sigma^{n\rho-1} H\underline{\pi}_{n\rho-1} E.$$

Corollary 2.10 *If R is a C_2–spectrum such that $\underline{\pi}_{n\rho-1} R = 0$, then all odd slices of R vanish.*

For the even slices, there is a similar formula involving homotopy Mackey functors of E.

Definition 2.11 If $\underline{M}$ is a C_2 Mackey functor, let $P^0\underline{M}$ denote the maximal quotient of $\underline{M}$ in which the restriction map $\underline{M}(C_2/C_2) \to \underline{M}(C_2/e)$ is injective.

There are several equivalent formulations, one of which is to notice that we can build a Mackey functor out of the kernel of the restriction by declaring that the value at C_2/C_2 is the kernel of the restriction map and that the value at $C_2/\{e\}$ is trivial. The functor $P^0\underline{M}$ is then the quotient of $\underline{M}$ by this sub-Mackey functor.

The second reformulation requires an endofunctor on Mackey functors.

Definition 2.12 If T is a finite C_2–set and $\underline{M}$ is a Mackey functor, then let $\underline{M}_T$ be the Mackey functor defined by

$$S \mapsto \underline{M}(T \times S).$$

The restriction map defines a map of Mackey functors

$$\underline{M} \to \underline{M}_{C_2},$$

and $P^0\underline{M}$ is simply the image of this map.

Proposition 2.13 [27, Corollary 2.16] *For any C_2–equivariant spectrum E, the even slices are determined by the formula*

$$P^{2n}_{2n}(E) = \Sigma^{n\rho} H P^0 \underline{\pi}_{n\rho}(E).$$

In particular, if $\underline{\pi}_{n\rho}(E)$ is constant, we have

$$P^{2n}_{2n}(E) = \Sigma^{n\rho} H\underline{\pi_{2n}(E)}.$$

Knowledge of the slices is important because of the slice spectral sequence

$$E_2^{s,t} = \underline{\pi}_{t-s} P^t_t X \implies \underline{\pi}_{t-s} X,$$

which we will always depict in Adams notation, where $E_2^{s,t}$ is in the spot $(t-s, s)$.

We need several Mackey functors. We will define them via a Lewis diagram, stacking the value of the Mackey functor at C_2/C_2 over that of $C_2/\{e\}$ and then drawing in the restriction map, the transfer map, and the action of the nontrivial element of the Weyl group.

Definition 2.14 Let $\underline{G}$, $\underline{\mathbb{Z}}_-$ and $\underline{\mathbb{Z}}^*$ be the Mackey functors defined by the following:

$$\begin{tikzcd} \underline{M}(C_2/C_2)\colon \arrow[d, bend right, "\mathrm{res}"'] \\ \underline{M}(C_2/e)\colon \arrow[u, bend right, "\mathrm{tr}"'] \arrow[loop below, "\gamma"] \end{tikzcd} \qquad \begin{tikzcd} \mathbb{Z}/2 \arrow[d, bend right] \\ 0 \arrow[u, bend right] \arrow[loop below] \end{tikzcd} \qquad \begin{tikzcd} 0 \arrow[d, bend right] \\ \mathbb{Z} \arrow[u, bend right] \arrow[loop below, "-"] \end{tikzcd} \qquad \begin{tikzcd} \mathbb{Z} \arrow[d, bend right, "2"'] \\ \mathbb{Z} \arrow[u, bend right, "1"'] \arrow[loop below, "1"] \end{tikzcd}$$

$$\underline{M}\colon \qquad\qquad \underline{G} \qquad\qquad \underline{\mathbb{Z}}_- \qquad\qquad \underline{\mathbb{Z}}^*$$

Lemma 2.15 *If X is a C_2–spectrum such that*

(1) $\underline{\pi}_{n\rho-1}X = 0$ *for all n, and*

(2) $\underline{\pi}_{n\rho}X = \underline{\mathbb{Z}} \otimes \pi_{2n}X$, *where $\pi_{2n}X$ has no 2–torsion,*

then we have

$$\underline{\pi}_{k\rho+1}X = \underline{G} \otimes_{\mathbb{Z}} \pi_{2k+2}X, \qquad \underline{\pi}_{k\rho}X = \underline{\mathbb{Z}} \otimes_{\mathbb{Z}} \pi_{2k}X,$$
$$\underline{\pi}_{k\rho-1}X = 0, \qquad \underline{\pi}_{k\rho-2}X = \underline{\mathbb{Z}}_- \otimes_{\mathbb{Z}} \pi_{2k-2}X.$$

Proof To simplify notation, let $A_k = \pi_{2k}X$, let $\underline{A}_k = \underline{\mathbb{Z}} \otimes A_k$, let $\underline{A}_k^- = \underline{\mathbb{Z}}_- \otimes A_k$, and let $B_k = \underline{G} \otimes A_k$. By assumption, we have $P^{2k-1}_{2k-1}X \simeq *$ and

$$P^{2k}_{2k}X \simeq S^{k\rho} \wedge H(\underline{A}_k).$$

Smashing the slice tower for X with $S^{-k\rho}$ gives the slice tower for $\Sigma^{-k\rho} \wedge X$, and this again has the property that the odd slices vanish and the even ones are of the above form. It therefore suffices to prove this for $k = 0$. The homotopy Mackey functors in question are all especially simple, as they are in the region where the can be no differentials in the slice spectral sequence, as we will see.

By the connectivity of the regular representation spheres, the $(2m)^{\text{th}}$ slice does not contribute to $\underline{\pi}_i X$ for $i = -2, -1, 0, 1$ and $m < -2$ or $m > 1$. Similarly, $\underline{H}_{-2}(S^{-2\rho}; \underline{\mathbb{Z}} \otimes A_k) = 0$ for any abelian group A_k (this is the essential part of the gap theorem in [28]), so the $(-4)^{\text{th}}$ slice does not contribute to these homotopy Mackey functors either. The cell structures for representation spheres then show that the slice E_2–term has the form depicted in Figure 1.

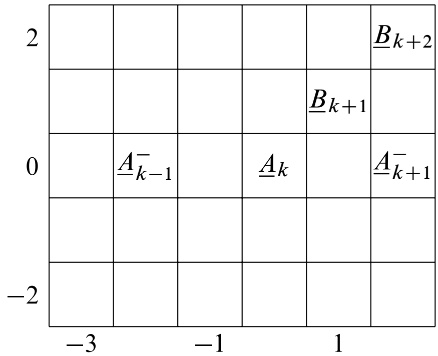


Figure 1: The slice E_2–term for $S^{-k\rho} \wedge X$, which is Adams graded, with a group in position $(t-s, s)$ recording $\underline{\pi}_{t-s}(P^t_t S^{-k\rho} \wedge X)$

In particular, there is no room for differentials or extensions in the range considered, and the result follows. □

3 Real orientations and real Landweber exactness

In this section, we will first treat some basics about real orientations. Then we will prove a real version of the Landweber exact functor theorem, both in classical and in stack language. In the last subsection, we define what we mean by forms of $BP\mathbb{R}\langle n\rangle$ and $E\mathbb{R}(n)$ and apply the real Landweber exact functor theorem to the latter.

3.1 Basics

Given a C_2–spectrum $E\mathbb{R}$, we denote by $E\mathbb{R}_\star(X)$ the value of the associated RO(C_2)–graded homology theory on a C_2–spectrum X and we set $E\mathbb{R}_\star = E\mathbb{R}_\star(\mathrm{pt})$. This is the value at C_2/C_2 of the associated Mackey functor valued homology.

Definition 3.1 A C_2–spectrum $E\mathbb{R}$ is *even* if $\underline{\pi}_{k\rho-1} E\mathbb{R} = 0$ for all $k \in \mathbb{Z}$. It is called *strongly even* if additionally $\underline{\pi}_{k\rho} E\mathbb{R}$ is a constant Mackey functor for all $k \in \mathbb{Z}$, ie if the restriction

$$\pi^{C_2}_{k\rho} E\mathbb{R} \to \pi^e_{k\rho} E\mathbb{R} \cong \pi^e_{2k} E\mathbb{R}$$

is an isomorphism.

For example, by Hu and Kriz [32, Theorem 4.11], the real bordism spectra $M\mathbb{R}$ and $BP\mathbb{R}$ are strongly even (see also Appendix A of [22] for an alternative exposition). These C_2–spectra were introduced by Landweber [36] and Araki [3] and modern treatments can be found in [32, Section 2] and [58, Example 2.14].

Recall the following definition:

Definition 3.2 Let X be a C_2–spectrum. A *real orientation* for $E\mathbb{R}$ is a class

$$x \in E\mathbb{R}^{\rho}(\mathbb{CP}^{\infty}) = [\mathbb{CP}^{\infty}, S^{\rho} \wedge E\mathbb{R}]^{C_2},$$

restricting to the class in $E\mathbb{R}^{\rho}(\mathbb{CP}^1) \cong [\mathbb{CP}^1, S^{\rho} \wedge E\mathbb{R}]^{C_2}$ corresponding to

$$1 \in [S^0, E\mathbb{R}]^{C_2} \cong [S^{\rho}, S^{\rho} \wedge E\mathbb{R}]^{C_2}$$

under the (chosen) isomorphism $S^{\rho} = \mathbb{CP}^1$. Here, we view $\mathbb{CP}^n$ as a C_2–space via complex conjugation.

By [32, Theorem 2.25], real orientations of commutative C_2–ring spectra are in one-to-one correspondence with homotopy classes of maps $M\mathbb{R} \to E\mathbb{R}$ of C_2–ring spectra, where ring spectra are understood to be up to homotopy. Another point of view uses the notion of a real vector bundle, ie a complex vector bundle $p\colon V \to X$ on a C_2–space together with an antilinear involution such that p is C_2–equivariant. If E is real oriented, then every real vector bundle carries a canonical E–orientation.

Lemma 3.3 *Every even C_2–spectrum $E\mathbb{R}$ is real orientable.*

Proof We have cofiber sequences

$$S^{(n+1)\rho-1} \to \mathbb{CP}^n \to \mathbb{CP}^{n+1}.$$

The long exact sequence in cohomology then shows that the map

$$E\mathbb{R}^{\rho}(\mathbb{CP}^{n+1}) \to E\mathbb{R}^{\rho}(\mathbb{CP}^n)$$

is surjective. The Milnor sequence gives the result. □

It is part of our philosophy that the Mackey functor $\underline{\pi}_{k\rho}$ behaves often much better than the integral Mackey functor $\underline{\pi}_{2k}$. The following is a weak version of a Whitehead theorem using $\underline{\pi}_{k\rho}$. We will formulate it in the language of equivariant homology theories as this will be convenient for our use in the real Landweber exact functor theorem.

Lemma 3.4 *Let $f\colon E\mathbb{R} \to F\mathbb{R}$ be a natural transformation of C_2–equivariant homology theories. Denote the underlying homology theories by E and F. Assume that f induces isomorphisms*

$$E\mathbb{R}_{k\rho} \to F\mathbb{R}_{k\rho} \quad \text{and} \quad E_k \to F_k$$

for all $k \in \mathbb{Z}$. Assume furthermore that $E\mathbb{R}_{k\rho-1} \to F\mathbb{R}_{k\rho-1}$ is mono for all $k \in \mathbb{Z}$ (this is the case, for example, if $E\mathbb{R}_{k\rho-1} = 0$). Then f is a natural isomorphism.

Proof It is well known that it is enough to show that $E_k \to F_k$ and $E\mathbb{R}_k \to E\mathbb{R}_k$ are isomorphisms for all $k \in \mathbb{Z}$. As the former is true by assumption, it is in particular enough to show that $f_{a+b\sigma}\colon E\mathbb{R}_{a+b\sigma} \to F\mathbb{R}_{a+b\sigma}$ is an isomorphism for all $a, b \in \mathbb{Z}$. This is true for $a = b$ again by assumption.

Smashing the cofiber sequence

$$(C_2)_+ \to S^0 \to S^\sigma$$

with $S^{a+b\sigma}$ gives the cofiber sequence

$$(C_2)_+ \wedge S^{a+b\sigma} \to S^{a+b\sigma} \to S^{a+(b+1)\sigma}.$$

We have a map between the associated long exact sequences:

$$\begin{tikzcd}
E_{a+b+1} \arrow[r] \arrow[d, "\cong"] & E\mathbb{R}_{a+(b+1)\sigma} \arrow[r] \arrow[d, "f_{a+(b+1)\sigma}"] & E\mathbb{R}_{a+b\sigma} \arrow[r] \arrow[d, "f_{a+b\sigma}"] & E_{a+b} \arrow[r] \arrow[d, "\cong"] & E\mathbb{R}_{(a-1)+(b+1)\sigma} \arrow[d, "f_{(a-1)+(b+1)\sigma}"] \\
F_{a+b+1} \arrow[r] & F\mathbb{R}_{a+(b+1)\sigma} \arrow[r] & F\mathbb{R}_{a+b\sigma} \arrow[r] & F_{a+b} \arrow[r] & F\mathbb{R}_{(a-1)+(b+1)\sigma}
\end{tikzcd}$$

The weak five lemma implies the following statements:

(M1) If $f_{a+(b+1)\sigma}$ is mono, then $f_{a+b\sigma}$ is mono.

(M2) If $f_{(a+1)+b\sigma}$ is epi and $f_{a+b\sigma}$ is mono, then $f_{a+(b+1)\sigma}$ is mono.

(E1) If $f_{a+b\sigma}$ is epi, then $f_{a+(b+1)\sigma}$ is epi.

(E2) If $f_{(a-1)+(b+1)\sigma}$ is mono and $f_{a+(b+1)\sigma}$ is epi, then $f_{a+b\sigma}$ is epi.

These imply the following four statements in turn:

(1) By hypothesis $f_{a+a\sigma} = f_{a\rho}$ is epi for all a, and hence repeated application of E1 shows that $f_{a+b\sigma}$ is epi for $b \geq a$.

(2) By hypothesis $f_{(a-1)+a\sigma} = f_{a\rho-1}$ is mono for all a, and hence $f_{a+b\sigma}$ is mono for $b \leq a+1$ by repeated application of M1.

Note that the regions in which $f_{a+b\sigma}$ is epi and mono overlap in two diagonals, allowing us to proceed.

(3) By repeated application of E2 we conclude that $f_{a+b\sigma}$ is epi for all a, b.

(4) By repeated application of M2 we conclude that $f_{a+b\sigma}$ is mono for all a, b.

Accordingly $f_{a+b\sigma}$ is both epi and mono for all a, b and the proof is complete. □

3.2 Real Landweber exactness

In this section, we want to prove a version of the Landweber exact functor theorem using the real bordism spectrum $M\mathbb{R}$.

The restriction maps $M\mathbb{R}_{k\rho} \to MU_{2k}$ are isomorphisms by [32, Theorem 2.28]. This defines a graded ring morphism $MU_{2*} \to M\mathbb{R}_\star$ along the morphism

$$2\mathbb{Z} \to \mathrm{RO}(C_2), \quad 2k \mapsto k\rho$$

of the monoids indexing the grading. In particular, $M\mathbb{R}_\star$ becomes a graded MU_{2*}–module in a suitable sense.

Definition 3.5 Let $E\mathbb{R}$ be a strongly even C_2–spectrum with underlying spectrum E. Then $E\mathbb{R}$ is called *real Landweber exact* if for every real orientation $M\mathbb{R} \to E\mathbb{R}$ the induced map

$$M\mathbb{R}_\star(X) \otimes_{MU_{2*}} E_{2*} \to E\mathbb{R}_\star(X)$$

is an isomorphism for every C_2–spectrum X.

Here, the gradings can be parsed in the following way: For every $k \in \mathbb{Z}$, we have a $2\mathbb{Z}$–graded MU_{2*}–module $M\mathbb{R}_{k+*\rho}(X)$ in the way described above so that the expression $M\mathbb{R}_{k+*\rho}(X) \otimes_{MU_{2*}} E_{2*}$ makes sense in the world of $2\mathbb{Z} = \rho\mathbb{Z}$–graded MU_{2*}–modules. Now observe that $\mathrm{RO}(C_2)$ is a free abelian group generated by ρ and 1; thus an $\mathrm{RO}(C_2)$–graded abelian group is an equivalent datum to a $\mathbb{Z}$–graded $\mathbb{Z}\rho$–graded abelian group and this expresses what $M\mathbb{R}_\star(X) \otimes_{MU_{2*}} E_{2*}$ means.

Theorem 3.6 (real Landweber exact functor theorem) (a) *Let E_{2*} be a graded Landweber exact MU_{2*}–algebra, concentrated in even degrees. Then*

$$X \mapsto M\mathbb{R}_\star(X) \otimes_{MU_{2*}} E_{2*}$$

is a C_2–equivariant homology theory.

(b) *Let $E\mathbb{R}$ be a strongly even C_2–spectrum whose underlying spectrum E is Landweber exact. Then $E\mathbb{R}$ is real Landweber exact.*

Let us shortly recall how Landweber exactness is treated nonequivariantly from the stacky point of view. Good sources are, for example, [20], [30] or Lectures 11 and 15 of [41].

The stack associated to the graded Hopf algebroid $(MU_{2*}, MU_{2*}MU)$ is $\mathcal{M}_{FG}$, the moduli stack of formal groups. This implies that the category of quasicoherent sheaves on $\mathcal{M}_{FG}$ is equivalent to that of evenly graded $(MU_{2*}, MU_{2*}MU)$–comodules (see for example [54, Remark 34]). The graded comodule MU_{*+2} corresponds to a line bundle ω on $\mathcal{M}_{FG}$. This allows us to define the *graded global sections* $\Gamma_{2*}(\mathcal{F})$ of a quasicoherent sheaf $\mathcal{F}$ on $\mathcal{M}_{FG}$ as $\Gamma(\mathcal{F}\otimes\omega^{\otimes *})$. Likewise, the category of quasicoherent sheaves on $(\mathrm{Spec}\ E_{2*})/\mathbb{G}_m$ is equivalent to that of evenly graded modules over E_{2*}; more precisely, a quasicoherent sheaf $\mathcal{F}$ on $(\mathrm{Spec}\ E_{2*})/\mathbb{G}_m$ corresponds to the graded module $\Gamma_{2*}(\mathcal{F}) = \Gamma(\mathcal{F}\otimes\omega_E^{\otimes *})$, where ω_E corresponds to the graded module E_{2*+2}. We remind the reader here that $(\mathrm{Spec}\ E_{2*})/\mathbb{G}_m$ denotes (as always) the stack quotient.

An MU_{2*}–algebra E_{2*} is Landweber exact if and only if the composite

$$f\colon \mathrm{Spec}\, E_{2*}/\mathbb{G}_m \to \mathrm{Spec}\, MU_{2*}/\mathbb{G}_m \to \mathcal{M}_{FG}$$

is flat (if the Landweber exactness criterion is phrased classically using the v_i, this is the nonformal part of the proof). Given a spectrum X, define quasicoherent sheaves $\mathcal{F}_i^X$ for $i=0,1$ on $\mathcal{M}_{FG}$ corresponding to the graded $(MU_{2*}, MU_{2*}MU)$–comodules $MU_{2*+i}X$. These are functors in X and define ungraded homology theories on spectra with values in quasicoherent sheaves on $\mathcal{M}_{FG}$. Because f is flat and thus f^* is exact, the functors

$$X \mapsto f^*\mathcal{F}_i^X$$

define homology theories with values in quasicoherent sheaves on $(\mathrm{Spec}\ E_{2*})/\mathbb{G}_m$. We want to identify $\Gamma_{2*}(f^*\mathcal{F}_i^X)$ with $MU_{2*+i}(X)\otimes_{MU_{2*}} E_{2*}$. The following lemma provides this identification and thus completes the proof of nonequivariant Landweber exactness.

Lemma 3.7 *Let $\mathcal{F}$ be a quasicoherent sheaf on $\mathcal{M}_{FG}$. Then*

$$\mathcal{F}(\mathrm{Spec}\, MU_{2*})\otimes_{MU_{2*}} E_{2*} \cong \Gamma_{2*}((\mathrm{Spec}\, E_{2*})/\mathbb{G}_m; f^*\mathcal{F}),$$

where we view $\mathcal{F}(\mathrm{Spec}\, MU_{2*})$ as an evenly graded MU_{2*}–module. These isomorphisms are natural in $\mathcal{F}$.

Proof We have a commutative diagram:

$$\begin{tikzcd} & \mathrm{Spec}\, MU_{2*}/\mathbb{G}_m \arrow[d, "q"] \\ \mathrm{Spec}\, E_{2*}/\mathbb{G}_m \arrow[ur, "g"] \arrow[r, "f"] & \mathcal{M}_{FG} \end{tikzcd}$$

By definition, $q^*\mathcal{F}$ corresponds to the evenly graded MU_{2*}–module $\mathcal{F}(\mathrm{Spec}\, MU_{2*})$. Therefore, we have that $f^*\mathcal{F}\cong g^*q^*\mathcal{F}$ corresponds to the evenly graded E_{2*}–module $\mathcal{F}(\mathrm{Spec}\, MU_{2*})\otimes_{MU_{2*}} E_{2*}$, proving the lemma. $\square$

Now we turn to the proof of Theorem 3.6. Part (a) of it can be proven analogously to Landweber exactness in the motivic setting as in [55], though we follow their approach only loosely. The crucial fact about $M\mathbb{R}$ is the following lemma, which was already implicitly treated in [32] and can also be found in [28].

Lemma 3.8 *The restriction*

$$(M\mathbb{R}_{*\rho}, M\mathbb{R}_{*\rho}M\mathbb{R}) \to (MU_{2*}, MU_{2*}MU)$$

defines an isomorphism of Hopf algebroids.

Proof It is clear that restriction defines a morphism of Hopf algebroids. It is left to show that $M\mathbb{R}_{*\rho}M\mathbb{R} \to MU_{2*}MU$ is an isomorphism.

Let $\mathcal{C}$ be the class of all pointed C_2–spaces and $\mathcal{C}^{\mathrm{st}}$ the class of all (genuine) C_2–spectra X such that

$$M\mathbb{R}_{*\rho}(X) \to MU_{2*}(X)$$

is an isomorphism and

$$M\mathbb{R}_{*\rho-1}(X) \to MU_{2*-1}(X)$$

is a monomorphism, where homology is understood to be reduced in the unstable case. Observe first that $S^0 \in \mathcal{C}$ and $X \in \mathcal{C}$ if and only if $\Sigma^\infty X \in \mathcal{C}^{\mathrm{st}}$. Furthermore, we have the following closure properties:

- Both $\mathcal{C}$ and $\mathcal{C}^{\mathrm{st}}$ are closed under weak equivalences and filtered homotopy colimits.
- If $X \in \mathcal{C}^{\mathrm{st}}$ and $S^{k\rho-1} \to X$ is a map, then its cofiber is also in $\mathcal{C}^{\mathrm{st}}$ as follows by the five lemma and from $M\mathbb{R}$ being strongly even.
- If $X \in \mathcal{C}$ and $V \to X$ is a real vector bundle, then the Thom space X^V is also in $\mathcal{C}$, as $M\mathbb{R}$ is real-oriented.
- If $X \in \mathcal{C}^{\mathrm{st}}$, then $\Sigma^{k\rho}X \in \mathcal{C}^{\mathrm{st}}$ for every $k \in \mathbb{Z}$ as well.

We will demonstrate that these properties imply that $M\mathbb{R} \in \mathcal{C}^{\mathrm{st}}$.

Depending on the model of $M\mathbb{R}$ of choice it is either easy to see or a theorem [28, (B.252)] that we can write $M\mathbb{R}$ as a directed homotopy colimit over $\Sigma^{-n\rho}MU(n)$, where $MU(n)$ is the suspension spectrum of the Thom space $BU(n)_+^{\gamma_n}$ with the C_2–action by complex conjugation (which gives the universal bundle γ_n the structure of a real bundle). The Grassmannian $BU(n)$ is a directed homotopy colimit of finite dimensional Grassmannians, which are built of cells of dimension $k\rho = k\mathbb{C}$ by the theory of Schubert cells. Thus, $M\mathbb{R}$ is in $\mathcal{C}^{\mathrm{st}}$. □

Proof of Theorem 3.6 We begin by defining, given a C_2–spectrum X, quasicoherent sheaves $\mathcal{F}_i^X$ for $i \in \mathbb{Z}$ on $\mathcal{M}_{FG}$ corresponding to the graded $(MU_{2*}, MU_{2*}MU) \cong (M\mathbb{R}_{*\rho}, M\mathbb{R}_{*\rho}M\mathbb{R})$–comodules $M\mathbb{R}_{*\rho+i}X$. As above, the $\mathcal{F}_i^X$ are C_2–equivariant homology theories with values in quasicoherent sheaves on $\mathcal{M}_{FG}$. Thus, the pullbacks $f^*\mathcal{F}_i^X$ are homology theories with values in quasicoherent sheaves on Spec $E_{2*}/\mathbb{G}_m$. By Lemma 3.7, the associated graded module is $M\mathbb{R}_{*\rho+i}(X) \otimes_{MU_{2*}} E_{2*}$, which is thus a homology theory as well; this proves part (a) of real Landweber exactness. Note that as $M\mathbb{R}_{*\rho+i}(X) \otimes_{MU_{2*}} E_{2*}$ has suspension isomorphisms for arbitrary (virtual) representations, it is isomorphic to the $\mathrm{RO}(C_2)$–graded theory associated to its degree-0 part.

For the proof of (b), choose a real orientation $M\mathbb{R} \to E\mathbb{R}$, which exists by Lemma 3.3. By Lemma 3.4 it is now enough to show that the induced maps

$$M\mathbb{R}_{*\rho} \otimes_{MU_{2*}} E_{2*} \to E\mathbb{R}_{*\rho}$$

and

$$MU_{2*} \otimes_{MU_{2*}} E_{2*} \to E_{2*}$$

are isomorphisms (as the odd groups are zero anyhow). The latter is clear and the former is true since both $\underline{\pi}_{*\rho}M\mathbb{R}$ and $\underline{\pi}_{*\rho}E\mathbb{R}$ are constant. □

By the following proposition, the real Landweber exact functor theorem can actually be used to produce C_2–spectra.

Theorem 3.9 *Any (ungraded) G–equivariant homology theory can be represented by a G–spectrum, ie for every G–equivariant homology theory h_0, there is a G–spectrum E such that there are isomorphisms $\pi_0^G(X \wedge E) \cong h_0(X)$, natural in a G–spectra X. Note that this implies natural isomorphisms $\pi_\star^G(X \wedge E) \cong h_\star(X)$ as well.*

Moreover, any transformation of G–equivariant homology theories can be represented by a map of G–spectra.

Proof By [31, Corollary 9.4.4], the homotopy category of genuine G–spectra is a Brown category, which means exactly the statement of our proposition. □

In the rest of the section, we will give some reformulations of the stacky point of view on Landweber exactness to show that two real Landweber exact spectra are equivalent if and only if their underlying spectra are equivalent. The following easy lemma will be useful.

Lemma 3.10 *Let $f\colon \mathcal{X} \to \mathcal{Y}$ be an affine morphism of algebraic stacks (in the sense of [54, Definition 6]) and $\mathcal{F}$ be a quasicoherent sheaf on $\mathcal{X}$ and $\mathcal{G}$ be a quasicoherent sheaf on $\mathcal{Y}$. Then the canonical homomorphism*

$$f_*\mathcal{F} \otimes_{\mathcal{O}_\mathcal{Y}} \mathcal{G} \to f_*(\mathcal{F} \otimes_{\mathcal{O}_\mathcal{X}} f^*\mathcal{G})$$

is an isomorphism.

Proof We can assume that $\mathcal{Y}$ is affine and hence also $\mathcal{X}$. Then it is clear. □

Recall the notation $\mathcal{F}_i^X$ from the proof of (real) Landweber exactness above.

Proposition 3.11 (a) *Let E be an even Landweber exact spectrum. The associated graded formal group on E_{2*} defines a map*

$$f\colon (\operatorname{Spec} E_{2*})/\mathbb{G}_m \to (\operatorname{Spec} MU_{2*})/\mathbb{G}_m \to \mathcal{M}_{FG}.$$

Then given a spectrum X, we have

$$E_*(X) \cong \Gamma_{2*}(\mathcal{M}_{FG}; \mathcal{F}_*^X \otimes_{\mathcal{O}_{\mathcal{M}_{FG}}} f_*\mathcal{O}_{(\operatorname{Spec} E_{2*})/\mathbb{G}_m}).$$

(b) *Let $E\mathbb{R}$ be an even real Landweber exact spectrum. The associated graded formal group on $E\mathbb{R}_{*\rho} \cong E_{2*}$ (for E the underlying spectrum) defines a map*

$$f\colon (\operatorname{Spec} E_{2*})/\mathbb{G}_m \to \mathcal{M}_{FG}$$

as above. Then given a C_2–spectrum X, we have

$$E\mathbb{R}_\bigstar(X) \cong \Gamma_{2*}(\mathcal{M}_{FG}; \mathcal{F}_*^X \otimes_{\mathcal{O}_{\mathcal{M}_{FG}}} f_*\mathcal{O}_{(\operatorname{Spec} E_{2*})/\mathbb{G}_m}),$$

where $\Gamma_{2n}(\mathcal{M}_{FG}; \mathcal{F}_i^X \otimes \cdots)$ is in degree $n\rho + i$.

Proof We will prove only part (a); the proof of part (b) only needs change in notation. As in the proof of Landweber exactness, the left-hand side decomposes into two pieces of the form $\Gamma_{2*}(f^*\mathcal{F}_i^X) \cong \Gamma_{2*}(f_*f^*\mathcal{F}_i^X)$. Thus, we only have to show that

$$f_*f^*\mathcal{F}_i^X \cong \mathcal{F}_i^X \otimes_{\mathcal{O}_{\mathcal{M}_{FG}}} f_*\mathcal{O}_{(\operatorname{Spec} E_{2*})/\mathbb{G}_m},$$

which follows directly from Lemma 3.10 with $\mathcal{F} = \mathcal{O}_{(\operatorname{Spec} E_{2*})/\mathbb{G}_m}$ and $\mathcal{G} = \mathcal{F}_i^X$. □

In particular, we see that the values of a (real) Landweber exact theory do not depend on the MU_{2*}–module structure of E_{2*}, but only on the graded quasicoherent sheaf $f_*\mathcal{O}_{(\operatorname{Spec} E_{2*})/\mathbb{G}_m}$ on $\mathcal{M}_{FG}$ defined by E_{2*}. This sheaf has an alternative description:

Lemma 3.12 *Let E be an even Landweber exact spectrum and $f\colon \operatorname{Spec} E_{2*}/\mathbb{G}_m \to \mathcal{M}_{FG}$ as above. Then we have an isomorphism $f_*\mathcal{O}_{(\operatorname{Spec} E_{2*})/\mathbb{G}_m} \cong \mathcal{F}_0^E$.*

Proof This was proven in the even-periodic context in the proof of [50, Proposition 2.4]. The general case is similar. ☐

Proposition 3.13 *Let $E\mathbb{R}$ and $F\mathbb{R}$ be two (strongly even) real Landweber exact C_2–spectra, whose underlying spectra E and F are equivalent. Then $E\mathbb{R}$ and $F\mathbb{R}$ are equivalent.*

Proof Assume $E \simeq F$. Then E and F define isomorphic graded quasicoherent sheaves $\mathcal{F}_0^E$ and $\mathcal{F}_0^F$ on $\mathcal{M}_{FG}$. Since $E\mathbb{R}$ and $F\mathbb{R}$ are real Landweber exact, Proposition 3.11 and Lemma 3.12 imply the following chain of isomorphisms, natural in a C_2–spectrum X:

$$\begin{aligned} E\mathbb{R}_0(X) &\cong \Gamma(\mathcal{M}_{FG}; \mathcal{F}_0^X \otimes_{\mathcal{O}_{\mathcal{M}_{FG}}} \mathcal{F}_0^E) \\ &\cong \Gamma(\mathcal{M}_{FG}; \mathcal{F}_0^X \otimes_{\mathcal{O}_{\mathcal{M}_{FG}}} \mathcal{F}_0^F) \\ &\cong F\mathbb{R}_0(X). \end{aligned}$$

Thus, the (ungraded) C_2–equivariant homology theories defined by $E\mathbb{R}$ and by $F\mathbb{R}$ are isomorphic. By Theorem 3.9, a natural isomorphism of C_2–equivariant homology theories induces an equivalence of the representing C_2–spectra. ☐

3.3 Forms of $BP\mathbb{R}\langle n\rangle$ and $E\mathbb{R}(n)$

Fix a prime p.

Definition 3.14 Let E be a complex oriented p–local commutative and associative ring spectrum (up to homotopy). The p–typification of its formal group law defines a ring morphism $BP_* \to E_*$.

(a) We call E a *form of $BP\langle n\rangle$* if the map

$$\mathbb{Z}_{(p)}[v_1, \dots, v_n] \subset BP_* \to E_*$$

is an isomorphism. This does not depend on the choice of v_i.

(b) We call E a *form of $E(n)$* if there is a choice of indecomposables $v_1, \dots, v_n \in BP_*$ with $|v_i| = 2(p^i - 1)$ such that the image of v_n under the homomorphism

$$\mathbb{Z}_{(p)}[v_1, \dots, v_n] \subset BP_* \to E_*$$

is invertible and the induced morphism $\mathbb{Z}_{(p)}[v_1, \dots, v_n, v_n^{-1}] \to E_*$ is an isomorphism.

Spectra as in (a) are also sometimes called *generalized $BP\langle n\rangle$* (see [37, Definition 4.1]). There is a real analogue, where we specialize to $p = 2$:

Definition 3.15 Let $E\mathbb{R}$ be an even real oriented 2–local commutative and associative C_2–ring spectrum (up to homotopy). This induces a formal group law on $E\mathbb{R}_\rho$ [32, Theorem 2.10]; its 2–typification defines a map $BP_{2*} \cong BP\mathbb{R}_{*\rho} \to E\mathbb{R}_{*\rho}$.

(a) We call $E\mathbb{R}$ a *form of* $BP\mathbb{R}\langle n\rangle$ if the map

$$\underline{\mathbb{Z}}_{(2)}[\bar{v}_1, \dots, \bar{v}_n] \subset \underline{\pi}_{*\rho} BP\mathbb{R} \to \underline{\pi}_{*\rho} E\mathbb{R}$$

is an isomorphism of constant Mackey functors. This does not depend on the choice of $\bar{v}_i$.

(b) We call $E\mathbb{R}$ a *form of* $E\mathbb{R}(n)$ if there is a choice of indecomposables $\bar{v}_1, \dots, \bar{v}_n \in BP\mathbb{R}_{*\rho}$ with $|\bar{v}_i| = (2^i - 1)\rho$ such that the image of $\bar{v}_n$ under the homomorphism

$$\mathbb{Z}_{(2)}[\bar{v}_1, \dots, \bar{v}_n] \subset BP\mathbb{R}_{*\rho} \to E\mathbb{R}_{*\rho}$$

is invertible and the induced morphism $\underline{\mathbb{Z}}_{(2)}[\bar{v}_1, \dots, \bar{v}_n, \bar{v}_n^{-1}] \to \underline{\pi}_{*\rho} E\mathbb{R}$ is an isomorphism of constant Mackey functors.

Note that a form $E\mathbb{R}(n)$ is always real Landweber exact by Theorem 3.6 as it is strongly even and its underlying spectrum is Landweber exact.

Proposition 3.16 *If for two forms of $E\mathbb{R}(n)$ their underlying spectra are equivalent, then they are equivalent as C_2–spectra.*

Proof As every form of $E\mathbb{R}(n)$ is real Landweber exact, this follows directly from Proposition 3.13. □

4 $\mathrm{TMF}_1(3)$ and friends

In this section, we will first define the versions of TMF we are after and compute $\pi_* \mathrm{Tmf}_1(3)$. In Section 4.2, we will run the homotopy fixed point spectral sequence for $\mathrm{tmf}_1(3)^{hC_2}$ and apply this to see that $\mathrm{tmf}_1(3)$ is a form of $BP\mathbb{R}\langle 2\rangle$. In Section 4.3, we will discuss the relationship between the C_2–spectra $\mathrm{tmf}_1(3)$, $\mathrm{Tmf}_1(3)$ and $\mathrm{TMF}_1(3)$. In particular, we will show that $\mathrm{TMF}_1(3) \simeq \mathrm{tmf}_1(3)[\overline{\Delta}^{-1}]$ and how this implies the real Landweber exactness of $\mathrm{TMF}_1(3)$.

4.1 Basics

Denote by $\mathcal{M}_{\mathrm{ell}}$ the moduli stack of elliptic curves and by $\overline{\mathcal{M}}_{\mathrm{ell}}$ its compactification. Mapping an elliptic curve to its formal group defines a flat map $\overline{\mathcal{M}}_{\mathrm{ell}} \to \mathcal{M}_{FG}$ to the moduli stack of formal groups. By [29] (extending earlier work by Goerss, Hopkins and Miller), the induced presheaf of even-periodic Landweber exact homology theories refines to a sheaf of E_∞–ring spectra $\mathcal{O}^{\mathrm{top}}$ on the log-étale site of $\overline{\mathcal{M}}_{\mathrm{ell}}$.

Denote by $\mathcal{M}_1(n)$ the moduli stack of elliptic curves with one chosen point of exact order n and by $\overline{\mathcal{M}}_1(n)$ its compactification, whose definition we will now review. In the compactification we have to allow not only (smooth) elliptic curves, but *generalized elliptic curves*, which can have as fibers also Néron m–gons for $m \mid n$. These are obtained by gluing m copies of $\mathbb{P}^1$, where 0 in the i^{th} $\mathbb{P}^1$ (for $i \in \mathbb{Z}/m\mathbb{Z}$) is attached to ∞ in the $(i+1)^{\mathrm{st}}$. For precise definitions see Deligne and Rapoport [17, Section II.1].

Definition 4.1 [17, IV.4.11–4.15; 16] We define the stack $\overline{\mathcal{M}}_1(n)$ to classify generalized elliptic curves $p\colon \mathcal{E} \to S$ over a base S with n invertible, together with an injection of group schemes $\mathbb{Z}/n\mathbb{Z} \to \mathcal{E}^\circ$ from the constant group scheme $\mathbb{Z}/n\mathbb{Z}$ over S into the smooth locus of $\mathcal{E}$ such that

(1) each geometric fiber $\operatorname{Spec} \bar{k} \times_S \mathcal{E}$ of p is either smooth or a Néron m–gon for some $m \mid n$, and
(2) the image of $\mathbb{Z}/n\mathbb{Z}$ intersects each irreducible component in every geometric fiber of $\mathcal{E}$ nontrivially.

We define

$$\mathrm{TMF}_1(n) = \mathcal{O}^{\mathrm{top}}(\mathcal{M}_1(n)),$$
$$\mathrm{Tmf}_1(n) = \mathcal{O}^{\mathrm{top}}(\overline{\mathcal{M}}_1(n)),$$
$$\mathrm{tmf}_1(n) = \tau_{\geq 0}\,\mathrm{Tmf}_1(n).$$

We remark that the last definition should only be considered appropriate for $n \geq 2$ if $\mathrm{tmf}_1(n)$ is even and $\pi_{2n}\,\mathrm{tmf}_1(n)$ is isomorphic to the ring of integral holomorphic modular forms $H^0(\overline{\mathcal{M}}_1(n); \omega^{\otimes *})$. The second assumption is always fulfilled, but in general there can be a nontrivial $\pi_1\,\mathrm{tmf}_1(n)$, which is isomorphic to $H^1(\overline{\mathcal{M}}_1(n); \omega)$ (as already remarked in [29, Remark 6.4]). Luckily, there are no such problems for $\mathrm{tmf}_1(3)$ as we will see at the end of this subsection.

The following lemma is well known:

Lemma 4.2 *The spectrum* $\mathrm{TMF}_1(n)$ *is Landweber exact for* $n \geq 2$.

Proof Throughout the proof, we will use the notations ω and $\mathcal{F}_i^X$ (for $i = 0$, 1 and X a spectrum) from Section 3.2.

First, we prove that for $\operatorname{Spec} A \to \mathcal{M}_{\mathrm{ell}}$ étale such that the pullback of ω to $\operatorname{Spec} A$ is trivial, $E = \mathcal{O}^{\mathrm{top}}(\operatorname{Spec} A)$ is Landweber exact: By the descent spectral sequence, E is even periodic. Thus, we can choose a complex orientation $MU \to E$. This defines a formal group law on $A = \pi_0 E$. By construction (see Behrens [8]), the composite morphism $g\colon \operatorname{Spec} A \to \mathcal{M}_{\mathrm{ell}} \to \mathcal{M}_{FG}$ classifies the underlying formal group. As g is flat, a version of the Landweber exact functor theorem (see Lemma 3.7 or [41, Lecture 15]) implies that the source of

$$(X \mapsto MU_*(X) \otimes_{MU_*} E_*) \to (X \mapsto E_*(X))$$

is a homology theory and thus the depicted morphism is a natural isomorphism. This proves that E is Landweber exact. Furthermore, it provides a natural isomorphism between $\pi_{2k-i}(\mathcal{O}^{\mathrm{top}} \wedge X)$ and the pullback of $\mathcal{F}_i^X \otimes \omega^{\otimes k}$ to $\mathcal{M}_{\mathrm{ell}}$ for $i = 0, 1$.

For $n \geq 4$, the stack $\mathcal{M}_1(n)$ is represented by an affine scheme [34, Corollary 2.7.3 and Scholie (4.7.0)]. For $n = 2, 3$ we have the slightly weaker statement that only $\mathcal{M}_1^1(n)$ is of the form $\operatorname{Spec} A$, where $\mathcal{M}_1^1(n)$ classifies elliptic curves where we choose not only a point of order n, but also a nowhere vanishing invariant differential; we recover $\mathcal{M}_1(n)$ as $\operatorname{Spec} A/\mathbb{G}_m$. This can either be shown along the same lines as the previous statement or deduced from concrete presentations (see eg [46, Proposition 3.2] and [7, Section 1.3]). In particular, the global sections functor

$$\Gamma\colon \operatorname{QCoh}(\mathcal{M}_1(n)) \to \mathrm{AbelianGroups}$$

on quasicoherent sheaves is exact. Indeed, $\operatorname{QCoh}(\operatorname{Spec} A/\mathbb{G}_m)$ is by Galois descent equivalent to the category of graded A–modules (where the grading comes from the $\mathbb{G}_m$–action). The global sections functor corresponds to $M_* \mapsto M_0$, which is clearly exact.

In particular, we see that the descent spectral sequence

$$H^s(\mathcal{M}_1(n); h^*\omega^{\otimes t}) \implies \pi_{2t-s}\,\mathrm{TMF}_1(n)$$

is concentrated in the 0–line, where $h\colon \mathcal{M}_1(n) \to \mathcal{M}_{FG}$ classifies the formal group. Thus, $\mathcal{M}_1(n) \simeq (\operatorname{Spec} \pi_{2*}\,\mathrm{TMF}_1(n))/\mathbb{G}_m$. By the same argument we get a natural isomorphism

$$\mathrm{TMF}_1(n)_{2k-i}(X) \cong H^0(\mathcal{M}_1(n); h^*(\mathcal{F}_i^X \otimes \omega^{\otimes k}))$$

for spectra X. Now Lemma 3.7 implies that we have isomorphisms

$$MU_*(X) \otimes_{MU_*} \mathrm{TMF}_1(n) \cong \mathrm{TMF}_1(n)_*(X),$$

again natural in X. $\square$

Sending the point x of order n to $[k]x$ for $k \in (\mathbb{Z}/n)^\times$ defines a $(\mathbb{Z}/n)^\times$–action on $\overline{\mathcal{M}}_1(n)$. In particular, this induces $(\mathbb{Z}/3)^\times = C_2$–actions on $\mathrm{TMF}_1(3)$ and $\mathrm{Tmf}_1(3)$.

Thus we will view these spectra as cofree C_2–spectra as in Section 2.2. We define the C_2–spectrum $\mathrm{tmf}_1(3)$ as the C_2–equivariant connective cover of $\mathrm{Tmf}_1(3)$ so that

$$\mathrm{tmf}_1(3)^{C_2} = \tau_{\geq 0}(\mathrm{Tmf}_1(3)^{hC_2}).$$

Note that this spectrum is not cofree as

$$\tau_{\geq 0}(\mathrm{Tmf}_1(3)^{hC_2}) \simeq \tau_{\geq 0}(\mathrm{tmf}_1(3)^{hC_2}),$$

which follows formally from the C_2–homotopy fixed point spectral sequence, and $\mathrm{tmf}_1(3)^{hC_2}$ has negative homotopy groups as we will see in the next subsection.

Denote by $\mathcal{M}_0(n)$ the moduli stack of elliptic curves with a chosen subgroup of order n and by $\overline{\mathcal{M}}_0(n)$ its compactification, defined as follows:

Definition 4.3 We define $\overline{\mathcal{M}}_0(n)$ for n squarefree[1] to classify generalized elliptic curves $p\colon \mathcal{E} \to S$ over a base S with n invertible, together with a subgroup $G \subset \mathcal{E}^\circ[n]$ such that

(1) each geometric fiber of p is either smooth or a Néron m–gon for some $m \mid n$,

(2) G is étale locally isomorphic to $\mathbb{Z}/n\mathbb{Z}$, and

(3) G intersects each irreducible component in every geometric fiber of $\mathcal{E}$ nontrivially.

We define

$$\mathrm{TMF}_0(3) = \mathcal{O}^{\mathrm{top}}(\mathcal{M}_0(3)) \quad \text{and} \quad \mathrm{Tmf}_0(3) = \mathcal{O}^{\mathrm{top}}(\overline{\mathcal{M}}_0(3)).$$

The forgetful maps

$$\mathcal{M}_1(n) \to \mathcal{M}_0(n) \quad \text{and} \quad \overline{\mathcal{M}}_1(n) \to \overline{\mathcal{M}}_0(n)$$

are $(\mathbb{Z}/n)^\times$–Galois coverings for n squarefree, as checked in [50, Theorem 7.12]. In particular, this implies that $\mathrm{Tmf}_0(3) \simeq \mathrm{Tmf}_1(3)^{hC_2}$ and $\mathrm{TMF}_0(3) \simeq \mathrm{Tmf}_1(3)^{hC_2}$. If we define $\mathrm{tmf}_0(3) = \tau_{\geq 0}\,\mathrm{Tmf}_0(3)$, then it follows that $\mathrm{tmf}_0(3) \simeq \mathrm{tmf}_1(3)^{C_2}$.

Next, we will study $\overline{\mathcal{M}}_1(3)$ in more detail. The following lemma essentially says that there can be only one reasonable compactification of our moduli stacks.

Lemma 4.4 *Let $f\colon Y \to X$ be a map of Deligne–Mumford stacks (over some base scheme S) and assume that X is locally noetherian. Let $\bar{f}_1, \bar{f}_2\colon \overline{Y}_1, \overline{Y}_2 \to X$ be finite morphisms from normal Deligne–Mumford stacks (over S) such that Y sits inside $\overline{Y}_1$ and $\overline{Y}_2$ as a dense open substack and $\bar{f}_i|_Y = f$ for $i = 1, 2$. Then $\overline{Y}_1 \simeq \overline{Y}_2$ as stacks over X.*

[1] For the subtleties for nonsquarefree n see [15].

Proof The same argument as in [21, Corollary 12.2] shows that it is enough to show that $(\overline{f}_1)_*\mathcal{O}_{\overline{Y}_1} \cong (\overline{f}_2)_*\mathcal{O}_{\overline{Y}_2}$ as both $\overline{f}_1$ and $\overline{f}_2$ are affine. In particular, it is enough to show the existence of a natural isomorphism between $\overline{Y}_1$ and $\overline{Y}_2$ if $X = \operatorname{Spec} A$ is affine. Then $\overline{Y}_1 = \operatorname{Spec} B_1$ and $\overline{Y}_2 = \operatorname{Spec} B_2$ are also affine. The inclusions $Y \subset \operatorname{Spec} B_1$ and $Y \subset \operatorname{Spec} B_2$ induce bijections of the sets of connected components. As all these schemes are normal and locally noetherian, all connected components are irreducible [21, Remark 6.3.7]. Thus, we can assume that $\overline{Y}_1$ and $\overline{Y}_2$ are irreducible and hence B_1 and B_2 are normal integral domains. Write $C = \Gamma(\mathcal{O}_Y)$. As Y is open in $\overline{Y}_i$, the map from B_i into its fraction field factors over C. In particular, B_i injects into C and is integrally closed in it. As it is also finite and thus integral over A, it consists exactly of those elements in C that are integral over A. In particular, we have a canonical isomorphism $B_1 \cong B_2$ of A–algebras. □

Note that $\overline{\mathcal{M}}_1(n) \to \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{n}\big]$ and $\overline{\mathcal{M}}_0(n) \to \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{n}\big]$ (if n is squarefree) are finite morphisms from normal (even regular) Deligne–Mumford stacks and $\mathcal{M}_1(n) \subset \overline{\mathcal{M}}_1(n)$ and $\mathcal{M}_0(n) \subset \overline{\mathcal{M}}_0(n)$ are open dense inclusions (see [17, IV.3.4] or [16, Theorem 4.1.1]; note that the complement of an effective Cartier divisor is open and dense). Thus, we can apply the previous lemma to approach the following well-known result (see eg [37]) that has to the knowledge of the authors not appeared with full proof in print.

Proposition 4.5 *We have equivalences*

$$\mathcal{M}_1(3) \simeq \operatorname{Spec}\big(\mathbb{Z}\big[\tfrac{1}{3}\big][a_1, a_3][\Delta^{-1}]\big)/\mathbb{G}_m,$$
$$\overline{\mathcal{M}}_1(3) \simeq \big(\operatorname{Spec}\big(\mathbb{Z}\big[\tfrac{1}{3}\big][a_1, a_3]\big) \setminus \{0\}\big)/\mathbb{G}_m =: \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3),$$

where:

- *The $\mathbb{G}_m$–action on $\operatorname{Spec}\big(\mathbb{Z}\big[\frac{1}{3}\big][a_1, a_3]\big)$ is induced by the grading with $|a_1| = 1$ and $|a_3| = 3$.*
- $\Delta = a_3^3(a_1^3 - 27a_3)$*.*
- $\{0\}$ *denotes the common vanishing locus of a_1 and a_3.*
- $\mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3)$ *is often called the weighted (stacky) projective line with weights 1 and 3.*

Proof The first equivalence follows from [46, Proposition 3.2].

Set $A = \mathbb{Z}\big[\frac{1}{3}\big][a_1, a_3]$. The equality $\mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3) = (\operatorname{Spec} A \setminus \{0\})/\mathbb{G}_m$ is just the definition of the weighted projective line. This is a proper and smooth Deligne–Mumford stack over $\operatorname{Spec} \mathbb{Z}\big[\frac{1}{3}\big]$ by [53, Proposition 2.1, Remark 2.2]. Note furthermore that $\mathcal{M}_1(3) \subset \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3)$ is a dense open substack.

To apply Lemma 4.4, we need to construct a finite morphism

$$\mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3) \to \overline{\mathcal{M}}\big[\tfrac{1}{3}\big]$$

that extends the morphism

$$\mathcal{M}_1(3) \to \mathcal{M}_{\mathrm{ell}}\big[\tfrac{1}{3}\big] \subset \overline{\mathcal{M}}_{\mathrm{ell}}\big[\tfrac{1}{3}\big].$$

The equation $y^2+a_1xy+a_3y=x^3$ defines a cubic curve over $\operatorname{Spec} A/\mathbb{G}_m$. We want to show that this equation actually defines a generalized elliptic curve E over $\mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3)$. For this, we have to check that for no map $f\colon \operatorname{Spec} k \to \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3)$ for k a field (of characteristic $\neq 3$), the pullback f^*E has a cusp. Equivalently, we have to show that for any values $a_1, a_3 \in k$ for which $c_4 = a_1^4 - 24a_1a_3$ and $\Delta = a_3^3(a_1^3 - 27a_3)$ vanish, a_1 and a_3 also vanish. First observe that if $c_4 = \Delta = 0$, then $a_1 = 0$ implies $a_3 = 0$ and vice versa. If $\Delta = 0$, either $a_3 = 0$ or $a_1^3 = 27a_3$. In the second case, $27a_1a_3 = a_1^4 = 24a_1a_3$ and thus $a_1 = 0$ or $a_3 = 0$.

Therefore, we obtain a map $p\colon \operatorname{Spec} A/\mathbb{G}_m \to \mathcal{M}_{\mathrm{cub}}\big[\frac{1}{3}\big]$ to the moduli stack of cubic curves that restricts to a map $\mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3) \to \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{3}\big]$, which in turn extends the map $\mathcal{M}_1(3) \to \mathcal{M}_{\mathrm{ell}}\big[\frac{1}{3}\big] \subset \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{3}\big]$.

As computed in the beginning of Section 7 of [6], the map p is surjective and we have $\operatorname{Spec} A/\mathbb{G}_m \times_{\mathcal{M}_{\mathrm{cub}}} \operatorname{Spec} A/\mathbb{G}_m \simeq (\operatorname{Spec} A[s,t]/(f,g))/\mathbb{G}_m$, where f and g are polynomials in s and t such that $A[s,t]/(f,g)$ is a finite flat A–module. As finiteness can be checked after fpqc-base change, the map p is finite and hence so is its restriction $\mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3) \to \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{3}\big]$, which is the base change $p \times_{\mathcal{M}_{\mathrm{cub}}[\frac{1}{3}]} \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{3}\big]$. Thus, the result follows by Lemma 4.4. □

By checking the gradings, we see that $p^*\omega \cong \mathcal{O}(1)$ for $p\colon \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3) \to \overline{\mathcal{M}}_{\mathrm{ell}}\big[\frac{1}{3}\big]$ the restriction of the morphism constructed in the proof above. (Here, ω denotes the line bundle $\pi_2\mathcal{O}^{\mathrm{top}}$ on $\overline{\mathcal{M}}_{\mathrm{ell}}$, which is also the pullback of the line bundle on $\mathcal{M}_{FG}$ we have denoted before by ω.) Thus, we have

$$H^s(\overline{\mathcal{M}}_1(3);\omega^{\otimes *}) \cong \begin{cases} \mathbb{Z}\big[\frac{1}{3}\big][a_1,a_3] & \text{for } s=0, \\ \mathbb{Z}\big[\frac{1}{3}\big][a_1,a_3]/(a_1^\infty, a_3^\infty) & \text{for } s=1, \\ 0 & \text{for } s \geq 2, \end{cases}$$

as shown, for example, in [53, Proposition 2.5]. Here, $\mathbb{Z}\big[\frac{1}{3}\big][a_1,a_3]/(a_1^\infty,a_3^\infty)$ denotes the $\mathbb{Z}\big[\frac{1}{3}\big][a_1,a_3]$–torsion module with $\mathbb{Z}\big[\frac{1}{3}\big]$–basis given by the monomials $1/a_1^ia_3^j$, where $i, j \geq 1$. Thus, the descent spectral sequence for $\mathrm{Tmf}_1(3)$ collapses. In particular, we see that $\pi_* \,\mathrm{tmf}_1(3) = \mathbb{Z}\big[\frac{1}{3}\big][a_1,a_3]$.

4.2 RO(C_2)–graded homotopy of $\mathrm{tmf}_1(3)$

Our goal in this subsection is to understand the C_2–equivariant RO(C_2)–graded homotopy groups of $\mathrm{tmf}_1(3)$. We will compute this via an RO(C_2)–graded homotopy fixed point spectral sequence, as described for general G in Section 2.3. When $G = C_2$ there are two important simplifications. The first allows us to identify the E_2–term more transparently:

Lemma 4.6 *Let E be a C_2–spectrum. Then*

$$\pi_*(E \wedge S^{\sigma-1}) \cong \pi_* E \otimes \mathrm{sgn}$$

as C_2–modules.

Proof This follows from the fact that the action map $t\colon S^\sigma \to S^\sigma$ has degree -1. □

Corollary 4.7 *If E is a C_2–spectrum, then the* RO(C_2)*–graded homotopy fixed point spectral sequence has the form*

$$H^s(C_2; \pi_t(E) \otimes \mathrm{sgn}^{\otimes r}) \implies \pi^{C_2}_{t-s+(\sigma-1)r} F(EC_{2+}, E).$$

The differential d_i goes from degree (r,s,t) to $(r, s+i, t+i-1)$. The tridegree (r,s,t) corresponds to the bidegree $((t-r)+r\sigma, s)$ in representation grading.

If E is even with π_{2n} flat over $\mathbb{Z}$ and the group C_2 acts on $\pi_{2n}E$ via $(-1)^n$, then the E^2–term is isomorphic to

$$\overline{\pi_{2*}E} \otimes \mathbb{Z}[u_{2\sigma}^{\pm 1}, a_\sigma]/2a_\sigma$$

with $|u_{2\sigma}| = (2-2\sigma, 0)$ and $|a_\sigma| = (-\sigma, 1)$. Here, $\overline{\pi_{2n}E}$ is the group $\pi_{2n}E$, but not in degree $2n$, but in degree $n + n\sigma$.

Proof The first part is clear. For the second, note that the RO(C_2)–graded C_2–representation $\pi_\star E$ is isomorphic to $\overline{\pi_{2*}E} \otimes \bigoplus_{r\in\mathbb{Z}} \mathrm{sgn}^{\otimes r}$ with $\mathrm{sgn}^{\otimes r}$ in degree $r(1-\sigma)$. The first tensor factor is invariant under the C_2–action and can thus be pulled out of the cohomology group. For the second one, we have $H^*(C_2; \bigoplus_{r\in\mathbb{Z}} \mathrm{sgn}^{\otimes r}) \cong \mathbb{Z}[u^{\pm 1}, a]/2a$ with $u \in H^0(C_2; \mathrm{sgn}^{\otimes 2})$ and $a \in H^1(C_2; \mathrm{sgn})$. □

The second C_2 simplification is a recasting of the RO(C_2)–graded homotopy fixed points spectral sequence in a way that allows us to read off permanent cycles. Recall that there is a C_2–equivariant map

$$a_\sigma\colon S^0 \to S^\sigma$$

which is essential but for which the restriction is null. The following is undoubtedly well known to experts.

Lemma 4.8 *The* $RO(C_2)$*–graded homotopy fixed points spectral sequence for a* C_2*–spectrum* X *coincides with the* a_σ*–Bockstein spectral sequence for* X*.*

Proof The map a_σ^n fits in a cofiber sequence

$$S(n\sigma)_+ \to S^0 \xrightarrow{a_\sigma^n} S^{n\sigma},$$

where $S(n\sigma)$ is the unit sphere in the representation $n\sigma$. Applying $F(-,X)$, we deduce a cofiber sequence of spectra

$$\Sigma^{-n\sigma} X \xrightarrow{a_\sigma^n} X \to F(S(n\sigma)_+, X).$$

The space $S(n\sigma)_+$ is also the $(n-1)$–skeleton of the standard model for EC_{2+} as the infinite sign sphere, and the map on function spectra induced by the inclusion of the $(n-1)$–skeleton into the n–skeleton coincides with the obvious map of cofibers:

$$\begin{tikzcd}
\Sigma^{-(n+1)\sigma} X \arrow[r, "a_\sigma^{n+1}"] \arrow[d, "a_\sigma"'] & X \arrow[r] \arrow[d, "1"] & F(S((n+1)\sigma)_+, X) \arrow[d] \\
\Sigma^{-n\sigma} X \arrow[r, "a_\sigma^n"'] & X \arrow[r] & F(S(n\sigma)_+, X).
\end{tikzcd}$$

Thus the filtration by powers of a_σ and the filtration by the skeleton of EC_{2+} coincide. □

Recall now from Section 4.1 that nonequivariantly

$$\pi_* \mathrm{tmf}_1(3) \cong \mathbb{Z}[\tfrac{1}{3}][a_1, a_3] \quad \text{and} \quad \pi_* \mathrm{TMF}_1(3) \cong \mathbb{Z}[\tfrac{1}{3}][a_1, a_3, \Delta^{-1}],$$

with $|a_1| = 2$ and $|a_3| = 6$. By Mahowald and Rezk [46, Proposition 3.4], the group C_2 acts by -1 on a_1 and a_3 in $\pi_* \mathrm{TMF}_1(3)$ and hence also in $\pi_* \mathrm{tmf}_1(3)$, as $\pi_* \mathrm{tmf}_1(3)$ sits inside $\pi_* \mathrm{TMF}_1(3)$.

By Corollary 4.7, the $RO(C_2)$–graded homotopy fixed point spectral sequence E_2–term for $\mathrm{tmf}_1(3)^{hC_2}$ can be written as

$$E_2^{*,*} = \mathbb{Z}[\tfrac{1}{3}][a_\sigma, u_{2\sigma}^{\pm 1}, \bar{a}_1, \bar{a}_3]/(2a_\sigma) \tag{1}$$

with degrees

$$|a_\sigma| = (-\sigma, 1) = (1-\rho, 1), \quad |u_{2\sigma}| = (2-2\sigma, 0) = (4-2\rho, 0),$$
$$|\bar{a}_1| = (1+\sigma, 0) = (\rho, 0), \qquad |\bar{a}_3| = (3+3\sigma, 0) = (3\rho, 0).$$

We start by identifying the permanent cycles corresponding to η and ν in the Hurewicz image in $\pi_* \mathrm{tmf}_1(3)^{hC_2}$. By [29, Theorem 6.2], there is a C_2–equivariant map

$$\mathrm{Tmf}_1(3) \to \mathrm{KU}$$

of E_∞–ring spectra into K-theory, inducing a map between the homotopy fixed point spectral sequences for $\mathrm{Tmf}_1(3)^{hC_2}$ and $\mathrm{KO} \simeq \mathrm{KU}^{hC_2}$. In the latter, η is of filtration 1, so it has to be of filtration ≤ 1 in the former. As the homotopy fixed point spectral sequences of $\mathrm{Tmf}_1(3)^{hC_2}$ and $\mathrm{tmf}_1(3)^{hC_2}$ agree in nonnegative degrees, η is also of filtration 1 in the homotopy fixed point spectral sequence for $\mathrm{tmf}_1(3)^{hC_2}$ and is thus detected by $a_\sigma \bar{a}_1$.

To identify ν, we observe the following lemma:

Lemma 4.9 *The composite* $\mathrm{Tmf}\left[\frac{1}{3}\right] \xrightarrow{\mathrm{res}} \mathrm{Tmf}_0(3) \xrightarrow{\mathrm{tr}} \mathrm{Tmf}\left[\frac{1}{3}\right]$ *is multiplication by* 4.

Proof This is true on the level of E_2–terms of homotopy fixed point spectral sequences, expressing $\mathrm{Tmf}_0(3)$ and $\mathrm{Tmf}\left[\frac{1}{3}\right]$ as homotopy fixed points of $\mathrm{Tmf}(3)$ (as the index in $\mathrm{GL}_2(\mathbb{Z}/3)$ of the subgroup of matrices of the form $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ is 4). The $\mathrm{Tmf}\left[\frac{1}{3}\right]$–linear self-maps of $\mathrm{Tmf}\left[\frac{1}{3}\right]$ are in one-to-one correspondence to elements in $\pi_0 \mathrm{Tmf}\left[\frac{1}{3}\right]$. These are all of filtration 0 in the descent spectral sequence by [35, Figure 26] and thus detected by their action on

$$\pi_0 \mathrm{Tmf}\left[\tfrac{1}{3}\right] = H^0(\mathrm{GL}_2(\mathbb{Z}/3); \pi_0 \mathrm{Tmf}(3)).$$

(As the arguments in [35] are computationally involved, we also sketch another way to arrive at this last result. If there were contributions of positive filtration to $\pi_0 \mathrm{Tmf}\left[\frac{1}{3}\right]$ in the descent spectral sequence, this group would contain torsion. Because $\pi_0 \mathrm{Tmf}\left[\frac{1}{3}\right] \cong \pi_0 \mathrm{tmf}\left[\frac{1}{3}\right]$, it suffices to show that $\pi_0 \mathrm{tmf} \cong \mathbb{Z}$. It was known by Hopkins and Miller and is shown in [49, Corollary 5.3] that the Adams–Novikov spectral sequence for tmf has as E^2–term the cohomology of the graded Weierstrass Hopf algebroid

$$(A = \mathbb{Z}[a_1, a_2, a_3, a_4, a_6], \Gamma = A[r, s, t]).$$

Here, $|r| = 4$, $|s| = 2$, and $|t| = 6$. It follows formally from the gradings in the cobar complex that $H^i(A, \Gamma) = 0$ in degrees smaller than $2i$ and that $H^0(A, \Gamma) \cong \mathbb{Z}$. The result follows.) $\square$

As 4ν in $\pi_3 \mathrm{Tmf}\left[\frac{1}{3}\right]$ is nonzero and of filtration 3, we know $\nu = \mathrm{res}(\nu) \in \pi_3 \mathrm{Tmf}_0(3)$ is of filtration ≤ 3 and nonzero. For degree reasons, it has to be detected by the image of $a_\sigma^3 \bar{a}_3$. As the homotopy fixed point spectral sequences for $\mathrm{tmf}_1(3)^{hC_2}$ and $\mathrm{Tmf}_1(3)^{hC_2}$ agree in this range, the same is true for $\mathrm{tmf}_1(3)^{hC_2}$.

Corollary 4.10 *The classes $\bar{a}_1$ and $\bar{a}_3$ are permanent cycles in the* $RO(C_2)$*–graded homotopy fixed point spectral sequence for* $\mathrm{tmf}_1(3)$.

Proof Since the homotopy fixed point spectral sequence and a_σ–Bockstein spectral sequences coincide, we learn that if an a_σ–multiple of a class is a permanent cycle, then the class is a permanent cycle. This in particular applies to $\eta = \bar{a}_1 a_\sigma$ and $\nu = \bar{a}_3 a_\sigma^3$. □

Corollary 4.11 *The class $u_{2\sigma}$ is the only generator of the E_2–term for the* $\mathrm{RO}(C_2)$*–graded homotopy fixed point spectral sequences for* $\mathrm{tmf}_1(3)$ *(as listed in Equation (1)) that is not a permanent cycle.*

Furthermore, the transfer of any element in the underlying homotopy is a permanent cycle. In particular, we conclude immediately that the classes

$$v_0(k) := 2u_{2\sigma}^k$$

for $k \in \mathbb{Z}$ are all permanent cycles which generate copies of $\mathbb{Z}$. These satisfy an obvious multiplicative relation

$$v_0(k)v_0(j) = 2v_0(j+k).$$

Next, we will determine the differentials. Note first that for degree reasons all d_{2k} are 0 for $k \geq 1$. While the other differentials could be deduced from [46], we will derive them independently.

Proposition 4.12 *We have the differential*

$$d_3(u_{2\sigma}) = a_\sigma^3 \bar{a}_1.$$

Proof Because $\bar{a}_1, \bar{a}_3$ and a_σ are permanent cycles, $d_3(u_{2\sigma}) = 0$ would imply that $E_2 = E_5$. On the other hand, we know that η is detected by $a_\sigma \bar{a}_1$. As $\eta^4 = 0$, the class $(a_\sigma \bar{a}_1)^4$ must be hit by a differential, which necessarily must be a d_3. Therefore, $d_3(u_{2\sigma}) \neq 0$. For degree reasons we get that $d_3(u_{2\sigma}) = a_\sigma^3 \bar{a}_1$. □

There is no room for a d_5–differential; indeed, a nontrivial d_5–differential would imply a differential of the form $d_5(u_{2\sigma}^2) = a_\sigma^5 y$ with y in the 0–line of degree $3+\sigma$, which is impossible. Thus, $E_7 = E_4$.

Proposition 4.13 *We have the differential*

$$d_7(u_{2\sigma}^2) = a_\sigma^7 \bar{a}_3.$$

Proof If $d_n(u_{2\sigma}^2) = a_\sigma^n x$, then x is in degree $(7-n) + (n-4)\rho$. As x can be written as $u_{2\sigma}^{2m}$ times a polynomial in $\bar{a}_1$ and $\bar{a}_3$, we see that $7-n$ must be divisible by 8. As $\bar{a}_1, \bar{a}_3$, a_σ and $2u_{2\sigma}$ are permanent cycles, $d_7(u_{2\sigma}^2) = 0$ would thus imply that

$E_7 = E_{15}$. On the other hand, we know that ν is detected by $a_\sigma^3\bar{a}_3$. As $\nu^4 = 0$, the class $(a_\sigma^3\bar{a}_3)^4$ must be hit by a differential, which necessarily must be a d_n with $n \leq 12$. Therefore, $d_7(u_{2\sigma}^2) \neq 0$. For degree reasons we get that $d_3(u_{2\sigma}^2) = a_\sigma^7\bar{a}_3$ (as $a_\sigma^7\bar{a}_1^3 = 0$ in E_7). □

The torsion produced by the first differential yields new d_7–cycles:

$$\bar{a}_1(k) := \bar{a}_1 u_{2\sigma}^{2k},$$

for $k \in \mathbb{Z}$. These also participate in the expected multiplicative relations:

$$\bar{a}_1(k)\bar{a}_1(j) = \bar{a}_1 \cdot \bar{a}_1(j+k) \quad \text{and} \quad \bar{a}_1(j)v_0(k) = \bar{a}_1 \cdot v_0(k+2j).$$

Remark 4.14 The classes $v_0(k)$ and $\bar{a}_1(j)$ form families exactly like the families $v_0(k)$ and $v_1(j)$ described by Hu and Kriz is the computation of the homotopy of $BP\mathbb{R}$.

There is no room for further differentials in E_8, which is the subalgebra of

$$\mathbb{Z}[\tfrac{1}{3}][a_\sigma, u_{2\sigma}, \bar{a}_1, \bar{a}_3]/(2a_\sigma, \bar{a}_1 a_\sigma^3, \bar{a}_3 a_\sigma^7)$$

generated by a_σ, $\bar{a}_1$, $\bar{a}_3$, $v_0(1)$, $v_0(2)$, $v_0(3)$, $\bar{a}_1(1)$ and $u_{2\sigma}^{\pm 4}$. Indeed, a nontrivial d_k–differential for $k \geq 8$ would imply a nontrivial differential of the form $d_k(\bar{a}_1(1)) = a_\sigma^k x$ or $d_k(u_{2\sigma}^4) = a_\sigma^k y$ for some x or y in the 0–line of degree $4+(k-3)\sigma$ or $7+(k-8)\sigma$, respectively; but the only of our generators of the 0–line not killed by a_σ^k is $u_{2\sigma}^{\pm 4}$, whose powers cannot be in degree $4+(k-3)\sigma$ or $7+(k-8)\sigma$. Therefore $E_8 = E_\infty$.

Theorem 4.15 *We have*

$$\pi_\star^{C_2} F(EC_{2+}, \mathrm{tmf}_1(3)) \cong \mathbb{Z}[\tfrac{1}{3}][a_\sigma, u_{2\sigma}^{\pm 4}, \bar{a}_1, \bar{a}_3, v_0(k), \bar{a}_1(1)]/R,$$

where the ideal R of relations is generated by

$$\begin{aligned} a_\sigma v_0(k) &= 0, & v_0(k+4) &= v_0(k)u_{2\sigma}^4, & \bar{a}_1(1)v_0(k) &= \bar{a}_1 v_0(k+2), \\ a_\sigma^3(\bar{a}_1, \bar{a}_1(1)) &= 0, & v_0(k)v_0(j) &= 2v_0(j+k), & \bar{a}_1(1)^2 &= \bar{a}_1 u_{2\sigma}^4, \\ a_\sigma^7\bar{a}_3 &= 0. \end{aligned}$$

Proof The presentation given was already shown to be a presentation of the E_∞–term. We just have to check all the relations also to hold in $\pi_\star^{C_2} F(EC_{2+}, \mathrm{tmf}_1(3))$. Observe first that no two torsion classes in different filtrations can converge to the same bidegree. This implies the first three relations must hold. In the next three relations, both sides are in the image of the transfer and thus these relations can be checked on underlying homotopy groups. The last relation holds again since there is no element of filtration ≥ 1 in this bidegree. □

Remark 4.16 We have $\pi^{C_2}_{a+b\sigma} F(EC_{2+}, \mathrm{tmf}_1(3)) \cong \pi^{C_2}_{a+b\sigma} \mathrm{tmf}_1(3)$ for all $a, b \geq 0$ and $\pi^{C_2}_{a+b\sigma} \mathrm{tmf}_1(3) = 0$ if $a < 0$ and $a + b < 0$; this follows from the cofiber sequence

$$S^{a+(b-1)\sigma} \to S^{a+b\sigma} \to S^{a+b} \wedge (C_2)_+.$$

Is it possible, but more complicated, to describe also the other groups in $\pi^{C_2}_\star \mathrm{tmf}_1(3)$. Note that $\pi^{C_2}_i F(EC_{2+}, \mathrm{tmf}_1(3)) \cong \pi_i \mathrm{tmf}_1(3)^{hC_2}$ can certainly be nontrivial for $i < 0$ (eg we have $a^8_\sigma u^{-4}_{2\sigma} \in \pi_{-8} \mathrm{tmf}_1(3)^{hC_2}$) and thus $\mathrm{tmf}_1(3) \not\simeq F(EC_{2+}, \mathrm{tmf}_1(3))$.

Corollary 4.17 *The spectrum* $\mathrm{tmf}_1(3)$ *is strongly even as a* C_2*–spectrum. In particular, it is real orientable and thus* $\mathrm{tmf}_1(3)_{(2)}$ *is a form of* $BP\mathbb{R}\langle 2\rangle$*. Furthermore,* $\mathrm{tmf}_1(3)_{(2)}[\bar{a}_3^{-1}]$ *is a form of* $E\mathbb{R}(2)$*.*

Proof It follows from Theorem 4.15 and the remark thereafter that $\mathrm{tmf}_1(3)$ is even as a C_2–spectrum and also that the Mackey functor $\underline{\pi}_{k\rho} \mathrm{tmf}_1(3)$ is constant for all $k \in \mathbb{Z}$. We present the argument for evenness and leave the other part to the reader. Let $y = a^l_\sigma x$ be a nonzero class in degree $k\rho - 1$ with x of filtration 0 and degree $(k-1) + (k+l)\sigma$. Clearly $l \geq 1$. In E_2, we can write x as $\bar{a}^i_1 \bar{a}^j_3 u^{2m}_{2\sigma}$. We see that $(k-1) - (k+l) = -(l+1)$ is divisible by 8; in particular, $l \geq 7$. This implies $i, j = 0$ and leads to a contradiction.

In the following, we localize everywhere implicitly at 2. The map $BP_* \to \mathrm{tmf}_1(3)_*$ induced by the 2–typification of the formal group law associated to the Weierstrass equation $y^2 + a_1 xy + a_3 y = x^3$ sends the Hazewinkel generators v_1 and v_2 exactly to a_1 and a_3. This implies together with $\mathrm{tmf}_1(3)$ being strongly even that $\mathrm{tmf}_1(3)$ is a form of $BP\mathbb{R}\langle 2\rangle$ and that $\mathrm{tmf}_1(3)[\bar{a}_3^{-1}]$ is a form of $E\mathbb{R}(2)$. □

Corollary 4.18 *There exists forms of* $BP\mathbb{R}\langle 2\rangle$ *and* $E\mathbb{R}(2)$ *that are strictly commutative* C_2*–ring spectrum.*

Proof By Theorem 2.7, the spectrum $\mathrm{tmf}_1(3)$ has the structure of a strictly commutative C_2–ring spectrum. By the last corollary, it is a form of $BP\mathbb{R}\langle 2\rangle$.

As shown in the next section, the spectrum $\mathrm{tmf}_1(3)[\bar{a}_3^{-1}]$ is equivalent to $\mathrm{Tmf}_1(3)[\bar{a}_3^{-1}]$ and thus cofree. Thus, we see by Theorem 2.4 that it has the structure of a strictly commutative C_2–spectrum. □

Remark 4.19 We do not know whether the forms of $BP\mathbb{R}\langle 2\rangle$ and $E\mathbb{R}(2)$ exhibited here are equivalent as C_2–spectra to other known forms, as for example defined via the Hazewinkel generators. Note though that two forms of $E\mathbb{R}(n)$ are equivalent if and only if their underlying spectra are equivalent by Proposition 3.16. Note further that in contrast to our result, the existence of any kind of (homotopy unital) multiplication seems to be unknown for general forms of $BP\mathbb{R}\langle n\rangle$, even for $n = 2$.

4.3 The relationship between $\mathrm{tmf}_1(3)$, $\mathrm{Tmf}_1(3)$ and $\mathrm{TMF}_1(3)$

The following is proven in [50, Theorem 7.12].

Proposition 4.20 *The map* $\mathrm{Tmf}_0(3) \to \mathrm{Tmf}_1(3)$ *is a faithful* C_2*–Galois extension in the sense of Rognes.*

Lemma 4.21 *Let* $\bar{f}$ *be a nonconstant homogeneous polynomial in* $\bar{a}_1$ *and* $\bar{a}_3$*. Then*

$$\mathrm{tmf}_1(3)[\bar{f}^{-1}] \to \mathrm{Tmf}_1(3)[\bar{f}^{-1}]$$

is an equivalence.

Proof For some $k > 0$, we have that $a_\sigma^7 \bar{f}^k = 0$ in $\pi_\star^{C_2} F((EC_2)_+, \mathrm{tmf}_1(3))$ and $|a_\sigma^7 \bar{f}^k| = r + s\sigma$ with $r, s \geq 0$. Thus we also have $a_\sigma^7 \bar{f}^k = 0$ in $\pi_\star^{C_2} \mathrm{tmf}_1(3)$ and therefore $\Phi^{C_2}(\mathrm{tmf}_1(3)[\bar{f}^{-1}]) = 0$ by Lemma 2.2. By [28, Corollary 10.6], $\mathrm{tmf}_1(3)[\bar{f}^{-1}]$ is then cofree. Thus, we have only to show that $\mathrm{tmf}_1(3)[\bar{f}^{-1}] \to \mathrm{Tmf}_1(3)[\bar{f}^{-1}]$ is an equivalence of underlying spectra. As every element of negative degree in $\pi_*^e \mathrm{Tmf}_1(3)$ is killed by a_1 and a_3, the result follows. □

Lemma 4.22 *Let* $\bar{f}$ *be a nonconstant homogeneous polynomial in* $\bar{a}_1$ *and* $\bar{a}_3$*. Denote by* $D(f)$ *the nonvanishing locus of the underlying element* $f \in H^0(\overline{\mathcal{M}}_1(3); \omega^*)$*. Then there is an equivalence*

$$\mathrm{Tmf}_1(3)[\bar{f}^{-1}] \to \mathcal{O}^{\mathrm{top}}(D(f))$$

of C_2*–spectra.*

Proof Note that the pullback of $D(f)$ along

$$\operatorname{Spec} \mathbb{Z}\left[\tfrac{1}{3}\right][a_1, a_3] - \{0\} \quad \to \quad \overline{\mathcal{M}}_1(3) \simeq \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3)$$

is $\operatorname{Spec} \mathbb{Z}\left[\tfrac{1}{3}\right][a_1, a_3][f^{-1}]$. By the same argument as in Lemma 4.2, the global sections functor

$$\Gamma\colon \mathrm{QCoh}(D(f)) = \mathrm{QCoh}\big(\operatorname{Spec}\big(\mathbb{Z}\left[\tfrac{1}{3}\right][a_1, a_3][f^{-1}]\big)/\mathbb{G}_m\big) \to \mathrm{AbelianGroups}$$

is exact. Therefore, the descent spectral sequence for $\mathcal{O}^{\mathrm{top}}(D(f))$ collapses and we have $\pi_* \mathcal{O}^{\mathrm{top}}(D(f)) \cong \mathbb{Z}\left[\tfrac{1}{3}\right][a_1, a_3][f^{-1}]$.

Note furthermore that $D(f)$ is C_2–invariant as f^2 is an invariant section. This induces a C_2–map of ring spectra $\mathrm{Tmf}_1(3) = \mathcal{O}^{\mathrm{top}}(\overline{\mathcal{M}}_1(3)) \to \mathcal{O}^{\mathrm{top}}(D(f))$. We want to show that the image of $\bar{f}$ is invertible in $\pi_\star^{C_2} \mathcal{O}^{\mathrm{top}}(D(f))$. It is detected in the homotopy fixed point spectral sequence HFPSS(f) for $\mathcal{O}^{\mathrm{top}}(D(f))^{hC_2}$ by $f u_{2\sigma}^k$ for some k. As f and $u_{2\sigma}$ are invertible, there exists an element

$$\overline{g} \in \pi^{C_2}_{-|\overline{f}|} \mathcal{O}^{\mathrm{top}}(D(f))$$

detected by $f^{-1} u_{2\sigma}^{-k}$. Clearly, the underlying class $\mathrm{res}(\overline{f}\,\overline{g}) \in \pi_0 \mathcal{O}^{\mathrm{top}}(D(f))$ equals 1. As HFPSS(f) receives a multiplicative map from the homotopy fixed point spectral sequence HFPSS for $\mathrm{tmf}_1(3)^{hC_2}$, the identification of $\pi_* \mathcal{O}^{\mathrm{top}}(D(f))$ above implies that HFPSS(f) $\cong$ HFPSS$[\overline{f}^{-1}]$. In particular, we can deduce that $\mathcal{O}^{\mathrm{top}}(D(f))$ is strongly even as a (cofree) C_2–spectrum. This implies that $\overline{f}\,\overline{g} = 1 \in \pi_0^{C_2} \mathcal{O}^{\mathrm{top}}(D(f))$ so that $\overline{f}$ is invertible. Thus, we get an induced map

$$\mathrm{Tmf}_1(3)[\overline{f}^{-1}] \to \mathcal{O}^{\mathrm{top}}(D(f))$$

of C_2–spectra.

By [50, Theorem 7.2] and the proof of [50, Theorem 7.12], the global sections functor

$$\Gamma\colon \mathrm{QCoh}(\overline{\mathcal{M}}_1(3), \mathcal{O}^{\mathrm{top}}) \to \mathrm{Tmf}_1(3)\text{–mod}$$

is an equivalence.[2] Thus, we can apply [50, Lemma 3.20] to see that

$$\mathrm{Tmf}_1(3)[\overline{f}^{-1}] \to \mathcal{O}^{\mathrm{top}}(D(f))$$

is an equivalence of underlying spectra. As both spectra are cofree, the result follows. □

This applies in particular to $\overline{f} = \overline{\Delta}$. Thus,

$$\mathrm{tmf}_1(3)[\overline{\Delta}^{-1}] \simeq \mathrm{Tmf}_1(3)[\overline{\Delta}^{-1}] \simeq \mathrm{TMF}_1(3)$$

as C_2–spectra (with $\overline{\Delta} = \bar{a}_3^3(\bar{a}_1^3 - 27\bar{a}_3)$). In particular, $\mathrm{TMF}_1(3)$ is strongly even. Thus, Theorem 3.6 implies:

Proposition 4.23 *The C_2–spectrum $\mathrm{TMF}_1(3)$ is real Landweber exact in the sense that there is a natural isomorphism*

$$M\mathbb{R}_\star(X) \otimes_{MU_{2*}} \mathrm{TMF}_1(3)_{2*} \to \mathrm{TMF}_1(3)_\star(X)$$

for all C_2–spectra X.

Note that the equivalence $\mathrm{tmf}_1(3)[\overline{\Delta}^{-1}] \simeq_{C_2} \mathrm{TMF}_1(3)$ also directly implies together with the computations from the previous sections that $\pi_* \mathrm{TMF}_0(3)$ has torsion and thus $\mathrm{TMF}_0(3)$ cannot be Landweber exact.

[2] We only really need that Γ commutes with homotopy colimits. As observed in the proof of [50, Proposition 3.8], this is automatic when the stack has finite cohomological dimension as $\overline{\mathcal{M}}_1(3)$ does. This circumvents the use of most of the heavy machinery in [50].

The following fiber square will be useful later.

Proposition 4.24 *We have a fiber square:*

$$\begin{tikzcd} \mathrm{Tmf}_1(3) \arrow[r] \arrow[d] & \mathrm{tmf}_1(3)[\bar{a}_1^{-1}] \arrow[d] \\ \mathrm{tmf}_1(3)[\bar{a}_3^{-1}] \arrow[r] & \mathrm{tmf}_1(3)[(\bar{a}_1\bar{a}_3)^{-1}] \end{tikzcd}$$

Proof The square

$$\begin{tikzcd} \overline{\mathcal{M}}_1(3) & D(a_1) \arrow[l] \\ D(a_3) \arrow[u] & D(a_1a_3) \arrow[l] \arrow[u] \end{tikzcd}$$

induces a fiber square

$$\begin{tikzcd} \mathcal{O}^{\mathrm{top}}(\overline{\mathcal{M}}_1(3)) \arrow[r] \arrow[d] & \mathcal{O}^{\mathrm{top}}(D(a_1)) \arrow[d] \\ \mathcal{O}^{\mathrm{top}}(D(a_3)) \arrow[r] & \mathcal{O}^{\mathrm{top}}(D(a_1a_3)) \end{tikzcd} \tag{2}$$

as

$$\overline{\mathcal{M}}_1(3) \simeq \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3) = D(a_1) \cup D(a_3),$$

and $\mathcal{O}^{\mathrm{top}}$ is a sheaf (see [50, Appendix A] for why the sheaf condition implies this).

By the last two lemmas, this is equivalent to

$$\begin{tikzcd} \mathrm{Tmf}_1(3) \arrow[r] \arrow[d] & \mathrm{tmf}_1(3)[\bar{a}_1^{-1}] \arrow[d] \\ \mathrm{tmf}_1(3)[\bar{a}_3^{-1}] \arrow[r] & \mathrm{tmf}_1(3)[(\bar{a}_1\bar{a}_3)^{-1}] \end{tikzcd}$$

as a square of C_2–spectra. □

5 Slices and Anderson duals

In this section, we will compute the slices of $\mathrm{TMF}_1(3)$ and $\mathrm{Tmf}_1(3)$ and apply this to compute the Anderson dual of $\mathrm{Tmf}_1(3)$.

5.1 Slices

We can apply the computations of the regular representation homotopy groups of $\mathrm{tmf}_1(3)$ and its localizations to determine their slices.

Since all of the odd slices vanish and the even slices are regular representation suspensions of $H\underline{\mathbb{Z}}\big[\frac{1}{3}\big]$ by Section 2.4 and Corollary 4.17, the homotopy groups "near multiples of regular representations" are easy to compute since the slice spectral sequence is especially simple here.

To ensure transparency with later notation and gradings, we introduce some notation. Let $R = \mathbb{Z}\big[\frac{1}{3}\big][a_1, a_3][f^{-1}]$ with f homogeneous. If $S \subset R_{2n}$ is any subset of homogeneous rational functions of degree $2n$, then let $\bar{S}$ denote the same rational functions, but with every instance of a_1 and a_3 replaced with $\bar{a}_1$ and $\bar{a}_3$ respectively. This is a notational device to ensure that the reader keep track of the $\mathrm{RO}(C_2)$–grading of barred elements, compared to the underlying $\mathbb{Z}$–grading of unbarred ones. Lemma 2.15 now gives us a description of the homotopy groups of the localizations of $\mathrm{tmf}_1(3)$:

Corollary 5.1 *Let M be one of*

$$\mathrm{tmf}_1(3), \quad \mathrm{tmf}_1(3)[\bar{a}_1^{-1}], \quad \mathrm{tmf}_1(3)[\bar{a}_3^{-1}] \quad \textit{or} \quad \mathrm{tmf}_1(3)[(\bar{a}_1\bar{a}_3)^{-1}].$$

For all k, we have

$$\underline{\pi}_{k\rho+1} M = \underline{G} \otimes \overline{\pi^u_{2k+2} M}, \qquad \underline{\pi}_{k\rho} M = \underline{\mathbb{Z}}\big[\tfrac{1}{3}\big] \otimes \overline{\pi^u_{2k} M},$$
$$\underline{\pi}_{k\rho-1} M = 0, \qquad \underline{\pi}_{k\rho-2} M = \underline{\mathbb{Z}}\big[\tfrac{1}{3}\big]_- \otimes \overline{\pi^u_{2k-2} M}.$$

Similarly, naturality of the slice spectral sequence implies that we understand the effect of the localization maps on homotopy groups in dimensions $k\rho-2, \dots, k\rho+1$.

Corollary 5.2 *For $k \in \mathbb{Z}$ and for $j = -2, -1, 0, 1$, the localization maps*

$$\underline{\pi}_{k\rho+j}\, \mathrm{tmf}_1(3)[\bar{a}_i^{-1}] \to \underline{\pi}_{k\rho+j}\, \mathrm{tmf}_1(3)[(\bar{a}_1\bar{a}_3)^{-1}]$$

are induced by the obvious inclusions of graded pieces of these graded rings.

Remark 5.3 We could also have read off these results from the homotopy fixed point spectral sequence, but the slice spectral sequence approach is both more conceptual and is easier for Mackey functor computations.

We want now to compute the slices of $\mathrm{Tmf}_1(3)$. To that end, we denote by $M[\bar{a}_1, \bar{a}_3]$ the monic monomials in $\mathbb{Z}\big[\frac{1}{3}\big][\bar{a}_1, \bar{a}_3]$.

Proposition 5.4 *The associated graded C_2–spectrum for the slice filtration of* $\mathrm{Tmf}_1(3)$ *is*

$$\bigvee_{P\in M[\bar{a}_1,\bar{a}_3]}\big(S^{|P|}\wedge H\underline{\mathbb{Z}}[\tfrac{1}{3}]\big)\vee\bigvee_{P\in M[\bar{a}_1,\bar{a}_3]}\big(S^{-|P|-4\rho-1}\wedge H\underline{\mathbb{Z}}[\tfrac{1}{3}]\big).$$

Proof We use Propositions 2.9 and 2.13 to read the slices out of the $RO(C_2)$–graded homotopy groups. The long exact sequence in homotopy associated to the fiber square in Proposition 4.24 and Corollary 5.1 identify the needed homotopy groups. For $k<0$, let R_k denote the degree-$2k$ piece of

$$\mathbb{Z}[\tfrac{1}{3}][a_1^{\pm 1},a_3^{\pm 1}]/\big(\mathbb{Z}[\tfrac{1}{3}][a_1^{\pm 1},a_3]+\mathbb{Z}[\tfrac{1}{3}][a_1,a_3^{\pm 1}]\big).$$

We then have isomorphisms

$$\underline{\pi}_{k\rho}\,\mathrm{Tmf}_1(3)=\underline{G}\otimes R_{k+1}\quad\text{and}\quad\underline{\pi}_{k\rho-1}\,\mathrm{Tmf}_1(3)=\underline{\mathbb{Z}}\otimes R_k.$$

The functor P^0 applied to the Mackey functor $\underline{G}$ yields zero, so we conclude by Proposition 2.13 that there are no negative even slices, and by Proposition 2.9 that all of the negative odd slices are of the desired form. □

This allows us to compute the E_2–term of the slice spectral sequence

$$E_2^{s,t}=\pi_{t-s}^{C_2}P_t^t\,\mathrm{Tmf}_1(3)\implies\pi_{t-s}^{C_2}\,\mathrm{Tmf}_1(3),$$

where P_t^t denotes the t–slice of $\mathrm{Tmf}_1(3)$. For $t=2k\geq 0$, we get

$$\begin{aligned}\pi_{2k-s}^{C_2}P_{2k}^{2k}\,\mathrm{Tmf}_1(3)&=\bigoplus_{P\in M[\bar{a}_1,\bar{a}_3]_{k\rho}}\pi_{2k-s}^{C_2}S^{k\rho}\wedge H\underline{\mathbb{Z}}[\tfrac{1}{3}]\\&=\bigoplus_{P\in M[\bar{a}_1,\bar{a}_3]_{k\rho}}H_{2k-s}^{C_2}\big(S^{k\rho},\underline{\mathbb{Z}}[\tfrac{1}{3}]\big)\\&=\bigoplus_{P\in M[\bar{a}_1,\bar{a}_3]_{k\rho}}H_{k-s}^{C_2}\big(S^{k\sigma},\underline{\mathbb{Z}}[\tfrac{1}{3}]\big).\end{aligned}$$

By [28, Example 3.16], we have

$$H_{k-s}^{C_2}\big(S^{k\sigma},\underline{\mathbb{Z}}[\tfrac{1}{3}]\big)=\begin{cases}\mathbb{Z}[\tfrac{1}{3}] & \text{if } 2k-s \text{ is divisible by 4 and } s=0,\\ \mathbb{Z}/2 & \text{if } 0<s\leq 2k-s \text{ and } (2k-s)-s \text{ is divisible by 4},\\ 0 & \text{otherwise.}\end{cases}$$

Similarly, one can reduce the computation for $t<0$ to Bredon *co*homology and use that

$$H_{C_2}^{k}\big(S^{d\sigma},\underline{\mathbb{Z}}[\tfrac{1}{3}]\big)=\begin{cases}\mathbb{Z}[\tfrac{1}{3}] & \text{if } d \text{ is even and } k=d,\\ \mathbb{Z}/2 & \text{if } k \text{ is odd and } 1<k\leq d,\\ 0 & \text{otherwise.}\end{cases}$$

We depict the slice spectral sequence in Figure 2. Here, an unboxed number n denotes n copies of $\mathbb{Z}/2$, a box denotes a copy of $\mathbb{Z}[\frac{1}{3}]$ and a boxed n denotes n copies of $\mathbb{Z}[\frac{1}{3}]$. The vertical coordinate is s and the horizontal one is $t-s$. In positive degrees, the differentials follow from those for $MU\mathbb{R}$ via the real orientation map, and these were determined in [28] and in [32]. For the differentials in negative degrees, we can use that this is a spectral sequence of algebras, so in particular, we have an action of the slice spectral sequence for $\mathrm{tmf}_1(3)$ on that of $\mathrm{Tmf}_1(3)$. This reduces the problem to understanding the differentials on the line L of slope one in Figure 2 passing through the "1" in $(-8,-1)$. This class is infinitely divisible by $\eta = \bar{a}_1 a_\sigma$ and $\nu = \bar{a}_3 a_\sigma^3$. The classes η^3 and ν^3 are hit by a d_3 and a d_7 respectively in the slice spectral sequence for $\mathrm{tmf}_1(3)$. As a class x on L is not hit by any differential for degree reasons, $\eta^{-3}x$ has thus to support a d_3–differential and $\nu^{-3}x$ a d_7–differential (if it does not support a d_3–differential). This forces the negative differentials.

5.2 Anderson duality

Let G be a finite group. For an injective abelian group J, the functor

$$\text{(genuine) } G\text{–Spectra} \to \text{graded abelian groups}, \qquad X \mapsto \mathrm{Hom}_{\mathbb{Z}}(\pi^G_{-*}X, J)$$

is representable by a G–spectrum I_J, as follows from Brown representability. If A is an abelian group and $A \to J^0 \to J^1$ an injective resolution, we define the G–spectrum I_A to be the fiber of $I_{J^0} \to I_{J^1}$. Given a G–spectrum X, we define its *A–Anderson dual* $I_A X$ by $F(X, I_A)$. It satisfies for all $k \in \mathbb{Z}$ the following functorial short exact sequence:

$$0 \to \mathrm{Ext}^1_{\mathbb{Z}}(\pi^G_{-k-1}X, A) \to \pi^G_k I_A X \to \mathrm{Hom}_{\mathbb{Z}}(\pi^G_{-k}X, A) \to 0.$$

For $G = \{e\}$ we get nonequivariant Anderson duality as explored in [1] and [60]. If G is (possibly) nontrivial, denote by $\mathcal{A}_G$ the *stable Burnside category*, by which we mean the full subcategory of $\mathrm{Ho}(\mathrm{Sp}_G)$ on the cosets $\Sigma^\infty(G/H)_+$. Given again a G–spectrum X, we see by precomposing with the functor

$$\mathcal{A}_G \to \mathrm{Sp}_G, \quad \Sigma^\infty(G/H)_+ \mapsto \Sigma^\infty(G/H)_+ \wedge X$$

that the short exact sequence above refines to a short exact sequence of Mackey functors

$$0 \to \mathrm{Ext}^1_{\mathbb{Z}}(\underline{\pi}_{-k-1}X, A) \to \underline{\pi}_k I_A X \to \mathrm{Hom}_{\mathbb{Z}}(\underline{\pi}_{-k}X, A) \to 0.$$

By smashing X with representation spheres, we see that it even refines to an $RO(G)$–graded sequence. Equivariant Anderson duality in the case $G = C_2$ has been explored in some detail in [56].

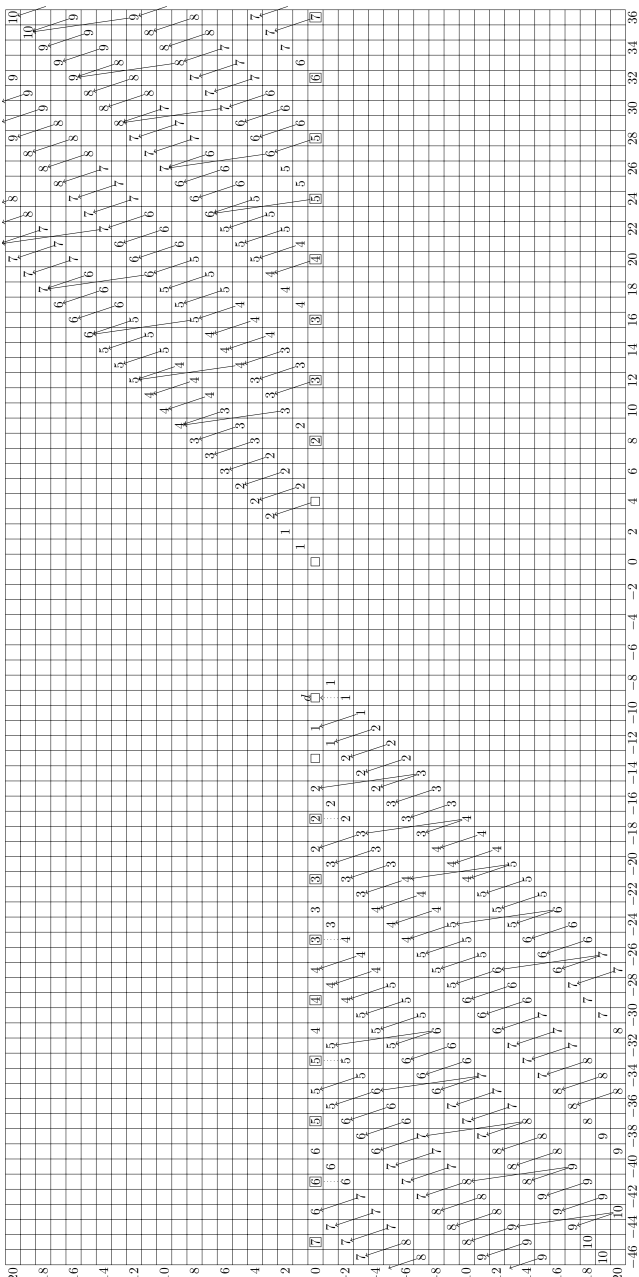


Figure 2: The slice spectral sequence for $\mathrm{Tmf}_1(3)$

One reason to be interested in Anderson (self) duality is the universal coefficient sequence, relating homology and cohomology. Let E be a G–spectrum, X be another G–spectrum and A be an abelian group. Then $I_A(X \wedge E) \simeq_G F(X, I_A E)$ implies the short exact sequence

$$0 \to \mathrm{Ext}^1_{\mathbb{Z}}(E_{V-1}X, A) \to (I_A E)^V X \to \mathrm{Hom}_{\mathbb{Z}}(E_V X, A) \to 0$$

for a real G–representation V. In particular, Anderson self-duality implies useful universal coefficient theorems; for example, $I_{\mathbb{Z}}\,\mathrm{KO} \simeq \Sigma^4\,\mathrm{KO}$ implies one of the main theorems of [1].

Our goal in this section is to compute the $\mathbb{Z}[\frac{1}{3}]$–Anderson dual of $\mathrm{Tmf}_1(3)$ as a C_2–spectrum and then deduce a computation of the $\mathbb{Z}[\frac{1}{3}]$–Anderson dual of $\mathrm{Tmf}_0(3)$.

Observe that $H\underline{\mathbb{Z}}^* \simeq S^{4-2\rho} \wedge H\underline{\mathbb{Z}}$ as $H^0(S^{4-2\rho}; \underline{\mathbb{Z}}) \cong \underline{\mathbb{Z}}^*$, where $\underline{\mathbb{Z}}^*$ is as in Definition 2.14. Thus, Proposition 5.4 implies that the associated graded C_2–spectrum for the slice filtration of $\mathrm{Tmf}_1(3)$ is

$$\bigvee_{P \in M[\bar{a}_1, \bar{a}_3]} (S^{|P|} \wedge H\underline{\mathbb{Z}}[\tfrac{1}{3}]) \vee \bigvee_{P \in M[\bar{a}_1, \bar{a}_3]} (S^{-|P|-2\rho-5} \wedge H\underline{\mathbb{Z}}[\tfrac{1}{3}]^*).$$

This suggests the following theorem:

Theorem 5.5 *There is a C_2–equivariant equivalence $I_{\mathbb{Z}[\frac{1}{3}]}\,\mathrm{Tmf}_1(3) \simeq \Sigma^{5+2\rho}\,\mathrm{Tmf}_1(3)$.*

Note that this theorem implies the universal coefficient sequence claimed in the introduction. To prove the theorem, we will start with two lemmas.

Lemma 5.6 *We have nonequivariantly $I_{\mathbb{Z}[\frac{1}{3}]}\,\mathrm{Tmf}_1(3) \simeq \Sigma^9\,\mathrm{Tmf}_1(3)$.*

Proof By Proposition 4.5, the moduli stack $\overline{\mathcal{M}}_1(3)$ is equivalent to the weighted projective stack $\mathcal{P}(1,3) = \mathcal{P}_{\mathbb{Z}[\frac{1}{3}]}(1,3)$ and the sheaf ω on $\overline{\mathcal{M}}_1(3)$ corresponds to $\mathcal{O}(1)$ on $\mathcal{P}(1,3)$. This weighted projective stack has Serre duality in the sense that there is a class

$$D = \frac{1}{a_1 a_3} \in H^1(\mathcal{P}(1,3); \mathcal{O}(-4))$$

such that

$$H^s(\mathcal{P}(1,3); \mathcal{F}) \otimes H^{1-s}(\mathcal{P}(1,3); \mathcal{F}^* \otimes \mathcal{O}(-4)) \to H^1(\mathcal{P}(1,3); \mathcal{O}(-4)) \cong \mathbb{Z}[\tfrac{1}{3}] \cdot D$$

is a perfect pairing for $s = 0, 1$ for an arbitrary coherent sheaf $\mathcal{F}$.

Let us write for brevity $R = \mathrm{Tmf}_1(3)$. As $\mathcal{P}(1,3)$ has cohomological dimension 1, the element D is a permanent cycle in the descent spectral sequence for R and is represented by a unique element in $\pi_{-9}R \cong \mathbb{Z}\big[\frac{1}{3}\big]$, which we will also denote by D. Denote by δ the element in $\pi_9 I_{\mathbb{Z}[\frac{1}{3}]}R$ with $\phi(\delta)(D) = 1$, where

$$\phi\colon \pi_9 I_{\mathbb{Z}[\frac{1}{3}]}R \xrightarrow{\cong} \mathrm{Hom}\big(\pi_{-9}R, \mathbb{Z}\big[\tfrac{1}{3}\big]\big).$$

The element δ induces a R–linear map $\widehat{\delta}\colon \Sigma^9 R \to I_{\mathbb{Z}[\frac{1}{3}]}R$.

We obtain a commutative diagram:

$$\begin{tikzcd}
\pi_{k-9}R \otimes \pi_{-k}R \arrow[r, "\widehat{\delta}_*\otimes\mathrm{id}"] \arrow[d] & \pi_k I_{\mathbb{Z}[\frac{1}{3}]}R \otimes \pi_{-k}R \arrow[r, "\phi\otimes\mathrm{id}", "\cong"'] & \mathrm{Hom}\big(\pi_{-k}R, \mathbb{Z}\big[\tfrac{1}{3}\big]\big) \otimes \pi_{-k}R \arrow[d] \\
\pi_{-9}R \arrow[rr, "\phi(\delta)", "\cong"'] & & \mathbb{Z}\big[\tfrac{1}{3}\big]
\end{tikzcd}$$

The left vertical map is a perfect pairing because of Serre duality (as described above), as is the right vertical map by definition. Thus, the map $\widehat{\delta}_*\colon \pi_{k-9}R \to \pi_k I_{\mathbb{Z}[\frac{1}{3}]}R$ is an isomorphism for all k. This shows that $\widehat{\delta}$ is an equivalence. $\square$

The following key lemma uses our information about the slices of $\mathrm{Tmf}_1(3)$:

Lemma 5.7 *The transfer*

$$\pi_{-9}\,\mathrm{Tmf}_1(3) = \pi^{e}_{-5-2\rho}\,\mathrm{Tmf}_1(3) \to \pi^{C_2}_{-5-2\rho}\,\mathrm{Tmf}_1(3)$$

is an isomorphism.

Proof The slice spectral sequence for $\Sigma^{2\rho}\,\mathrm{Tmf}$ (as shown in Figure 3, where dots stand for the Mackey functor G and a box with a cross stands for $\underline{\mathbb{Z}}^*$) gives an isomorphism of Mackey functors

$$\underline{\pi}_{-5-2\rho}\,\mathrm{Tmf}_1(3) \cong \underline{\pi}_{-5-2\rho}S^{-4\rho-1} \wedge H\underline{\mathbb{Z}}\big[\tfrac{1}{3}\big] \cong H^2\big(S^{2\sigma}; \underline{\mathbb{Z}}\big[\tfrac{1}{3}\big]\big) \cong \underline{\mathbb{Z}}\big[\tfrac{1}{3}\big]^*. \quad \square$$

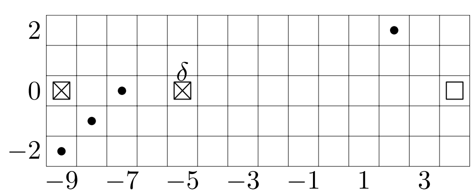


Figure 3: The E^2–term of the slice spectral sequence for $\pi_{k-2\rho}\,\mathrm{Tmf}_1(3)$

Proof of Theorem 5.5 Consider the following commutative diagram:

$$\begin{tikzcd}
\pi^{C_2}_{5+2\rho} I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3) \arrow[r, two heads] \arrow[d, "\mathrm{res}"] & \mathrm{Hom}\big(\pi^{C_2}_{-5-2\rho} \mathrm{Tmf}_1(3), \mathbb{Z}\big[\tfrac{1}{3}\big]\big) \arrow[d, "\cong"', "\mathrm{res}=\mathrm{tr}^*"] \\
\pi_9 I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3) \arrow[r, "\cong"] & \mathrm{Hom}\big(\pi_{-9} \mathrm{Tmf}_1(3), \mathbb{Z}\big[\tfrac{1}{3}\big]\big)
\end{tikzcd}$$

By the last lemma, tr^* is an isomorphism. This implies that we can refine the element $\delta \in \pi_9 I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3)$ corresponding to the equivalence $\Sigma^9 \mathrm{Tmf}_1(3) \to I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3)$ from Lemma 5.6 to an element $\widetilde{\delta} \in \pi^{C_2}_{5+2\rho} I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3)$. This induces a C_2–equivariant $\mathrm{Tmf}_1(3)$–linear map

$$\Sigma^{5+2\rho} \mathrm{Tmf}_1(3) \to I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3)$$

that is an equivalence of underlying spectra. By Proposition 4.20 and [57, Proposition 6.3.3], we know that $\mathrm{Tmf}_1(3)^{tC_2} \simeq \Phi^{C_2} \mathrm{Tmf}_1(3)$ vanishes and thus $\mathrm{Tmf}_1(3)$ and $I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_1(3)$ are cofree C_2–spectra by [28, Corollary 10.6]. Thus, the theorem follows. □

This allows us also to compute the Anderson dual of $\mathrm{Tmf}_0(3)$. As in [60], we will use the following lemma:

Lemma 5.8 *Let A be an abelian group and X be a spectrum with an action by a finite group G. Assume that the norm map $X_{hG} \to X^{hG}$ is an equivalence. Then there is an equivalence $(I_A X)^{hG} \simeq I_A(X^{hG})$.*

Proof We have the following chain of equivalences:

$$(I_A X)^{hG} \simeq F(X, I_A)^{hG} \simeq F(X_{hG}, I_A) \simeq F(X^{hG}, I_A) \simeq I_A(X^{hG}) \qquad \square$$

As noted in the proof of Theorem 5.5, $\mathrm{Tmf}_1(3)^{tC_2}$ vanishes. Thus, we get:

Corollary 5.9 *There is an equivalence $I_{\mathbb{Z}[\frac{1}{3}]} \mathrm{Tmf}_0(3) \simeq (\Sigma^{5+2\rho} \mathrm{Tmf}_1(3))^{hC_2}$.*

6 The Picard groups

In this section we will compute the Picard groups of $\mathrm{TMF}_0(3)$, $\mathrm{Tmf}_0(3)$ and related spectra. We recommend Mathew and Stojanoska [51] for a good introduction to Picard groups and our techniques are very similar to theirs.

6.1 Generalities

In the following, we will often use the language of ∞–categories. We choose the same model as [33] and [39], namely quasicategories. For the theory of (symmetric) monoidal ∞–categories see either [44] or [23] for a shorter introduction.

If $\mathcal{C}$ is a monoidal category, we denote by $\mathrm{Pic}(\mathcal{C})$ the group of isomorphism classes of invertible spectra; note that this is a priori a proper class (or large set, depending on set-theoretic conventions), but will always be a (small) set in our situation. If $\mathcal{C}$ is a monoidal ∞–category, we denote by $\mathcal{P}ic(\mathcal{C})$ the maximal ∞–subgroupoid (Kan complex) of the full subcategory of invertible objects. Clearly, $\pi_0\,\mathcal{P}ic(\mathcal{C}) \cong \mathrm{Pic}(\mathrm{Ho}(\mathcal{C}))$. If $\mathcal{C}$ is a symmetric monoidal ∞–category, $\mathcal{P}ic(\mathcal{C})$ inherits the structure of a group-like E_∞–space; indeed, $\mathcal{P}ic(\mathcal{C})$ is a symmetric monoidal ∞–category and thus by [44, Example 2.1.2.18, Remark 2.4.2.6, Corollary 5.1.1.5] a $\mathrm{Comm} = E_\infty$–algebra in the ∞–category of ∞–groupoids, which agrees with that of spaces. Thus, there is a connective spectrum $\mathfrak{pic}(\mathcal{C})$ with $\Omega^\infty\mathfrak{pic}(\mathcal{C}) \simeq \mathcal{P}ic(\mathcal{C})$ by a result of Boardman and Vogt and of May (see [44, Remark 5.2.6.26] for an ∞–categorical treatment). Note that we have $\pi_i\mathfrak{pic}(\mathcal{C}) \cong \pi_i\,\mathcal{P}ic(\mathcal{C})$ in this situation.

Given an E_2–ring spectrum R, its ∞–category R–mod of (left) R–modules has the structure of a monoidal ∞–category [44, Proposition 7.1.2.6]. We define the *Picard group* $\mathrm{Pic}(R)$ of R to be $\mathrm{Pic}(\mathrm{Ho}(R\text{–mod}))$ and the *Picard space* $\mathcal{P}ic(R)$ to be $\mathcal{P}ic(R\text{–mod})$. If R is an E_∞–ring spectrum, then R–mod is even a *symmetric* monoidal ∞–category. We define then $\mathfrak{pic}(R)$ to be $\mathfrak{pic}(R\text{–mod})$.

For us, a *derived stack* will be a pair $\mathcal{X} = (X, \mathcal{O}^{\mathrm{top}})$, where X is a Deligne–Mumford stack and $\mathcal{O}^{\mathrm{top}}$ is a sheaf of even-periodic E_∞–ring spectra with $\pi_0\mathcal{O}^{\mathrm{top}}$ isomorphic to the structure sheaf $\mathcal{O}_X$ of X. For example, X might be a moduli stack of elliptic curves. For a derived stack $\mathcal{X} = (X, \mathcal{O}^{\mathrm{top}})$, we write $\mathrm{Pic}(\mathcal{X})$ etc for the Picard group, space or spectrum of the symmetric monoidal ∞–category of quasicoherent $\mathcal{O}^{\mathrm{top}}$–modules $\mathrm{QCoh}(\mathcal{X})$ on $\mathcal{X}$. For a short treatment of quasicoherent sheaves in this context see [50, Section 2.3] and for a full-blown treatment see [42].

Definition 6.1 We call a derived stack $\mathcal{X} = (X, \mathcal{O}^{\mathrm{top}})$ 0–*affine* if the global sections functor

$$\Gamma\colon \mathrm{QCoh}(\mathcal{X}) \to \mathcal{O}^{\mathrm{top}}(X)\text{–mod}$$

is an equivalence of symmetric monoidal ∞–categories.

Clearly, $\mathfrak{pic}(\mathcal{X}) \simeq \mathfrak{pic}(\mathcal{O}^{\mathrm{top}}(X))$ if $\mathcal{X}$ is 0–affine. It was shown in [50] that the (compactified) moduli stack of elliptic curves with arbitrary level structure together with its derived structure sheaf $\mathcal{O}^{\mathrm{top}}$ is 0–affine.

The following Mayer–Vietoris principle will be useful later.

Lemma 6.2 *Let $\mathcal{X} = (X, \mathcal{O}_{\mathcal{X}})$ be a 0–affine derived stack and $U, V \subset X$ be a covering by open substacks. Then we have a long exact sequence*

$$\cdots \to \mathrm{GL}_1\, \pi_0 \mathcal{O}_{\mathcal{X}}(U \cap V) \xrightarrow{\partial} \mathrm{Pic}(\mathcal{O}_{\mathcal{X}}(X)) \to \mathrm{Pic}(\mathcal{O}_{\mathcal{X}}(U)) \times \mathrm{Pic}(\mathcal{O}_{\mathcal{X}}(V)) \to \mathrm{Pic}(\mathcal{O}_{\mathcal{X}}(U \cap V))$$

of abelian groups.

Proof As shown in [51, Section 3.1], the presheaf $\mathcal{P}ic$ defined by

$$\mathcal{P}ic(W \to X) = \mathcal{P}ic(\mathcal{O}^{\mathrm{top}}(W \to X))$$

(where $W \to X$ is étale) is actually a sheaf. Thus, we have a homotopy pullback square:

$$\begin{tikzcd} \mathcal{P}ic(X, \mathcal{O}_{\mathcal{X}}) \arrow[r] \arrow[d] & \mathcal{P}ic(U, \mathcal{O}_{\mathcal{X}}|_U) \arrow[d] \\ \mathcal{P}ic(V, \mathcal{O}_{\mathcal{X}}|_V) \arrow[r] & \mathcal{P}ic(U \cap V, \mathcal{O}_{\mathcal{X}}|_{U \cap V}) \end{tikzcd}$$

The identification of these Picard spaces with those of $\mathcal{O}_{\mathcal{X}}(X)$ etc follows from the fact that X, U, V and $U \cap V$ are 0–affine (see [50, Proposition 3.27]). This fiber square induces the long exact sequence in the lemma. □

Remark 6.3 By the last proof the boundary map

$$\mathrm{GL}_1\, \pi_0 \mathcal{O}_{\mathcal{X}}(U \cap V) \to \mathrm{Pic}(\mathcal{O}_{\mathcal{X}}(X))$$

is induced by the map

$$\mathrm{GL}_1\, \mathcal{O}_{\mathcal{X}}(U \cap V) \simeq \Omega\, \mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(U \cap V)) \to \mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(U)) \times^h_{\mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(U \cap V))} \mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(V))$$

of spaces. Thus, it can be described as follows: An element $g \in \mathrm{GL}_1\, \pi_0 \mathcal{O}_{\mathcal{X}}(U \cap V)$ induces an $\mathcal{O}_{\mathcal{X}}$–linear self-equivalence f of $\mathcal{O}_{\mathcal{X}}|_{U \cap V}$. The triple $(\mathcal{O}_{\mathcal{X}}|_U, \mathcal{O}_{\mathcal{X}}|_V, f)$ defines an element of the homotopy fiber product $\mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(U)) \times^h_{\mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(U \cap V))} \mathcal{P}ic(\mathcal{O}_{\mathcal{X}}(V))$. As noted above, this gluing datum defines an invertible $\mathcal{O}_{\mathcal{X}}$–module on $\mathcal{X}$ and this invertible module represents $\partial(g)$.

Let now $A \to B$ be a faithful G–Galois extension in the sense of Rognes [57]. Then by [51, Section 3.3], we have the following theorem:

Theorem 6.4 *There is an equivalence $\mathfrak{pic}(A) \simeq \tau_{\geq 0} \mathfrak{pic}(B)^{hG}$.*

There is also another equivariant interpretation of the Picard group of A if $A \to B$ is a faithful G–Galois extension. View $B \simeq F(EG_+, B)$ as a cofree G–spectrum. Denote the category of equivariant B–modules by G-B–mod. As B is cofree and $A \to B$ is a faithful Galois extension, $\Phi^G B \simeq B^{tG}$ is contractible. By [28, Corollary 10.6] every (equivariant) B–module is thus cofree again. Therefore, a map in G-B–mod is a weak equivalence if it is an underlying weak equivalence. It is then a consequence of Galois descent of the form in [52, Lemma 6.1.4, Proposition 6.2.6] that there is a monoidal equivalence $\mathrm{Ho}(A\text{–mod}) \simeq \mathrm{Ho}(G\text{-}B\text{–mod})$. Thus, $\mathrm{Pic}(R) \cong \mathrm{Pic}(\mathrm{Ho}(G\text{-}B\text{–mod}))$, the group of equivariant invertible B–modules. We will denote the latter group by $\mathrm{Pic}_G(B)$.

6.2 A generalized Baker–Richter theorem

Baker and Richter proved in [5] that the Picard group of an E_∞–ring spectrum R is completely algebraic if R is even periodic and $\pi_0 R$ is a regular complete local ring. This applies, for example, to the Lubin–Tate spectra E_n. Mathew and Stojanoska generalized this in [51] by dropping the condition that $\pi_0 R$ is complete and local (and also weakened the periodicity requirement). The main purpose of this subsection is to show that the assumption of periodicity is superfluous.

Let R be an E_2–ring spectrum. Let $\overline{L}$ be an invertible $\pi_* R$–module. Then $\overline{L}$ is projective over $\pi_* R$. Thus, there is an R–module L with $\pi_* L \cong \overline{L}$ and this module L is well defined up to isomorphism in $\mathrm{Ho}(R\text{–mod})$. This defines a map $\mathrm{Pic}(\pi_* R) \to \mathrm{Pic}(R)$. By the degenerate Künneth spectral sequence, this is a homomorphism.

Let R_* be a commutative graded ring. By an element $x \in R_*$ we will always mean a *homogeneous* element and by an ideal $I \subset R_*$ we will always mean a *homogeneous* ideal. We call R_* *local* if it has a unique maximal ideal $\mathfrak{m}$. We call a graded local ring *regular* if the maximal ideal is generated by a finite regular sequence. We call a graded local ring *complete* if the map $R_* \to \lim_k R_*/\mathfrak{m}^k$ is an isomorphism. We call an arbitrary commutative graded ring *regular* if every localization of it at a prime ideal is regular.

We have the following generalization of [5, Theorem 38].

Theorem 6.5 *Let R be an E_2–ring spectrum. Assume that $\pi_* R$ is concentrated in even degrees and regular. Then the morphism* $\mathrm{Pic}(\pi_* R) \to \mathrm{Pic}(R)$ *is an isomorphism.*

This is not really new as this generalization is just a combination of [5, Remark 39] and [51, Theorem 2.4.6]. We will sketch a proof anyhow as we introduce one simplification, avoiding the use of obstruction theory for A_∞–structures.

Let M be an invertible R–module with $M \wedge_R N \simeq R$ for some R–module N. It is enough to show that $\pi_* M$ is a projective $\pi_* R$–module. For this, it is enough to show that the completion $(\widehat{\pi_* M})_{\mathfrak{m}}$ is a projective $\widehat{\pi_* R}_{\mathfrak{m}}$–module for every maximal ideal $\mathfrak{m} \subset \pi_* R$.

The theory from [43, Section 4.2] implies that there is actually an R–module $\widehat{M}_{\mathfrak{m}}$ with $\pi_* \widehat{M}_{\mathfrak{m}} \cong (\widehat{\pi_* M})_{\mathfrak{m}}$.[3] We have $\widehat{M}_{\mathfrak{m}} \wedge_{\widehat{R}_{\mathfrak{m}}} \widehat{N}_{\mathfrak{m}} \simeq \widehat{R}_{\mathfrak{m}}$ by [43, Remark 4.2.6] and because N is a finite R–module. Note here that $\widehat{R}_{\mathfrak{m}}$ also inherits an E_2–structure.

Let $x_1, \dots, x_n$ be a regular sequence generating $\mathfrak{m}$. Consider the $\widehat{R}_{\mathfrak{m}}$–module

$$\widehat{R}_{\mathfrak{m}}/\underline{x} = \widehat{R}_{\mathfrak{m}}/x_1 \wedge_{\widehat{R}_{\mathfrak{m}}} \cdots \wedge_{\widehat{R}_{\mathfrak{m}}} \widehat{R}_{\mathfrak{m}}/x_n,$$

obtained by killing the regular sequence $x_1, \dots, x_n$. Because $\widehat{R}_{\mathfrak{m}}$ is even, every x_i acts trivially on $\widehat{R}_{\mathfrak{m}}/x_i$ and hence on $\widehat{R}_{\mathfrak{m}}/\underline{x}$. Indeed, the composite

$$\Sigma^{|x_i|} \widehat{R}_{\mathfrak{m}} \to \Sigma^{|x_i|} \widehat{R}_{\mathfrak{m}}/x_i \xrightarrow{\cdot x_i} \widehat{R}_{\mathfrak{m}}/x_i$$

is zero and thus the second map factors over an $\widehat{R}_{\mathfrak{m}}$–linear map $\Sigma^{2|x_i|+1} \widehat{R}_{\mathfrak{m}} \to \widehat{R}_{\mathfrak{m}}/x_i$, which must be zero as well.

By [19, Theorem V.2.6][4] $\widehat{R}_{\mathfrak{m}}/\underline{x}$ has the structure of an $\widehat{R}_{\mathfrak{m}}$–ring spectrum in the sense that there exists a map

$$\widehat{R}_{\mathfrak{m}}/\underline{x} \wedge_{\widehat{R}_{\mathfrak{m}}} \widehat{R}_{\mathfrak{m}}/\underline{x} \to \widehat{R}_{\mathfrak{m}}/\underline{x}$$

that is unital up to homotopy.[5] For an arbitrary $\widehat{R}_{\mathfrak{m}}$–module X, set $X/\underline{x} = X \wedge_{\widehat{R}_{\mathfrak{m}}} \widehat{R}_{\mathfrak{m}}/\underline{x}$.

Claim 6.6 *The map*

$$\pi_*(X_1/\underline{x}) \otimes_{\pi_* \widehat{R}_{\mathfrak{m}}} \pi_*(X_2/\underline{x}) \to \pi_*(X_1/\underline{x} \wedge_{\widehat{R}_{\mathfrak{m}}} X_2/\underline{x}) \to \pi_*((X_1 \wedge_{\widehat{R}_{\mathfrak{m}}} X_2)/\underline{x})$$

factors over a map

$$\pi_*(X_1/\underline{x}) \otimes_{\pi_* \widehat{R}_{\mathfrak{m}}/\underline{x}} \pi_*(X_2/\underline{x}) \to \pi_*((X_1 \wedge_{\widehat{R}_{\mathfrak{m}}} X_2)/\underline{x}),$$

which is an isomorphism for all $\widehat{R}_{\mathfrak{m}}$–modules X_1 and X_2.

Proof It factors as every x_i acts trivially on $X_1/\underline{x} = X_1 \wedge_{\widehat{R}_{\mathfrak{m}}} \widehat{R}_{\mathfrak{m}}/\underline{x}$.

The map is clearly an isomorphism if $X_1 = \widehat{R}_{\mathfrak{m}}$. Both sides are homological in X_1 — since $\pi_*(\widehat{R}_{\mathfrak{m}}/\underline{x})$ is a graded field — and compatible with arbitrary coproducts. Thus, it is an isomorphism for all $X_1 \in \widehat{R}_{\mathfrak{m}}$–mod. □

[3] Lurie only considers ideals in $\pi_0 R$, but the theory also works for homogeneous ideals in $\pi_* R$ under our assumptions.

[4] While the source states the result only for E_∞–ring spectra, the same proof works also for E_2–ring spectra.

[5] For our argument, this naive result suffices, while Baker and Richter use that $\widehat{R}_{\mathfrak{m}}/\underline{x}$ has an A_∞–structure.

In particular $\pi_*(\widehat{M}_{\mathfrak{m}}/\underline{x})$ is in the Picard group of $\pi_*(\widehat{R}_{\mathfrak{m}}/\underline{x})$. Thus, $\pi_*(\widehat{M}_{\mathfrak{m}}/\underline{x})$ is a free $\pi_*(\widehat{R}_{\mathfrak{m}}/\underline{x})$–module of rank 1.

As in [5], one can show that $\pi_*(\widehat{M}_{\mathfrak{m}}/(x_1^{i_1},\dots,x_n^{i_n}))$ is a cyclic $\pi_*\widehat{R}_{\mathfrak{m}}$–module for $i_1,\dots,i_n \geq 1$, using the Nakayama lemma for graded rings. Using the completeness of $\pi_*\widehat{R}_{\mathfrak{m}}$, one can show as in [5] that $\pi_*\widehat{M}_{\mathfrak{m}}$ is a shift of $\pi_*\widehat{R}_{\mathfrak{m}}$. In particular, $\pi_*\widehat{M}_{\mathfrak{m}}$ is projective over $\pi_*\widehat{R}_{\mathfrak{m}}$ as we wanted to show.

6.3 The case of $\mathbf{TMF_1(3)}$ and $\mathbf{Tmf_1(3)}$

Lemma 6.7 *We have isomorphisms*

$$\operatorname{Pic}\mathrm{TMF}_1(3)\cong\mathbb{Z}/6,\qquad \operatorname{Pic}\mathrm{tmf}_1(3)[a_1^{-1}]\cong\mathbb{Z}/2,$$
$$\operatorname{Pic}\mathrm{tmf}_1(3)[a_3^{-1}]\cong\mathbb{Z}/6,\quad \operatorname{Pic}\mathrm{tmf}_1(3)[a_1^{-1}\bar{a}_3^{-1}]\cong\mathbb{Z}/2.$$

In all the cases, all the invertible modules are equivalent to suspensions of the ground ring spectrum.

Proof We will just prove the lemma for $\mathrm{TMF}_1(3)$, as the other cases are analogous. By Theorem 6.5,

$$\operatorname{Pic}\mathrm{TMF}_1(3)\cong\operatorname{Pic}(\pi_*\mathrm{TMF}_1(3)).$$

An evenly graded $\pi_{2*}\mathrm{TMF}_1(3)$–module is an equivalent datum to a quasicoherent sheaf on $\mathcal{M}_1(3)\simeq\operatorname{Spec}\mathbb{Z}\big[\tfrac{1}{3}\big][a_1,a_3][\Delta^{-1}]/\mathbb{G}_m$. Furthermore, an arbitrary graded $\pi_*\mathrm{TMF}_1(3)$–module splits into an even and an odd part. Therefore, an invertible $\pi_*\mathrm{TMF}_1(3)$–module has to be either completely even or completely odd. We hence have a short exact sequence

$$0\to\operatorname{Pic}(\mathcal{M}_1(3))\to\operatorname{Pic}(\pi_*\mathrm{TMF}_1(3))\to\mathbb{Z}/2\to 0,$$

where the first map corresponds to the inclusion of the even part and the map to $\mathbb{Z}/2$ indicates whether the invertible module is even or odd.

Given a line bundle $\mathcal{L}$ on $\mathcal{M}_1(3)$, we can extend it to the weighted projective stacky line $\overline{\mathcal{M}}_1(3)$. Indeed, by [53, Lemma 3.2], we can extend $\mathcal{L}$ to a reflexive sheaf on $\overline{\mathcal{M}}_1(3)$ and every reflexive sheaf of rank 1 is a line bundle by [25, Proposition 1.9]. Every line bundle on a weighted projective stacky line is of the form $\mathcal{O}(k)$ for some $k\in\mathbb{Z}$ as can be seen, for example, along the lines of [53, Proposition 3.4]. As noted after Proposition 4.5, the line bundle $\mathcal{O}(k)$ restricts to the (pullback of) the line bundle $\omega^{\otimes k}$ on $\mathcal{M}_1(3)$. Thus, the map $\phi\colon\mathbb{Z}\to\operatorname{Pic}(\mathcal{M}_1(3))$ sending k to $\omega^{\otimes k}$ is surjective.

It follows from the identification of $\mathcal{M}_1(3)$ above that

$$H^0(\mathcal{M}_1(3);\omega^{\otimes *})\cong\mathbb{Z}\big[\tfrac{1}{3}\big][a_1,a_3,\Delta^{-1}]$$

with $\Delta = a_3^3(a_1^3 - 27a_3)$. As $a_3 \in H^0(\mathcal{M}_1(3); \omega^{\otimes 3})$ is thus invertible on $\mathcal{M}_1(3)$, it defines a trivialization of $\omega^{\otimes 3}$ and thus $\phi(3) = 0$. The resulting morphism

$$\bar{\phi}\colon \mathbb{Z}/3 \to \mathrm{Pic}(\mathcal{M}_1(3))$$

is an isomorphism as there is no invertible section of $H^0(\mathcal{M}_1(3); \omega^{\otimes i})$ for $i = 1$ or 2. As the subgroup of $\mathrm{Pic}\,\mathrm{TMF}_1(3)$ spanned by the $\Sigma^k\,\mathrm{TMF}_1(3)$ is isomorphic to $\mathbb{Z}/6$, the lemma follows. □

Proposition 6.8 *The extensions*

$$\begin{gathered}
\mathrm{TMF}_0(3) \to \mathrm{TMF}_1(3),\\
(\mathrm{tmf}_1(3)[\bar{a}_1^{-1}])^{hC_2} \to \mathrm{tmf}_1(3)[\bar{a}_1^{-1}],\\
(\mathrm{tmf}_1(3)[\bar{a}_3^{-1}])^{hC_2} \to \mathrm{tmf}_1(3)[\bar{a}_3^{-1}],\\
(\mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}])^{hC_2} \to \mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}]
\end{gathered}$$

are faithful C_2–Galois extensions in the sense of Rognes.

Proof We obtain these maps of E_∞–ring spectra by applying $\mathcal{O}^{\mathrm{top}}$ to the C_2–Galois covers of stacks

$$\begin{gathered}
\mathcal{M}_1(3) \to \mathcal{M}_0(3),\\
D(a_1) \to D(a_1)/C_2,\\
D(a_3) \to D(a_3)/C_2,\\
D(a_1a_3) \to D(a_1a_3)/C_2,
\end{gathered}$$

as follows from the results in Section 4.3. Here, D denotes the nonvanishing locus. By the main result of [50], the derived stack $(\mathcal{M}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}})$ is 0–affine and by [50, Proposition 3.29] the same is true for the targets of the above four Galois covers. Then [50, Theorem 5.6] implies the result. □

Theorem 6.9 *We have isomorphisms*

$$\begin{array}{rcll}
\mathrm{Pic}_{C_2}\,\mathrm{TMF}_1(3) & \cong & \mathrm{Pic}(\mathrm{TMF}_0(3)) & \cong \mathbb{Z}/48,\\
\mathrm{Pic}_{C_2}\,\mathrm{tmf}_1(3)[\bar{a}_1^{-1}] & \cong & \mathrm{Pic}((\mathrm{tmf}_1(3)[\bar{a}_1^{-1}])^{hC_2}) & \cong \mathbb{Z}/8,\\
\mathrm{Pic}_{C_2}\,\mathrm{tmf}_1(3)[\bar{a}_3^{-1}] & \cong & \mathrm{Pic}((\mathrm{tmf}_1(3)[\bar{a}_3^{-1}])^{hC_2}) & \cong \mathbb{Z}/48,\\
\mathrm{Pic}_{C_2}\,\mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}] & \cong & \mathrm{Pic}((\mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}])^{hC_2}) & \cong \mathbb{Z}/8.
\end{array}$$

In all the cases, all the (equivariant) invertible modules are equivalent to (integer) suspensions of the ground ring spectrum.

Proof We will only prove this in the first case. The other cases are similar. The first isomorphism follows directly from Proposition 6.8 and the discussion at the end of the previous subsection.

In the following, we will denote by *HFPSS* the homotopy fixed point spectral sequence for the C_2–action on $\mathrm{TMF}_1(3)$ and differentials in it will be denoted by d^{HF}. We will always use the Adams convention that the k^{th} column consists of the groups $H^s(C_2;\pi_t\,\mathrm{TMF}_1(3))$ with $k=t-s$.

We have $\mathrm{TMF}_1(3)\simeq_{C_2}\mathrm{tmf}_1(3)[\overline{\Delta}^{-1}]$ with $\overline{\Delta}=\bar{a}_3^3(\bar{a}_1^3-27\bar{a}_3)$ by the results of Section 4.3. As $\overline{\Delta}$ is a permanent cycle, this allows us to deduce from the results of Section 4.2 all differentials in the HFPSS. For example, $\gamma=\bar{a}_3^4/\overline{\Delta}$ is a permanent cycle.

It is easy to see that the $(-1)^{\mathrm{st}}$ column of the HFPSS for $\mathrm{TMF}_1(3)$ is in cohomological degrees ≤ 7 isomorphic to $\mathbb{F}_2[\gamma]\cdot b_3\oplus\mathbb{F}_2[\gamma]\cdot b_7$ with $b_3=a_\sigma^3\bar{a}_1u_{2\sigma}^{-1}$ of cohomological degree 3 and $b_7=a_\sigma^7\bar{a}_3u_{2\sigma}^{-2}$ of degree 7. Recall from Section 4.2 that $\bar{a}_1,\bar{a}_3$ and a_σ are permanent cycles while $d_3^{\mathrm{HF}}(u_{2\sigma})=a_\sigma^3\bar{a}_1$ and $d_7^{\mathrm{HF}}(u_{2\sigma}^2)=a_\sigma^7\bar{a}_3$. We thus have the differentials

$$d_3^{\mathrm{HF}}(\gamma^k b_3)=\gamma^k a_\sigma^3\bar{a}_1u_{2\sigma}^{-2}\,d_3^{\mathrm{HF}}(u_{2\sigma})=\gamma^k b_3^2$$

and

$$d_7^{\mathrm{HF}}(\gamma^k b_7)=\gamma^k a_\sigma^7\bar{a}_3u_{2\sigma}^{-4}\,d_7^{\mathrm{HF}}(u_{2\sigma}^2)=\gamma^k b_7^2$$

in the HFPSS.

As $\mathrm{TMF}_0(3)\to\mathrm{TMF}_1(3)$ is a faithful C_2–Galois extension, Theorem 6.4 implies an equivalence $\mathfrak{pic}(\mathrm{TMF}_0(3))\simeq\tau_{\ge 0}(\mathfrak{pic}(\mathrm{TMF}_1(3)))^{hC_2}$. This gives the *Picard spectral sequence*

$$H^s(C_2;\pi_t\,\mathfrak{pic}\;\mathrm{TMF}_1(3))$$

that converges to $\pi_{t-s}\mathfrak{pic}\;\mathrm{TMF}_0(3)$ for $t-s\ge 0$. Differentials in it will be denoted by d^{Pic}.

The Picard group of $\mathrm{TMF}_1(3)$ is $\mathbb{Z}/6$ by Lemma 6.7 and $\mathrm{GL}_1\,\pi_0\,\mathrm{TMF}_1(3)$ is isomorphic to $\mathbb{Z}\times\mathbb{Z}/2$, generated by $\frac{1}{3}$ and -1. Thus,

$$\pi_t\,\mathfrak{pic}\;\mathrm{TMF}_1(3)=\begin{cases}\mathbb{Z}/6 & \text{for } t=0,\\ \mathbb{Z}\times\mathbb{Z}/2 & \text{for } t=1,\\ \pi_{t-1}\,\mathrm{TMF}_1(3) & \text{for } t\ge 2.\end{cases}$$

We are interested in the 0^{th} column of the Picard spectral sequence. We have

$$H^0(C_2;\mathbb{Z}/6)=\mathbb{Z}/6\quad\text{and}\quad H^1(C_2;\mathbb{Z}\times\mathbb{Z}/2)\cong\mathbb{Z}/2;$$

for $s \geq 2$ the 0^{th} column of the Picard spectral sequence agrees with the $(-1)^{\text{st}}$ column of the HFPSS. For an element x in the $(-1)^{\text{st}}$ column of the HFPSS, denote the corresponding element in the 0^{th} column of the Picard spectral sequence by $\underline{x}$.

If $x \in E^s_{*,s}$ is in cohomological degree s, then, by [51, Theorem 6.1.1], we have $d_s^{\mathrm{Pic}}(\underline{x}) = \underline{d_s^{\mathrm{HF}}(x) + x^2}$. For degree reasons, the first possible differential for $\underline{\gamma^k b_3}$ is a d_3^{Pic} and this equals $\underline{(\gamma^k + \gamma^{2k}) b_3^2}$. This is zero only if $k = 0$. Likewise for degree reasons, the first possible differential for $\underline{\gamma^k b_7}$ is a d_7^{Pic} and this equals $\underline{(\gamma^k + \gamma^{2k}) b_7^2}$. This is again zero only if $k = 0$, so that $\underline{b_3}$ and $\underline{b_7}$ are the only permanent cycles in the 0^{th} column of the Picard spectral sequence in cohomological degrees $2 \leq s \leq 7$.

It is easy to check that each element in the $(-1)^{\text{st}}$ column of the HFPSS of cohomological degree ≥ 8 either supports a d_3– or d_7–differential or is hit by a d_3– or d_7–differential from an element of degree ≥ 8. By [51, Comparison Tool 5.2.4], this implies that all nontrivial elements in the 0^{th} column of the Picard spectral sequence in cohomological degrees ≥ 8 support nontrivial differentials or are hit by differentials.

Thus, $\mathrm{Pic}(\mathrm{TMF}_0(3))$ has at most $6 \cdot 2 \cdot 2 \cdot 2 = 48$ elements. We just need to show that the image of the morphism

$$\mathbb{Z} \to \mathrm{Pic}(\mathrm{TMF}_1(3)), \quad k \mapsto \Sigma^k \,\mathrm{TMF}_0(3)$$

has order 48. This follows easily from the fact that 48 is the smallest period of $\pi_* \,\mathrm{TMF}_0(3)$ as Δ is not a permanent cycle in the HFPSS. $\square$

Lemma 6.10 *Let E be a strongly even C_2 E_2–ring spectrum. Then every even projective $\pi_* E$ module can be realized by a strongly even E–module in a unique way up to homotopy, giving in particular a well-defined homomorphism*

$$\mathrm{Pic}_{\mathrm{even}}(\pi_*^e E) \to \mathrm{Pic}^{C_2}(E).$$

Proof Let P be an even projective $\pi_*^e E$–module. We can write P as the image of an idempotent endomorphism f of a free even $\pi_* E$–module F. We can write $F = \bigoplus_I \pi_*^e \Sigma^{2n_i} E$. Define a free E–module $\mathbb{F}$ by $\mathbb{F} = \bigoplus \Sigma^{n_i \rho} E$. Because E is strongly even, we have $\pi_*^e \mathbb{F} \cong F$ and we can lift f to an idempotent endomorphism of $\mathbb{F}$, whose mapping telescope we denote by $\mathbb{P}$. This is the required realization of P.

If we have another strongly even E–module $\mathbb{P}'$ with $\pi_* \mathbb{P}' \cong P$ as $\pi_*^e E$–modules, we can lift the morphism $F \to P$ to an E–module morphism $\mathbb{F} \to \mathbb{P}'$ and further to a morphism $\mathbb{P} \to \mathbb{P}'$ that induces an isomorphism on π_*^e. By Lemma 3.4, this is an equivalence.

Thus, we get a well-defined map

$$\mathrm{Pic}_{\mathrm{even}}(\pi_*^e E) \to \mathrm{Pic}^{C_2}(E).$$

To show that it is an homomorphism, we have to show that for strongly even projective E–modules $\mathbb{P}_1$ and $\mathbb{P}_2$, the smash product $\mathbb{P}_1 \wedge_E \mathbb{P}_2$ is still strongly even and has underlying homotopy groups $\pi^e_* \mathbb{P}_1 \otimes_{\pi^e_* E} \pi^e_* \mathbb{P}_2$. This is clear by a retraction argument from the corresponding statement for free modules of the form $\bigoplus_{i \in I} \Sigma^{n_i \rho} E$. $\square$

Question 6.11 Let E be a C_2 E_2–ring spectrum. Assume that E is strongly even and that $\pi^u_* E$ is a regular graded ring and an integral domain. Is every invertible E–module of the form $S^V \wedge L$, where $V \in \mathrm{RO}(C_2)$ and L is a strongly even E–module with $\pi^e_* L \in \mathrm{Pic}(\pi^e_* E)$?

Using the lemma above, the question can be restated as asking for the surjectivity of the homomorphism

$$\mathrm{RO}(C_2) \oplus \mathrm{Pic}_{\mathrm{even}}(\pi^e_* E) \to \mathrm{Pic}^{C_2}(E).$$

A positive answer to this question would be a real generalization of the theorem by Baker and Richter given here as Theorem 6.5.

We could provide a similar spectral sequence argument as above for the computation of $\mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3))$, but we prefer to use a Mayer–Vietoris style argument instead. This will demonstrate how the computation of $\mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3))$ follows essentially formally from the fact that the Picard groups $\mathrm{Pic}_{C_2}(\mathrm{tmf}_1(3)[\bar{a}_1^{-1}])$ and $\mathrm{Pic}_{C_2}(\mathrm{tmf}_1(3)[\bar{a}_3^{-1}])$ are generated by the suspension of the ground ring spectrum.

Theorem 6.12 *The morphism*

$$\mathrm{RO}(C_2) \to \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)), \quad V \mapsto S^V \wedge \mathrm{Tmf}_1(3)$$

is surjective. Its kernel is generated by $8 - 8\sigma$. Thus,

$$\mathrm{Pic}(\mathrm{Tmf}_0(3)) \cong \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)) \cong \mathbb{Z} \oplus \mathbb{Z}/8.$$

Proof By Lemmas 4.21, 4.22 and 6.2, we have an exact sequence:

$$\begin{aligned} \mathrm{GL}_1\, \pi_0^{C_2} \mathrm{tmf}_1(3)[\bar{a}_1^{-1}] \times \mathrm{GL}_1\, \pi_0^{C_2} \mathrm{tmf}[\bar{a}_3^{-1}] &\xrightarrow{f} \mathrm{GL}_1\, \pi_0^{C_2} \mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}] \\ \xrightarrow{\delta} \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)) &\to \mathrm{Pic}_{C_2}(\mathrm{tmf}_1(3)[\bar{a}_1^{-1}]) \times \mathrm{Pic}_{C_2}(\mathrm{tmf}_1(3)[\bar{a}_3^{-1}]) \\ &\xrightarrow{g} \mathrm{Pic}_{C_2}(\mathrm{tmf}_1(3)[\bar{a}_1^{-1}, \bar{a}_3^{-1}]). \end{aligned}$$

By Corollary 5.1, we have $\pi_0^{C_2} \mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}] \cong \mathbb{Z}\left[\tfrac{1}{3}\right][(\bar{a}_1^3\bar{a}_3^{-1})^{\pm 1}]$. Thus,

$$\mathrm{GL}_1\, \pi_0^{C_2} \mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}] \cong \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}/2,$$

generated by $\frac{1}{3}$, $\bar{a}_1^3\bar{a}_3^{-1}$ and -1, and $\mathrm{coker}(f) \cong \mathbb{Z}$, generated by $[\bar{a}_1^3\bar{a}_3^{-1}]$.

We claim that $\partial(\bar{a}_1^3\bar{a}_3^{-1}) \simeq S^{3\rho} \wedge \mathrm{Tmf}_1(3)$. Indeed, we have trivializations

$$\bar{a}_3\colon S^{3\rho} \wedge \mathrm{tmf}_1(3)[\bar{a}_3^{-1}] \to \mathrm{tmf}_1(3)[\bar{a}_3^{-1}]$$

and

$$\bar{a}_1^3\colon S^{3\rho} \wedge \mathrm{tmf}_1(3)[\bar{a}_1^{-1}] \to \mathrm{tmf}_1(3)[\bar{a}_1^{-1}].$$

Therefore, we get $S^{3\rho} \wedge \mathrm{Tmf}_1(3)$ by gluing $\mathrm{tmf}_1(3)[\bar{a}_3^{-1}]$ and $\mathrm{tmf}_1(3)[\bar{a}_1^{-1}]$ by the map $\bar{a}_1^3\bar{a}_3^{-1}$ on $\mathrm{tmf}_1(3)[\bar{a}_1^{-1}\bar{a}_3^{-1}]$.

By Theorem 6.9, $\ker(g) \cong \mathbb{Z}/48$. Furthermore, $\Sigma^{8-8\sigma}\,\mathrm{Tmf}_1(3) \simeq_{C_2} \mathrm{Tmf}_1(3)$ as $u_{2\sigma}^4$ is a permanent cycle. Thus, we get a commutative diagram:

$$\begin{tikzcd}
0 \arrow[r] & \mathbb{Z} \arrow[r, "3\rho"] \arrow[d, "\cong"] & RO(C_2)/(8-8\sigma) \arrow[r] \arrow[d] & RO(C_2)/(8-8\sigma, 3\rho) \cong \mathbb{Z}/48 \arrow[r] \arrow[d, "\cong"] & 0 \\
0 \arrow[r] & \mathrm{coker}(f) \arrow[r] & \mathrm{Pic}_{C_2}(\mathrm{Tmf}_1(3)) \arrow[r] & \ker(g) \arrow[r] & 0
\end{tikzcd}$$

Thus, the middle map is also an isomorphism. □

Remark 6.13 The map $\mathrm{Pic}(\mathrm{Tmf}) \to \mathrm{Pic}(\mathrm{Tmf}_0(3))$ is *not* surjective. The former has been identified with $\mathbb{Z} \oplus \mathbb{Z}/24$ in [51, Theorem B, Construction 8.4.2], where the summands are generated by $\Sigma\,\mathrm{Tmf}$ and by the global sections $\mathcal{O}^{\mathrm{top}}(\mathcal{I})$. Here, $\mathcal{I}$ is a line bundle on the derived compactified moduli stack of elliptic curves $(\overline{\mathcal{M}}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}})$ obtained by gluing $\Sigma^{24}\mathcal{O}^{\mathrm{top}}$ on $\mathcal{M}_{\mathrm{ell}}$ and $\Sigma^{24}\mathcal{O}^{\mathrm{top}}$ on $\overline{\mathcal{M}}_{\mathrm{ell}}[c_4^{-1}]$ via the clutching function $j = c_4^3/\Delta$.

We claim that the module $\mathcal{O}^{\mathrm{top}}(\mathcal{I}) \wedge_{\mathrm{Tmf}} \mathrm{Tmf}_0(3)$ is 2–torsion in $\mathrm{Pic}(\mathrm{Tmf}_0(3))$. Indeed, for $p\colon \overline{\mathcal{M}}_0(3) \to \overline{\mathcal{M}}_{\mathrm{ell}}$, we have for an arbitrary locally free sheaf $\mathcal{F}$ on $(\overline{\mathcal{M}}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}})$ an equivalence

$$\begin{aligned}
\Gamma(\mathcal{F}) \wedge_{\mathrm{Tmf}} \mathrm{Tmf}_0(3) &\simeq \Gamma\big(\overline{\mathcal{M}}_{\mathrm{ell}}; \mathcal{F} \wedge_{\mathcal{O}^{\mathrm{top}}} p_*\mathcal{O}^{\mathrm{top}}_{\overline{\mathcal{M}}_0(3)}\big) \\
&\simeq \Gamma\big(\overline{\mathcal{M}}_{\mathrm{ell}}; p_*\big(p^*\mathcal{F} \wedge_{\mathcal{O}^{\mathrm{top}}_{\overline{\mathcal{M}}_0(3)}} \mathcal{O}^{\mathrm{top}}_{\overline{\mathcal{M}}_0(3)}\big)\big) \\
&\simeq \Gamma(\overline{\mathcal{M}}_0(3); p^*\mathcal{F}).
\end{aligned}$$

In the first equivalence, we use that $(\overline{\mathcal{M}}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}})$ is 0–affine and in the second we use the projection formula (see [24, Exercise II.5.1d] for the algebraic statement, from which the topological can be deduced). Thus, $\mathcal{O}^{\mathrm{top}}(\mathcal{I}) \wedge_{\mathrm{Tmf}} \mathrm{Tmf}_0(3)$ can be constructed as $\mathcal{O}^{\mathrm{top}}(p^*\mathcal{I})$, where $p^*\mathcal{I}$ can be constructed by an analogous gluing construction on $\overline{\mathcal{M}}_0(3)$, gluing $\Sigma^{24}\mathcal{O}^{\mathrm{top}}$ on $\mathcal{M}_0(3)$ and $\Sigma^{24}\mathcal{O}^{\mathrm{top}}$ on $\overline{\mathcal{M}}_0(3)[c_4^{-1}]$ via the clutching function $j = c_4^3/\Delta$ with $c_4 = a_1^4 - 24a_1a_3$. There is an equivalence of gluing data

$$(\mathcal{O}^{\mathrm{top}}, \mathcal{O}^{\mathrm{top}}, \mathrm{id}) \to (\Sigma^{48}\mathcal{O}^{\mathrm{top}}, \Sigma^{48}\mathcal{O}^{\mathrm{top}}, j^2)$$

given by $\Delta^2\colon \mathcal{O}^{\mathrm{top}} \to \mathcal{O}^{\mathrm{top}}$ on $\mathcal{M}_0(3)$ and $c_4^6\colon \mathcal{O}^{\mathrm{top}} \to \Sigma^{48}\mathcal{O}^{\mathrm{top}}$ on $\overline{\mathcal{M}}_0(3)[c_4^{-1}]$. Note here that $\Delta^2 = \overline{\Delta}^2 u_{2\sigma}^{12}$ is a permanent cycle for $\mathrm{TMF}_0(3)$. Thus, $2 \cdot [\mathcal{I}] = 0 \in \mathrm{Pic}(\overline{\mathcal{M}}_0(3), \mathcal{O}^{\mathrm{top}}) \cong \mathrm{Pic}(\mathrm{Tmf}_0(3))$.

As not every torsion in $\mathrm{Pic}(\mathrm{Tmf}_0(3))$ is 2–torsion, $\mathrm{Pic}(\mathrm{Tmf}) \to \mathrm{Pic}(\mathrm{Tmf}_0(3))$ is indeed not surjective.

References

[1] **D W Anderson**, *Universal coefficient theorems for K-theory*, preprint (1970) Available at http://faculty.tcu.edu/gfriedman/notes/

[2] **M Ando**, **M J Hopkins**, **C Rezk**, *Multiplicative orientations of KO–theory and of the spectrum of topological modular forms*, preprint (2010) Available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.128.1530

[3] **S Araki**, *Orientations in τ–cohomology theories*, Japan. J. Math. 5 (1979) 403–430 MR

[4] **M F Atiyah**, *K–theory and reality*, Quart. J. Math. Oxford Ser. 17 (1966) 367–386 MR

[5] **A Baker**, **B Richter**, *Invertible modules for commutative $\mathbb{S}$–algebras with residue fields*, Manuscripta Math. 118 (2005) 99–119 MR

[6] **T Bauer**, *Computation of the homotopy of the spectrum* `tmf`, from "Groups, homotopy and configuration spaces" (N Iwase, T Kohno, R Levi, D Tamaki, J Wu, editors), Geom. Topol. Monogr. 13, Geom. Topol. Publ., Coventry (2008) 11–40 MR

[7] **M Behrens**, *A modular description of the $K(2)$–local sphere at the prime 3*, Topology 45 (2006) 343–402 MR

[8] **M Behrens**, *The construction of tmf*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) 131–188 MR

[9] **M Behrens**, **M Hill**, **M J Hopkins**, **M Mahowald**, *On the existence of a v_2^{32}–self map on $M(1,4)$ at the prime 2*, Homology Homotopy Appl. 10 (2008) 45–84 MR

[10] **M Behrens**, **K Ormsby**, *On the homotopy of $Q(3)$ and $Q(5)$ at the prime 2*, Algebr. Geom. Topol. 16 (2016) 2459–2534 MR

[11] **M Behrens**, **K Ormsby**, **N Stapleton**, **V Stojanoska**, *On the ring of cooperations for 2–primary connective topological modular forms*, preprint (2015) arXiv

[12] **M Behrens**, **S Pemmaraju**, *On the existence of the self map v_2^9 on the Smith–Toda complex $V(1)$ at the prime 3*, from "Homotopy theory: relations with algebraic geometry, group cohomology, and algebraic K–theory" (P Goerss, S Priddy, editors), Contemp. Math. 346, Amer. Math. Soc., Providence, RI (2004) 9–49 MR

[13] **A J Blumberg**, **M A Hill**, *Operadic multiplications in equivariant spectra, norms, and transfers*, Adv. Math. 285 (2015) 658–708 MR

[14] **U Bunke**, **T Nikolaus**, *Twisted differential cohomology*, preprint (2014) arXiv

[15] **K Cesnavicius**, *A modular description of* $\mathscr{X}_0(n)$, preprint (2015) arXiv

[16] **B Conrad**, *Arithmetic moduli of generalized elliptic curves*, J. Inst. Math. Jussieu 6 (2007) 209–278 MR

[17] **P Deligne**, **M Rapoport**, *Les schémas de modules de courbes elliptiques*, from "Modular functions of one variable, II", Lecture Notes in Math. 349, Springer (1973) 143–316 MR

[18] **D Dugger**, *Multiplicative structures on homotopy spectral sequences, II*, preprint (2003) arXiv

[19] **A D Elmendorf**, **I Kriz**, **M A Mandell**, **J P May**, *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs 47, Amer. Math. Soc., Providence, RI (1997) MR

[20] **P G Goerss**, *Realizing families of Landweber exact homology theories*, from "New topological contexts for Galois theory and algebraic geometry" (A Baker, B Richter, editors), Geom. Topol. Monogr. 16, Geom. Topol. Publ., Coventry (2009) 49–78 MR

[21] **U Görtz**, **T Wedhorn**, *Algebraic geometry, I: Schemes with examples and exercises*, Vieweg + Teubner, Wiesbaden (2010) MR

[22] **J P C Greenlees**, **L Meier**, *Gorenstein duality for Real spectra*, preprint (2016) arXiv

[23] **M Groth**, *A short course on ∞–categories*, preprint (2010) arXiv

[24] **R Hartshorne**, *Algebraic geometry*, Graduate Texts in Mathematics 52, Springer (1977) MR

[25] **R Hartshorne**, *Stable reflexive sheaves*, Math. Ann. 254 (1980) 121–176 MR

[26] **M A Hill**, *The String bordism of BE_8 and $BE_8 \times BE_8$ through dimension* 14, Illinois J. Math. 53 (2009) 183–196 MR

[27] **M A Hill**, *The equivariant slice filtration: a primer*, Homology Homotopy Appl. 14 (2012) 143–166 MR

[28] **M A Hill**, **M J Hopkins**, **D C Ravenel**, *On the nonexistence of elements of Kervaire invariant one*, Ann. of Math. 184 (2016) 1–262 MR

[29] **M Hill**, **T Lawson**, *Topological modular forms with level structure*, Invent. Math. 203 (2016) 359–416 MR

[30] **H Hohnhold**, *The Landweber exact functor theorem*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) 35–46 MR

[31] **M Hovey**, **J H Palmieri**, **N P Strickland**, *Axiomatic stable homotopy theory*, Mem. Amer. Math. Soc. 610, Amer. Math. Soc., Providence, RI (1997) MR

[32] **P Hu**, **I Kriz**, *Real-oriented homotopy theory and an analogue of the Adams–Novikov spectral sequence*, Topology 40 (2001) 317–399 MR

[33] **A Joyal**, *The theory of quasi-categories and its applications*, lecture notes (2008)

[34] **N M Katz**, **B Mazur**, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies 108, Princeton University Press (1985) MR

[35] **J Konter**, *The homotopy groups of the spectrum Tmf*, preprint (2012) arXiv

[36] **P S Landweber**, *Conjugations on complex manifolds and equivariant homotopy of MU*, Bull. Amer. Math. Soc. 74 (1968) 271–274 MR

[37] **T Lawson**, **N Naumann**, *Strictly commutative realizations of diagrams over the Steenrod algebra and topological modular forms at the prime 2*, Int. Math. Res. Not. 2014 (2014) 2773–2813 MR

[38] **L G Lewis, Jr**, **M A Mandell**, *Equivariant universal coefficient and Künneth spectral sequences*, Proc. London Math. Soc. 92 (2006) 505–544 MR

[39] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton University Press (2009) MR

[40] **J Lurie**, *A survey of elliptic cohomology*, from "Algebraic topology" (N A Baas, E M Friedlander, B Jahren, P A Østvær, editors), Abel Symp. 4, Springer (2009) 219–277 MR

[41] **J Lurie**, *Chromatic homotopy theory*, lecture notes (2010) Available at `http://math.harvard.edu/~lurie/252x.html`

[42] **J Lurie**, *Derived algebraic geometry, VIII: Quasi-coherent sheaves and Tannaka duality theorems*, unpublished manuscript (2011) Available at `http://math.harvard.edu/~lurie`

[43] **J Lurie**, *Derived algebraic geometry, XII: Proper morphisms, completions, and the Grothendieck existence theorem*, unpublished manuscript (2011) Available at `http://math.harvard.edu/~lurie`

[44] **J Lurie**, *Higher algebra*, unpublished manuscript (2014) Available at `http://math.harvard.edu/~lurie`

[45] **M Mahowald**, **M Hopkins**, *The structure of 24 dimensional manifolds having normal bundles which lift to $BO[8]$*, from "Recent progress in homotopy theory" (D M Davis, J Morava, G Nishida, W S Wilson, N Yagita, editors), Contemp. Math. 293, Amer. Math. Soc., Providence, RI (2002) 89–110 MR

[46] **M Mahowald**, **C Rezk**, *Topological modular forms of level 3*, Pure Appl. Math. Q. 5 (2009) 853–872 MR

[47] **M A Mandell**, **J P May**, *Equivariant orthogonal spectra and S–modules*, Mem. Amer. Math. Soc. 755, Amer. Math. Soc., Providence, RI (2002) MR

[48] **M A Mandell**, **J P May**, **S Schwede**, **B Shipley**, *Model categories of diagram spectra*, Proc. London Math. Soc. 82 (2001) 441–512 MR

[49] **A Mathew**, *The homology of tmf*, Homology Homotopy Appl. 18 (2016) 1–29 MR

[50] **A Mathew**, **L Meier**, *Affineness and chromatic homotopy theory*, J. Topol. 8 (2015) 476–528 MR

[51] **A Mathew**, **V Stojanoska**, *The Picard group of topological modular forms via descent theory*, Geom. Topol. 20 (2016) 3133–3217 MR

[52] **L Meier**, *United elliptic homology*, PhD thesis, University of Bonn (2012) Available at `http://hss.ulb.uni-bonn.de/2012/2969/2969.htm`

[53] **L Meier**, *Vector bundles on the moduli stack of elliptic curves*, J. Algebra 428 (2015) 425–456 MR

[54] **N Naumann**, *The stack of formal groups in stable homotopy theory*, Adv. Math. 215 (2007) 569–600 MR

[55] **N Naumann**, **M Spitzweck**, **P A Østvær**, *Motivic Landweber exactness*, Doc. Math. 14 (2009) 551–593 MR

[56] **N Ricka**, *Equivariant Anderson duality and Mackey functor duality*, Glasg. Math. J. 58 (2016) 649–676 MR

[57] **J Rognes**, *Galois extensions of structured ring spectra, Stably dualizable groups*, Mem. Amer. Math. Soc. 898, Amer. Math. Soc., Providence, RI (2008) MR

[58] **S Schwede**, *Lectures on equivariant stable homotopy theory*, lecture notes (2014) Available at `http://www.math.uni-bonn.de/people/schwede/`

[59] **M Spitzweck**, *Operads, algebras and modules in general model categories*, preprint (2001) arXiv

[60] **V Stojanoska**, *Duality for topological modular forms*, Doc. Math. 17 (2012) 271–311 MR

Department of Mathematics, University of California, Los Angeles
Box 951555, Los Angeles, CA 90095-1555, United States

Mathematisches Institut, University of Bonn
Endenicher Allee 60, D-53115 Bonn, Germany

`mikehill@math.ucla.edu`, `lmeier@math.uni-bonn.de`

`http://math.ucla.edu/~mikehill`,
`http://www.math.uni-bonn.de/people/lmeier/`

Received: 3 August 2015 Revised: 4 November 2016

Algebraic & Geometric Topology 17 (2017) 2013–2038

An algebraic model for commutative $H\mathbb{Z}$–algebras

BIRGIT RICHTER
BROOKE SHIPLEY

We show that the homotopy category of commutative algebra spectra over the Eilenberg–Mac Lane spectrum of an arbitrary commutative ring R is equivalent to the homotopy category of E_∞–monoids in unbounded chain complexes over R. We do this by establishing a chain of Quillen equivalences between the corresponding model categories. We also provide a Quillen equivalence to commutative monoids in the category of functors from the category of finite sets and injections to unbounded chain complexes.

55P43

1 Introduction

Let R be an arbitrary commutative ring. In Shipley [29] it was shown that the model category of algebra spectra over the Eilenberg–Mac Lane spectrum, HR, is connected to the model category of differential graded R–algebras via a chain of Quillen equivalences. In this paper we extend this result to the case of commutative HR–algebra spectra. As a guiding example we consider the function spectrum $F(X, HR)$ from a space X to the Eilenberg–Mac Lane spectrum of a commutative ring R. As R is commutative, $F(X, HR)$ is a commutative HR–algebra spectrum whose homotopy groups are the cohomology groups of the space X with coefficients in R:

$$\pi_{-n} F(X, HR) \cong H^n(X; R).$$

The singular cochains on X with coefficients in R, denoted by $S^*(X; R)$, give a chain model of the cohomology of X by regrading. We set

$$C_{-*}(X; R) := S^*(X; R).$$

Note that for $R = \mathbb{F}_p$ the Steenrod operations on $H^*(X; R)$ can be constructed from the $\cup_i$–products. These are chain homotopies that measure the failure of the cup-product to produce a strictly graded commutative product of cochains. Thus, in general, one cannot expect to find a model of the singular cochains of a space that is a differential graded commutative R–algebra. Instead, one must work with E_∞–algebra structures.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2013

See also Cenkl [3, Theorem 2]. A notable exception are rational cochains of a space with the Sullivan cochains as a strictly differential graded commutative model.

We establish a chain of Quillen equivalences between commutative HR–algebra spectra, $C(HR\text{–mod})$, and differential graded E_∞–R–algebras, $E_\infty\mathsf{Ch}_R$:

$$\begin{tikzcd}
C(HR\text{–mod}) \arrow[r, shift left, "Z"] & C(\mathsf{Sp}^\Sigma(s\mathsf{mod}_R)) \arrow[l, shift left, "U"] \arrow[r, shift right, leftarrow, "L_N"'] \arrow[r, shift left, "\Phi^* N"'] & C(\mathsf{Sp}^\Sigma(\mathsf{ch}_R)) \arrow[r, shift left, "i"] & C(\mathsf{Sp}^\Sigma(\mathsf{Ch}_R)) \arrow[l, shift left, "C_0"] \arrow[d, shift left, "R_\varepsilon"] \\
 & & E_\infty\mathsf{Ch}_R \arrow[r, shift left, "F_0"] & E_\infty(\mathsf{Sp}^\Sigma(\mathsf{Ch}_R)) \arrow[l, shift left, "\mathrm{Ev}_0"] \arrow[u, shift left, "L_\varepsilon"]
\end{tikzcd}$$

Here, our intermediary categories include symmetric spectra (Sp^Σ) over the categories of simplicial R–modules ($s\mathsf{mod}_R$), nonnegatively graded chain complexes over R (ch_R), and unbounded chain complexes over R (Ch_R). The functors will be introduced in the sections below.

The fact that there is such an equivalence should not be surprising, but to our knowledge, no explicit chain of Quillen equivalences can be found in the literature.

In the context of infinite loop space theory, E_∞–ring spectra, and their units, the theory of $\mathcal{I}$–spaces is important; see Sagave and Schlichtkrull [22]. Here $\mathcal{I}$ is the category of finite sets and injections and $\mathcal{I}$–spaces are functors from $\mathcal{I}$ to simplicial sets. More generally, functor categories from $\mathcal{I}$ to categories of modules feature as FI–modules in the work of Church, Ellenberg and Farb [6] and others. We relate symmetric spectra in unbounded chain complexes over R via a chain of Quillen equivalences to the category of unbounded $\mathcal{I}$–chain complexes and prove that commutative monoids in this category, $C(\mathsf{Ch}_R^{\mathcal{I}})$, provide an alternative model for commutative HR–algebra spectra. In fact, there is a chain of Quillen equivalences between $C(HR\text{–mod})$ and $E_\infty(\mathsf{Ch}_R^{\mathcal{I}})$, the E_∞–monoids in unbounded $\mathcal{I}$–chain complexes over R, that passes via $E_\infty(\mathsf{Sp}^\Sigma(\mathsf{Ch}_R))$ and $E_\infty\mathsf{Ch}_R$. The rigidification result of Pavlov and Scholbach [20, Theorem 3.4.4] for symmetric spectra implies that the model category $E_\infty(\mathsf{Ch}_R^{\mathcal{I}})$ is Quillen equivalent to the one of commutative monoids in $\mathsf{Ch}_R^{\mathcal{I}}$, that is, $C(\mathsf{Ch}_R^{\mathcal{I}})$. Taking these results together we obtain a chain of Quillen equivalences between commutative HR–algebra spectra and commutative monoids in $\mathcal{I}$–chain complexes over R. See Theorem 9.5. We expect that our comparison result makes it possible to find explicit commutative $\mathcal{I}$–chain models for certain commutative HR–algebras and there is ongoing work on this by Richter, Sagave and Schulz with applications to logarithmic structures on commutative ring spectra in mind.

If $R=\mathbb{Q}$ is the field of rational numbers we can extend our chain of Quillen equivalences and obtain a comparison (Corollary 8.4) between commutative $H\mathbb{Q}$–algebra spectra and differential graded commutative $\mathbb{Q}$–algebras.

Mike Mandell showed in [16, Theorem 7.11] that for every commutative ring R the homotopy categories of E_∞–HR–algebra spectra and of E_∞–monoids in the category of unbounded R–chain complexes are equivalent. He also claims in loc. cit. that he can improve this equivalence of homotopy categories to an actual chain of Quillen equivalences. He suggests using the methods of Schwede and Shipley [27], but only associative monoids are treated there.

Our approach is different from Mandell's because we work in the setting of symmetric spectra. The idea to integrate the symmetric groups into the monoidal structure to construct a symmetric monoidal category of spectra is due to Jeff Smith. Our arguments heavily rely on combinatorial and monoidal features of the category of symmetric spectra in the categories of simplicial sets, simplicial R–modules, nonnegatively graded chain complexes (ch_R) and unbounded chain complexes (Ch_R).

The structure of the paper is as follows: We recall some basic facts and some model categorical features of symmetric spectra in Section 2. In Section 3 we recall results from Pavlov and Scholbach [19; 20] that establish model structures on commutative ring spectra in the cases that arise as intermediate steps in our chain of Quillen equivalences and we also recall their rigidification result. We sketch how to use methods from Chadwick and Mandell [4] for an alternative proof. The Quillen equivalence between commutative HR–algebra spectra and commutative symmetric ring spectra in simplicial R–modules can be found in Section 4 as Theorem 4.1. The Quillen equivalence between the latter model category and commutative symmetric ring spectra in nonnegatively graded chain complexes is based on the Dold–Kan correspondence and is stated as Theorem 6.6 in Section 6. There is a natural inclusion functor $i\colon \mathsf{ch} \to \mathsf{Ch}$ and the Quillen equivalence between commutative symmetric ring spectra in ch and in Ch (see Corollary 7.3) is based on this functor. In Section 8 we establish a Quillen equivalence between E_∞–monoids in symmetric spectra in unbounded chain complexes and E_∞–monoids in unbounded chain complexes. The link with E_∞–monoids and commutative monoids in the diagram category of chain complexes indexed by the category of finite sets and injections is worked out in Section 9.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant No. 0932078000 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the spring 2014 program on algebraic topology. Shipley was also supported during this project by the NSF under Grants No. 1104396 and 1406468. We are grateful to Dmitri Pavlov and Jakob Scholbach for sharing draft versions of Pavlov and Scholbach [19; 20] with us. We thank Benjamin Antieau and Steffen Sagave for helpful comments on an earlier version of this paper.

2 Background

In the following we will consider model category structures that are transferred by an adjunction. Given an adjunction

$$\mathcal{C} \underset{R}{\overset{L}{\rightleftarrows}} \mathcal{D}$$

where $\mathcal{C}$ is a model category and $\mathcal{D}$ is a bicomplete category, we call a model structure on $\mathcal{D}$ *right-induced* if the weak equivalences and fibrations in $\mathcal{D}$ are determined by the right adjoint functor R.

We use the general setting of symmetric spectra as in [11]. Let $(\mathcal{C}, \otimes, \mathbf{1})$ be a bicomplete closed symmetric monoidal category and let K be an object of $\mathcal{C}$. A symmetric sequence in $\mathcal{C}$ is a family of objects $X(n) \in \mathcal{C}$ with $n \in \mathbb{N}_0$ such that the n^{th} level $X(n)$ carries an action of the symmetric group Σ_n. Symmetric sequences form a category $\mathcal{C}^\Sigma$ whose morphisms are given by families of Σ_n–equivariant morphisms $f(n)$, $n \geqslant 0$. For every $r \geqslant 0$ there is a functor $G_r \colon \mathcal{C} \to \mathcal{C}^\Sigma$ with

$$G_r(C)(n) = \begin{cases} \Sigma_n \times C & \text{for } n = r, \\ \varnothing & \text{for } n \neq r, \end{cases}$$

where $\varnothing$ denotes the initial object of $\mathcal{C}$. Here $\Sigma_n \times C = \bigsqcup_{\Sigma_n} C$ carries the Σ_n–action that permutes the summands.

We consider the symmetric sequence $\mathsf{Sym}(K)$ whose n^{th} level is $K^{\otimes n}$. Here we follow the usual convention that $K^{\otimes 0}$ is the unit $\mathbf{1}$. The category $\mathcal{C}^\Sigma$ inherits a symmetric monoidal structure from $\mathcal{C}$: for X, $Y \in \mathcal{C}^\Sigma$ we set

$$(X \odot Y)(n) = \bigsqcup_{p+q=n} \Sigma_n \times_{\Sigma_p \times \Sigma_q} X(p) \otimes Y(q).$$

It is straightforward to show (see for instance [11, Section 7]) that $\mathsf{Sym}(K)$ is a commutative monoid in $\mathcal{C}^\Sigma$.

The category of *symmetric spectra* (in $\mathcal{C}$ with respect to K), $\mathsf{Sp}^\Sigma(\mathcal{C}, K)$, is the category of right $\mathsf{Sym}(K)$–modules in $\mathcal{C}^\Sigma$. Explicitly, a symmetric spectrum is a family of Σ_n–objects $X(n) \in \mathcal{C}$ together with Σ_n–equivariant maps

$$X(n) \otimes K \to X(n+1)$$

for all $n \geqslant 0$ such that the composites

$$X(n) \otimes K^{\otimes p} \to X(n+1) \otimes K^{\otimes p-1} \to \cdots \to X(n+p)$$

are $\Sigma_n \times \Sigma_p$–equivariant for all $n, p \geqslant 0$. Morphisms in $\mathsf{Sp}^\Sigma(\mathcal{C}, K)$ are morphisms of symmetric sequences that are compatible with the right $\mathsf{Sym}(K)$–module structure.

There is an evaluation functor Ev_n that maps an $X \in \mathsf{Sp}^\Sigma(\mathcal{C}, K)$ to $X(n) \in \mathcal{C}$. This functor has a left adjoint

$$F_n\colon \mathcal{C} \to \mathsf{Sp}^\Sigma(\mathcal{C}, K)$$

such that $F_n(C)(m)$ is the initial object for $m < n$ and

$$F_n(C)(m) \cong \Sigma_m \times_{\Sigma_{m-n}} C \otimes K^{\otimes m-n} \quad \text{if } m \geqslant n.$$

Note that $F_n(C) \cong G_n(C) \odot \mathsf{Sym}(K)$.

Symmetric spectra form a symmetric monoidal category $(\mathsf{Sp}^\Sigma(\mathcal{C}, K), \wedge, \mathsf{Sym}(K))$ such that for $X, Y \in \mathsf{Sp}^\Sigma(\mathcal{C}, K)$,

$$X \wedge Y = X \odot_{\mathsf{Sym}(K)} Y.$$

Here $X \odot_{\mathsf{Sym}(K)} Y$ denotes the coequalizer of the diagram

$$X \odot \mathsf{Sym}(K) \odot Y \rightrightarrows X \odot Y$$

where we use the right action of $\mathsf{Sym}(K)$ on X and we use the right action of $\mathsf{Sym}(K)$ on Y after applying the twist-map in the symmetric monoidal structure on $\mathcal{C}^\Sigma$.

A crucial map is

$$\lambda\colon F_1 K \to F_0 \mathbf{1}; \tag{1}$$

it is given as the adjoint to the identity map $K \to \mathrm{Ev}_1 F_0 \mathbf{1} = K$.

We recall the basics about model category structures on symmetric spectra from [11]: If $\mathcal{C}$ is a closed symmetric monoidal model category which is left proper and cellular and if K is a cofibrant object of $\mathcal{C}$, then there is a *projective model structure on the category* $\mathsf{Sp}^\Sigma(\mathcal{C}, K)$ [11, Theorem 8.2], $\mathsf{Sp}^\Sigma(\mathcal{C}, K)_{\mathrm{proj}}$, such that the fibrations and weak equivalences are levelwise fibrations and weak equivalences in $\mathcal{C}$ and such that the cofibrations are determined by the left lifting property with respect to the class of acyclic fibrations.

This model structure has a Bousfield localization with respect to the set of maps

$$\{\zeta_n^{QC}\colon F_{n+1}(QC \otimes K) \to F_n(QC) \mid n \geqslant 0\},$$

where $Q(-)$ is a cofibrant replacement and C runs through the domains and codomains of the generating cofibrations of $\mathcal{C}$. The map ζ_n^{QC} is adjoint to the inclusion map into the component of $F_n(QC)(n+1)$ corresponding to the identity permutation. We call the Bousfield localization of $\mathsf{Sp}^\Sigma(\mathcal{C}, K)_{\mathrm{proj}}$ at this set of maps the *stable model structure on* $\mathsf{Sp}^\Sigma(\mathcal{C}, K)$ and denote it by $\mathsf{Sp}^\Sigma(\mathcal{C}, K)^s$.

As we are interested in commutative monoids in symmetric spectra, we use positive variants of the above mentioned model structures: Let $\mathsf{Sp}^\Sigma(\mathcal{C}, K)^+_{\mathrm{proj}}$ be the model structure where fibrations are maps that are fibrations in each level $n \geqslant 1$ and weak equivalences are levelwise weak equivalences for positive levels. The cofibrations are again determined by their lifting property and they turn out to be isomorphisms in level zero (compare [17, Section 14]). By adapting the localizing set and considering only positive n, we get the positive stable model structure on $\mathsf{Sp}^{\Sigma}(\mathcal{C}, K)$ and we denote it by $\mathsf{Sp}^{\Sigma}(\mathcal{C}, K)^{s,+}$.

Remark 2.1 We consider several examples of categories $\mathcal{C}$ with different choices of objects $K \in \mathcal{C}$. Despite the name, the stable model structure on $\mathsf{Sp}^{\Sigma}(\mathcal{C}, K)$ does not have to define a stable model category in the sense that the category is pointed with a homotopy category that carries an invertible suspension functor. Proposition 9.1 for instance makes this explicit in the case when K is the unit of the symmetric monoidal structure on $\mathcal{C}$.

3 Model structures on algebras over an operad over $\mathsf{Sp}^{\Sigma}(\mathcal{C})$ for $\mathcal{C} = \mathsf{ch}, \mathsf{sAb}, \mathsf{Ch}$

From now on we restrict to the case $R = \mathbb{Z}$ in order to ease notation. The proofs work in general.

Establishing right-induced model structures for commutative monoids in model categories is hard. Sometimes it is not possible, for instance there is no right-induced model structure on differential graded commutative rings, because the free functor does not respect acyclicity. However, if the underlying model category is nice enough, then such model structures can be established. In broader generality, one might ask whether algebras over operads possess a right-induced model structure. In our setting we will apply the results of Pavlov and Scholbach. They show in [19, Theorem 5.10] and [20, Theorem 3.4.1] that for a tractable, pretty small, left proper, h–monoidal, flat symmetric monoidal model category $\mathcal{C}$ the category of $\mathcal{O}$–algebras in $\mathsf{Sp}^{\Sigma}(\mathcal{C}, K)^{s,+}$ has a right-induced model structure. Here $\mathcal{O}$ is an operad in $\mathcal{C}$. See loc. cit. for an explanation of the assumptions. These conditions are satisfied for the model categories of simplicial abelian groups and both nonnegatively graded and unbounded chain complexes. Hence, using their results, we obtain:

Theorem 3.1 *The category of $\mathcal{O}$–algebras in $\mathsf{Sp}^{\Sigma}(\mathcal{C}, K)^{s,+}$ has a right-induced model structure for $\mathcal{C} = \mathsf{sAb}, \mathsf{ch}, \mathsf{Ch}$, any K and any operad $\mathcal{O}$ in $\mathcal{C}$.*

We follow the convention that an E_∞–operad $\mathcal{P}$ in Ch (or ch, sAb) is a symmetric unital operad whose augmentation induces a weak equivalence to the operad that describes commutative monoids. For the sake of brevity we call algebras over an E_∞–operad *E_∞–monoids*. Pavlov and Scholbach also prove a rigidification theorem

[19, Theorem 7.5; 20, Theorem 3.4.4]. We apply this to the case of E_∞–monoids and in this case it provides a Quillen equivalence between the model category of E_∞–monoids in $\mathsf{Sp}^\Sigma(\mathcal{C}, K)^{s,+}$ and commutative monoids in $\mathsf{Sp}^\Sigma(\mathcal{C}, K)^{s,+}$. Related rectification results in the setting of spaces instead of chain complexes are due to [8] and [22]. Berger and Moerdijk obtain general results about rectifications of homotopy algebra structures in [2].

Other approaches to model structures for commutative monoids in symmetric spectra and rigidification results can be found for instance in work by John Harper [9], David White [31], and Steven Chadwick and Michael Mandell [4].

In the following we sketch an alternative proof of the existence of a positive stable right-induced model structure for the category of symmetric spectra in the category of unbounded chain complexes, $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$, where $\mathbb{Z}[1]$ denotes the chain complex which is concentrated in chain degree one with chain group $\mathbb{Z}$. This proof uses a modification of the methods used by Chadwick and Mandell [4]. A similar proof works for the categories of symmetric spectra in simplicial abelian groups, $\mathsf{Sp}^\Sigma(\mathsf{sAb}, \widetilde{\mathbb{Z}}(\mathbb{S}^1))$, with $K = \widetilde{\mathbb{Z}}(\mathbb{S}^1)$ the reduced free abelian simplicial group generated by the simplicial 1–sphere, and for symmetric spectra in the category of nonnegatively graded chain complexes, $\mathsf{Sp}^\Sigma(\mathsf{ch}, \mathbb{Z}[1])$.

A reader who is just interested in the application of these results is invited to resume reading in Section 4.

Theorem 3.2 *Let $\mathcal{O}$ be an operad in Ch. Then the category $\mathcal{O}(\mathsf{Sp}^\Sigma(\mathsf{Ch}))$ of $\mathcal{O}$–algebras over $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ is a model category with fibrations and weak equivalences created in the positive stable model structure on $\mathsf{Sp}^\Sigma(\mathsf{Ch})$.*

Theorem 3.3 *Let $\phi\colon \mathcal{O} \to \mathcal{O}'$ be a map of operads. The induced adjoint functors*

$$\mathcal{O}(\mathsf{Sp}^\Sigma(\mathsf{Ch})) \underset{R_\phi}{\overset{L_\phi}{\rightleftarrows}} \mathcal{O}'(\mathsf{Sp}^\Sigma(\mathsf{Ch}))$$

form a Quillen adjunction. This is a Quillen equivalence if $\phi(n)\colon \mathcal{O}(n) \to \mathcal{O}'(n)$ is a (nonequivariant) weak equivalence for each n.

In particular, if ε is the augmentation from any E_∞–operad to the commutative operad, then it induces a Quillen equivalence between the categories of E_∞–monoids and of commutative monoids in $\mathsf{Sp}^\Sigma(\mathsf{Ch})$.

The proofs of both of these theorems use the following statement, which is a translation of [17, Lemma 15.5] to $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ with a slight generalization based on [4, Remark 8.3(i)]. As a model for $E\Sigma_n$ in the category $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ we take F_0 applied to the normalization of the free simplicial abelian group generated by the nerve of the translation category of the symmetric group Σ_n.

Proposition 3.4 *Let X and Z be objects in $\mathsf{Sp}^\Sigma(\mathsf{Ch})$.*

(1) *Let K be a chain complex, assume X has a Σ_i–action, and let $n > 0$. Then the quotient map*
$$q\colon E\Sigma_{i+} \wedge_{\Sigma_i} ((F_n K)^{\wedge i} \wedge X) \to ((F_n K)^{\wedge i} \wedge X)/\Sigma_i$$
is a level homotopy equivalence.

(2) *For any positive cofibrant object X and any Σ_i–equivariant object Z,*
$$q\colon E\Sigma_{i+} \wedge_{\Sigma_i} (Z \wedge X^{\wedge i}) \to (Z \wedge X^{\wedge i})/\Sigma_i$$
is a π_*–isomorphism.

Proof First, the proof of [17, Lemma 15.5] easily translates to the setting of $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ from $\mathsf{Sp}^\Sigma(\mathcal{S})$ considered there. The key point is that if $q \geqslant ni$, then $E\Sigma_i \times \Sigma_q \to \Sigma_q$ is a $(\Sigma_i \times \Sigma_{q-ni})$–equivariant homotopy equivalence. As mentioned in [4, Remark 8.3(i)], the proof of the first statement in [17, Lemma 15.5] still works when X has a Σ_i–action because the Σ_i–action remains free on Σ_q (or $\mathcal{O}(q)$ in the explicit case there). Similarly the second statement here follows by the same cellular filtration of X as in [17, Lemma 15.5]. □

The proofs of both of the theorems above also require the following definition and statement of properties.

Definition 3.5 A chain map $i\colon A \to B$ in Ch is an *h–cofibration* if each homomorphism $i_n\colon A_n \to B_n$ has a section (or splitting). These are the cofibrations in a model structure on Ch; see [5, Example 3.4], [24, Proposition 4.6.2], or [18, Theorem 18.3.1]. We say a map $i\colon X \to Y$ in $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ is an *h–cofibration* if each level $i_n\colon X_n \to Y_n$ is an h–cofibration as a chain map.

Below we refer to Σ_n–equivariant h–cofibrations. These are Σ_n–equivariant maps for which the underlying nonequivariant map is an h–cofibration. We use the following properties of h–cofibrations below.

Proposition 3.6 (1) *The generating cofibrations and acyclic cofibrations in Ch are h–cofibrations.*

(2) *Sequential colimits and pushouts preserve h–cofibrations.*

(3) *If f and g are two h–cofibrations in Ch, then their pushout product $f \Box g$ is also an h–cofibration.*

(4) *If f is an h–cofibration in Ch, then $F_i f$ is an h–cofibration in $\mathsf{Sp}^\Sigma(\mathsf{Ch})$.*

(5) *For every Σ_n–equivariant object Z, subgroup H of Σ_n, Σ_n–equivariant h–cofibration f, and $i \geqslant n$, the map $Z \wedge_H F_i(f)$ is an h–cofibration.*

We write $\mathcal{O}I$ and $\mathcal{O}J$ for the sets of maps in $\mathcal{O}(\mathsf{Sp}^\Sigma(\mathsf{Ch}))$ obtained by applying the free $\mathcal{O}$–algebra functor to the generating cofibrations I and generating acyclic cofibrations J from [29]. Since $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ is a combinatorial model category and the free functor $\mathcal{O}$ commutes with filtered direct limits, to prove Theorem 3.2 it is enough to prove the following lemma by [26, Lemma 2.3].

Lemma 3.7 *Every sequential composition of pushouts in $\mathcal{O}(\mathsf{Sp}^\Sigma(\mathsf{Ch}))$ of maps in $\mathcal{O}J$ is a stable equivalence.*

Proof of Lemma 3.7 This follows as in [4, 8.7–8.10]. Chadwick and Mandell consider pushouts of algebras over an operad $\mathcal{O}$ for three different symmetric monoidal categories of spectra simultaneously (including $\mathsf{Sp}^\Sigma(\mathcal{S})$); all of their arguments hold as well for $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ using the properties of h–cofibrations listed in Proposition 3.6 and the generalization of [17, Lemma 15.5] given in Proposition 3.4(2). □

Proof of Theorem 3.3 This follows as in [4, Theorem 8.2] again using Proposition 3.6 and Proposition 3.4 . □

4 Commutative $H\mathbb{Z}$–algebras and $\mathsf{Sp}^\Sigma(\mathsf{sAb})$

In this section we consider the Quillen equivalence between $H\mathbb{Z}$–module spectra and $\mathsf{Sp}^\Sigma(\mathsf{sAb})$ and show that it also induces an equivalence on the associated categories of commutative monoids. Recall the functor Z from $H\mathbb{Z}$–modules to $\mathsf{Sp}^\Sigma(\mathsf{sAb})$ from [29] which is given by $Z(M) = \widetilde{\mathbb{Z}}(M) \wedge_{\widetilde{\mathbb{Z}}H\mathbb{Z}} H\mathbb{Z}$ where $\widetilde{\mathbb{Z}}$ is the free abelian group on the nonbasepoint simplices on each level. The right adjoint of Z is given by recognizing that $\mathsf{Sym}(\widetilde{\mathbb{Z}}(\mathbb{S}^1))$, the unit in $\mathsf{Sp}^\Sigma(\mathsf{sAb})$, is isomorphic to $\widetilde{\mathbb{Z}}(\mathbb{S}) \cong H\mathbb{Z}$. The right adjoint is labeled U for underlying. In [29, Proposition 4.3], the pair (Z, U) was shown to induce a Quillen equivalence on the standard model structures. Since Z is strong symmetric monoidal, (Z, U) also induces an adjunction between the commutative monoids. We use the right-induced model structure on commutative monoids in $\mathsf{Sp}^\Sigma(\mathsf{sAb})$ and $H\mathbb{Z}$–module spectra [20, Theorem 3.4.1].

Theorem 4.1 *The functors Z and U induce a Quillen equivalence between commutative $H\mathbb{Z}$–algebra spectra and commutative symmetric ring spectra over* sAb*:*

$$Z\colon C(H\mathbb{Z}\text{–mod}) \rightleftarrows C(\mathsf{Sp}^\Sigma(\mathsf{sAb})) : U$$

Proof It follows from [29, Proof of Proposition 4.3] that U preserves and detects all weak equivalences and fibrations since weak equivalences and fibrations are determined on the underlying category of symmetric spectra in pointed simplicial sets, $\mathsf{Sp}^\Sigma(\mathcal{S}_*)$.

To show that (Z, U) is a Quillen equivalence, by [17, Lemma A.2(iii)] it is enough to show that for all cofibrant commutative $H\mathbb{Z}$ algebras A, the map $A \to UZA$ is a stable equivalence. If A were in fact cofibrant as an $H\mathbb{Z}$ module spectrum, this would follow from the Quillen equivalence on the module level [29]. In the standard model structure on commutative algebra spectra though, cofibrant objects are not necessarily cofibrant as modules. The positive flat model (or R–model) structures from [28, Theorem 3.2] were developed for just this reason. In Lemma 4.2 we show that for positive flat cofibrant commutative $H\mathbb{Z}$ algebras B, the map $B \to UZB$ is a stable equivalence. It follows from Lemma 4.2 that $A \to UZA$ is a stable equivalence for all standard (positive) cofibrant commutative $H\mathbb{Z}$ algebras A, since such A are also positive flat cofibrant by [28, Proposition 3.5]. See also [19, Theorem 8.10] for an alternative approach to this theorem. □

As discussed in the proof above, we next consider the flat model (or R–model) structures from [28, Theorem 3.2]; see also [25, III, Sections 2 and 3].

Lemma 4.2 *For positive flat cofibrant commutative $H\mathbb{Z}$ algebras B, the map $B \to UZB$ is a stable equivalence.*

Proof The crucial property for positive flat cofibrant ($H\mathbb{Z}$–cofibrant) commutative monoids is that they are also (absolute) flat cofibrant as underlying modules. Thus, if B is a positive flat cofibrant commutative $H\mathbb{Z}$–algebra, then it is also an (absolute) flat cofibrant $H\mathbb{Z}$–module by [28, Corollary 4.3]. (In fact B is also a positive flat cofibrant $H\mathbb{Z}$–module by [28, Corollary 4.1], but we do not use that here.) Since the Quillen equivalence in [29, Proposition 4.3] is with respect to the standard model structures [29, Proposition 2.9], we next translate to that setting. Consider a cofibrant replacement $p\colon cB \to B$ in the standard model structure on $H\mathbb{Z}$–modules; the map p is a trivial fibration and hence a level equivalence. Consider the commuting diagram:

$$\begin{array}{ccc} cB & \xrightarrow{p} & B \\ \downarrow & & \downarrow \\ UZcB & \longrightarrow & UZB \end{array}$$

The left map is a stable equivalence by [29, Proposition 4.3]. In Lemma 4.3 below we show that Z takes level equivalences between flat cofibrant objects to level equivalences. By [28, Proposition 2.8], cB is flat cofibrant, so it follows that the bottom map is also a stable equivalence. Thus, the right map is a stable equivalence as well. □

Lemma 4.3 *The functor Z takes level equivalences between flat cofibrant objects to level equivalences.*

Proof Here we will consider Z as a composite of two functors and we will always work over symmetric spectra in pointed simplicial sets, $\mathsf{Sp}^{\Sigma}(\mathcal{S}_*)$, by forgetting from sAb to $\mathcal{S}_*$ wherever necessary. The first component is $\widetilde{\mathbb{Z}}$ from $H\mathbb{Z}$–modules to $\widetilde{\mathbb{Z}}H\mathbb{Z}$–modules, and the second component is the extension of scalars functor μ_* associated to the ring homomorphism $\mu\colon \widetilde{\mathbb{Z}}H\mathbb{Z} \to H\mathbb{Z}$ induced by recognizing $H\mathbb{Z}$ as isomorphic to $\widetilde{\mathbb{Z}}\mathbb{S}$ and using the monad structure on $\widetilde{\mathbb{Z}}$.

First, note that $\widetilde{\mathbb{Z}}$ is applied to each level and preserves level equivalences as a functor from simplicial sets to simplicial abelian groups. The functor $\widetilde{\mathbb{Z}}$ also preserves flat cofibrations, and hence flat cofibrant objects. The generating flat cofibrations ($H\mathbb{Z}$–cofibrations) are of the form $H\mathbb{Z} \otimes M$ where M is the class of monomorphisms of symmetric sequences. Since $\widetilde{\mathbb{Z}}$ is strong symmetric monoidal, these maps are taken to maps of the form $\widetilde{\mathbb{Z}}(H\mathbb{Z}) \otimes \widetilde{\mathbb{Z}}(M)$. Since $\widetilde{\mathbb{Z}}$ preserves monomorphisms, these are contained in the generating flat ($\widetilde{\mathbb{Z}}H\mathbb{Z}$–) cofibrations, which are of the form $\widetilde{\mathbb{Z}}H\mathbb{Z} \otimes M$.

Next, note that restriction of scalars, μ^*, preserves level equivalences and level fibrations since they are determined as maps on the underlying flat (S–) model structure; see the paragraph above [28, Theorem 2.6] and [28, Proposition 2.2]. It follows by adjunction that μ_* preserves the flat cofibrations and level equivalences between flat cofibrant objects. □

Remark 4.4 In the proof of Theorem 4.1 we use a reduction argument that allows us to establish the desired Quillen equivalence by checking that the unit map of the adjunction is a weak equivalence on flat cofibrant objects in the flat model structure on commutative $H\mathbb{Z}$–algebras. This approach avoids a discussion of a flat model structure on commutative symmetric ring spectra in simplicial abelian groups.

5 Dold–Kan correspondence for commutative monoids

The classical Dold–Kan correspondence is an equivalence of categories between the category of simplicial abelian groups, sAb, and the category of nonnegatively graded chain complexes of abelian groups, ch. In this section we establish a Quillen equivalence between categories of commutative monoids in symmetric sequences of simplicial abelian groups, $C(\mathsf{sAb}^{\Sigma})$, and nonnegatively graded chain complexes, $C(\mathsf{ch}^{\Sigma})$, carrying positive model structures. In the special case of pointed commutative monoids in symmetric sequences of simplicial modules and nonnegatively graded chain complexes, such a Quillen equivalence is established in [21, Theorem 6.5].

In the next section we extend this equivalence from symmetric sequences to symmetric spectra. We first define the relevant model structures on the categories of symmetric sequences in simplicial abelian groups, sAb^{Σ}, and chain complexes, ch^{Σ}.

Definition 5.1
- Let $f\colon A \to B$ be a morphism in ch^Σ. Then f is a positive weak equivalence, if $H_*(f)(\ell)$ is an isomorphism for positive levels $\ell > 0$. It is a positive fibration, if $f(\ell)$ is a fibration in the projective model structure on nonnegatively graded chain complexes for all $\ell > 0$.
- A morphism $g\colon C \to D$ in sAb^Σ is a positive fibration if $g(\ell)$ is a fibration of simplicial abelian groups in positive levels and it is a positive weak equivalence if $g(\ell)$ is a weak equivalence for all $\ell > 0$.

In both cases, the positive cofibrations are determined by their left lifting property with respect to positive acyclic fibrations. Positive cofibrations are cofibrations that are isomorphisms in level zero. One can check directly that the above definitions give model category structures or use Hirschhorn's criterion [10, Theorem 11.6.1] and restrict to the diagram category whose objects are natural numbers greater than or equal to one and then use the trivial model structure in level zero with cofibrations being isomorphisms and weak equivalences and fibrations being arbitrary. The generating cofibrations are maps of the form $G_r(i)$ for r positive and such that i is a generating cofibration in chain complexes (simplicial modules). The generating acyclic cofibrations are maps of the form $G_r(j)$ for r positive and where j is a generating acyclic cofibration in chain complexes (simplicial modules).

We also get the corresponding right-induced model structures on commutative monoids:

Definition 5.2 An f in $C(\mathsf{ch}^\Sigma)(A, B)$ is a positive weak equivalence (fibration) if the map on underlying symmetric sequences, $U(f)$ in $\mathsf{ch}^\Sigma(U(A), U(B))$, is a positive weak equivalence (fibration). Similarly, g in $C(\mathsf{sAb}^\Sigma)(C, D)$ is a positive weak equivalence (fibration) if the map on underlying symmetric sequences, $U(g) \in \mathsf{sAb}^\Sigma(U(C), U(D))$ is a positive weak equivalence (fibration).

In [21, Corollary 5.8, Definition 6.2] these model structures were established for *pointed* commutative monoids in symmetric sequences of simplicial modules and nonnegatively graded chain complexes. An object A in $C(\mathsf{ch}^\Sigma)$ or $C(\mathsf{sAb}^\Sigma)$ is called pointed, if its zeroth level is the unit of the monoidal structure of the base category. We recall the key points of the argument in the proof below. This also makes it clear that the results of [21] can be adapted to the setting of Definition 5.1.

Lemma 5.3 *The structures defined in Definition 5.2 yield cofibrantly generated model categories where the generating cofibrations are $C(G_r(i))$ and the generating acyclic cofibrations are $C(G_r(j))$ with i, j as above and r positive.*

Proof Adjunction gives us that the maps with the right lifting property with respect to all $C(G_r(j))$, $r > 0$, are precisely the positive fibrations and the ones with the RLP with respect to all $C(G_r(i))$, $r > 0$, are the positive acyclic fibrations. Performing the small object argument based on the $C(G_r(j))$ for all positive r yields a factorization of any map as a positive acyclic cofibration and a fibration whereas the small object argument based on the $C(G_r(i))$ for positive r gives the other factorization. □

Let $\underline{\mathbb{Z}}$ denote the constant simplicial abelian group with value $\mathbb{Z}$. In the positive model structures cofibrant objects are commutative monoids whose zeroth level is isomorphic to $\underline{\mathbb{Z}}$ in $C(\mathsf{sAb}^\Sigma)$ or to $\mathbb{Z}[0]$ in $C(\mathsf{ch}^\Sigma)$. In particular, such objects are pointed in the sense of [21, Definition 5.1].

Let Γ denote the functor from nonnegatively graded chain complexes to simplicial abelian groups that is the inverse of the normalization functor. We can extend Γ to a functor from ch^Σ to sAb^Σ by applying Γ in every level. As the category of symmetric sequences of abelian groups is an abelian category, the pair (N, Γ) is still an equivalence of categories.

In the following we extend the result [21, Theorem 6.5] in the pointed setting, to the setting of positive model structures.

Theorem 5.4 *Let $C(\mathsf{sAb}^\Sigma)$ and $C(\mathsf{ch}^\Sigma)$ carry the positive model structures. Then the normalization functor $N\colon C(\mathsf{sAb}^\Sigma) \to C(\mathsf{ch}^\Sigma)$ is the right adjoint in a Quillen equivalence and its left adjoint is denoted L_N.*

Proof A left adjoint L_N to N is constructed in [21, Lemma 6.4]. As positive fibrations and weak equivalences are defined via the forgetful functors to sAb^Σ and ch^Σ, the functor N is a right Quillen functor and N also detects weak equivalences. Every object is fibrant, so we have to show that the unit of the adjunction

$$\eta\colon A \to NL_N(A)$$

is a weak equivalence for all cofibrant $A \in C(\mathsf{ch}^\Sigma)$. But cofibrant objects are pointed and for these it is shown in [21, Proof of Theorem 6.5] that the unit map is a weak equivalence. □

6 Extension to commutative ring spectra

We will show that the pair (L_N, N) gives rise to a Quillen equivalence $(L_N, \phi^* N)$ on the level of commutative symmetric ring spectra.

Lemma 6.1 *The Quillen pair* (L_N, N) *satisfies*

$$L_N(\mathsf{Sym}\, X_*) \cong \mathsf{Sym}(\Gamma(X_*))$$

for all nonnegatively graded chain complexes X_*.

Proof We can identify $\mathsf{Sym}(C_*)$ with the free commutative monoid generated by $G_1 X_*$, $C(G_1 X_*)$. Then, by definition of L_N, we obtain

$$L_N(C(G_1 X_*)) \cong C(\Gamma(G_1 X_*)) \cong C(G_1 \Gamma(X_*)) \cong \mathsf{Sym}(\Gamma(X_*)). \qquad \square$$

Let $\mathcal{C}$ be a category and let A be an object of $\mathcal{C}$. Then we denote by $A \downarrow \mathcal{C}$ the category of objects under A.

Corollary 6.2 *Let* $C(\mathsf{ch}^\Sigma)$ *and* $C(\mathsf{sAb}^\Sigma)$ *carry the positive model category structures and consider the induced model structures on the categories under a specific object. Then the model categories* $\mathsf{Sym}(\mathbb{Z}[0]) \downarrow C(\mathsf{ch}^\Sigma)$ *and* $\mathsf{Sym}(\underline{\mathbb{Z}}) \downarrow C(\mathsf{sAb}^\Sigma)$ *are Quillen equivalent.*

Proof By Lemma 6.1 we know that

$$L_N\, \mathsf{Sym}(\mathbb{Z}[0]) \cong \mathsf{Sym}(\underline{\mathbb{Z}}).$$

A direct calculation shows that $N(\mathsf{Sym}(\underline{\mathbb{Z}}))$ is isomorphic to $\mathsf{Sym}(\mathbb{Z}[0])$. Therefore the Quillen equivalence (L_N, N) passes to a Quillen adjunction on the under categories. As the classes of fibrations, weak equivalences and cofibrations in the under categories are determined by the ones in the ambient category, this adjunction is a Quillen equivalence. $\square$

Note that there is an isomorphism of categories between the category of commutative monoids in $\mathsf{Sp}^\Sigma(\mathsf{sAb}, \widetilde{\mathbb{Z}}(\mathbb{S}^1))$ and the category $\mathsf{Sym}(\widetilde{\mathbb{Z}}(\mathbb{S}^1)) \downarrow C(\mathsf{sAb}^\Sigma)$. A similar isomorphism of categories compares commutative monoids in $\mathsf{Sp}^\Sigma(\mathsf{ch}, \mathbb{Z}[1])$ and objects in $\mathsf{Sym}(\mathbb{Z}[1]) \downarrow C(\mathsf{ch}^\Sigma)$. We can extend the Quillen equivalence from Corollary 6.2 to these under categories. Recall from [29, page 358] that $\mathcal{N}$ is the symmetric sequence in chain complexes with $N(\widetilde{\mathbb{Z}}(\mathbb{S}^\ell))$ in level ℓ. We denote by $\mathbb{1}$ the unit of the symmetric monoidal category ch^Σ. This is the symmetric sequence with $\mathbb{Z}[0]$ in level zero and zero in all positive levels.

Proposition 6.3 *The functors* $(L_N, \Phi^* N)$ *induce a Quillen equivalence on the model categories* $\mathsf{Sym}(\mathbb{Z}[1]) \downarrow C(\mathsf{ch}^\Sigma)$ *and* $\mathsf{Sym}(\widetilde{\mathbb{Z}}(\mathbb{S}^1)) \downarrow C(\mathsf{sAb}^\Sigma)$ *where* $C(\mathsf{ch}^\Sigma)$ *and* $C(\mathsf{sAb}^\Sigma)$ *carry the positive model structures. Here,* Φ^* *is a suitable change-of-rings functor.*

Proof As $\Gamma(\mathbb{Z}[1])$ is isomorphic to $\widetilde{\mathbb{Z}}(\mathbb{S}^1)$ we obtain with Lemma 6.1 that

$$L_N(\mathsf{Sym}(\mathbb{Z}[1])) \cong \mathsf{Sym}(\widetilde{\mathbb{Z}}(\mathbb{S}^1)).$$

Therefore, if A is an object in $\mathsf{Sym}(\mathbb{Z}[1]) \downarrow C(\mathsf{ch}^\Sigma)$, then $L_N(A)$ is an object of $\mathsf{Sym}(\widetilde{\mathbb{Z}}(\mathbb{S}^1)) \downarrow C(\mathsf{sAb}^\Sigma)$. We consider the functors

$$\begin{tikzcd}
\mathsf{Sym}(\mathbb{Z}[1]) \downarrow C(\mathsf{ch}^\Sigma) \arrow[rr, "L_N"] & & \mathsf{Sym}(\widetilde{\mathbb{Z}}(\mathbb{S}^1)) \downarrow C(\mathsf{sAb}^\Sigma) \arrow[dl, "N"] \\
& \mathcal{N} \downarrow C(\mathsf{ch}^\Sigma) \arrow[ul, "\Phi^*"] &
\end{tikzcd}$$

where $\Phi\colon \mathsf{Sym}(\mathbb{Z}[1]) \to \mathcal{N}$ is induced by the shuffle transformation (see [29, page 358]) and Φ^* is the associated change-of-rings map. Note that $NL_N \,\mathsf{Sym}\, \mathbb{Z}[1] \cong \mathcal{N}$. Both functors N and Φ^* preserve and detect level and stable weak equivalences [29, Proof of Proposition 4.4], therefore they preserve and detect positive weak equivalences and hence it suffices to show that

$$A \to \Phi^* N L_N A$$

is a weak equivalence in the model category $\mathsf{Sym}(\mathbb{Z}[1]) \downarrow C(\mathsf{ch}^\Sigma)$ for all cofibrant objects $\alpha\colon \mathsf{Sym}(\mathbb{Z}[1]) \to A$. There is a map of commutative monoids $\gamma\colon \mathbb{1} \to \mathsf{Sym}(\mathbb{Z}[1])$ which is given by the identity in level zero and by the zero map in higher levels. Let γ^* be the associated change-of-rings functor:

$$\begin{tikzcd}
\mathbb{1} \arrow[r, "\gamma"] & \mathsf{Sym}(\mathbb{Z}[1]) \arrow[r, "\alpha"] \arrow[d, "\Phi"'] & A \arrow[d, "\eta_A"] \\
& \mathcal{N} \arrow[r, "NL_N\alpha"] & NL_N A
\end{tikzcd}$$

Note that $\eta_A \circ \alpha \circ \gamma = NL_N\alpha \circ \Phi \circ \gamma$. As A is cofibrant, we know that the map $\alpha(0)\colon \mathbb{Z}[0] = \mathsf{Sym}(\mathbb{Z}[1])(0) \to A(0)$ is an isomorphism. Therefore $\gamma^*(A)$ is positively cofibrant as an object in $C(\mathsf{ch}^\Sigma)$. Hence the map

$$\gamma^*(A) \to \gamma^*\Phi^* N L_N(A)$$

is a positive weak equivalence in $C(\mathsf{ch}^\Sigma)$, ie a level equivalence in all positive levels (it is also a weak equivalence in level zero). As γ^* is the identity on objects and only changes the module structure we get that

$$A \to \Phi^* N L_N(A)$$

is a level equivalence in $\mathsf{Sym}(\mathbb{Z}[1]) \downarrow C(\mathsf{ch}^\Sigma)$. $\square$

Remark 6.4 With the positive model structure, $C(\mathsf{ch}^{\Sigma})$ is not left proper. Consider for instance the map $CG_r(0) = \mathbb{1} \to CG_r(\mathbb{Z}[0])$. This map is a cofibration for positive r in the positive model structure. On the other hand, take the projection map from $\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z}$. This yields a map π in $C(\mathsf{ch}^{\Sigma})$ from the initial object $\mathbb{1}$ to $\mathbb{1}/2\mathbb{1}$ (where the latter object is concentrated in level zero with value $\mathbb{Z}/2\mathbb{Z}[0]$). As we work in the positive model structure, this map is actually a weak equivalence. If we push out π along the cofibration $\mathbb{1} \to CG_r(\mathbb{Z}[0])$ we get

$$g\colon CG_r(\mathbb{Z}[0]) \to CG_r(\mathbb{Z}[0]) \odot \mathbb{1}/2\mathbb{1}.$$

In level r this is the chain map

$$\begin{aligned} g(r)\colon G_r(\mathbb{Z}[0])(r) &\cong \mathbb{Z}[\Sigma_r] \otimes \mathbb{Z}[0] \cong \mathbb{Z}[\Sigma_r][0] \\ &\to \mathbb{Z}[\Sigma_r] \otimes \mathbb{Z}[0] \otimes \mathbb{Z}/2\mathbb{Z}[0] \cong \mathbb{Z}/2\mathbb{Z}[\Sigma_r][0]. \end{aligned}$$

Therefore we do not get an isomorphism for positive r and the pushout of the weak equivalence π is not a weak equivalence.

We want to transfer our results to a comparison of commutative monoids in symmetric spectra of simplicial abelian groups and nonnegatively graded chain complexes where we consider the positive stable model structure.

Lemma 6.5 *Cofibrant objects in $C(\mathsf{Sp}^{\Sigma}(\mathsf{ch}, \mathbb{Z}[1]))$ in the positive stable model structure are cofibrant in $C(\mathsf{ch}^{\Sigma})$.*

Proof We can express the map $\mathbb{1} \to \mathsf{Sym}(\mathbb{Z}[1])$ as

$$\mathbb{1} \cong C(G_1(0)) \to C(G_1(\mathbb{Z}[1])) = \mathsf{Sym}(\mathbb{Z}[1]).$$

Therefore the unit of $\mathsf{Sym}(\mathbb{Z}[1])$ is $C(G_1(i))$ with $i\colon 0 \to \mathbb{Z}[1]$ and hence it is a cofibration and therefore the initial object $\mathsf{Sym}(\mathbb{Z}[1])$ of $C(\mathsf{Sp}^{\Sigma}(\mathsf{ch}, \mathbb{Z}[1]))$ is cofibrant in $C(\mathsf{ch}^{\Sigma})$.

As usual, let $\mathbb{S}^n$ denote the chain complex whose only nontrivial chain group is $\mathbb{Z}$ in degree n and let $\mathbb{D}^n$ denote the chain complex with $\mathbb{D}^n_n = \mathbb{D}^n_{n-1} = \mathbb{Z}$ and $\mathbb{D}^n_i = 0$ for all $i \neq n,\ n-1$ whose only nontrivial boundary map is the identity. The cofibrant generators of the positive stable model structure are the maps

$$\mathsf{Sym}(\mathbb{Z}[1]) \odot G_m(\mathbb{S}^{n-1}) \xrightarrow{\mathsf{Sym}(\mathbb{Z}[1]) \odot G_m(i_n)} \mathsf{Sym}(\mathbb{Z}[1]) \odot G_m(\mathbb{D}^n), \tag{2}$$

where i_n is the cofibration of chain complexes $i_n\colon \mathbb{S}^{n-1} \to \mathbb{D}^n$ and $m \geqslant 1$. The $\odot$–product is the coproduct in the category $C(\mathsf{ch}^{\Sigma})$ and thus the map $\mathsf{Sym}(\mathbb{Z}[1]) \odot G_m(i_n)$

is the coproduct of the identity map on $\mathsf{Sym}(\mathbb{Z}[1])$ and the map $G_m(i_n)$ and hence a cofibration in $C(\mathsf{ch}^\Sigma)$.

Coproducts of generators as in (2) are cofibrations in $C(\mathsf{ch}^\Sigma)$ as well, because the coproduct in $C(\mathsf{Sp}^\Sigma(\mathsf{ch},\mathbb{Z}[1]))$ is given by the $\odot_{\mathsf{Sym}(\mathbb{Z}[1])}$–product.

Every cofibrant object is a retract of a cell-object and these are sequential colimits of pushout diagrams of the form

$$\begin{tikzcd} X \arrow[r] \arrow[d, "f"'] & A^{(n)} \arrow[d, dotted] \\ Y \arrow[r, dotted] & A^{(n+1)} \end{tikzcd}$$

where f is a coproduct of maps like in (2) and $A^{(n)}$ is inductively constructed such that $A^{(0)}$ is $\mathsf{Sym}(\mathbb{Z}[1]))$. We can inductively assume that X, Y and $A^{(n)}$ are cofibrant in $C(\mathsf{ch}^\Sigma)$. The pushout in $C(\mathsf{Sp}^\Sigma(\mathsf{ch},\mathbb{Z}[1]))$ is the pushout in $C(\mathsf{ch}^\Sigma)$ and hence the pushout $A^{(n+1)}$ is cofibrant in $C(\mathsf{ch}^\Sigma)$ as well. Sequential colimits and retracts of cofibrant objects are cofibrant. □

Theorem 6.6 *The Quillen pair $(L_N, \Phi^* N)$ induces a Quillen equivalence between $C(\mathsf{Sp}^\Sigma(\mathsf{ch},\mathbb{Z}[1]))$ and $C(\mathsf{Sp}^\Sigma(\mathsf{sAb},\widetilde{\mathbb{Z}}(\mathbb{S}^1)))$ with the model structures that are right-induced from the positive stable model structures on the underlying categories of symmetric spectra.*

Proof We have to show that the unit of the adjunction

$$A \to \Phi^* N L_N A$$

is a stable equivalence for all cofibrant $A \in C(\mathsf{Sp}^\Sigma(\mathsf{ch},\mathbb{Z}[1]))$. Lemma 6.5 ensures that A is cofibrant as an object in $C(\mathsf{ch}^\Sigma)$. Both A and $\Phi^* N L_N A$ receive a unit map from $\mathsf{Sym}(\mathbb{Z}[1])$. As in the proof of Proposition 6.3 we get that

$$\gamma^* A \to N L_N \gamma^* A$$

is a level equivalence in $C(\mathsf{ch}^\Sigma)$ and therefore the map $A \to \Phi^* N L_N A$ is a level equivalence in $C(\mathsf{Sp}^\Sigma(\mathsf{ch},\mathbb{Z}[1]))$ and hence a stable equivalence. □

7 Comparison of spectra in bounded and unbounded chain complexes

Recall that ch denotes the category of nonnegatively graded chain complexes and Ch is the category of unbounded chain complexes of abelian groups. There is a canonical inclusion functor $i\colon \mathsf{ch} \to \mathsf{Ch}$ and a good truncation functor $C_0\colon \mathsf{Ch} \to \mathsf{ch}$ which assigns to an unbounded chain complex X_* the nonnegatively graded chain complex $C_0(X_*)$ with

$$C_0(X_*)_m = \begin{cases} X_m & \text{for } m > 0, \\ \text{cycles}(X_0) & \text{for } m = 0. \end{cases}$$

We denote the induced functors on the corresponding categories of symmetric spectra again by i and C_0. In this section we consider the Quillen equivalence

$$i\colon \mathsf{Sp}^\Sigma(\mathsf{ch}) \rightleftarrows \mathsf{Sp}^\Sigma(\mathsf{Ch}) :C_0$$

and show that it extends to a Quillen equivalence of categories of commutative monoids. The original Quillen equivalence is established in [29, Proposition 4.9] for the usual stable model structures. Here we consider instead the positive stable model structures from [17, Section 14] and then consider the right-induced model structures on commutative monoids where f is a weak equivalence or fibration if it is an underlying positive weak equivalence or fibration. Note that the weak equivalences of the stable model structure agree with the weak equivalences of the positive stable model structure in $\mathsf{Sp}^\Sigma(\mathsf{ch}, \mathbb{Z}[1])$ and $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$. For this reason the positive and stable model structures are Quillen equivalent; see also [17, Proposition 14.6]. It follows that the Quillen equivalence induced by i and C_0 on the usual stable model structures also induces a Quillen equivalence on the positive stable model structures.

Proposition 7.1 *The adjoint functors i and C_0 form a Quillen equivalence between the positive stable model structures on $\mathsf{Sp}^\Sigma(\mathsf{ch}, \mathbb{Z}[1])$ and $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$.*

Corollary 7.2 *Let f be a positive stably fibrant replacement functor in $\mathsf{Sp}^\Sigma(\mathsf{Ch})$ and let $\eta\colon X \to C_0 iX$ be the unit of the adjunction. The composite $X \to C_0 iX \to C_0 fiX$ is a stable equivalence for all objects X in $\mathsf{Sp}^\Sigma(\mathsf{ch}, \mathbb{Z}[1])$.*

Proof It follows from the proof of Proposition 7.1 that the derived unit of the adjunction is a weak equivalence whenever X is positive cofibrant. Since positive trivial fibrations are positive levelwise weak equivalences and a positive cofibrant replacement $cX \to X$ is a positive trivial fibration, we only need to show that $C_0 fi$ preserves positive levelwise equivalences. The inclusion i preserves positive levelwise equivalences and f preserves stable equivalences. Any stable equivalence between positive stably fibrant objects is a positive levelwise equivalence, so fi preserves positive levelwise equivalences. Since C_0 preserves positive levelwise equivalences between positive stably fibrant objects, the corollary follows. □

Corollary 7.3 *The adjoint functors i and C_0 induce a Quillen equivalence between the commutative monoids in $\mathsf{Sp}^\Sigma(\mathsf{ch}, \mathbb{Z}[1])$ and $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$.*

Proof Since the weak equivalences and fibrations are determined on the underlying positive stable model structures, C_0 still preserves fibrations and weak equivalences between positive stably fibrant objects. By [12, Lemma 4.1.7] it is then enough to check the derived composite $C_0 i$ is a stable equivalence for all cofibrant commutative monoids. This is shown for all objects in Corollary 7.2. The fibrant replacement functor for commutative monoids will be different, but the properties used in the proof of that corollary still hold, so we conclude. □

8 Quillen equivalence between E_∞–monoids in Ch and $\mathsf{Sp}^\Sigma(\mathsf{Ch})$

We fix a cofibrant E_∞–operad $\mathcal{O}$ in Ch (in the model structure on operads as in [30, Section 2, Remark 2]) and we consider the operad $F_0\mathcal{O}$ in symmetric spectra in chain complexes.

Let Ch carry the projective model structure and let $E_\infty\mathsf{Ch}$ denote the category of $\mathcal{O}$–algebras in Ch with its right-induced model structure [30, Section 4, Theorem 4]. This model structure exists because Ch is a cofibrantly generated monoidal model category, it satisfies the monoid axiom [29, Corollary 3.4] and $\mathcal{O}$ is cofibrant. Alternatively, we could work with Mandell's model structure on E_∞–monoids in Ch using the operad of the chains on the linear isometries operad [15]. See also [1] for general existence results of model structures for categories of algebras over operads.

Similarly, $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$ with the stable model structure is a cofibrantly generated monoidal model category satisfying the monoid axiom [29, Corollary 3.4], and as the set of generating acyclic cofibrations for the positive stable model structure on $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$ is a subset of the ones for the stable structure, the positive stable model category also satisfies the monoid axiom. We consider two model structures for $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$, the E_∞–monoids in $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$:

- We denote by $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^{s,+}$ the model structure in which the forgetful functor to the positive stable model category structure on $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$ determines the fibrations and weak equivalences.
- Let $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^s$ denote the model category whose fibrations and weak equivalences are determined by the forgetful functor to the stable model structure on $\mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$.

Proposition 8.1 *The model structure $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^{s,+}$ is Quillen equivalent to the model structure $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^s$.*

Proof We consider the adjunction

$$(E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^s) \underset{R}{\overset{L}{\leftrightarrows}} (E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^{s,+}),$$

where R and L are both the identity functor. If p is a fibration in the a positive stable fibration in $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$. Therefore R preserves fibrations. As the weak equivalences in both model structures agree, R is a right Quillen functor and it preserves and reflects weak equivalences. Hence the unit of the adjunction is a weak equivalence. □

In the following we use Hovey's comparison result [11, Theorem 9.1]: Tensoring with $\mathbb{Z}[1]$ induces a Quillen autoequivalence on the category of unbounded chain complexes, so we get that the pair (F_0, Ev_0) induces a Quillen equivalence

$$\mathsf{Ch} \underset{\mathrm{Ev}_0}{\overset{F_0}{\rightleftarrows}} \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^s.$$

We can then transfer this Quillen equivalence to the corresponding categories of E_∞–monoids: Both F_0 and Ev_0 are strong symmetric monoidal functors. Fix a cofibrant E_∞–operad $\mathcal{O}$ in Ch as above. As $\mathrm{Ev}_0 \circ F_0$ is the identity, Ev_0 maps $F_0\mathcal{O}$–algebras in $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$ to $\mathcal{O}$–algebras in unbounded chain complexes.

Theorem 8.2 *The functors* (F_0, Ev_0) *induce a Quillen equivalence*

$$F_0\colon E_\infty \mathsf{Ch} \rightleftarrows E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])^s :\mathrm{Ev}_0.$$

Proof The proof follows Hovey's proof of [11, Theorem 5.1]. It is easy to see that Ev_0 reflects weak equivalences between stably fibrant objects: If $f\colon X \to Y$ is such a map and $f(0)$ is a weak equivalence, then $f(\ell)$ is a weak equivalence for all $\ell \geqslant 0$, because X and Y are fibrant and $(-) \otimes \mathbb{Z}[1]$ is a Quillen equivalence.

In our case $(-) \otimes \mathbb{Z}[1]$ is an equivalence of categories with inverse the functor $\mathrm{Hom}(\mathbb{Z}[1], -)$, where $\mathrm{Hom}(-,-)$ is the internal homomorphism bifunctor.

Therefore, for any X in $E_\infty \mathsf{Ch}$, we have that $F_0 X$ is stably fibrant because

$$(F_0 X)_n = X \otimes \mathbb{Z}[n] \cong \mathrm{Hom}(\mathbb{Z}[1], X \otimes \mathbb{Z}[n+1])$$

and as every object in Ch is fibrant, $F_0 X$ is always fibrant in the projective model structure on $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$.

Let A be a cofibrant object in $E_\infty \mathsf{Ch}$. We have to show that

$$\eta\colon A \to \mathrm{Ev}_0 W(F_0 A)$$

is a weak equivalence, for $W(-)$ the fibrant replacement in $E_\infty \mathsf{Sp}^\Sigma(\mathsf{Ch}, \mathbb{Z}[1])$. But we saw that $F_0 A$ is fibrant and $A \to \mathrm{Ev}_0 F_0 A = A$ is the identity map, thus η is a weak equivalence. See also [19, Theorem 8.10] for an alternative approach to this theorem. □

Observe that all of the Quillen equivalences that we have established so far did not use any particular properties of $\mathbb{Z}$. We can therefore generalize our results as follows.

Corollary 8.3 *Let R be a commutative ring with unit. There is a chain of Quillen equivalences between the model category of commutative HR–algebra spectra and E_∞–monoids in the category of unbounded R–chain complexes.*

For $R=\mathbb{Q}$ we can strengthen the result:

Corollary 8.4 *There is a chain of Quillen equivalences between the model category of commutative $H\mathbb{Q}$–algebra spectra and differential graded commutative $\mathbb{Q}$–algebras.*

Proof It is well known that the category of differential graded commutative algebras and E_∞–monoids in $\mathsf{Ch}(\mathbb{Q})$ possess a right-induced model category structure and that there is a Quillen equivalence between them. For a proof of these facts see for instance [14, Section 7.1.4]. □

Remark 8.5 Note that the proof of Theorem 8.2 applies in broader generality: If $\mathcal{O}$ is an arbitrary operad in the category of chain complexes such that right-induced model structures on $\mathcal{O}$–algebras in Ch and on $F_0(\mathcal{O})$–algebras in $\mathsf{Sp}^\Sigma(\mathsf{Ch},\mathbb{Z}[1])^s$ exist, then the pair (F_0,Ev_0) yields a Quillen equivalence between the model category of $\mathcal{O}$–algebras in Ch and the model category of $F_0(\mathcal{O})$–algebras in $\mathsf{Sp}^\Sigma(\mathsf{Ch},\mathbb{Z}[1])^s$.

9 Symmetric spectra and $\mathcal{I}$–chain complexes

Let $\mathcal{I}$ denote the skeleton of the category of finite sets and injective maps with objects the sets $\boldsymbol{n}=\{1,\dots,n\}$ for $n\geqslant 0$ with the convention that $\mathbf{0}=\varnothing$. The set of morphisms $\mathcal{I}(\boldsymbol{p},\boldsymbol{n})$ consists of all injective maps from $\boldsymbol{p}$ to $\boldsymbol{n}$. In particular, this set is empty if n is smaller than p. The category $\mathcal{I}$ is a symmetric monoidal category under disjoint union of sets.

For any category $\mathcal{C}$ we consider the diagram category $\mathcal{C}^{\mathcal{I}}$ of functors from $\mathcal{I}$ to $\mathcal{C}$. If $(\mathcal{C},\otimes,e)$ is symmetric monoidal, then $\mathcal{C}^{\mathcal{I}}$ inherits a symmetric monoidal structure: For $A,B\in\mathcal{C}^{\mathcal{I}}$ we set

$$(A\boxtimes B)(\boldsymbol{n})=\operatorname{colim}_{\boldsymbol{p}\sqcup\boldsymbol{q}\to\boldsymbol{n}} A(\boldsymbol{p})\otimes B(\boldsymbol{q}).$$

For details about $\mathcal{I}$–diagrams see [22]. The following fact is folklore; it was pointed out to Shipley by Jeff Smith in 2006 at the Mittag-Leffler Institute.

Proposition 9.1 *Let $\mathcal{C}$ be any closed symmetric monoidal category with unit e. Then the category $\mathsf{Sp}^{\Sigma}(\mathcal{C}, e)$ is isomorphic to the diagram category $\mathcal{C}^{\mathcal{I}}$.*

Proof Let $X \in \mathsf{Sp}^{\Sigma}(\mathcal{C}, e)$. Then $X(n) \in \mathcal{C}^{\Sigma_n}$ and we have Σ_n–equivariant maps $X(n) \cong X(n) \otimes e \to X(n+1)$, such that the composite

$$\sigma_{n,p}\colon X(n) \cong X(n) \otimes e^{\otimes p} \to X(n+1) \otimes e^{\otimes p-1} \to \cdots \to X(n+p)$$

is $\Sigma_n \times \Sigma_p$–equivariant for all $n, p \geqslant 0$.

We send X to $\phi(X) \in \mathcal{C}^{\mathcal{I}}$ with $\phi(X)(\boldsymbol{n}) = X(n)$. If $i = i_{p,n-p} \in \mathcal{I}(\boldsymbol{p}, \boldsymbol{n})$ is the standard inclusion, then we let $\phi(i)\colon \phi(X)(\boldsymbol{p}) \to \phi(X)(\boldsymbol{n})$ be $\sigma_{p,n-p}$. Every morphism $f \in \mathcal{I}(\boldsymbol{p}, \boldsymbol{n})$ can be written as $\xi \circ i$ where i is the standard inclusion and $\xi \in \Sigma_n$. For such ξ, the map $\phi(\xi)$ is given by the Σ_n–action on $X(n) = \phi(X)(\boldsymbol{n})$.

If $f = \xi' \circ i$ is another factorization of f into the standard inclusion followed by a permutation, then ξ and ξ' differ by a permutation $\tau \in \Sigma_n$ which maps all j with $1 \leqslant j \leqslant p$ identically, ie τ is of the form $\tau = \mathrm{id}_{\boldsymbol{p}} \oplus \tau'$ with $\tau' \in \Sigma_{n-p}$. As the structure maps $\sigma_{p,n-p}$ are $\Sigma_p \times \Sigma_{n-p}$–equivariant, the induced map $\phi(f) = \phi(\xi') \circ \phi(i)$ agrees with $\phi(\xi) \circ \phi(i)$.

The inverse of ϕ, denoted by ψ, sends A, an $\mathcal{I}$–diagram in $\mathcal{C}$, to the symmetric spectrum $\psi(A)$ whose n^{th} level is $\psi(A)(n) = A(\boldsymbol{n})$. The Σ_n–action on $\psi(A)(n)$ is given by the corresponding morphisms $\Sigma_n \subset \mathcal{I}(\boldsymbol{n}, \boldsymbol{n})$ and the structure maps of the spectrum are defined as

$$\psi(A)(n) \otimes e^{\otimes p} = A(\boldsymbol{n}) \otimes e^{\otimes p} \xrightarrow{\cong} A(\boldsymbol{n}) \xrightarrow{A(i_{n,p})} A(\boldsymbol{n} + \boldsymbol{p}) = \psi(A)(n+p).$$

The functors ϕ and ψ are well-defined and inverse to each other. □

Lemma 9.2 *The functors ϕ and ψ are strong symmetric monoidal.*

Proof Consider two free objects $F_s C_*$ and $F_t D_*$ in $\mathsf{Sp}^{\Sigma}(\mathcal{C}, e)$ for two chain complexes C_* and D_*. We know in general [11, Section 7] that

$$F_s C_* \wedge F_t D_* \cong F_{s+t}(C_* \otimes D_*). \tag{3}$$

Note that as an object in $\mathcal{C}^{\mathcal{I}}$ we have for $\boldsymbol{n} \in \mathcal{I}$

$$\phi(F_s C_*)(\boldsymbol{n}) = \mathbb{Z}\Sigma_n \otimes_{\mathbb{Z}\Sigma_{n-s}} C_*$$

for $n \geqslant s$ and zero otherwise. This coincides with the value of the free $\mathcal{I}$–diagram on $\boldsymbol{n}$,

$$F_s^{\mathcal{I}}(C_*)(\boldsymbol{n}) = \mathbb{Z}\mathcal{I}(\boldsymbol{s}, \boldsymbol{n}) \otimes C_*,$$

and in fact this yields an isomorphism of functors. Similarly, $\psi(F_s^{\mathcal{I}}(C_*)) \cong F_s C_*$.

As the symmetric monoidal product in $\mathcal{C}^{\mathcal{I}}$ is given by left Kan extension along the exterior product using the monoidal structure of $\mathcal{C}$ we get

$$F_s^{\mathcal{I}}(C_*) \boxtimes F_t^{\mathcal{I}}(D_*) \cong F_{s+t}^{\mathcal{I}}(C_* \otimes D_*). \tag{4}$$

From (3) we obtain that

$$\psi(F_s^{\mathcal{I}}(C_*)) \wedge \psi(F_t^{\mathcal{I}}(D_*)) \cong \psi(F_{s+t}^{\mathcal{I}}(C_* \otimes D_*)) \cong \psi(F_s^{\mathcal{I}}(C_*) \boxtimes F_t^{\mathcal{I}}(D_*))$$

and (4) yields

$$\phi(F_s C_*) \boxtimes \phi(F_t D_*) \cong \phi(F_{s+t}(C_* \otimes D_*)) \cong \phi(F_s C_* \wedge F_t D_*).$$

The used isomorphisms are associative and compatible with the symmetry isomorphisms. Every object in $\mathsf{Sp}^{\Sigma}(\mathcal{C}, e)$ and $\mathcal{C}^{\mathcal{I}}$ can be written as a colimit of free objects and as $\mathcal{C}$ is closed, the general case follows from the free case. □

Remark 9.3 In [20, Proposition 3.3.9] Pavlov and Scholbach describe explicitly (for a well-behaved symmetric monoidal model category $\mathcal{C}$) how the unstable and stable model structures on $\mathsf{Sp}^{\Sigma}(\mathcal{C}, e)$ transfer to $\mathcal{C}^{\mathcal{I}}$ under the above mentioned isomorphism of categories. If $\mathcal{C}$ is Ch, their assumptions are satisfied.

Note that the weak equivalences in $\mathsf{Ch}^{\mathcal{I}}$ have an explicit description: they are the maps that become weak equivalences after applying a corrected homotopy colimit [7, Definition 5.1]. This is the homotopy colimit of the diagram where every node is functorially replaced by a cofibrant object first. To see this, consider Dugger's Bousfield localizations of diagram categories in [7, Section 5]. As the cofibrations and the fibrant objects in his model structure in [7, Theorem 5.2] agree with ours, an argument due to Joyal [13, Proposition E.1.10] ensures that we have the same class of weak equivalences as well.

Taking a cofibrant E_∞–operad $\mathcal{O}$ in Ch then ensures that $\mathcal{O}$–algebras in $\mathsf{Sp}^{\Sigma}(\mathsf{Ch}, \mathbb{Z}[0])^s$ and in $\mathsf{Ch}^{\mathcal{I}}$ carry a model category structure such that the forgetful functor determines fibrations and weak equivalences

Since tensoring with the unit $\mathbb{Z}[0]$ is isomorphic to the identity, we can repeat all of the arguments in the previous section with $\mathbb{Z}[1]$ replaced by $\mathbb{Z}[0]$. Thus we also obtain that the model category $E_\infty \mathsf{Sp}^{\Sigma}(\mathsf{Ch}, \mathbb{Z}[0])^s$ is Quillen equivalent to the model category of E_∞–monoids in Ch. Summarizing:

Theorem 9.4 *There is a chain of Quillen equivalences*

$$E_\infty \mathsf{Sp}^{\Sigma}(\mathsf{Ch}, \mathbb{Z}[1])^s \underset{\mathrm{Ev}_0}{\overset{F_0}{\leftrightarrows}} E_\infty \mathsf{Ch} \underset{\mathrm{Ev}_0}{\overset{F_0}{\rightleftarrows}} E_\infty \mathsf{Sp}^{\Sigma}(\mathsf{Ch}, \mathbb{Z}[0])^s$$

and the rightmost model category is isomorphic to $E_\infty \mathsf{Ch}^{\mathcal{I}}$.

Last but not least we can connect commutative HR–algebras to commutative $\mathcal{I}$–chain complexes. The positive stable model structure on $\mathsf{Sp}^{\Sigma}(\mathsf{Ch}(R), R[0])$ satisfies the assumptions of [19, Theorem 5.10] and hence commutative monoids and E_∞–monoids in $\mathsf{Sp}^{\Sigma}(\mathsf{Ch}(R), R[0])^{s,+}$ carry model category structures and there is a Quillen equivalence between them [20, Theorem 3.4.1, Theorem 3.4.4]. This yields that the model categories of commutative $\mathcal{I}$–chain complexes, $C(\mathsf{Ch}(R)^{\mathcal{I},+})$, and E_∞ $\mathcal{I}$–chain complexes, $E_\infty(\mathsf{Ch}(R)^{\mathcal{I},+})$ are Quillen equivalent, if we take the model structure that is right-induced from the positive model structure on $\mathsf{Ch}(R)^{\mathcal{I},+}$.

Theorem 9.5 *There is a chain of Quillen equivalences between the model categories of commutative HR–algebra spectra, $C(HR\text{–mod})$, and commutative monoids in the category $\mathsf{Ch}(R)^{\mathcal{I}}$ where the latter carries the right-induced model structure from the positive model structure on $\mathsf{Ch}(R)^{\mathcal{I}}$, $\mathsf{Ch}(R)^{\mathcal{I},+}$.*

We close with an important example of a commutative $\mathcal{I}$–chain complex. Consider a chain complex C_* together with a 0–cycle, ie with a map $\eta\colon \mathbb{Z}[0] \to C_*$. The assignment $\boldsymbol{n} \mapsto C_*^{\otimes n}$ defines a functor sym from $\mathcal{I}$ to the category of unbounded chain complexes (namely $\mathsf{Sym}(C_*)$). Schlichtkrull shows in [23] that sym is the algebraic analogue of the symmetric product in the category of spaces.

References

[1] **C Berger**, **I Moerdijk**, *Axiomatic homotopy theory for operads*, Comment. Math. Helv. 78 (2003) 805–831 MR

[2] **C Berger**, **I Moerdijk**, *Resolution of coloured operads and rectification of homotopy algebras*, from "Categories in algebra, geometry and mathematical physics" (A Davydov, M Batanin, M Johnson, S Lack, A Neeman, editors), Contemp. Math. 431, Amer. Math. Soc. (2007) 31–58 MR

[3] **B Cenkl**, *Cohomology operations from higher products in the de Rham complex*, Pacific J. Math. 140 (1989) 21–33 MR

[4] **S G Chadwick**, **M A Mandell**, *E_n genera*, Geom. Topol. 19 (2015) 3193–3232 MR

[5] **J D Christensen**, **M Hovey**, *Quillen model structures for relative homological algebra*, Math. Proc. Cambridge Philos. Soc. 133 (2002) 261–293 MR

[6] **T Church**, **J S Ellenberg**, **B Farb**, *FI-modules and stability for representations of symmetric groups*, Duke Math. J. 164 (2015) 1833–1910 MR

[7] **D Dugger**, *Replacing model categories with simplicial ones*, Trans. Amer. Math. Soc. 353 (2001) 5003–5027 MR

[8] **P G Goerss**, **M J Hopkins**, *Moduli spaces of commutative ring spectra*, from "Structured ring spectra" (A Baker, B Richter, editors), London Math. Soc. Lecture Note Ser. 315, Cambridge Univ. Press (2004) 151–200 MR

[9] **J E Harper**, *Homotopy theory of modules over operads in symmetric spectra*, Algebr. Geom. Topol. 9 (2009) 1637–1680 MR

[10] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc. (2003) MR

[11] **M Hovey**, *Spectra and symmetric spectra in general model categories*, J. Pure Appl. Algebra 165 (2001) 63–127 MR

[12] **M Hovey**, **B Shipley**, **J Smith**, *Symmetric spectra*, J. Amer. Math. Soc. 13 (2000) 149–208 MR

[13] **A Joyal**, *The theory of quasi-categories and its applications*, lecture notes (2008) Available at `http://tinyurl.com/joyalquasi`

[14] **J Lurie**, *Higher algebra*, unpublished manuscript (2014) Available at `http://math.harvard.edu/~lurie/papers/HA.pdf`

[15] **M A Mandell**, *Flatness for the E_∞ tensor product*, from "Homotopy methods in algebraic topology" (J P C Greenlees, R R Bruner, N Kuhn, editors), Contemp. Math. 271, Amer. Math. Soc. (2001) 285–309 MR

[16] **M A Mandell**, *Topological André–Quillen cohomology and E_∞ André–Quillen cohomology*, Adv. Math. 177 (2003) 227–279 MR

[17] **M A Mandell**, **J P May**, **S Schwede**, **B Shipley**, *Model categories of diagram spectra*, Proc. London Math. Soc. 82 (2001) 441–512 MR

[18] **J P May**, **K Ponto**, *More concise algebraic topology: localization, completion, and model categories*, Univ. of Chicago Press (2012) MR

[19] **D Pavlov**, **J Scholbach**, *Admissibility and rectification of colored symmetric operads*, preprint (2014) arXiv

[20] **D Pavlov**, **J Scholbach**, *Symmetric operads in abstract symmetric spectra*, preprint (2014) arXiv

[21] **B Richter**, *On the homology and homotopy of commutative shuffle algebras*, Israel J. Math. 209 (2015) 651–682 MR

[22] **S Sagave**, **C Schlichtkrull**, *Diagram spaces and symmetric spectra*, Adv. Math. 231 (2012) 2116–2193 MR

[23] **C Schlichtkrull**, *The homotopy infinite symmetric product represents stable homotopy*, Algebr. Geom. Topol. 7 (2007) 1963–1977 MR

[24] **R Schwänzl**, **R M Vogt**, *Strong cofibrations and fibrations in enriched categories*, Arch. Math. (Basel) 79 (2002) 449–462 MR

[25] **S Schwede**, *An untitled book project about symmetric spectra* Available at http://www.math.uni-bonn.de/people/schwede/SymSpec.pdf

[26] **S Schwede**, **B E Shipley**, *Algebras and modules in monoidal model categories*, Proc. London Math. Soc. 80 (2000) 491–511 MR

[27] **S Schwede**, **B Shipley**, *Equivalences of monoidal model categories*, Algebr. Geom. Topol. 3 (2003) 287–334 MR

[28] **B Shipley**, *A convenient model category for commutative ring spectra*, from "Homotopy theory: relations with algebraic geometry, group cohomology, and algebraic K–theory" (P Goerss, S Priddy, editors), Contemp. Math. 346, Amer. Math. Soc. (2004) 473–483 MR

[29] **B Shipley**, *$H\mathbb{Z}$–algebra spectra are differential graded algebras*, Amer. J. Math. 129 (2007) 351–379 MR

[30] **M Spitzweck**, *Operads, algebras and modules in general model categories*, preprint (2001) arXiv

[31] **D White**, *Model structures on commutative monoids in general model categories*, preprint (2014) arXiv

Department Mathematik, Universität Hamburg
Hamburg, Germany

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago
Chicago, IL, United States

richter@math.uni-hamburg.de, shipleyb@uic.edu

http://www.math.uni-hamburg.de/home/richter/,
http://homepages.math.uic.edu/~bshipley/

Received: 29 September 2015 Revised: 9 December 2016

Algebraic & Geometric Topology 17 (2017) 2039–2050

Eigenvalue varieties of Brunnian links

FRANÇOIS MALABRE

In this article, it is proved that the eigenvalue variety of the exterior of a nontrivial, non-Hopf, Brunnian link in $\mathbb{S}^3$ contains a nontrivial component of maximal dimension. Eigenvalue varieties were first introduced to generalize the A–polynomial of knots in $\mathbb{S}^3$ to manifolds with nonconnected toric boundary. The result presented here generalizes, for Brunnian links, the nontriviality of the A–polynomial of nontrivial knots in $\mathbb{S}^3$.

57M25; 57M27

The A–polynomial of a knot in $\mathbb{S}^3$ is a two-variable polynomial constructed from the $\mathrm{SL}_2\mathbb{C}$–character variety of the knot exterior. Let K be a knot in $\mathbb{S}^3$ and let $\pi_1 K$ denote the fundamental group of the exterior of K; the peripheral subgroup $\mathbb{Z}^2$ is generated by a meridian μ and a longitude λ, and the zero-set of the A–polynomial A_K is the locus of eigenvalues for a common eigenvector of $\rho(\mu)$ and $\rho(\lambda)$ of representations ρ from $\pi_1 K$ to $\mathrm{SL}_2\mathbb{C}$. It was first introduced by Cooper, Culler, Gillet, Long and Shalen [2], where it is also proved that the A–polynomial of any knot contains the A–polynomial of the unknot as a factor. The A–polynomial of a knot is said to be *nontrivial* if it contains other factors, and it was also proved in the same article [2] that hyperbolic knots and nontrivial torus knots always have a nontrivial A–polynomial. This was later established in full generality for all nontrivial knots by Dunfield and Garoufalidis [4], and independently by Boyer and Zhang [1]; both proofs use a theorem by Kronheimer and Mrowka [5] on Dehn fillings on knots and representations in SU_2.

The notion of A–polynomial can be generalized to any 3–manifold M with connected toric boundary by specifying a *peripheral system* (generators of $\pi_1\partial M \hookrightarrow \pi_1 M$). Stimulated by the work of Lash in [6], it was then extended to manifolds with nonconnected boundary by Tillmann. In his Ph D thesis [11] and the subsequent article [12], Tillmann presented the *eigenvalue variety* $\mathfrak{E}(M)$ associated to a 3–manifold M with toric boundary. If the boundary of M consists of n tori, the associated eigenvalue variety $\mathfrak{E}(M)$ is an algebraic subspace of $\mathbb{C}^{2n}$ corresponding to the closure of peripheral eigenvalues taken by representations (or equivalently, characters) of $\pi_1 M$ in $\mathrm{SL}_2\mathbb{C}$. Under these assumptions, Tillmann proved in [12] that the dimension of any component of $\mathfrak{E}(M)$ is at most n.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2039

In the same way as any A–polynomial is divisible by the A–polynomial of the unknot, any eigenvalue variety $\mathfrak{E}(M)$ contains components $\mathfrak{E}^{\mathrm{red}}(M)$ corresponding to reducible characters. Components of $\mathfrak{E}^{\mathrm{red}}(M)$ have maximal dimension, and any other component of $\mathfrak{E}(M)$ with maximal dimension is called a *nontrivially maximal* component of $\mathfrak{E}(M)$.

If M is hyperbolic, its character variety contains a family of distinguished components $Y_1, \dots, Y_k$ called the *geometric components*, each one containing the character of a discrete faithful representation. Using Thurston's results [10], Tillmann proved that each geometric component produces a nontrivially maximal component in $\mathfrak{E}(M)$, generalizing the result of [2] on hyperbolic knots. However, the question of classifying 3–manifolds M for which $\mathfrak{E}(M)$ contains a nontrivially maximal component, or even determining whether nontrivial exteriors of links in $\mathbb{S}^3$ have this property, remains open.

In this article, we answer this matter for a family of links in $\mathbb{S}^3$, the *Brunnian links*. A link in $\mathbb{S}^3$ is called *Brunnian* if any of its proper sublinks is trivial, and we prove:

Theorem 1 *The eigenvalue variety of any nontrivial non-Hopf Brunnian link contains a nontrivially maximal component.*

The defining property of Brunnian links makes them stable under $1/q$–Dehn fillings, which permits us to apply the Kronheimer–Mrowka theorem [5, Theorem 1] to produce irreducible characters in a similar fashion as in [4] and [1]. Then, an induction on the number of components of the links produces nontrivially maximal components of their eigenvalue varieties.

This article is divided into two sections: First we recall the construction of the eigenvalue variety $\mathfrak{E}(L)$ for a link L in $\mathbb{S}^3$, its defining ideal $\mathcal{A}(L)$ and some of its properties, as presented in [12], to introduce notation for the following section. Then we study the family of Brunnian links in $\mathbb{S}^3$ and prove the main result of this article.

Acknowledgements The content of this paper forms part of the author's Ph D thesis [7]. He would like to express his gratitude to his advisors Michel Boileau and Joan Porti for their constant support during the realization of his Ph D, and also to thank Stephan Tillmann and Julien Marché for their valuable inputs on character and eigenvalue varieties.

1 Eigenvalue varieties of links in $\mathbb{S}^3$

First we briefly review the notion of *eigenvalue variety* associated to a link in $\mathbb{S}^3$, first introduced by Tillmann in [11, Section 3.2.4], and we reproduce the construction here (with a slightly different vocabulary) in order to set the notation for the next section.

1.1 Character varieties

Let π be a finitely generated group; the $\mathrm{SL}_2\mathbb{C}$–*representation variety* of π is the algebraic affine set $\hom(\pi, \mathrm{SL}_2\mathbb{C})$ and is denoted by $R(\pi)$. The algebraic Lie group $\mathrm{SL}_2\mathbb{C}$ acts on $R(\pi)$ by conjugation, and the algebraic quotient under this action is the $\mathrm{SL}_2\mathbb{C}$–*character variety* of π, denoted by $X(\pi)$. The ring $\mathbb{C}[X(\pi)]$ of regular functions on the character variety is equal to the subring $\mathbb{C}[R(\pi)]^{\mathrm{SL}_2\mathbb{C}}$ of invariant functions. Dually, the inclusion $\mathbb{C}[X(\pi)] \hookrightarrow \mathbb{C}[R(\pi)]$ induces a natural algebraic epimorphism $t\colon R(\pi) \to X(\pi)$, and any regular function on $R(\pi)$ factors through t if and only if it is invariant under the conjugation action of $\mathrm{SL}_2\mathbb{C}$. In particular, for any γ in π, the function $\tau_\gamma\colon R(\pi) \to \mathbb{C}$ mapping $\rho \mapsto \operatorname{tr}\rho(\gamma)$ defines a regular function I_γ on $X(\pi)$ called the *trace function at* γ; the trace functions finitely generate the ring $\mathbb{C}[X(\pi)]$; see [3] for example. Representation and character varieties are contravariant functors: any group morphism $\pi \to \pi'$ induces regular maps according to the following commutative diagram:

$$\begin{tikzcd} R(\pi') \arrow[r] \arrow[d, "t"'] & R(\pi) \arrow[d, "t"] \\ X(\pi') \arrow[r] & X(\pi) \end{tikzcd}$$

In the case where the group π is the fundamental group of a manifold M (resp. the exterior of a link L in $\mathbb{S}^3$), the representation and character varieties will be denoted by $R(M)$ and $X(M)$ (resp. $R(L)$ and $X(L)$).

1.2 Abelian characters

Any group π has an abelianization π^{ab} and a canonical projection $\pi \to \pi^{\mathrm{ab}}$ which induces regular maps:

$$\begin{tikzcd} R(\pi^{\mathrm{ab}}) \arrow[r] \arrow[d, "t"'] & R(\pi) \arrow[d, "t"] \\ X(\pi^{\mathrm{ab}}) \arrow[r] & X(\pi) \end{tikzcd}$$

The image of $R(\pi^{\mathrm{ab}})$ in $R(\pi)$ is precisely the closed set $R^{\mathrm{ab}}(\pi)$ of abelian representations of π, and the image of $X(\pi^{\mathrm{ab}})$ is a closed subset of $X(\pi)$ called the set of *abelian characters* of π and denoted by $X^{\mathrm{ab}}(\pi)$.

Remark In $\mathrm{SL}_2\mathbb{C}$, characters of reducible representations are characters of abelian representations. If $R^{\mathrm{red}}(\pi)$ is the closed set of reducible representations and $X^{\mathrm{red}}(\pi)$ is its image in $X(\pi)$, then $X^{\mathrm{red}}(\pi) = X^{\mathrm{ab}}(\pi)$.

Let Δ denote the map from $\mathbb{C}^*$ to $\mathrm{SL}_2\mathbb{C}$ mapping $z \mapsto \left[\begin{smallmatrix} z & 0 \\ 0 & z^{-1} \end{smallmatrix}\right]$; by composition, Δ defines maps:

$$\begin{tikzcd} \hom(\pi, \mathbb{C}^*) \arrow[r, "\Delta_*"] \arrow[dr, "d"'] & R^{\mathrm{ab}}(\pi) \arrow[d, "t"] \\ & X^{\mathrm{ab}}(\pi) \end{tikzcd}$$

The map d is two-to-one onto $X^{\mathrm{ab}}(\pi)$ and invariant under inversion in $\hom(\pi, \mathbb{C}^*)$; for any φ in $\hom(\pi, \mathbb{C}^*)$ and γ in π,

$$I_\gamma \circ d(\varphi) = \varphi(\gamma) + \varphi(\gamma)^{-1}.$$

1.3 Eigenvalue varieties

Let L be a link in $\mathbb{S}^3$, let $|L|$ denote its number of components and let $\pi_1 L$ be the fundamental group of its exterior; the boundary of the exterior of L is a disjoint union of $|L|$ tori T_K, one for each component K of the link L. Each inclusion $\pi_1 T_K \hookrightarrow \pi_1 L$ induces a regular map $r_K\colon X(L) \to X(T_K)$. Since $\pi_1 T_K$ is abelian, $X(T_K) = X^{\mathrm{ab}}(T_K)$, and denoting $\hom(\pi_1 T_K, \mathbb{C}^*)$ by $E(T_K)$, we obtain the following diagram:

$$\begin{tikzcd} & \prod_{K\subset L} E(T_K) \arrow[d, "d"] \\ X(L) \arrow[r, "r"'] & \prod_{K\subset L} X(T_K) \end{tikzcd}$$

Following Tillmann [11; 12], the *eigenvalue variety* of L is defined as the Zariski closure of the preimage by d of the image of r:

$$\mathfrak{E}(L) = \overline{d^{-1}(r(X(L)))}.$$

Dually, there are ring-maps

$$\begin{tikzcd} & \bigotimes_{K\subset L} \mathbb{C}[E(T_K)] \\ \mathbb{C}[X(L)] & \bigotimes_{K\subset L} \mathbb{C}[X(T_K)] \arrow[l, "r^*"] \arrow[u, "d^*"'] \end{tikzcd}$$

and the defining ideal $\mathcal{A}(L)$ of $\mathfrak{E}(L)$ is called the $\mathcal{A}$*–ideal* of L and is the radical of the image by d^* of the kernel of r^*:

$$\mathcal{A}(L) = \sqrt{d^*(\ker\ r^*)}.$$

Each torus T_K is equipped with a *standard peripheral system* (μ_K, λ_K) of meridian and longitude of each component. This produces canonical coordinates (m_K, ℓ_K) in $\mathbb{C}^* \times \mathbb{C}^*$ for $E(T_K)$, and $\mathfrak{E}(L)$ is naturally a subset of $(\mathbb{C}^*)^{2|L|}$; dually, $\mathbb{C}[E(T_K)]$ is isomorphic to $\mathbb{C}[\mathfrak{m}_K^{\pm 1}, \mathfrak{l}_K^{\pm 1}]$, and $\mathcal{A}(L)$ is an ideal of $\mathbb{C}[\mathfrak{m}^{\pm}, \mathfrak{l}^{\pm}] = \bigotimes_{K\subset L} \mathbb{C}[\mathfrak{m}_K^{\pm 1}, \mathfrak{l}_K^{\pm 1}]$.

Proposition 2 *Let $\mathfrak{E}^{\mathrm{ab}}(L)$ denote the part of $\mathfrak{E}(L)$ corresponding to abelian characters and $\mathcal{A}^{\mathrm{ab}}(L)$ the corresponding defining ideal; $\mathfrak{E}^{\mathrm{ab}}(L)$ is a union of copies of $(\mathbb{C}^*)^{|L|}$, and $\mathcal{A}^{\mathrm{ab}}(L)$ is given in $\mathbb{C}[\mathfrak{m}^{\pm}, \mathfrak{l}^{\pm}]$ by*

$$\mathcal{A}^{\mathrm{ab}}(L) = \left\langle \mathfrak{l}_K - \prod_{K' \neq K} \mathfrak{m}_{K'}^{\pm \operatorname{lk}(K,K')} \right\rangle,$$

where $\operatorname{lk}(K, K')$ denotes the linking number of the components K and K'.

Proof The meridians form a basis of the homology group of the link exterior, and each longitude is given by the linking numbers

$$\lambda_K = \sum_{K' \neq K} \operatorname{lk}(K, K')\mu_{K'}.$$

Therefore, any morphism from $\pi_1 L$ to $\mathbb{C}^*$ is determined by the images of the meridians, and for any φ in $\hom(\pi_1 L, \mathbb{C}^*)$ and each longitude λ_K,

$$\varphi(\lambda_K) = \prod_{K \neq K'} \varphi(\mu_{K'})^{\operatorname{lk}(K,K')}.$$

By the invariance under inversion, any point $(m_K, \ell_K)_{K \subset L}$ of $\mathfrak{E}^{\mathrm{ab}}(L)$ then satisfies

$$\ell_K = \prod_{K \neq K'} m_K^{\pm \operatorname{lk}(K,K')}.$$

Conversely, for any $\xi = (m_K, \ell_K)_{K \subset L}$ satisfying these equations, there exists φ in $\hom(\pi_1 L, \mathbb{C}^*)$ such that $d(\xi) = r(\Delta_* \varphi)$, so $\mathcal{A}^{\mathrm{ab}}(L)$ is given by

$$\mathcal{A}^{\mathrm{ab}}(L) = \left\langle \mathfrak{l}_K - \prod_{K' \neq K} \mathfrak{m}_{K'}^{\pm \operatorname{lk}(K,K')} \right\rangle. \qquad \square$$

Remark For links with one component (knots), the $\mathcal{A}$–ideal is generated by the A–polynomial of the knot, and A^{ab} is the $\mathfrak{l} - 1$ factor corresponding to abelian characters.

By the defining equations of $\mathcal{A}^{\mathrm{ab}}(L)$, we have that $\mathfrak{E}^{\mathrm{ab}}(L)$ always has dimension $|L|$. As a matter of fact, by Tillmann [11, Proposition 3.10; 12, Proposition 13], any component of $\mathfrak{E}(L)$ has dimension at most $|L|$, which leads to the following definition:

Definition 3 A component of $\mathfrak{E}(L)$ is called *nontrivially maximal* if it has dimension $|L|$ and is not contained in $\mathfrak{E}^{\mathrm{ab}}(L)$.

Using Thurston's results on hyperbolic manifolds, Tillmann showed the following:

Theorem 4 [12, Proposition 13] *If L is a hyperbolic link in $\mathbb{S}^3$, then any geometric component of the character variety produces a nontrivially maximal component in the eigenvalue variety of L.*

Besides these cases, it is not known whether the eigenvalue variety of all (nontrivial) links admits a maximal nontrivial component. For knots, this is equivalent to the nontriviality of the A–polynomial (besides the $\mathfrak{l}-1$ factor) and was proven independently by Dunfield and Garoufalidis in [4], and Boyer and Zhang in [1]. In the next section, we answer this matter for Brunnian links in $\mathbb{S}^3$.

2 Characters of Brunnian links

In this section, we prove Theorem 1. First we recall some basic facts on $1/q$–Dehn fillings on links in $\mathbb{S}^3$; then we present Brunnian links and, after having studied their stability under these Dehn fillings, we use the Kronheimer–Mrowka theorem to create families of characters of exteriors of Brunnian links. Finally, we prove that these characters span a nontrivially maximal component in the eigenvalue varieties of nontrivial, non-Hopf, Brunnian links.

2.1 Dehn fillings

Any $1/q$–Dehn filling on the unknot in $\mathbb{S}^3$ produces $\mathbb{S}^3$ again; therefore, the $1/q$–Dehn filling over an unknotted component of a link in $\mathbb{S}^3$ produces the exterior of another link in $\mathbb{S}^3$.

Let $L = K \sqcup L'$ be a link with K an unknotted component of L, and let L_q denote the link obtained by $1/q$–surgery on K (so, in particular, $L' = L_0$). Any sublink L'' of L_q is obtained by $1/0$–Dehn filling along the other components. Because the meridians are unchanged by $1/q$–Dehn fillings, any proper sublink L'' of L_q is obtained by $1/q$–Dehn filling along K in the sublink $L'' \sqcup K$ of L.

Remark With this notation, if $L'' \sqcup K$ is trivial in $\mathbb{S}^3$, then so is L''.

The meridians are unchanged by $1/q$–Dehn fillings, but the longitudes are changed according to the linking numbers. With the same notation as above, if (μ, λ) is a standard peripheral system for a component J of L, then the new longitude λ_q of J in L_q is

$$\lambda_q = \lambda + q\,\mathrm{lk}(K, J)^2 \mu,$$

and the linking number $\mathrm{lk}_q(J, J')$ of any two components J and J' of L_q is given by

$$\mathrm{lk}_q(J, J') = \mathrm{lk}(J, J') - q\,\mathrm{lk}(K, J)\,\mathrm{lk}(K, J').$$

A link is called *homologically trivial* if all the linking numbers between components vanish. By the previous discussion, the link obtained by $1/q$–Dehn fillings on an unknotted component of a homologically trivial link is still homologically trivial and has the same longitudes.

The proof of Theorem 1 uses Dehn fillings to produce closed 3–manifolds which admit irreducible representations; this will be done by iterating $1/q$–Dehn fillings along the components of the link. However, even if all the components of a link L in $\mathbb{S}^3$ are unknotted, a $1/q$–Dehn filling along a component generally knots the other components, thus making impossible to continue the process while remaining in $\mathbb{S}^3$. In other words, to achieve this goal, we need a family of links $\mathcal{L}$ satisfying

- if $L \in \mathcal{L}$ has two or more components, each is individually unknotted;
- for any $K \sqcup L_0$ in $\mathcal{L}$, we have that L_q is also in $\mathcal{L}$.

In the next section, we show that the family of *Brunnian links* in $\mathbb{S}^3$ satisfies these conditions. Moreover, nontriviality can be preserved in the process, making it possible to reason by induction on the number of components of the link.

2.2 Brunnian links

Definition 5 A link is called *Brunnian* if any of its proper sublinks is trivial.

Remark Any knot is considered Brunnian; for links with more components, we have:

- If a Brunnian link has two or more components, they are individually unknotted.
- Any Brunnian link with three or more components is homologically trivial.
- By Section 2.1, if $L = K \sqcup L_0$ is Brunnian, L_q is also Brunnian for any integer q.

Given $L = K \sqcup L_0$ Brunnian, we can perform a $1/p$–surgery on a component of L_q to obtain another Brunnian link, and so on, until obtaining a knot in $\mathbb{S}^3$. However, any $1/q$–Dehn filling on a component of the Hopf link or the unlink produces the unlink. Therefore, given a Brunnian link $L = K \sqcup L_0$, we need to prevent L_q from being the Hopf link or the unlink in order to obtain, in fine, a nontrivial knot in $\mathbb{S}^3$.

If $L = K \sqcup K'$ is a Brunnian link with two components, this is a special case of Mathieu's theorem from [9]. This more general result on knots in a solid torus (links with one unknotted component) asserts that, besides the Hopf link, for any $|q| \geqslant 2$, any $1/q$–Dehn filling on an unknotted component of a 2–component link in $\mathbb{S}^3$ produces a nontrivial knot. For our concern, this implies that, for any $|q| \geqslant 2$, the $1/q$–Dehn filling on any component of a Brunnian, non-Hopf, nontrivial 2–link may never produce the trivial knot.

On the other hand, if L has three components or more, it is homologically trivial, and the work of Mangum and Stanford [8, Theorem 2 and its proof] ensures that, for any integer q and any homologically trivial Brunnian link $L = K \sqcup L_0$, if L is nontrivial, then L_q is trivial if and only if $q = 0$. Otherwise, it is a nontrivial, homologically trivial Brunnian link (in particular, it is never the Hopf link).

Therefore, we obtain the following result for the stability of nontrivial non-Hopf Brunnian links under $1/q$–Dehn fillings:

Proposition 6 *Let $L = K \sqcup L_0$ be a nontrivial, non-Hopf, Brunnian link in $\mathbb{S}^3$. Then for any $|q| \geqslant 2$, the link L_q is a Brunnian link in $\mathbb{S}^3$, nontrivial and non-Hopf.*

We will use the stability of nontrivial non-Hopf Brunnian links to apply the Kronheimer–Mrowka theorem on some Dehn fillings of the link exteriors to produce nontrivially maximal components in the eigenvalue varieties. On the other hand, for the Hopf link and the trivial link, no such component exists:

Proposition 7 *The eigenvalue varieties of the Hopf link and the trivial link do not admit any nontrivially maximal component.*

Proof The fundamental group of the exterior of the Hopf link is abelian, so all the characters are abelian, and $\mathfrak{E} = \mathfrak{E}^{\mathrm{ab}}$.

On the other hand, for the trivial link, all the longitudes are nullhomotopic and are therefore trivialized by any representation, so $\mathcal{A} = \langle \mathfrak{l}_K - 1,\ K \subset L \rangle = \mathcal{A}^{\mathrm{ab}}$. □

2.3 Kronheimer–Mrowka characters

By the Kronheimer–Mrowka theorem from [5], any nontrivial $1/q$–Dehn filling along a nontrivial knot in $\mathbb{S}^3$ produces a closed 3–manifold which admits an irreducible representation in SU_2. By Proposition 6, if $L = K \sqcup L_0$ is a nontrivial Brunnian link in $\mathbb{S}^3$, L_q is nontrivial for any $|q| \geqslant 2$. Performing another $1/p$–Dehn filling on a component of L_q (in the new standard peripheral system if the link is not homologically trivial) will produce again a nontrivial Brunnian link; this process may be continued until a nontrivial knot is produced, on which a final $1/k$–Dehn filling may be performed to obtain a closed 3–manifold which admits an irreducible representation in SU_2.

For any Brunnian link $L = K_1 \sqcup \cdots \sqcup K_n$ in $\mathbb{S}^3$, and any $\underline{q} = (q_1, \ldots, q_k)$ in $\mathbb{Z}^k$ for $k \leqslant n$, we denote by $L(\underline{q})$ the 3–manifold obtained by performing $1/q_i$–Dehn fillings on the components of L, where each $1/q_i$–Dehn filling is performed in the standard peripheral system given after the Dehn fillings $1/q_j$ for $j < i$.

Remark As already pointed out, the meridians never change, and since L is assumed Brunnian, longitudes change only if L is a Brunnian link with two components $L = K_1 \sqcup K_2$ with nonzero linking number α; in that case, denoting by $(\mu_i, \lambda_i)_{i=1,2}$ the respective standard peripheral systems, any $1/q_1$–Dehn filling on K_1 changes the longitude λ_2 into $\lambda_2 + q_1\alpha^2\mu_2$. Therefore, a $1/q_2$–Dehn filling on K_2 is performed along the slope

$$(1 + q_1q_2\alpha^2)\mu_2 + q_2\lambda_2 \in H_1(T_{K_2}).$$

Proposition 8 *Let $L = K_1 \sqcup \cdots \sqcup K_n$ be a nontrivial Brunnian link in $\mathbb{S}^3$ different from the Hopf link, and let $\underline{q} = (q_1, \ldots, q_n)$ be a family of integers.*

- *If $q_i = 0$ for some $1 \leqslant i \leqslant n$, then $L_{\underline{q}} = \mathbb{S}^3$.*
- *If $|q_i| \geqslant 2$ for all $1 \leqslant i \leqslant n$, then there exists an irreducible representation*

$$\rho_{\underline{q}} \colon \pi_1 L_{\underline{q}} \to \mathrm{SU}_2.$$

Proof First, if one of the q_i is zero, the link $L_{(q_1,\ldots,q_i)}$ is trivial, so performing $1/q_k$–Dehn fillings for $i < k \leqslant n$ produces the standard 3–sphere.

On the other hand, if all the $|q_i|$ are greater than 1, each $L_{(q_1,\ldots,q_k)}$ for $k \leqslant n$ is nontrivial by Proposition 6, so $L_{(q_1,\ldots,q_{n-1})}$ is a nontrivial knot in $\mathbb{S}^3$. The Kronheimer–Mrowka theorem concludes the proof. □

By inclusion of SU_2 in $\mathrm{SL}_2\mathbb{C}$, we can consider $\rho_{\underline{q}}$ as an irreducible representation of $R(L_{\underline{q}})$ (with no nontrivial parabolic image). Moreover, composing with the group homomorphism $\pi_1 L \to \pi_1 L_{\underline{q}}$, we may also consider $\rho_{\underline{q}}$ as an irreducible representation of $R(L)$. The irreducible characters $\chi_{\underline{q}} = t(\rho_{\underline{q}})$ obtained this way are called *Kronheimer–Mrowka characters*, and we denote by $X_{\mathrm{KM}}(L)$ the Zariski closure in $X(L)$ of all Kronheimer–Mrowka characters:

$$X_{\mathrm{KM}}(L) = \overline{\{\chi_{\underline{q}},\ \underline{q} \in (\mathbb{Z} \setminus \{-1, 0, 1\})^{|L|}\}}.$$

Remark The subset $X_{\mathrm{KM}}(L)$ of $X(L)$ may contain several algebraic components.

Remark For any nontrivial, non-Hopf, Brunnian link $L = K \sqcup L_0$, the group homomorphism $i_q \colon \pi_1 L \to \pi_1 L_q$ induces an algebraic map

$$i_q^* \colon X(L_q) \to X(L),$$

and if $|q| \geqslant 2$, then $i_q^* X_{\mathrm{KM}}(L_q) \subset X_{\mathrm{KM}}(L)$.

Any representation $\rho_{\underline{q}}$ satisfies the $1/q_K$–Dehn filling relations for each component K of L. On the other hand, no $\rho_{\underline{q}}(\mu_K)$ is trivial since, otherwise, it would satisfy the $1/0$ relation on K; it would then factor as a representation of $\mathbb{S}^3$ and therefore be trivial. Since $\rho_{\underline{q}}$ factors in SU_2, this is equivalent to $\operatorname{tr}\rho_{\underline{q}}(\mu_K\lambda_K^{q_K}) = 2$ and $\operatorname{tr}\rho_{\underline{q}}(\mu_K) \neq 2$.

It follows that any Kronheimer–Mrowka character $\chi_{\underline{q}}$ satisfies, for any $K \subset L$,

$$I_{\mu_K\lambda_K^{q_K}}(\chi_{\underline{q}}) = 2, \tag{1}$$

$$I_{\mu_K}(\chi_{\underline{q}}) \neq 2. \tag{2}$$

Finally, following Section 1, we denote by $\mathfrak{E}_{\mathrm{KM}}(L)$ the part corresponding to $X_{\mathrm{KM}}(L)$ in $\mathfrak{E}(L)$. For any $\xi_{\underline{q}} \in \mathfrak{E}_{\mathrm{KM}}(L)$ corresponding to a Kronheimer–Mrowka character $\chi_{\underline{q}}$ in $X_{\mathrm{KM}}(L)$, and any component K of L, (1) and (2) imply that

$$\mathfrak{m}_K\mathfrak{l}_K^{q_K}(\xi_{\underline{q}}) = 1, \tag{3}$$

$$\mathfrak{m}_K(\xi_{\underline{q}}) \neq 1. \tag{4}$$

Remark Together with the equations for $\mathcal{A}^{\mathrm{red}}(L)$, this implies that no such point $\xi_{\underline{q}}$ is in $\mathfrak{E}^{\mathrm{red}}(L)$, so no component of $\mathfrak{E}_{\mathrm{KM}}(L)$ is contained in $\mathfrak{E}^{\mathrm{red}}(L)$.

2.4 Maximal components

In this last section, we prove the following result which implies Theorem 1:

Theorem 9 *For any nontrivial Brunnian link L different from the Hopf link, $\mathfrak{E}_{\mathrm{KM}}(L)$ contains a maximal component.*

Proof This is proved by induction on the number of components of L.

For the base case, L is a knot K, and the proof is the same as the one for the nontriviality of the A–polynomial of nontrivial knots from Dunfield and Garoufalidis in [4] or Boyer and Zhang in [1].

For any $|q| \geqslant 2$, performing $1/q$–surgery produces an irreducible character χ_q in $X(K)$ and a point $\xi_q = (m_q, \ell_q)$ in $\mathfrak{E}(K)$. They show that there are infinitely many distinct ℓ_q obtained this way, so $\mathfrak{E}_{\mathrm{KM}}(K)$ contains a curve different from the line $\mathfrak{l}-1$. We do not reproduce this proof here, but very similar ideas are used in the induction step.

For the induction step, let $L = K \sqcup L_0$ be a nontrivial, non-Hopf, Brunnian link in $\mathbb{S}^3$. For any $|q| \geqslant 2$, L_q is nontrivial, non-Hopf and Brunnian, so we can assume, by induction, that $\mathfrak{E}_{\mathrm{KM}}(L_q)$ contains a maximal component.

We have the commutative diagram

$$\begin{tikzcd}
X_{\mathrm{KM}}(L_q) \arrow[r] \arrow[d, "r_q"'] & X_{\mathrm{KM}}(L) \arrow[d, "r"] \\
\prod_{J\neq K} X(T_J) & \prod_{J\subset L} X(T_J) \arrow[l] \\
\prod_{J\neq K} E(T_J) \arrow[u, "d"] & \prod_{J\subset L} E(T_J) \arrow[l] \arrow[u, "d"']
\end{tikzcd}$$

so there exists X_q in $X_{\mathrm{KM}}(L)$ corresponding to $\mathfrak{E}_q$ in $\mathfrak{E}_{\mathrm{KM}}(L)$ such that $\dim \mathfrak{E}_q \geqslant |L|-1$. If $\dim \mathfrak{E}_q = |L|$ for some q, then there is nothing more to prove.

Let us assume now that all the components $\mathfrak{E}_q$ have dimension $|L|-1$. We will show that $\mathfrak{E}_{\mathrm{KM}}(L)$ contains infinitely many different such subspaces $\mathfrak{E}_q$; by algebraicity, this means that $\mathfrak{E}_{\mathrm{KM}}(L)$ must contain a component of dimension $|L|$, which will conclude the proof of Theorem 9.

The subspaces $\mathfrak{E}_q$ will be separated using the following lemma:

Lemma 10 *For any integers q, q',*

$$\mathfrak{E}_q \subset \mathfrak{E}_{q'} \implies \mathfrak{l}_K^{q-q'}|_{\mathfrak{E}_q} \equiv 1.$$

Moreover, for any $|q| \geqslant 2$, the set $\{p \in \mathbb{Z} \mid \mathfrak{l}_K^p|_{\mathfrak{E}_q} \equiv 1\}$ is an ideal $d_q\mathbb{Z}$ with $q \notin d_q\mathbb{Z}$.

Proof For any ξ in $\mathfrak{E}_q$, we have $\mathfrak{m}_K \mathfrak{l}_K^q(\xi) = 1$ by (3), so if ξ also belongs to $\mathfrak{E}_{q'}$, then $\mathfrak{m}_K \mathfrak{l}_K^{q'}(\xi) = 1$ and $\mathfrak{l}_K^{q-q'}(\xi) = 1$. Therefore, if $\mathfrak{E}_q \subset \mathfrak{E}_{q'}$, then $\mathfrak{l}_K^{q-q'} \equiv 1$ on $\mathfrak{E}_q$.

If q is in the ideal $d_q\mathbb{Z}$, the surgery relation implies that $\mathfrak{m}_K|_{\mathfrak{E}_q} \equiv 1$, in contradiction with (4). □

If $S = \{q \in \mathbb{Z} \setminus \{-1, 0, 1\} \mid d_q = 0\}$ is infinite, then by Lemma 10, $\mathfrak{E}_q \neq \mathfrak{E}_{q'}$ for $q \neq q'$ in S, so $(\mathfrak{E}_q)_{q\in S}$ is a family of infinitely many distinct subspaces.

Otherwise, there exists N in $\mathbb{N}$ such that, for any $q \geqslant N$, $d_q \geqslant 2$. Let $(q_i)_{i\in\mathbb{N}}$ be a family of integers such that

- $q_0 \geqslant N$;
- for any j in $\mathbb{N}$, we have $q_{j+1} \geqslant q_j$ and $q_{j+1} \in \bigcap_{i=1}^{j} d_{q_i}\mathbb{Z}$.

Then $(\mathfrak{E}_{q_i})_{i\in\mathbb{N}}$ contains infinitely many different subspaces, since

$$\mathfrak{E}_{q_i} \neq \mathfrak{E}_{q_j} \quad \text{for all } i < j.$$

Indeed, for any j in $\mathbb{N}$, let us assume that $\mathfrak{E}_{q_i} = \mathfrak{E}_{q_j}$ for some $i < j$. By Lemma 10, this would imply that $q_j - q_i \in d_{q_i}\mathbb{Z}$. But $q_j \in d_{q_i}\mathbb{Z}$ by construction, so this would imply $q_i \in d_{q_i}\mathbb{Z}$, a contradiction.

We have proved that $\mathfrak{E}_{\mathrm{KM}}(L)$ contains infinitely many different subsets of dimension $|L|-1$; by algebraicity, it must contain a component of dimension $|L|$, which concludes the proof of Theorem 9. $\square$

References

[1] **S Boyer**, **X Zhang**, *Every nontrivial knot in S^3 has nontrivial A–polynomial*, Proc. Amer. Math. Soc. 133 (2005) 2813–2815 MR

[2] **D Cooper**, **M Culler**, **H Gillet**, **D D Long**, **P B Shalen**, *Plane curves associated to character varieties of 3–manifolds*, Invent. Math. 118 (1994) 47–84 MR

[3] **M Culler**, **P B Shalen**, *Varieties of group representations and splittings of 3–manifolds*, Ann. of Math. 117 (1983) 109–146 MR

[4] **N M Dunfield**, **S Garoufalidis**, *Non-triviality of the A–polynomial for knots in S^3*, Algebr. Geom. Topol. 4 (2004) 1145–1153 MR

[5] **P B Kronheimer**, **T S Mrowka**, *Dehn surgery, the fundamental group and* SU(2), Math. Res. Lett. 11 (2004) 741–754 MR

[6] **A E Lash**, *Boundary curve space of the Whitehead link complement*, PhD thesis, Univ. of Calif., Santa Barbara (1993) MR `http://tinyurl.com/LashThesis1993`

[7] **F Malabre**, *Eigenvalue varieties of abelian trees of groups and link-manifolds*, PhD thesis, Univ. Autònoma de Barcelona (2015) `http://tinyurl.com/MalabreTh15`

[8] **B Mangum**, **T Stanford**, *Brunnian links are determined by their complements*, Algebr. Geom. Topol. 1 (2001) 143–152 MR

[9] **Y Mathieu**, *Unknotting, knotting by twists on disks and property* (P) *for knots in S^3*, from "Knots 90" (A Kawauchi, editor), De Gruyter, Berlin (1992) 93–102 MR

[10] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes, Princeton University (1979) `http://msri.org/publications/books/gt3m`

[11] **S Tillmann**, *On character varieties*, PhD thesis, Univ. of Melbourne (2002) Available at `http://www.maths.usyd.edu.au/u/tillmann/papers/thesis.pdf`

[12] **S Tillmann**, *Boundary slopes and the logarithmic limit set*, Topology 44 (2005) 203–216 MR

Department of Mathematics, University of Barcelona
Gran Vía, 585, 08007 Barcelona, Spain

`malabre@ub.edu`

Received: 11 December 2015 Revised: 29 November 2016

Algebraic & Geometric Topology 17 (2017) 2051–2080

A refinement of Betti numbers and homology in the presence of a continuous function, I

DAN BURGHELEA

We propose a refinement of the Betti numbers and the homology with coefficients in a field of a compact ANR X, in the presence of a continuous real-valued function on X. The refinement of Betti numbers consists of finite configurations of points with multiplicities in the complex plane whose total cardinalities are the Betti numbers, and the refinement of homology consists of configurations of vector spaces indexed by points in the complex plane, with the same support as the first, whose direct sum is isomorphic to the homology. When the homology is equipped with a scalar product, these vector spaces are canonically realized as mutually orthogonal subspaces of the homology.

The assignments above are in analogy with the collections of eigenvalues and generalized eigenspaces of a linear map in a finite-dimensional complex vector space. A number of remarkable properties of the above configurations are discussed.

55N35; 46M20, 57R19

1 Introduction

The results of this paper and its subsequent part II, mostly obtained in collaboration with Stefan Haller, provide a shorter version of some results in [3], still unpublished, extend their generality based on the involvement of the topology of Hilbert cube manifolds and refine them as configurations of complex numbers and of vector spaces.

Precisely, for a fixed field κ and $r \geq 0$, one proposes a refinement of the Betti numbers $b_r(X)$ of a compact ANR X [1] and a refinement of the homology $H_r(X)$ with coefficients in the field κ in the presence of a continuous function $f\colon X \to \mathbb{R}$.

The refinements consists of finite configurations of points with multiplicity located in the plane $\mathbb{R}^2 = \mathbb{C}$, denoted by δ^f_r, equivalently of monic polynomials with complex coefficients $P^f_r(z)$, of degree the Betti numbers $b_r(X)$, and finite configurations of κ–vector spaces denoted by $\hat{\delta}^f_r$ with the same support and direct sum of all vector spaces isomorphic to $H_r(X)$; see Theorem 4.1. The points of the configurations δ^f_r,

[1] See the definition of an ANR in Section 2.2.

Published: 3 August 2017

DOI: 10.2140/agt.2017.17.2051

equivalently the zeros of the polynomials $P^f_r(z)$, are complex numbers $z=a+ib\in\mathbb{C}$ with both a,b critical values;[2] see Theorem 4.1. The two configurations are related by $\dim\hat\delta^f_r=\delta^f_r$.

We show the following:

(1) The assignment $f\rightsquigarrow P^f_r(z)$ is continuous when f varies in the space of continuous maps equipped with the compact open topology; see Theorem 4.2.

(2) For an open and dense subset of continuous maps (defined on X, an ANR satisfying some mild properties) the points of the configurations δ^f_r or the zeros of the polynomials $P^f_r(z)$ have multiplicity one; see Theorem 4.1.

(3) When X is a closed topological n–manifold, the Poincaré duality between the Betti numbers β_r and β_{n-r} gets refined to a Poincaré duality between configurations δ^f_r and δ^f_{n-r}, and the Poincaré duality between $H_r(X)$ and $H_{n-r}(X)^*$ to a Poincaré duality between configurations $\hat\delta^f_r$ and $(\hat\delta^f_{n-r})^*$; see Theorem 4.3.

(4) For each point of the configuration δ^f_r, equivalently zero z of the polynomial $P^f_r(z)$, the assigned vector space $\hat\delta^f_r(z)$ has dimension the multiplicity of z and is a quotient of vector subspaces $\hat\delta^f_r(z)=\mathbb{F}_r(z)/\mathbb{F}'_r(z)$, $\mathbb{F}'_r(z)\subseteq\mathbb{F}_r(z)\subseteq H_r(X)$. When $\kappa=\mathbb{R}$ or $\mathbb{C}$ and $H_r(X)$ is equipped with a Hilbert space structure $\hat\delta^f_r(z)$ identifies canonically to a subspace $\boldsymbol{H}_r(z)$ of $H_r(X)$ such that $\boldsymbol{H}_r(z)\perp\boldsymbol{H}_r(z')$ for $z\neq z'$ and $\bigoplus_z\boldsymbol{H}_r(z)=H_r(X)$; see Theorem 4.1. This provides an additional structure (direct sum decomposition of $H_r(X)$, which in view of Theorem 4.1, for a generic f, has all components of dimension 1).

We refer to the system $(H_r(X), P^f_r(z), \hat\delta^f_r)$ as the *r–homology spectral package* of (X,f), in analogy with the spectral package of (V,T), where V is a vector space and T a linear endomorphism, which consists of the characteristic polynomial $P^T(z)$ with its roots z_i, the eigenvalues of T and with their corresponding generalized eigenspaces V_{z_i}.

In case X is the underlying space of a closed oriented Riemannian manifold (M^n,g) and $\kappa=\mathbb{R}$ or $\mathbb{C}$, the vector space $H_r(M^n)$, via the identification with the harmonic r–forms, has a structure of a Hilbert space. The configuration $\hat\delta^f_r$, for f generic, provides a base in the space of harmonic forms.

All these results are collected in the main theorems below, Theorems 4.1–4.3, which were partially established in [3], not yet in print, but under more restrictive hypotheses like "X homeomorphic to a simplicial complex" or "f a tame map". In this paper, we removed these hypotheses using results on Hilbert cube manifolds reviewed in Section 2.3, and complete them with additional results.

[2] See Section 2.2 below for the definition of regular and critical value.

It is worth noting that the points of the configurations δ^f_r located above and on the diagonal in the plane $\mathbb{R}^2$ determine and are determined by the closed r–bar codes in the level persistence of f, while those below the diagonal are determined by and determine the open $(r-1)$–bar codes in the level persistence as observed in [3]. The algorithms proposed by Carlsson, de Silva and Morozov [4] and the author and Dey [2] can be used for their calculation.

Similar refinements hold for angle-valued maps and will be discussed in part II. In this case, the homology has to be replaced by either the Novikov homology of (X, ξ_f) which in our work is a finitely generated free module over the ring of Laurent polynomials $\kappa[t^{-1}, t]$ or, in case κ is $\mathbb{R}$ or $\mathbb{C}$, by the L_2–homology of the infinite cyclic cover defined by $\xi_f \in H^1(X:\mathbb{Z})$, determined by f. In this case, the L_2–homology is regarded as a Hilbert module over the von Neumann algebra associated to the group $\mathbb{Z}$, $\boldsymbol{H}_r(z)$ are Hilbert submodules and $\delta^f_r(x)$ is the von Neumann dimension of $\boldsymbol{H}_r(z)$. Note that the L_2–Betti numbers are actually the Novikov–Betti numbers of (X, ξ_f) (which agree with the rank of the corresponding free module).

Acknowledgements The author thanks S Ferry for help in clarifying a number of aspects about Hilbert cube manifolds and ANRs. The author is equally grateful to the referee for many suggestions, requests for clarifications and sometimes alternative arguments.

2 Preliminary definitions

2.1 Configurations

Let X be a topological space. A *finite configuration of points in* X is a map

$$\delta\colon X \to \mathbb{Z}_{\geq 0}$$

with finite support.

A *finite configuration of vector spaces indexed by points in* X is a map with finite support

$$\hat{\delta}\colon X \to \mathrm{Vect}$$

(ie $\hat{\delta}(x) = 0$ for all but finitely many $x \in X$), where Vect denotes the collection of κ–vector spaces.

For N a positive integer, denote by $\mathcal{C}_N(X)$ the set of configurations of points in X with total cardinality N:

$$\mathcal{C}_N(X) := \Big\{\delta\colon X \to \mathbb{Z}_{\geq 0} \mid \sum_{x\in X} \delta(x) = N\Big\}.$$

For V a finite-dimensional κ–vector space, denote by $\mathcal{P}(V)$ the set of subspaces of V and by $\mathcal{C}_V(X)$ the set

$$\mathcal{C}_V(X):=\Big\{\bar\delta\colon X\to\mathcal{P}(V)\,\Big|\,\sharp\{x\in X\mid \bar\delta(x)\neq 0\}<\infty,\ \bar\delta(x)\cap\sum_{y\neq x}\bar\delta(y)=0,\ \sum_{x\in X}\bar\delta(x)=V\Big\}.$$

Here $\sharp$ denotes cardinality of the set in braces.

Consider the map

$$e\colon \mathcal{C}_V(X)\to\mathcal{C}_{\dim V}(X)$$

defined by

$$e(\bar\delta)(x)=\dim\bar\delta(x),$$

and call the configuration $e(\bar\delta)$ the *dimension* of $\bar\delta$.

Both sets $\mathcal{C}_N(X)$ and $\mathcal{C}_V(X)$ can be equipped with natural topology (the *collision topology*). One way to describe these topologies is to specify for each δ or $\hat\delta$ a system of *fundamental neighborhoods*. If δ has as support the set of points $\{x_1,x_2,\dots,x_k\}$, a fundamental neighborhood $\mathcal{U}$ of δ is specified by a collection of k disjoint open neighborhoods $U_1,\dots,U_k$ of $x_1,\dots,x_k$, and consists of $\big\{\delta'\in\mathcal{C}_N(X)\mid\sum_{x\in U_i}\delta'(x)=\delta(x_i)\big\}$. Similarly if $\bar\delta$ has as support the set of points $\{x_1,x_2,\dots,x_k\}$ with $\bar\delta(x_i)=V_i\subseteq V$, a fundamental neighborhood $\mathcal{U}$ of $\bar\delta$ is specified by a collection of k disjoint open neighborhoods $U_1,U_2,\dots,U_k$ of $x_1,\dots,x_k$, and consists of

$$\Big\{\bar\delta'\in\mathcal{C}_V(X)\mid x\in U_i\Rightarrow\bar\delta'(x)\subset V_i,\ \bigoplus_{x\in U_i}\bar\delta'(x)=V_i\Big\}.$$

Clearly e is continuous.

When κ is an infinite field, the topology of $\mathcal{C}_V(X)$ has too many connected components to be useful unless the geometry forces the possible values of the configurations to be at most countable.

When $\kappa=\mathbb{R}$ or $\mathbb{C}$ and V is a Hilbert space, it is natural to consider the subset of $\mathcal{C}^O_V(X)\subset\mathcal{C}_V(X)$ consisting of configurations whose vector spaces $\bar\delta(x)$ are mutually orthogonal. In this case for $\bar\delta$ with support the set of points $\{x_1,x_2,\dots,x_k\}$ and $\bar\delta(x_i)=V_i\subseteq V$, one can consider a fundamental neighborhood $\mathcal{U}$ of $\bar\delta$ that is specified by a collection of k disjoint open neighborhoods $U_1,U_2,\dots,U_k$ of $x_1,\dots,x_k$ and open neighborhoods $O_1,O_2,\dots,O_k$ of V_i in $G_{\dim V_i}(V)$, and consists of

$$\Big\{\bar\delta'\in\mathcal{C}^O_V(X)\mid\bigoplus_{x\in U_i}\hat\delta'(x)\in O_i\Big\}.$$

Here $G_k(V)$ denotes the Grassmannian of k–dimensional subspaces of V.

With respect to this topology e is continuous, surjective and proper, with fiber above δ, the subset of $G_{n_1}(V) \times G_{n_2}(V) \times \cdots \times G_{n_k}(V)$ consisting of $(V'_1, V'_2, \dots, V'_k)$, $V'_i \in G_{n_i}(V)$ mutually orthogonal, where $n_i = \dim V_i$. This set is compact and is actually an algebraic variety.

Remark (1) $\mathcal{C}_N(X) = X^N/\Sigma_N$ is the so-called N–symmetric product, and if X is a metric space with distance D then the collision topology is the topology defined by the distance $\underline{D}$ on X^N/Σ_N induced from the distance on X^N given by $D(x_1, x_2, \dots, x_N; y_1, y_2, \dots, y_N) := \sup_{i=1,\dots,N}\{D(x_i, y_i)\}$.

(2) If $X = \mathbb{R}^2 = \mathbb{C}$ then $\mathcal{C}_N(X)$ identifies to the set of monic polynomials with complex coefficients. To the configuration δ whose support consists of the points $z_1, z_2, \dots, z_k$ with $\delta(z_i) = n_i$, one associates the monic polynomial $P^f(z) = \prod_i (z - z_i)^{n_i}$. Then $\mathcal{C}_N(X)$ and $\mathbb{C}^N$ are identified as metric spaces.

(3) The space $\mathcal{C}_V(X)$ and thus $\mathcal{C}_V(\mathbb{R}^2)$ can be equipped with a complete metric which induces the collision topology but this will not be used here.

2.2 Tame maps

Recall that a metrizable space X is an ANR if any closed subset A of a metrizable space B with A homeomorphic to X has a neighborhood U which retracts to A; see [7, Chapter 3]. Recall also that any space homeomorphic to a locally finite simplicial complex, a finite-dimensional topological manifold or an infinite-dimensional manifold (ie a paracompact Hausdorff space locally homeomorphic to the Hilbert space l_2 or the Hilbert cube I^∞) is an ANR; see [7].

Convention All maps $f\colon X \to \mathbb{R}$ in this paper are continuous proper maps defined on an ANR X, hence if such maps exists, X is locally compact. From now on the words "proper continuous" should always be assumed to precede the word "map" even if not specified.

The following concepts are consistent with the familiar terminology in topology:

- A map $f\colon X \to \mathbb{R}$ is *weakly tame* if for any $t \in \mathbb{R}$, the level $f^{-1}(t)$ is an ANR. Therefore, for any bounded or unbounded closed interval $I = [a, b]$, $a, b \in \mathbb{R} \sqcup \{\infty, -\infty\}$, $f^{-1}(I)$ is an ANR. Indeed if $I = [a, b]$, in view of the hypothesis that $f^{-1}(a)$ and $f^{-1}(b)$ are ANRs and of the definition of ANR, there exists an open set $U \subset X \setminus f^{-1}(a, b)$ which retracts to $f^{-1}(a) \sqcup f^{-1}(b)$. Then $U \cup f^{-1}[a, b]$ is an open set in X which retracts to $f^{-1}(I)$. Since X is an ANR this suffices to conclude that $f^{-1}(I)$is an ANR; see [7]. A similar argument can be used for $I = (-\infty, a]$ or $I = [b, \infty)$.

- The number $t \in \mathbb{R}$ is a *regular value* if there exists $\epsilon > 0$ such that for any $t' \in (t-\epsilon, t+\epsilon)$, the open set $f^{-1}(t-\epsilon, t+\epsilon)$ retracts by deformation to $f^{-1}(t')$. A number t which is not a regular value is a *critical value*. In view of the hypothesis on f a map (ie X locally compact and f proper), the requirement on t in the definition of *weakly tame* is satisfied for any regular value t. Informally, the critical values are the values t for which the topology of the level (= homotopy type) changes. One denotes by $\mathrm{Cr}(f)$ the collection of critical values of f.

- The map f is called *tame* if it is weakly tame and, in addition,

  (a) the set of critical values $\mathrm{Cr}(f) \subset \mathbb{R}$ is discrete, and

  (b) $\epsilon(f) := \inf\{|c-c'| : c, c' \in \mathrm{Cr}(f), c \neq c'\}$ satisfies $\epsilon(f) > 0$.

If X is compact then (a) implies (b).

- An ANR which has the tame maps dense in the set of all maps with respect to the fine C_0–topology is called a *good* ANR.

There exist compact ANRs (actually compact homological n–manifolds; see [6]) with no codimension-one subsets which are ANRs, hence compact ANRs which are not *good*.

The reader should be aware of the following rather obvious facts.

Observation 2.1 (1) If f is a weakly tame map then $f^{-1}([a,b])$ is a compact ANR and has the homotopy type of a finite simplicial complex (see [8]) and therefore has finite-dimensional homology with respect to any field κ.

(2) If X is a locally finite simplicial complex and f is a simplicial map, then f is weakly tame with the set of critical values discrete. Critical values are among the values of f on vertices. If in addition X is compact then f is tame.

(3) If X is homeomorphic to a finite simplicial complex then the set of tame maps is dense in the set of all continuous maps with the C_0–topology (ie compact open topology). The same remains true if X is a compact Hilbert cube manifold, defined in the next section. In particular all these spaces are good ANRs.

For the needs of this paper, weaker than usual concepts of regular or critical values and tameness, relative to homology with coefficients in the field κ, suffice. They are introduced in Section 3.

2.3 Compact Hilbert cube manifolds

Recall the following:

- The *Hilbert cube* Q is the infinite product $Q = I^\infty = \prod_{i\in\mathbb{Z}_{\geq 1}} I_i$ with $I_i = [0,1]$. The topology of Q is given by the distance $d(\bar{u},\bar{v}) = \sum_i |u_i - v_i|/2^i$ with $\bar{u} = \{u_i \in I, i \in \mathbb{Z}_{\geq 1}\}$ and $\bar{v} = \{v_i \in I, i \in \mathbb{Z}_{\geq 1}\}$.
- The space Q is a compact ANR and so is any $X \times Q$ for any compact ANR X.
- A *compact Hilbert cube manifold* is a compact Hausdorff space locally homeomorphic to the Hilbert cube Q.

For $f\colon X \to \mathbb{R}$ and $F\colon X \times Q \to \mathbb{R}$, denote by $\bar{f}_Q\colon X \times Q \to \mathbb{R}$ and $F_k\colon X \times Q \to \mathbb{R}$ the maps defined by

$$\bar{f}_Q(x,\bar{u}) = f(x) \quad \text{and} \quad F_k(x,\bar{u}) = F(x,u_1,u_2,\ldots,u_k,0,0,\ldots).$$

Observation 2.2 In view of the definition of $\bar{f}_Q$ and of the metric on Q, observe the following:

(1) If $f\colon X \to \mathbb{R}$ is a tame map, so is $\bar{f}_Q$.

(2) If X is compact then the sequence of maps F_n is uniformly convergent to the map F when $n \to \infty$.

The following are basic results about compact Hilbert cube manifolds whose proof can be found in [5].

Theorem 2.3 (1) (R Edwards) *If X is a compact ANR then $X \times Q$ is a compact Hilbert cube manifold.*

(2) (T Chapman) *Any compact Hilbert cube manifolds is homeomorphic to $K \times Q$ for some finite simplicial complex K.*

(3) (T Chapman) *If $\omega\colon X \to Y$ is a homotopy equivalence between two finite simplicial complexes with Whitehead torsion $\tau(\omega) = 0$ then the there exists a homeomorphism $\omega'\colon X \times Q \to Y \times Q$ such that ω' and $\omega \times \mathrm{id}_Q$ are homotopic. As a consequence of Observation 2.4 below, two compact Hilbert cube manifolds which are homotopy equivalent become homeomorphic after product with $\mathbb{S}^1$.*

Observation 2.4 (folklore) If ω is a homotopy equivalence between two finite simplicial complexes then $\omega \times \mathrm{id}_{\mathbb{S}^1}$ has the Whitehead torsion $\tau(\omega \times \mathrm{id}_{\mathbb{S}^1}) = 0$.

As a consequence of the above statements we have the following proposition.

Proposition 2.5 *Any compact Hilbert cube manifold M is a good ANR.*

Proof A map $f\colon M \to \mathbb{R}$, M a compact Hilbert cube manifold, is called *special* if there exists a finite simplicial complex K, a map $g\colon K \to \mathbb{R}$ and a homeomorphism $\theta\colon M \to K \times Q$ such that $\bar{g} \cdot \theta = f$, and a special map is PL[3] if in addition g is PL. By Observation 2.2 any map $f\colon M \to \mathbb{R}$ is $\epsilon/2$–close to a special map. Since any continuous real-valued map defined on a simplicial complex K is $\epsilon/2$–close to a PL map then any special map on M is $\epsilon/2$–close to a special PL map. Consequently f is ϵ–close to a special PL map which is tame in view of Observations 2.1 and 2.2. This implies that the set of tame maps is dense in the set of all continuous maps. □

3 The configurations δ_r^f and $\hat{\delta}_r^f$

In this paper we fix a field κ, and for a space X denote by $H_r(X)$ the homology of X with coefficients in the field κ. Let $f\colon X \to \mathbb{R}$ be a map. As in the previous section, f is proper continuous and X is a locally compact ANR. One defines

(1) the sublevel $X_a := f^{-1}(-\infty, a])$,

(2) the superlevel $X^b := f^{-1}([b, \infty))$,

(3) $\mathbb{I}_a^f(r) := \operatorname{img}(H_r(X_a) \to H_r(X)) \subseteq H_r(X)$,

(4) $\mathbb{I}_f^b(r) := \operatorname{img}(H_r(X^b) \to H_r(X)) \subseteq H_r(X)$,

(5) $\mathbb{F}_r^f(a,b) = \mathbb{I}_a^f(r) \cap \mathbb{I}_f^b(r) \subseteq H_r(X)$.

Clearly one has the following observation.

Observation 3.1 (1) For $a' \le a$ and $b \le b'$, one has $\mathbb{F}_r^f(a',b') \subseteq \mathbb{F}_r^f(a,b)$.

(2) For $a' \le a$ and $b \le b'$, one has $\mathbb{F}_r^f(a',b) \cap \mathbb{F}_r^f(a,b') = \mathbb{F}_r^f(a',b')$.

(3) $\sup_{x \in X} |f(x) - g(x)| < \epsilon$ implies $\mathbb{F}^g(a-\epsilon, b+\epsilon) \subseteq \mathbb{F}_r^f(a,b)$.

Note that we also have the following proposition.

Proposition 3.2 *If f is a map as above then* $\dim \mathbb{F}_r^f(a,b) < \infty$.

Proof If X is compact, there is nothing to prove since $H_r(X)$ has finite dimension. Suppose X is not compact. In view of Observation 3.1(1), it suffices to check the statement for $a > b$. If f is weakly tame, in view of Observation 2.1 X_a, X^b and

[3] PL stands for piecewise linear.

$X_a \cap X^b$ are ANRs, with $X_a \cap X^b$ compact and $X = X_a \cup X^b$, hence the Mayer–Vietoris long exact sequence in homology is valid. Denote by $i_a(r)\colon H_r(X_a) \to H_r(X)$ and $i^b(r)\colon H_r(X^b) \to H_r(X)$ the inclusion-induced linear maps and observe that $\mathbb{F}_r(a,b) := \mathbb{I}_a \cap \mathbb{I}^b \subseteq i_a(r)(\ker(i_a(r) - i^b(r)))$. In view of the Mayer–Vietoris sequence in homology, $\ker(i_a(r) - i^b(r))$ is isomorphic to a quotient of the vector space of $H_r(X_a \cap X^b)$, hence of finite dimension, and the result holds.

If f is not weakly tame, one argue as follows. It is known that any X a locally compact ANR is proper homotopy dominated with respect to any open cover by some locally finite simplicial complex K; see [1].[4] Choose such a cover, for example $f^{-1}(n-1, n+1)_{n\in\mathbb{Z}}$ and such a homotopy domination $X \xrightarrow{i} K \xrightarrow{\pi} X$ for this cover. Choose $g\colon K \to \mathbb{R}$ a proper simplicial approximation of $f \cdot \pi$ (hence tame) and $a' > a$ and $b' < b$ such that $i(X_a^f) \subset K_{a'}^g$ and $i(X_f^b) \subset K_g^{b'}$. Then $\mathbb{F}_r^f(a,b)$ is isomorphic to a subspace of $\mathbb{F}_r^g(a',b')$. Since the dimension of $\mathbb{F}_r^g(a',b')$ is finite, so is the dimension of $\mathbb{F}_r^f(a,b)$. □

Definition 3.3 We say a real number t is a *homologically regular value* if there exists $\epsilon(t) > 0$ such that for any $0 < \epsilon < \epsilon(t)$ the inclusions $\mathbb{I}_{t-\epsilon}^f(r) \subseteq \mathbb{I}_t^f(r) \subseteq \mathbb{I}_{t+\epsilon}^f(r)$ and $\mathbb{I}_f^{t-\epsilon}(r) \supseteq \mathbb{I}_f^t(r) \supseteq \mathbb{I}_f^{t+\epsilon}(r)$ are equalities, and a *homologically critical value* if it is not a homologically regular value.

Denote by $\mathrm{CR}(f)$ the set of all homologically critical values. If f is weakly tame then $\mathrm{CR}(f) \subseteq \mathrm{Cr}(f)$.

Proposition 3.4 *If $f\colon X \to \mathbb{R}$ is a map (hence X is ANR and f is proper) then $\mathrm{CR}(f)$ is discrete.*

Proof As pointed out above in the proof of Proposition 3.2, one can find a proper simplicial map $g\colon K \to \mathbb{R}$ and a proper homotopy domination $\alpha\colon K \to X$ such that $|f \cdot \alpha - g| < M$. If so, for any $a < b$ with $a, b \in \mathbb{R}$, one has $\dim(\mathbb{I}_b^f(r)/\mathbb{I}_a^f(r)) \le \dim(\mathbb{I}_{b+M}^g(r)/\mathbb{I}_{a-M}^g(r)) \le \dim(H_r(g^{-1}([a-M, b+M]), g^{-1}(a-M)) < \infty$, which implies that there are only finitely many changes in $\mathbb{I}_t^f(r)$ for t with $a \le t \le b$, Similar arguments show that there are only finitely many changes of $\mathbb{I}_f^t(r)$ for t with $a \le t \le b$. This suffices to have $\mathrm{CR}(f) \cap [a,b]$ a finite set for any $a < b$, hence $\mathrm{CR}(f)$ discrete. □

Definition 3.5 Define $\tilde{\epsilon}(f) := \inf |c' - c''|$ where $c', c'' \in \mathrm{CR}(f)$ and $c' \neq c''$, and call f *homologically tame* (with respect to κ) if $\tilde{\epsilon}(f) > 0$.

Clearly tame maps are homologically tame with respect to any field κ, and $\tilde{\epsilon}(f) > \epsilon(f)$.

[4] As a replacement for an argument based on an incorrect reference, the above argument and reference were proposed by the referee.

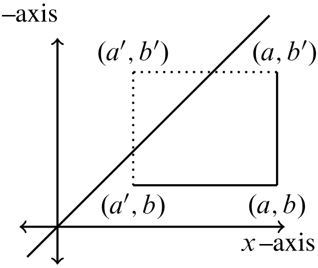


Figure 1: The box $B := (a', a] \times [b, b') \subset \mathbb{R}^2$

Consider the sets of the form $B = (a', a] \times [b, b')$ with $a' < a, b < b'$ and refer to B as a *box*; see Figure 1.

To a box B we assign the quotient of subspaces

$$\mathbb{F}_r^f(B) := \mathbb{F}_r^f(a, b)/(\mathbb{F}_r^f(a', b) + \mathbb{F}_r^f(a, b')),$$

and define

$$F_r^f(a, b) := \dim \mathbb{F}_r^f(a, b), \quad F_r^f(B) := \dim \mathbb{F}_r^f(B).$$

In view of Observation 3.1(2), one has

$$F_r^f(B) := F_r^f(a, b) + F^f(a', b') - F_r^f(a', b) - F^f(a, b').$$

It will also be convenient to define

$$(\mathbb{F}_r^f)'(B) := \mathbb{F}_r^f(a', b) + \mathbb{F}_r^f(a, b') \subseteq \mathbb{F}_r^f(a, b),$$

in which case

$$\mathbb{F}_r^f(B) = \mathbb{F}_r^f(a, b)/(\mathbb{F}_r^f)'(B).$$

We denote by $\pi^B_{ab,r}$ the obvious projection

$$\pi^B_{ab,r} \colon \mathbb{F}_r^f(a, b) \to \mathbb{F}_r^f(B). \tag{1}$$

To ease the writing, when no risk of ambiguity, one drops f from the notation.

If $\kappa = \mathbb{R}$ or $\mathbb{C}$ and $H_r(X)$ is equipped with an inner product (nondegenerate positive definite hermitian scalar product), one denotes by $\boldsymbol{H}_r(B)$ the orthogonal complement of $\mathbb{F}_r'(B) = (\mathbb{F}_r(a', b) + \mathbb{F}(a, b'))$ inside $\mathbb{F}_r(a, b)$, which is a finite-dimensional Hilbert space, and one has

$$\boldsymbol{H}_r(B) \subseteq \mathbb{F}_r(a, b) \subseteq H_r(X).$$

Proposition 3.6 *Let $a'' < a' < a$, $b < b'$ and B_1, B_2 and B the boxes $B_1 = (a'', a'] \times [b, b'')$, $B_2 = (a', a] \times [b, b')$ and $B = (a'', a] \times [b, b')$; see Figure 2 (left).*

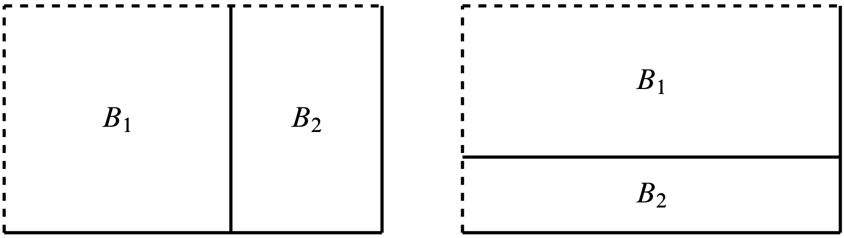


Figure 2

(a) *The inclusions* $B_1 \subset B$ *and* $B_2 \subset B$ *induce the linear maps*

$$i^B_{B_1,r}\colon \mathbb{F}_r(B_1) \to \mathbb{F}_r(B), \tag{2}$$

$$\pi^{B_2}_{B,r}\colon \mathbb{F}_r(B) \to \mathbb{F}_r(B_2) \tag{3}$$

such that the following sequence is exact:

$$0 \to \mathbb{F}_r(B_1) \xrightarrow{i^B_{B_1,r}} \mathbb{F}_r(B) \xrightarrow{\pi^{B_2}_{B,r}} \mathbb{F}_r(B_2) \to 0.$$

(b) *If* $H_r(X)$ *is equipped with a scalar product then*

$$\boldsymbol{H}_r(B_1) \perp \boldsymbol{H}_r(B_2) \quad \textit{and} \quad \boldsymbol{H}_r(B) = \boldsymbol{H}_r(B_1) \oplus \boldsymbol{H}_r(B_2).$$

Proposition 3.7 *Let* $a' < a$, $b < b' < b''$ *and* B_1, B_2 *and* B *the boxes* $B_1 = (a', a] \times [b', b'')$, $B_2 = (a', a] \times [b, b')$ *and* $B = (a', a] \times [b, b'')$; *see Figure 2 (right).*

(a) *The inclusions* $B_1 \subset B$ *and* $B_2 \subset B$ *induce the linear maps*

$$i^B_{B_1,r}\colon \mathbb{F}_r(B_1) \to \mathbb{F}_r(B), \tag{4}$$

$$\pi^{B_2}_{B,r}\colon \mathbb{F}_r(B) \to \mathbb{F}_r(B_2) \tag{5}$$

such that the following sequence is exact:

$$0 \to \mathbb{F}_r(B_1) \xrightarrow{i^B_{B_1,r}} \mathbb{F}_r(B) \xrightarrow{\pi^{B_2}_{B,r}} \mathbb{F}_r(B_2) \to 0.$$

(b) *If* $\kappa = \mathbb{R}$ *or* $\mathbb{C}$ *and* $H_r(X)$ *is equipped with a scalar product then*

$$\boldsymbol{H}_r(B_1) \perp \boldsymbol{H}_r(B_2) \quad \textit{and} \quad \boldsymbol{H}_r(B) = \boldsymbol{H}_r(B_1) \oplus \boldsymbol{H}_r(B_2).$$

Proof Item (a) in both Propositions 3.6 and 3.7 follows from Observation 3.1(1) and (2). To conclude item (b) note that $\boldsymbol{H}_r(B_2)$ as a subspace of $\mathbb{F}_r(a'', b)$ in Proposition 3.6 and as a subspace of $\mathbb{F}_r(a, b'')$ in Proposition 3.7 is orthogonal to a subspace which contains $\boldsymbol{H}_r(B_1)$. $\square$

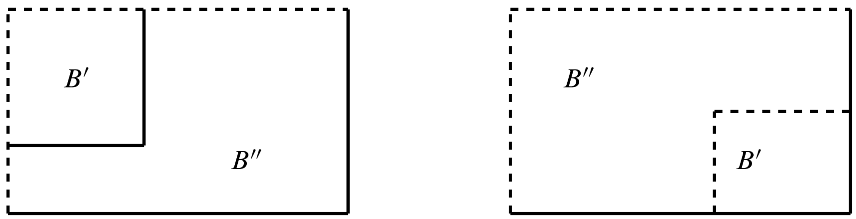


Figure 3

In view of Propositions 3.6 and 3.7, one has the following observation.

Observation 3.8 (1) If B' and B'' are two boxes with $B' \subseteq B''$ and B' is located in the upper left corner of B'' (see Figure 3 (left)) then the inclusion induces the canonical injective linear maps $i^{B''}_{B',r} : \mathbb{F}_r(B') \to \mathbb{F}_r(B'')$.

(2) If B' and B'' are two boxes with $B' \subseteq B''$ and B' is located in the lower right corner of B'' (see Figure 3 (right)) then the inclusion induces the canonical surjective linear maps $\pi^{B'}_{B'',r} : \mathbb{F}_r(B'') \to \mathbb{F}_r(B')$.

(3) If B is a finite disjoint union of boxes $B = \bigsqcup B_i$ then $\mathbb{F}_r(B)$ is isomorphic to $\bigoplus_i \mathbb{F}_r(B_i)$; the isomorphism is not canonical.

(4) If in addition $\kappa = \mathbb{R}$ or $\mathbb{C}$ and $H_r(X)$ is a Hilbert space then $\boldsymbol{H}_r(B) = \bigoplus_i \boldsymbol{H}_r(B_i)$.

In view of this observation, define $B(a,b:\epsilon) = (a-\epsilon, a] \times [b, b+\epsilon)$ and

$$\hat{\delta}^f_r(a,b) := \varinjlim_{\epsilon \to 0} \mathbb{F}_r(B(a,b;\epsilon)).$$

The limit refers to the direct system $\mathbb{F}_r(B(a,b;\epsilon')) \to \mathbb{F}_r(B(a,b;\epsilon''))$ whose arrows are the surjective linear maps induced by the inclusion of $B(a,b;\epsilon')$ as the lower right corner of $B(a,b;\epsilon'')$ for $\epsilon' < \epsilon''$.

Define also

$$\delta^f_r(a,b) := \lim_{\epsilon \to 0} F_r(B(a,b;\epsilon)).$$

Clearly one has $\dim \hat{\delta}^f_r(a,b) = \delta^f_r(a,b)$. Denote by $\operatorname{supp} \delta^f_r$ the set

$$\operatorname{supp} \delta^f_r := \{(a,b) \in \mathbb{R}^2 \mid \delta^f_r(a,b) \neq 0\}.$$

Observation 3.9 For any (a,b), $a,b \in \mathbb{R}$, the direct system stabilizes and $\hat{\delta}^f_r(a,b) = \mathbb{F}^f(B(a,b;\epsilon))$ for some ϵ small enough. Moreover $\delta^f_r(a,b) \neq 0$ implies that $a,b \in \mathrm{CR}(f)$. In particular $\operatorname{supp} \delta^f_r$ is a discrete subset of $\mathbb{R}^2$. If f is homologically tame then for any (a,b) with $a,b\in \mathrm{CR}(f)$, we have $\hat{\delta}^f_r(a,b)=\mathbb{F}^f(B(a,b;\epsilon))$ for any ϵ, $0<\epsilon<\tilde{\epsilon}(f)$.

Recall that for a box $B=(a',a]\times[b,b')$, we have denoted the canonical projection on $\mathbb{F}_r(B)=\mathbb{F}(a,b)/\mathbb{F}'(B)$ by $\pi^B_{ab,r}\colon \mathbb{F}_r(a,b)\to\mathbb{F}_r(B)$, and for $B'=(a'',a]\times[b,b'')$, $a''\le a'<a$, $b''\ge b'>b$, we have denoted by $\pi^B_{B',r}\colon \mathbb{F}_r(B')\to\mathbb{F}_r(B)$ the canonical surjective linear map between quotient spaces induced by $\mathbb{F}'(B')\subseteq\mathbb{F}'(B)\subseteq\mathbb{F}(a,b)$. Clearly

$$\pi^B_{ab,r}=\pi^B_{B',r}\cdot\pi^{B'}_{ab,r}.$$

Consider the surjective linear map

$$\pi_r(a,b)\colon \mathbb{F}(a,b)\to \varinjlim_{\epsilon\to 0}\mathbb{F}(B(a,b;\epsilon))=\hat{\delta}^f_r(a,b),$$
$$\pi_r(a,b):=\varinjlim_{\epsilon\to 0}\pi^{B(a,b;\epsilon)}_{ab,r}.$$

Definition 3.10 A *special splitting* is a linear map

$$s_r(a,b)\colon \hat{\delta}^f_r(a,b)\to\mathbb{F}_r(a,b)$$

which satisfies $\pi_r(a,b)\cdot s_r(a,b)=\mathrm{id}$. In particular, in view of Observation 3.1, for any $\alpha>a$ and $\beta<b$, we have $\mathrm{img}(s_r(a,b))\subseteq\mathbb{F}_r(\alpha,\beta)$.

We denote by $i_r(a,b)$ the composition of $s_r(a,b)$ with the inclusion $\mathbb{F}_r(a,b)\subseteq H_r(X)$.

The diagram

$$\begin{tikzcd}
H_r(X) & \mathbb{F}_r(a,b) \arrow[l, "\supseteq"'] \arrow[r, "\pi_r(a,b)"'] \arrow[d, "\pi^B_{ab,r}"'] & \hat{\delta}^{\tilde f}_r(a,b) \arrow[l, bend right, "s_r(a,b)"'] \arrow[ll, bend right, "i_r(a,b)"'] \arrow[dl, "i^B_r(a,b)"] \\
\mathbb{F}_r(B_1) \arrow[r, "i^B_{B',r}"'] & \mathbb{F}_r(B) \arrow[r, "\pi^{B_2}_{B,r}"'] & \mathbb{F}_r(B_2)
\end{tikzcd} \tag{6}$$

reviews for the reader the linear maps considered so far. In this diagram suppose $B=(\alpha',\alpha]\times[\beta,\beta')$ with $a\in(\alpha',\alpha]$ and $b\in[\beta,\beta')$ and $B=B_1\sqcup B_2$ as in Figure 2 (left). In view of Observations 3.8 and 3.9, one has the following.

Observation 3.11 (1) If $(a,b)\in B_2$ then $\pi^{B_2}_{B,r}\cdot i^B_r(a,b)$ is injective.

(2) If $(a,b)\in B_1$ then $\pi^{B_2}_{B,r}\cdot i^B_r(a,b)$ is zero.

Choose special splittings $\{s_r(a,b) \mid (a,b) \in \operatorname{supp}(\delta_r^{\tilde f})\}$, and consider the sum

$$I_r = \sum_{(a,b)\in \operatorname{supp}(\delta_r^{\tilde f})} i_r(a,b) : \bigoplus_{(a,b)\in \operatorname{supp}(\delta_r^{\tilde f})} \hat\delta_r^f(a,b) \to H_r(X),$$

and for a finite or infinite box B the sum

$$I_r^B = \sum_{(a,b)\in \operatorname{supp}(\delta_r^{\tilde f})\cap B} i_r^B(a,b) : \bigoplus_{(a,b)\in \operatorname{supp}(\delta_r^{\tilde f})\cap B} \hat\delta_r^f(a,b) \to \mathbb{F}_r(B).$$

For $\Sigma \subseteq \operatorname{supp}(\delta_r^f)$ denote by $I_r(\Sigma)$ the restriction of I_r to $\bigoplus_{(a,b)\in\Sigma} \hat\delta_r^f(a,b)$ and for $\Sigma \subseteq \operatorname{supp}(\delta_r^f) \cap B$ denote by $I_r^B(\Sigma)$ the restriction of I_r^B to $\bigoplus_{(a,b)\in\Sigma} \hat\delta_r^f(a,b)$. Note the following.

Observation 3.12 For $B = B_1 \sqcup B_2$ as in Figure 2 and $\Sigma \subseteq \operatorname{supp} \delta_r^{\tilde f}$ with $\Sigma = \Sigma_1 \sqcup \Sigma_2$, $\Sigma_1 \subseteq B_1$ and $\Sigma_2 \subseteq B_2$, the diagram

$$\begin{tikzcd}
\mathbb{F}_r(B_1) \arrow[r] & \mathbb{F}_r(B) \arrow[r] & \mathbb{F}_r(B_2) \\
\bigoplus_{(a,b)\in\Sigma_1} \hat\delta_r^{\tilde f}(a,b) \arrow[r] \arrow[u, "I_r^{B_1}(\Sigma_1)"'] & \bigoplus_{(a,b)\in\Sigma} \hat\delta_r^{\tilde f}(a,b) \arrow[r] \arrow[u, "I_r^{B}(\Sigma)"'] & \bigoplus_{(a,b)\in\Sigma_2} \hat\delta_r^{\tilde f}(a,b) \arrow[u, "I_r^{B_2}(\Sigma_2)"']
\end{tikzcd}$$

is commutative. In particular if $I_r^{B_1}(\Sigma_1)$ and $I_r^{B_2}(\Sigma_2)$ are injective then so is $I_r^B(\Sigma)$.

If $\kappa = \mathbb{R}$ or $\mathbb{C}$ and $H_r(X)$ is equipped with a Hilbert space structure, then the inverse of the restriction of $\pi_r(a,b)$ to the orthogonal complement of $\ker(\pi_r(a,b))$ provides a *canonical special splitting* $s_r(a,b)$. For these canonical special splittings, one denotes by $\hat{\hat\delta}_r^f$ the assignment

$$\hat{\hat\delta}_r^f(a,b) = \boldsymbol{H}_r(a,b) := \operatorname{img} s_r(a,b).$$

Then if X is compact in view of Observation 3.8(4) the assignment $\hat{\hat\delta}_r^f$ is a configuration $\mathcal{C}^O_{H_r(X)}(\mathbb{R}^2)$. The configuration $\hat{\hat\delta}_r^f(a,b)$ has the configuration $\delta_r^f \in \mathbb{C}^{\dim H_r(X)}$ as its dimension.

Let f be a map, and for any $(a,b) \in \mathbb{R}^2$ choose a special splitting $s_r(a,b)$: $\hat\delta_r^f(a,b) \to H_r(X)$.

Observation 3.13 (1) For any $\Sigma \subseteq \operatorname{supp}(\delta_r^f)$ (resp. $\Sigma \subseteq \operatorname{supp}(\delta_r^f) \cap B$), the linear maps $I_r(\Sigma)$ (resp. $I_r^B(\Sigma)$) are injective.

(2) For any box $B = (a', a] \times [b, b')$ the set $\delta_r^f \cap B$ is finite.

(3) For any box B, the linear map I_r^B is an isomorphism.

(4) If X compact, $m < \inf f$ and $M > \sup f$ then $H_r(X) = \mathbb{F}_r((m, M] \times [m, M))$ and I_r is an isomorphism. Therefore, for any special splittings, the collection of subspaces $\mathrm{img}(i_r(a, b))$ provide a configuration of subspaces of $H_r(X)$ hence and element in $\mathcal{C}_{H_r(X)}(\mathbb{R}^2)$.

Proof (1) If $\Sigma \subset B$ then in view of Observations 3.11 and 3.12, the injectivity of $I_r^B(\Sigma)$ implies the ineffectiveness of $I_r^{B'}(\Sigma)$ for any box $B' \supseteq B$, as well as the injectivity of $I_r(\Sigma)$. To check the injectivity of $I_r^B(\Sigma)$, one proceeds as follows:

- If the cardinality of Σ is one, then the statement follows from Observation 3.11.
- If all elements (α_i, β_i), $i = 1, \dots, k$, of Σ have the same first component $\alpha_i = a$, the statement follows by induction on k. One writes the box $B = B_1 \sqcup B_2$ as in Figure 2 (left) such that B_2 contains one element of Σ, say (α_1, β_1), and B_1 contains the remaining $k-1$ elements. The injectivity follows from Observation 3.12 in view of the injectivity of $I_r^{B_2}(\Sigma \cap B_2)$ and of $I_r^{B_1}(\Sigma \cap B_1)$, assumed by the induction hypothesis.
- In general, one writes Σ as the disjoint union $\Sigma = \Sigma_1 \sqcup \Sigma_2 \sqcup \cdots \sqcup \Sigma_k$ such that each Σ_i contains all points of Σ with the same first component a_i, and $a_k > a_{k-1} > \cdots > a_2 > a_1$. One proceeds again by induction on k. One decomposes the box B as in Figure 2 (right), $B = B_1 \sqcup B_2$ such that $\Sigma_1 \subset B_2$ and $(\Sigma \setminus \Sigma_1) \subset B_1$. The injectivity of $I_r^B(\Sigma)$ follows then using Observation 3.12 from the injectivity of $I_r^{B_2}(\Sigma_1)$ and the induction hypothesis which assumes the injectivity of $I_r^{B_1}(\Sigma \cap B_1)$.

(2) In view of (1), any subset of $\mathrm{supp}(\delta_r^f) \cap B$ with $B = (a', a] \times [b, b')$ has cardinality smaller than $\dim \mathbb{F}_r(a, b)$, which by Proposition 3.2 is finite. Hence Σ is finite.

(3) The injectivity of I_r^B is ensured by (1). The surjectivity follows from the equality of the dimension of the source and of the target implied by Observations 3.8 and 3.9.

(4) This follows from definitions and from (3). □

In case X is not compact, for the needs of part II of this paper it is useful to extend Observation 3.13(3) to the situation of an infinite box $B(a, b; \infty) := (-\infty, a] \times [b, \infty)$, and evaluate the image of I_r, which might not be a finite-dimensional space. For this purpose we introduce the following:

(1) $\mathbb{I}^f_{-\infty}(r) = \bigcap_{a \in \mathbb{R}} \mathbb{I}^f_a(r)$ and $\mathbb{I}^\infty_f(r) = \bigcap_{b \in \mathbb{R}} \mathbb{I}^b_f(r)$,

(2) $\mathbb{F}^f_r(-\infty, b) := \mathbb{I}^f_{-\infty}(r) \cap \mathbb{I}^b_f(r)$ and $\mathbb{F}^f_r(a, \infty) := \mathbb{I}^f_a(r) \cap \mathbb{I}^\infty_f(r)$,

(3) $(\mathbb{F}^f)'_r(B(a, b; \infty)) := \mathbb{F}^f_r(-\infty, b) + \mathbb{F}^f_r(a, \infty)$,

(4) $\mathbb{F}^f_r(B(a, b; \infty)) := \mathbb{F}^f_r(a, b) / (\mathbb{F}^f)'_r(B(a, b; \infty))$.

Observation 3.14 (1) In view of the finite-dimensionality of $\mathbb{F}_r(a,b)$, one has the following:

(i) For any a, there exists $b(a)$ such that
$$\mathbb{F}_r(a,b(a)) = \mathbb{F}_r(a,b') = \mathbb{F}_r(a,\infty)$$
provided that $b' \geq b(a)$.

(ii) For any b, there exists $a(b))$ such that
$$\mathbb{F}_r(-\infty,b) = \mathbb{F}_r(a',b) = \mathbb{F}_r(a(b),b)$$
provided that $a' \leq a(b)$.

(2) In view of (1), for $a' < a(b)$ and $b' > b(a)$, the canonical projections
$$\mathbb{F}_r(B(a,b;\infty)) \to \mathbb{F}_r((a',a]\times[b,b')) \to \mathbb{F}_r((a(b),a]\times[b,b(a)))$$
are isomorphisms.

Observation 3.15 (addendum to Observation 3.13(3)) The maps
$$\bigoplus_{(a',b')\in\operatorname{supp}(\delta^f_r)\cap B(a,b;\infty)} i_r^{B(a,b;\infty)}(a',b')\colon \bigoplus_{(a',b')\in\operatorname{supp}\delta^f_r\cap B(a,b;\infty)} \hat{\delta}^f_r(a',b') \to \mathbb{F}_r(B(a,b;\infty)),$$
$$\bigoplus_{(a,b)\in\operatorname{supp}(\delta^f_r)} i_r(a,b)\colon \bigoplus_{(a,b)\in\operatorname{supp}(\delta^f_r)} \hat{\delta}^f_r(a,b) \to H_r(X)/(\mathbb{I}^f_{-\infty}(r)+\mathbb{I}^\infty_f(r))$$
are isomorphisms.

Proof The first isomorphism follows from Observations 3.13 and 3.14.

For the second, note that for $k<k'$ (for simplicity in writing we drop f and r from the notation)
$$(\mathbb{I}_{-\infty}\cap\mathbb{I}^{-k'}+\mathbb{I}_{k'}\cap\mathbb{I}^{\infty})\cap\mathbb{I}^{-k}\cap\mathbb{I}_k = \mathbb{I}_{-\infty}\cap\mathbb{I}^{-k}+\mathbb{I}_k\cap\mathbb{I}^{\infty}$$
and that
$$H_r(X) = \varinjlim_{k\to\infty}\mathbb{F}_r(k,-k) = \varinjlim_{k\to\infty}\mathbb{I}^{-k} = \varinjlim_{k\to\infty}\mathbb{I}_k.$$
Then in view of stabilization properties,
$$\varinjlim \frac{\mathbb{F}(k,-k)}{\mathbb{I}_{-\infty}\cap\mathbb{I}^{-k}+\mathbb{I}_k\cap\mathbb{I}^{\infty}} = \frac{H_r(X)}{\mathbb{I}_{-\infty}+\mathbb{I}^{\infty}}. \qquad \square$$

Let $D(a,b;\epsilon) := (a-\epsilon,a+\epsilon]\times[b-\epsilon,b+\epsilon)$. If $x=(a,b)$, one also writes $D(x;\epsilon)$ for $D(a,b;\epsilon)$.

Proposition 3.16 (see [3, Proposition 5.6]) *Let $f\colon X \to \mathbb{R}$ be a tame map and $\epsilon < \epsilon(f)/3$. For any map $g\colon X \to \mathbb{R}$ which satisfies $\|f-g\|_\infty < \epsilon$ and $a, b \in \mathrm{Cr}(f)$ critical values, one has*

$$\sum_{x \in D(a,b;2\epsilon)} \delta^g_r(x) = \delta^f_r(a,b), \tag{7}$$

$$\operatorname{supp} \delta^g_r \subset \bigcup_{(a,b) \in \operatorname{supp} \delta^f_r} D(a,b;2\epsilon). \tag{8}$$

If in addition $H_r(X)$ is equipped with a Hilbert space structure ($\kappa = \mathbb{R}$ or $\mathbb{C}$), the above statement can be strengthened to

$$x \in D(a,b;2\epsilon) \Rightarrow \hat{\delta}^g_r(x) \subseteq \hat{\delta}^f_r(a,b), \qquad \bigoplus_{x \in D(a,b;2\epsilon)} \hat{\delta}^g_r(x) = \hat{\delta}^f_r(a,b). \tag{9}$$

Proposition 3.16 implies that in an ϵ–neighborhood of a tame map f (with respect to the $\|\cdot\|_\infty$ norm) any other map g has the support of δ^g_r in a 2ϵ–neighborhood of the support of δ^f_r and in case X compact is of cardinality counted with multiplicities equal to $\dim H_r(X)$.

Proof of Proposition 3.16 See [3]. Consider a collection of real numbers

$$C := \{\cdots < c_i < c_{i+1} < c_{i+2} < \cdots \mid i \in \mathbb{Z}\}$$

which satisfies the following properties:

(1) $\mathrm{Cr}(f) \subseteq C$,

(2) $c_{i+1} - c_i > \epsilon(f)$,

(3) $\lim_{i\to\infty} c_i = \infty$,

(4) $\lim_{i\to-\infty} c_i = -\infty$.

Next, one establishes two intermediate results.

Lemma 3.17 *For f as in Proposition 3.16 and $c_i, c_j \in C$, one has*

$$\begin{aligned}\hat{\delta}^f_r(c_i,c_j) &= \mathbb{F}^f_r((c_{i-1},c_i] \times [c_j,c_{j+1})) \\ &= \mathbb{F}^f_r(c_i,c_j)/\mathbb{F}^f_r(c_{i-1},c_j) + \mathbb{F}^f_r(c_i,c_{j+1}),\end{aligned} \tag{10}$$

and therefore

$$\begin{aligned}\delta^f_r(c_i,c_j) &= F^f_r((c_{i-1},c_i] \times [c_j,c_{j+1})) \\ &= F^f_r(c_{i-1},c_{j+1}) + F^f_r(c_i,c_j) - F^f_r(c_{i-1},c_j) - F^f_r(c_i,c_{j+1}).\end{aligned} \tag{11}$$

Proof It is known (see [7], for example) that X a closed subset of Y and X, Y ANRs implies that X is a neighborhood deformation retract [7]. Then in view of the tameness of f, for any $\epsilon', \epsilon'' \in (0, \epsilon(f))$ one has

$$\begin{aligned} &\mathbb{F}_r^f(c_i, c_j) = \mathbb{F}_r^f(c_i + \epsilon', c_j) = \mathbb{F}_r^f(c_{i+1} - \epsilon'', c_j) = \mathbb{F}_r^f(c_{i+1} - \epsilon'', c_{j-1} + \epsilon''), \\ &\mathbb{F}_r^f(c_i, c_j) = \mathbb{F}_r^f(c_i, c_j - \epsilon') = \mathbb{F}_r^f(c_i, c_{j-1} + \epsilon'') = \mathbb{F}_r^f(c_{i+1} - \epsilon', c_{j-1} + \epsilon''). \end{aligned} \tag{12}$$

Since $\epsilon < \epsilon(f)$, in view of the definition of $\hat{\delta}_r^f$ one has

$$\begin{aligned} \hat{\delta}_r^f(c_i, c_j) &= \mathbb{F}_r^f((c_i - \epsilon, c_i] \times [c_j, c_j + \epsilon)) \\ &= \mathbb{F}_r^f(c_i, c_j) / \mathbb{F}_r^f(c_i - \epsilon, c_j) + \mathbb{F}_r^f(c_i, c_j + \epsilon). \end{aligned} \tag{13}$$

Combining (13) with (12) one obtains the equality (10):

$$\hat{\delta}_r^f(c_i, c_j) = \mathbb{F}_r^f(c_i, c_j) / \mathbb{F}_r^f(c_{i-1}, c_j) + \mathbb{F}_r^f(c_i, c_{j+1}).$$

Since $\mathbb{F}^f(c_{i-1}, c_j) \cap \mathbb{F}^f(c_i, c_{j+1}) = \mathbb{F}^f(c_{i-1}, c_{j+1})$ one has

$$\begin{aligned} &\dim(\mathbb{F}_r^f(c_{i-1}, c_j) + \mathbb{F}_r^f(c_i, c_{j+1})) \\ &\qquad = \dim \mathbb{F}_r^f(c_{i-1}, c_j) + \dim \mathbb{F}_r^f(c_i, c_{j+1} - \dim \mathbb{F}^f(c_{i-1}, c_{j+1}) \end{aligned}$$

and the equality (11) follows. □

To simplify the notation, the index r in the following lemma will be dropped.

Lemma 3.18 *Suppose f is tame. Let $a = c_i$, $b = c_j$, $c_i, c_j \in C$ and $\epsilon < \epsilon(f)/3$. If g is a continuous map with $\|f - g\|_\infty < \epsilon$, then*

$$\begin{aligned} \mathbb{F}_r^g(a - 2\epsilon, b + 2\epsilon) &= \mathbb{F}_r^f(c_{i-1}, c_{j+1}), \\ \mathbb{F}_r^g(a + 2\epsilon, b - 2\epsilon) &= \mathbb{F}_r^f(c_i, c_j), \\ \mathbb{F}_r^g(a + 2\epsilon, b + 2\epsilon) &= \mathbb{F}_r^f(c_i, c_{j+1}), \\ \mathbb{F}_r^g(a - 2\epsilon, b - 2\epsilon) &= \mathbb{F}_r^f(c_{i-1}, c_j). \end{aligned} \tag{14}$$

Proof Since $\|f - g\|_\infty < \epsilon$, in view of Observation 3.1(3) one has

$$\begin{aligned} \mathbb{F}_r^f(a - 3\epsilon, b + 3\epsilon) \subseteq \mathbb{F}_r^g(a - 2\epsilon, b + 2\epsilon) &\subseteq \mathbb{F}_r^f(a - \epsilon, b + \epsilon), \\ \mathbb{F}_r^f(a + \epsilon, b - \epsilon) \subseteq \mathbb{F}_r^g(a + 2\epsilon, b - 2\epsilon) &\subseteq \mathbb{F}_r^f(a + 3\epsilon, b - 3\epsilon), \\ \mathbb{F}_r^f(a + \epsilon, b + 3\epsilon) \subseteq \mathbb{F}_r^g(a + 2\epsilon, b + 2\epsilon) &\subseteq \mathbb{F}_r^f(a + 3\epsilon, b + \epsilon), \\ \mathbb{F}_r^f(a - 3\epsilon, b - \epsilon) \subseteq \mathbb{F}_r^g(a - 2\epsilon, b - 2\epsilon) &\subseteq \mathbb{F}_r^f(a - \epsilon, b - 3\epsilon). \end{aligned} \tag{15}$$

Since $3\epsilon < \epsilon(f)$, one has

$$\begin{aligned} \mathbb{F}^f(a-3\epsilon, b+3\epsilon) &= \mathbb{F}^f(a-\epsilon, b+\epsilon),\\ \mathbb{F}^f(a+\epsilon, b-\epsilon) &= \mathbb{F}^f(a+3\epsilon, b-3\epsilon),\\ \mathbb{F}^f(a+\epsilon, b+3\epsilon) &= \mathbb{F}^f(a+3\epsilon, b+\epsilon),\\ \mathbb{F}^f(a-3\epsilon, b-\epsilon) &= \mathbb{F}^f(a-\epsilon, b-3\epsilon), \end{aligned} \tag{16}$$

which imply that in (15) the inclusion $\subseteq$ is actually equality.

Note that in view of the equalities (12), for $\epsilon', \epsilon'' < \epsilon(f)$ one has

$$\begin{aligned} \mathbb{F}^f(c_{i-1}, c_{j+1}) &= \mathbb{F}^f(a-\epsilon', b+\epsilon''),\\ \mathbb{F}^f(c_i, c_j) &= \mathbb{F}^f(a+\epsilon', b-\epsilon''),\\ \mathbb{F}^f(c_i, c_{j+1}) &= \mathbb{F}^f(a+\epsilon', b+\epsilon''),\\ \mathbb{F}^f(c_{i-1}, c_j) &= \mathbb{F}^f(a-\epsilon', b-\epsilon''). \end{aligned} \tag{17}$$

Then (15) and (17) imply (14) and hence the statement of Lemma 3.18. $\square$

Next observe that Lemma 3.18 gives (for $a = c_i$, $b = c_j$ with $c_i, c_j \in C$) the equality

$$\mathbb{F}^g((a-2\epsilon, a+2\epsilon] \times [b-2\epsilon, b+2\epsilon)) = \mathbb{F}^f((c_{i-1}, c_i] \times [c_j, c_{j+1})).$$

This combined with Lemma 3.17 implies $\mathbb{F}^g((a-2\epsilon,a+2\epsilon]\times[b-2\epsilon,b+2\epsilon)) = \hat{\delta}^f(a,b)$, which combined with Observation 3.13 implies the inclusion (7) and the equality (9), not only for critical values but for any $a, b \in C$.

To check inclusion (8) observe the following:

(a) $\|f-g\|_\infty < \epsilon$ implies $X_a^f \subseteq X_{a+\epsilon}^g \subseteq X_{a+2\epsilon}^f$ and $X_f^b \subseteq X_g^{b-\epsilon} \subseteq X_f^{b-2\epsilon}$, and when $a, b \in C$,

$$\mathbb{F}^f(a,b) \subseteq \mathbb{F}^g(a+\epsilon, b-\epsilon) \subseteq \mathbb{F}^f(a+2\epsilon, b-2\epsilon). \tag{18}$$

(b) When $\epsilon < \epsilon(f)/3$, the inclusions (18) imply

$$\mathbb{F}^f(a,b) = \mathbb{F}^g(a+\epsilon, b-\epsilon) = \mathbb{F}^f(a+2\epsilon, b-2\epsilon)$$

which in view of Observation 3.15 implies

$$\begin{aligned} \sum_{x\in(-\infty,a]\times(b,\infty)\cap \operatorname{supp}\delta_r^f} \delta_r^f(x) &= \sum_{y\in(-\infty,a+\epsilon]\times(b-\epsilon,\infty)\cap \operatorname{supp}\delta_r^g} \delta_r^g(y)\\ &= \sum_{x\in(-\infty,a+2\epsilon]\times(b-2\epsilon,\infty)\cap \operatorname{supp}\delta_r^f} \delta_r^f(x). \end{aligned} \tag{19}$$

Since $\mathbb{R}^2 = \bigcup_{i\in\mathbb{Z}} B(c_i, c_{-i}; \infty)$, (19) and (7) rule out the existence of an $x \in \operatorname{supp}(\delta^g_r)$ away from $\bigcup_{x\in\operatorname{supp}(\delta^f_r)} D(x; 2\epsilon)$, finishing the proof of Proposition 3.16. $\square$

Let K be a compact ANR and $f\colon X \to \mathbb{R}$ be a map. Denote by

$$\bar{f}_K; X \times K \to \mathbb{R}$$

the composition $f \cdot \pi_K$ with $\pi_K\colon X \times K \to X$ the first factor projection. If f is weakly tame then so is $\bar{f}_K$ and the set of critical values of f and of $\bar{f}_K$ are the same. Moreover in view of the Künneth theorem about the homology of the cartesian product of two spaces one can make the following observation.

Observation 3.19 (1) $\mathbb{F}^{\bar{f}_K}_r(a,b) = \bigoplus_{0\le k\le r} \mathbb{F}^f_k(a,b) \otimes H_{r-k}(K)$, and therefore

(2) $\hat{\delta}^{\bar{f}_K}_r(a,b) = \bigoplus_{0\le k\le r} \hat{\delta}^f_k(a,b) \otimes H_{r-k}(K)$, and

(3) $\hat{\delta}^{\bar{f}_K}_r(a,b) = \hat{\delta}^f_k(a,b)$ when K is acyclic.

Note that the embedding $I\colon C(X;\mathbb{R}) \to C(X \times K;\mathbb{R})$ defined by $I(f) = \bar{f}_K$ is an isometry when both spaces are equipped with the distance $\|\cdot\|_\infty$. Note also that when K is acyclic one has $\delta^f_r = \delta^{I(f)}_r$ and $\hat{\delta}^f_r = \hat{\delta}^{I(f)}_r$ provided that $H_r(X)$ is identified with $H_r(X \times K)$.

4 The main results

Theorem 4.1 (topological results) *Suppose X is compact and $f\colon X \to \mathbb{R}$ a map.*[5]

(1) $\delta^f_r(x) \ne 0$ *with* $x = (a,b)$ *implies that both* $a, b \in \mathrm{CR}(f)$.

(2) $\sum_{x\in\mathbb{R}^2} \delta^f_r(x) = \dim H_r(X)$ *and* $\bigoplus_{x\in\mathbb{R}^2} \hat{\delta}^f_r(x) = H_r(X)$. *In particular, we have* $\delta^f_r \in \mathcal{C}_{\dim H_r(X)}(\mathbb{R}^2)$.

(3) *If $H_r(X)$ is equipped with a Hilbert space structure then* $\hat{\hat{\delta}}^f \in \mathcal{C}^O_{H_r(X)}(\mathbb{R}^2)$.

(4) *If X is homeomorphic to a finite simplicial complex or a compact Hilbert cube manifold then for an open and dense set of maps f in the space of continuous maps with compact open topology,* $\delta^f_r(x) = 0$ *or* 1.

Statements (1) and (3) formulated in terms of bar codes (see [2]) were verified first in [3] under the hypothesis that f is a tame map.

[5] This means X is also ANR and f continuous.

Theorem 4.2 (stability) *Suppose X is a compact ANR.*

(1) *The assignment $f \rightsquigarrow \delta^f_r$ provides a continuous map from the space of real-valued maps $C(X;\mathbb{R})$ equipped with the compact open topology to the space of configurations $\mathcal{C}_{b_r}(\mathbb{R}^2)=\mathbb{C}^{b_r}$ and $b_r=\dim H_r(X)$, equipped with the collision topology (also regarded as the space of monic polynomials of degree b_r). Moreover, with respect to the canonical metric $\underline{D}$ on the space of configurations, which induces the collision topology, one has*

$$\underline{D}(\delta^f,\delta^g)<2D(f,g).$$

Recall that $D(f,g):=\|f-g\|_\infty=\sup_{x\in X}|f(x)-g(x)|$.

(2) *If $\kappa=\mathbb{R}$ or $\mathbb{C}$ then the assignment $f \rightsquigarrow \hat{\hat{\delta}}{}^f_r$ is continuous with respect to both collision topologies. (The continuity with respect to the first implies that with respect to the second.)*

Theorem 4.2(1) was first established in [3] under the hypothesis X homeomorphic to a finite simplicial complex.

Theorem 4.3 (Poincaré duality) (1) *Suppose X is a closed smooth κ–orientable manifold*[6] *of dimension n, and f a continuous map. Then $\delta^f_r(a,b)=\delta^f_{n-r}(b,a)$.*

(2) *In addition any collection of isomorphisms $H_r(X)\to H_r(X)^*$ induce the isomorphisms of the configuration $\hat{\delta}^f_r$ and $\hat{\delta}^f_{n-r}\cdot\tau$ with $\tau(a,b)=(b,a)$.*

Item (1) of the above theorem was established in [3] for f a tame map.

4.1 Proof of Theorem 4.1

Items (1)–(3) are contained in Observation 3.13 and Observation 3.9.

We first prove item (4). In view of Theorem 4.2, whose proof does not involve Theorem 4.1, it suffices to establish only the density in the space of all continuous functions of tame maps f with δ^f_r taking values only 0 and 1.

We say that a tame map $f\colon X\to\mathbb{R}$ satisfies Property G if the following holds.

Property G There exists a finite sequence of real numbers

$$a=a_0<a_1<\cdots<a_n<a_{n+1}=b$$

[6]The results probably remain true as stated for topological manifolds based essentially on the same arguments, but being unable to find appropriate references we formulate them under the hypothesis of smoothness.

such that

(1) $\mathbb{I}_a^f(r) = 0$ and $\mathbb{I}_b^f(r) = H_r(X)$,

(2) for any $i \geq 1$, $\dim(\mathbb{I}_{a_i}^f / \mathbb{I}_{a_{i-1}}^f) \leq 1$.

The verification of Theorem 4.1(4) is based on the Observations 4.4 and 4.5.

Observation 4.4 For any tame map f which satisfies Property G, the configuration δ_r^f takes only the values 0 and 1.

If f has Property G then it satisfies $\dim(\mathbb{I}_{a_i}^f / \mathbb{I}_{a_{i-1}}^f) \leq 1$ for $a_i = c_i, i = 1, \dots, n$; since for $\alpha < \beta$ with no critical value in the open interval (α, β) and β a regular value, the inclusion $X_\alpha^f \subset X_\beta^f$ induces isomorphism in homology and for any $a' \leq a \leq b \leq b'$, $\dim(\mathbb{I}_b^f(r)/\mathbb{I}_a^f(r)) \leq \dim(\mathbb{I}_{b'}^f(r)/\mathbb{I}_{a'}^f(r))$.

If so, then for any two consecutive critical values $c_{i-1} < c_i$ and any other critical value c_j, the inclusion $\mathbb{F}_r(c_{i-1}, c_j) \subseteq \mathbb{F}_r(c_i, c_j)$ has cokernel of dimension at most one, which by (10) in Lemma 3.17 implies that δ_r^f takes only the values 0 and 1. Based on this observation, if X is a compact smooth manifold (possibly with boundary), any Morse function $f\colon X \to \mathbb{R}$ which takes different values of different critical points has Property G.

Indeed if $\{\cdots < c_i < c_{i+1} < \cdots\}$ is the collection of all critical values, $X_{c_{i+1}}^f$ is homotopy equivalent to a space obtained from $X_{c_i}^f$ by adding a closed disk D^k along $\partial D^k = S^{k-1}$ or $\partial D_+^k = D^{k-1}$, which insures that Property G is satisfied. Since the set of such Morse functions is dense in the space of all continuous functions equipped with the C_0–topology, item (4) is verified (once Theorem 4.2 is established).

If X is a compact Hilbert cube manifold, then is homeomorphic to $M \times Q$ with M a compact smooth manifold (possible with boundary), and any continuous map $f\colon X \to \mathbb{R}$ is arbitrarily closed to $\bar{f}_Q$, with $f\colon M \to \mathbb{R}$ a Morse function. This observation establishes item (4) for compact Hilbert cube manifolds.

If X is a finite simplicial complex, one needs the following observation.

Observation 4.5 If X is a finite simplicial complex and $a < b$, one can construct a map $h\colon X \to \mathbb{R}$ simplicial on the barycentric subdivision of X with the following properties:

(1) $a < h(x) < b$;

(2) h takes different values on the barycenters of different simplices;

(3) the value of h on the barycenter of a simplex σ is strictly larger than the values of h on the barycenter of any of its faces.

Proof The construction is straightforward. Such a map satisfies Property G, since adding a simplex to a finite simplicial complex might change the dimension of the homology with at most one unit, and for any α, X^h_α retracts by deformation to the simplicial complex generated by the barycenters on which h takes value smaller or equal to α.

For $f\colon X\to\mathbb{R}$ a simplicial map, X a finite simplicial complex with critical values $\{\cdots<c_{i-1}<c_i<\cdots\}$, if for some i we have $\dim(\mathbb{I}^f_{c_i}/\mathbb{I}^f_{c_{i-1}})\geq 2$, one chooses $\epsilon<\epsilon(f)/2$ and a subdivision of X which makes $f^{-1}(c_i\pm\epsilon/2)$ and $f^{-1}(c_i))$, and thus $f^{-1}([c_i-\epsilon/2,c_i+\epsilon/2])$ and $f^{-1}([c_i,c_i+\epsilon])$, subcomplexes. One takes the barycentric subdivision of this subdivision and replaces f by g, the simplicial map for the new triangulation. We define the map g to take the same value as f on the barycenters of simplices not contained in $f^{-1}(c_i)$, and as h constructed using Observation 4.5 for $a=c_I-\epsilon/2$, $b=c_i+\epsilon/2$ on the barycenters of simplices contained in $f^{-1}(c_i)$. The map g gets as possible critical values, in addition to the critical values of f the critical values of $h=g|_{f^{-1}(c_i)}$. We leave the reader to check that g satisfies Property G in view of the fact that h does and $\epsilon<\epsilon(f)$. Clearly g differs from f by less than ϵ as it follows from construction.

Since simplicial maps (for some subdivision) are dense in the space of continuous maps and any simplicial map is arbitrarily close to one that satisfies Property G, item (4) follows. □

4.2 Stability: proof of Theorem 4.2

The stability theorem is a consequence of Proposition 3.16. In order to explain this we begin with a few observations:

(1) Consider the space of maps $C(X,\mathbb{R})$, X a compact ANR, equipped with the compact open topology which is induced from the metric

$$D(f,g):=\sup_{x\in X}|f(x)-g(x)|=\|f-g\|_\infty.$$

This metric is complete.

(2) Observe that if $f,g\in C(X,\mathbb{R})$, then for any $t\in[0,1]$,

$$h_t:=tf(x)+(1-t)g(x)\in C(X;\mathbb{R})$$

is continuous, and for any $0=t_0<t_1<\cdots<t_{N-1}<t_N=1$ one has the equality

$$D(f,g)=\sum_{0\leq i<N}D(h_{t_{i+1}},h_{t_i}). \tag{20}$$

(3) If X is a simplicial complex let $\mathcal{U} \subset C(X,\mathbb{R})$ denote the subset of PL maps. Then

(i) $\mathcal{U}$ is a dense subset in $C(X,\mathbb{R})$;

(ii) if $f,g \in \mathcal{U}$ then $h_t \in \mathcal{U}$, hence $\epsilon(h_t) > 0$, hence for any $t \in [0,1]$ there exists $\delta(t) > 0$ such that $t', t'' \in (t-\delta(t), t+\delta(t))$ implies $D(h_{t'}, h_t) < \epsilon(h_t)/3$.

These two statements are not hard to check. Recall the following:

- f is PL on X if with respect to some subdivision of X f is simplicial (ie the restriction of f to each simplex is linear), and
- for any two PL maps f, g, there exists a common subdivision of X which makes f and g simultaneously simplicial, hence h_t is a simplicial map for any t.

Item (i) follows from the fact that continuous maps can be approximated with arbitrary accuracy by PL maps and item (ii) follows from the continuity in t of the family h_t and from the compactness of X.

(4) Consider $\mathcal{C}_{b_r}(\mathbb{R}^2) = \mathbb{C}^{b_r}$, $b_r = \dim(H_r(X)$, with the canonical metric $\underline{D}$ which is complete. Since any map in $\mathcal{U}$ is tame, in view of Proposition 3.16, $f,g \in \mathcal{U}$ with $D(f,g) < \epsilon(f)/3$ implies

$$\underline{D}(\delta^f_r, \delta^g_r) \leq 2D(f,g). \tag{21}$$

To prove Theorem 4.2, first check that inequality (21) extends to all $f,g \in \mathcal{U}$. To do that we start with $f,g \in \mathcal{U}$ and consider the homotopy h_t, $t \in [0,1]$ defined above.

Choose a sequence $0 < t_1 < t_3 < \cdots < t_{2N-1} < 1$ such that for $i = 1, \dots, (2N-1)$, the intervals $(t_{2i-1} - \delta(t_{2i-1}), t_{2i-1} + \delta(t_{2i-1}))$ cover $[0,1]$ and

$$(t_{2i-1}, t_{2i-1} + \delta(t_{2i-1})) \cap (t_{2i+1} - \delta(t_{2i+1}), t_{2i+1}) \neq \varnothing.$$

This is possible in view of the compactness of $[0,1]$.

Take $t_0 = 0, t_{2N} = 1$ and $t_{2i} \in (t_{2i-1}, t_{2i-1} + \delta(t_{2i-1})) \cap (t_{2i+1} - \delta(t_{2i+1})$. To simplify the notation, abbreviate h_{t_i} to h_i.

In view of item (3)(ii) and item (4) (inequality (21)), one has

$$\begin{aligned} |t_{2i-1} - t_{2i}| < \delta(t_{2i-1}) \quad &\text{implies} \quad \underline{D}(\delta^{h_{2i-1}}, \delta^{h_{2i}}) < 2D(h_{2i-1}, h_{2i}), \\ |t_{2i} - t_{2i+1}| < \delta(t_{2i+1}) \quad &\text{implies} \quad \underline{D}(\delta^{h_{2i}}, \delta^{h_{2i+1}}) < 2D(h_{2i}, h_{2i+1}). \end{aligned}$$

Then we have

$$\underline{D}(\delta^f, \delta^g) \leq \sum_{0 \leq i < 2N-1} \underline{D}(\delta^{h_i}, \delta^{h_{i+1}}) \leq 2 \sum_{0 \leq i < 2N-1} D(h_i, h_{i+1}) = D(f,g).$$

In view of the density of $\mathcal{U}$ and the completeness of the metrics on $C(X;\mathbb{R})$ and $\mathcal{C}_{b_r}(\mathbb{R}^2)$, inequality (21) extends to the entire $C(X;\mathbb{R})$ when X is a simplicial complex. Indeed, the assignment $\mathcal{U} \ni f \rightsquigarrow \delta^f_r \in \mathcal{C}_{b_r}(\mathbb{R}^2)$ preserves the Cauchy sequences.

Next we verify (21) for $X = K \times Q$, K a simplicial complex and Q the Hilbert cube. For this purpose we write $Q := I^k \times Q^{\infty-k}$ and say that $f\colon K \times Q \to \mathbb{R}$ is an $(\infty{-}k)$–PL map if $f = \bar{g}_{Q^{\infty-k}}$ (see Section 2.3 for the definition of $\bar{g}_{Q^{\infty-k}}$) with $g\colon K \times I^k \to \mathbb{R}$ a PL map. Clearly an $(\infty{-}k)$–PL map is an $(\infty{-}k')$–PL map for $k' \geq k$.

Denote by $C_{\mathrm{PL}}(K \times Q;\mathbb{R})$ the set of maps in $C(K \times Q;\mathbb{R})$ which are $(\infty{-}k)$–PL for some k.

In view of Observation 2.2, $C_{\mathrm{PL}}(K \times Q;\mathbb{R})$ is dense in $C(K \times Q;\mathbb{R})$. To conclude that (21) holds for $K \times Q$, it suffices to check the inequality for $f_1 = (\bar{g}_1)_{Q^{\infty-k}}$, $f_2 = (\bar{g}_2)_{Q^{\infty-k}} \in C_{\mathrm{PL}}(K \times Q;\mathbb{R})$. The inequality holds since, in view of Observation 3.19, we have $\delta^{f_i} = \delta^{g_i}$.

Since by Theorem 2.3 any compact Hilbert cube manifold is homeomorphic to $K \times Q$ for some finite simplicial complex K, inequality (21) holds for X any compact Hilbert cube manifold. Since for any X a compact ANR, by Theorem 2.3, $X \times Q$ is a Hilbert cube manifold, $I\colon C(X;\mathbb{R}) \to C(X \times Q;\mathbb{R})$ defined by $I(f) = \bar{f}_Q$ is an isometric embedding and $\delta^f = \delta^{\bar{f}_Q}$, (21) holds for any X a compact ANR.

Both parts of Theorem 4.2 follow from inequality (21) and Proposition 3.16(9).

4.3 Poincaré duality: proof of Theorem 4.3

Before we proceed to the proof of Theorem 4.3, the following elementary observation on linear algebra, used also in part II, will be useful.

For the commutative diagram

$$E := \begin{array}{ccc} C & \xrightarrow{\gamma_2} & A_2 \\ \Big\downarrow{\scriptstyle \gamma_1} & & {\scriptstyle \alpha_2}\Big\downarrow \\ A_1 & \xrightarrow{\alpha_1} & B \end{array}$$

define

$$\ker(E) := \ker(C \xrightarrow{\gamma} A_1 \times_B A_2),$$
$$\operatorname{coker}(E) := \operatorname{coker}(A_1 \oplus_C A_2 \xrightarrow{\alpha} B)$$

with

$A_1 \times_B A_2 = \{(a_1, a_2) \in A_1 \times A_2 \mid \alpha_1(a_1) = \alpha_2(a_2)\},$
$A_1 \oplus_C A_2 = A_1 \oplus A_2 \big/ \big\{(a_1, a_2) \in A_1 \times A_2 \mid a_1 = \beta_1(c), a_2 = -\beta_2(c) \text{ for some } c \in C\big\}$
and with $\gamma(c) = (\gamma_1(c), \gamma_2(c))$ and $\alpha(a_1, a_2) = \alpha_1(a_1) + \alpha_2(a_2)$.

If one denotes by E^* the dual diagram

$$E^* := \begin{tikzcd} C^* & A_2^* \arrow[l, "\gamma_2^*"'] \\ A_1^* \arrow[u, "\gamma_1^*"] & B^* \arrow[l, "\alpha_1^*"] \arrow[u, "\alpha_2^*"'] \end{tikzcd}$$

then we have a canonical isomorphism

$$\ker(E) = (\operatorname{coker}(E^*))^*. \tag{22}$$

Note the following.

Proposition 4.6 *If in the diagram E either all arrows are injective and α is injective or all arrows are surjective and γ is surjective, then*

$$\dim(\operatorname{coker} E) = \dim C + \dim B - \dim A_1 - \dim A_2.$$

The proof is a straightforward calculation of dimensions.

For the proof of extended Poincaré duality claimed by Theorem 4.3 it is useful to provide an alternative definition of $\mathbb{F}_r(B)$ for a box B.

For this purpose introduce the quotient space

$$\mathcal{G}_r(a,b) = H_r(X)/(\mathbb{I}_a(r) + \mathbb{I}^b(r)).$$

Consider a box $B = (a', a] \times [b, b')$ and denote by $\mathcal{G}(B)$ and $\mathcal{F}(B)$ the diagrams

$$\mathcal{G}(B) := \begin{tikzcd} \mathcal{G}_r(a',b') \arrow[r] \arrow[d] & \mathcal{G}_r(a,b') \arrow[d] \\ \mathcal{G}_r(a',b) \arrow[r] & \mathcal{G}_r(a,b) \end{tikzcd} \qquad \mathcal{F}(B) := \begin{tikzcd} \mathbb{F}_r(a',b') \arrow[r] \arrow[d] & \mathbb{F}_r(a,b') \arrow[d] \\ \mathbb{F}_r(a',b) \arrow[r] & \mathbb{F}_r(a,b) \end{tikzcd}$$

whose arrows are induced by the inclusions $\mathbb{I}_{a'}(r) \subseteq \mathbb{I}_a(r)$ and $\mathbb{I}^{b'}(r) \subseteq \mathbb{I}^b(r)$. Let

$$\mathcal{G}_r^f(B) := \ker \mathcal{G}(B)$$

and recognize that

$$\mathbb{F}^f_r(B) = \operatorname{coker}\mathcal{F}(B).$$

Note that the hypotheses of Proposition 4.6 are verified, (1) for $\mathcal{G}(B)$ and (2) for $\mathcal{F}(B)$, and $\mathcal{G}_r(B)$ identifies to $\ker(\mathcal{G}(B))$ and $\mathbb{F}_r(B)$ to $\operatorname{coker}(\mathcal{F}(B))$.

Since $\mathcal{G}_r(a',b)\times_{\mathcal{G}_r(a,b)}\mathcal{G}_r(a,b') = H_r(X)/((\mathbb{I}_{a'}(r)+\mathbb{I}^b(r))\cap(\mathbb{I}_a(r)+\mathbb{I}^{b'}(r)))$, the vector space $\mathcal{G}_r(B)$ is canonically isomorphic to

$$\big((\mathbb{I}_{a'}(r)+\mathbb{I}^b(r))\cap(\mathbb{I}_a(r)+\mathbb{I}^{b'}(r))\big)/(\mathbb{I}_{a'}(r)+\mathbb{I}^{b'}(r)). \tag{23}$$

Similarly, since $\mathbb{F}_r(a',b)\oplus_{\mathbb{F}_r(a',b')}\mathbb{F}_r(a,b')) = (\mathbb{I}_{a'}(r)\cap\mathbb{I}^b(r)+\mathbb{I}_a(r)\cap\mathbb{I}^{b'}(r))$, the vector space $\mathbb{F}_r(B)$ is canonically isomorphic to

$$\mathbb{I}_a(r)\cap\mathbb{I}^b(r)/(\mathbb{I}_{a'}(r)\cap\mathbb{I}^b(r)+\mathbb{I}_a(r)\cap\mathbb{I}^{b'}(r)).$$

The obvious inclusion $\mathbb{I}_a(r)\cap\mathbb{I}^b(r)\subseteq(\mathbb{I}_{a'}(r)+\mathbb{I}^b(r))\cap(\mathbb{I}_a(r)+\mathbb{I}^{b'}(r))$ induces the linear map

$$\begin{aligned}\mathbb{F}_r(B) = \mathbb{I}_a(r)\cap\mathbb{I}^b(r)/(\mathbb{I}_{a'}(r)\cap\mathbb{I}^b(r)+\mathbb{I}_a(r)\cap\mathbb{I}^{b'}(r))&\\ \to(\mathbb{I}_{a'}(r)+\mathbb{I}^b(r))\cap(\mathbb{I}_a(r)+\mathbb{I}^{b'}(r))/(\mathbb{I}_{a'}(r)+\mathbb{I}^{b'}(r)&=\mathcal{G}_r(B).\end{aligned}$$

Proposition 4.7 *For any map $f\colon X\to\mathbb{R}$ and any box B the canonical linear map $\mathbb{F}_r(B)\to\mathcal{G}_r(B)$ defined above is an isomorphism: $\mathbb{F}^f_r(B)=\mathcal{G}^f_r(B)$.*

Proof Note that the injectivity is straightforward. Indeed, suppose

$$\mathbb{I}_a(r)\cap\mathbb{I}^b(r)\ni x = x_1+x_2$$

with $x_1\in\mathbb{I}_{a'}(r)$ and $x_2\in\mathbb{I}^{b'}(r)$. Then $x_1 = x-x_2\in\mathbb{I}^b(r)$ and $x_2\in(\mathbb{I}_a(r)\cap\mathbb{I}^{b'}(r))$.

To check the surjectivity, start with $x = x_1+y_1 = x_2+y_2$ such that $x_1\in\mathbb{I}_{a'}$, $y_1\in\mathbb{I}^b$, $x_2\in\mathbb{I}_a$, $y_2\in\mathbb{I}^{b'}$. Then $x-x_1-y_2$ is equivalent to x in $\mathcal{G}_r(B)$. But $x-x_1-y_2 = y_1-y_2 = x_2-x_1$ hence it belongs to $\mathbb{I}^b$ and to $\mathbb{I}^a$. □

Let $f\colon M^n\to\mathbb{R}$ be a map, M^n a κ–orientable closed topological manifold and a,b regular values such that the restriction of f to $f^{-1}(a-\epsilon,a+\epsilon)$ and $f^{-1}(b-\epsilon,b+\epsilon)$ for a small enough positive ϵ are topological submersions. This makes $f^{-1}(a)$ and $f^{-1}(b)$ codimension-one topological submanifolds of M.

Let $i_a\colon M_a\to M$, $i^b\colon M^b\to M$, $j_a\colon M\to(M,M_a)$, $j^b\colon M\to(M,M^b)$ denote the obvious inclusions, $i_a(k), i^b(k), j_a(k), j^b(k)$ denote the inclusion induced linear maps for homology in degree k, and $r_a(k), r^b(k), s_a(k), s^b(k)$ denote the inclusion induced linear maps in cohomology (with coefficients in the field κ), as indicated in diagrams (24) and (25) below. Poincaré duality provides the commutative diagrams (24) and (25) with all vertical arrows isomorphisms:

$$
\begin{tikzcd}
H_r(M_a) \arrow[r, "i_a(r)"] \arrow[d] & H_r(M) \arrow[r, "j_a(r)"] \arrow[d] & H_r(M, M_a) \arrow[d] \\
H^{n-r}(M, M^a) \arrow[r, "s^a(n-r)"] \arrow[d] & H^{n-r}(M) \arrow[r, "r^a(n-r))"] \arrow[d] & H^{n-r}(M^a) \arrow[d] \\
(H_{n-r}(M, M^a))^* \arrow[r, "(j^a(n-r))^*"] & (H_{n-r}(M))^* \arrow[r, "(i^a(n-r))^*"] & (H_{n-r}(M^a))^*
\end{tikzcd}
\tag{24}
$$

$$
\begin{tikzcd}
H_r(M^b) \arrow[r, "i^b(r)"] \arrow[d] & H_r(M) \arrow[r, "j^b(r)"] \arrow[d] & H_r(M, M^b) \arrow[d] \\
H^{n-r}(M, M_b) \arrow[r, "s_b(n-r)"] \arrow[d] & H^{n-r}(M)) \arrow[r, "r_b(n-r)"] \arrow[d] & H^{n-r}(M_b) \arrow[d] \\
(H_{n-r}(M, M_b))^* \arrow[r, "(j_b(n-r))^*"] & (H_{n-r}(M))^* \arrow[r, "(i_b(n-r))^*"] & (H_{n-r}(M_b))^*
\end{tikzcd}
\tag{25}
$$

As a consequence of these two diagrams, observe that Poincaré duality provides a canonical isomorphism

$$\mathbb{F}^f_r(a,b) = (\mathcal{G}^f_{n-r}(b,a))^*. \tag{26}$$

Indeed, observe the following:

- $\mathbb{F}_r(a,b) = \ker(j_a(r), j^b(r))$ by the exactness of the first rows in diagrams (24) and (25). Precisely $\ker(j_a(r), j^b(r)) = \ker j_a(r) \cap j^b(r) = \mathbb{I}_a(r) \cap \mathbb{I}^b(r)$.

- $\ker(j_a(r), j^b(r)) \equiv \ker(r^a(n-r), r_b(n-r))$ by the isomorphism of the upper vertical arrows in these diagrams.

- $\ker(r^a(n-r), r_b(n-r)) \equiv \ker\big((i^a(n-r))^*, (i_b(n-r))^*\big)$ by the isomorphism of the lower vertical arrow in these diagrams.

The isomorphisms above are induced by Poincaré duality and cohomology in terms of homology; their composition is still referred to as Poincaré duality.

- $\ker\big((i^a(n-r))^*, (i_b(n-r))^*\big) = \big(\operatorname{coker}\big(i^a(n-r) + i_b(n-r)\big)\big)^* = (\mathcal{G}^f_{n-r}(b,a))^*$ by standard finite-dimensional linear algebra duality.

Putting together these equalities one obtains (26).

Suppose M is a closed κ–orientable smooth manifold and $f\colon M \to \mathbb{R}$ a smooth map which is locally polynomial (ie in the neighborhood of any point, in some local coordinates, is a polynomial). Such a map is tame. For $(a,b) \in \mathbb{R}^2$ choose ϵ small enough so that the intervals $(a-\epsilon, a), (a, a+\epsilon)$ as well as $(a-\epsilon, a), (a, a+\epsilon)$ are contained in the set of regular values (in the sense of differential calculus). Such a choice is possible in view of the tameness of f.

To establish the result as stated for such a map we proceed as follows.

In view of the tameness of f,

$$\hat{\delta}^f_r(a,b) = \mathbb{F}^f_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon)). \tag{27}$$

By definition,

$$\mathbb{F}^f_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon)) = \operatorname{coker} \mathcal{F}_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon)). \tag{28}$$

By Proposition 4.7,

$$\operatorname{coker} \mathcal{F}_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon)) = \ker\bigl(\mathcal{G}_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon))\bigr). \tag{29}$$

By equality (22),

$$\begin{aligned}\ker\bigl(\mathcal{G}_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon))\bigr)& \\ = \bigl(\operatorname{coker}\bigl(\mathcal{G}_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon))^*\bigr)\bigr)^*&.\end{aligned} \tag{30}$$

By equality (26),

$$\begin{aligned}\bigl(\operatorname{coker}\bigl(\mathcal{G}_r((a-\epsilon, a+\epsilon] \times [b-\epsilon, b+\epsilon))^*\bigr)\bigr)^*& \\ = \bigl(\operatorname{coker}\bigl(\mathcal{F}_{n-r}((b-\epsilon, b+\epsilon] \times [a-\epsilon, a+\epsilon))\bigr)\bigr)^*&.\end{aligned} \tag{31}$$

In view of the equality $\mathbb{F}^f_r(B) = \operatorname{coker} \mathcal{F}(B)$,

$$\begin{aligned}\bigl(\operatorname{coker}\bigl(\mathcal{F}_{n-r}((b-\epsilon, b+\epsilon] \times [a-\epsilon, a+\epsilon))\bigr)\bigr)^*& \\ = \bigl(\mathbb{F}_{n-r}((b-\epsilon, b+\epsilon] \times [a-\epsilon, a+\epsilon))\bigr)^*&.\end{aligned} \tag{32}$$

In view of the tameness of f,

$$\bigl(\mathbb{F}^f_{n-r}((b-\epsilon, b+\epsilon] \times [a-\epsilon, a+\epsilon))\bigr)^* = (\hat{\delta}^f_{n-r}(b,a))^*. \tag{33}$$

Putting together equalities (27)–(33), one derives the result for f as above. In view of Theorem 4.2 and the fact that locally polynomial maps are dense in the space of all continuous maps when X is a smooth manifold, the result holds as stated.

A comment The hypothesis of compact ANR can be replaced by ANR with total homology of finite dimension and proper map by *homologically proper map*, which means that for I a closed interval, the total homology of $f^{-1}(I)$ has finite dimension. All results remain unchanged with essentially the same proof. An interesting situation when such a generalization is relevant is the case of the absolute value of the complex polynomial function f when restricted to the complement of its zeros, which will be treated in future work, but can be easily reduced to the case of a proper map considered above.

References

[1] **B J Ball**, *Alternative approaches to proper shape theory*, from "Studies in topology" (N M Stavrakas, K R Allen, editors), Academic Press (1975) 1–27 MR

[2] **D Burghelea**, **T K Dey**, *Topological persistence for circle-valued maps*, Discrete Comput. Geom. 50 (2013) 69–98 MR

[3] **D Burghelea**, **S Haller**, *Topology of angle valued maps, bar codes and Jordan blocks*, preprint (2013) arXiv

[4] **G Carlsson**, **V de Silva**, **D Morozov**, *Zigzag persistent homology and real-valued functions*, from "Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry", ACM, New York (2009) 247–256

[5] **T A Chapman**, *Lectures on Hilbert cube manifolds*, CBMS 28, Amer. Math. Soc., Providence, RI (1976) MR

[6] **R J Daverman**, **J J Walsh**, *A ghastly generalized n–manifold*, Illinois J. Math. 25 (1981) 555–576 MR

[7] **S-t Hu**, *Theory of retracts*, Wayne State Univ. Press, Detroit (1965) MR

[8] **J Milnor**, *On spaces having the homotopy type of a CW–complex*, Trans. Amer. Math. Soc. 90 (1959) 272–280 MR

Department of Mathematics, The Ohio State University
Columbus, OH, United States

burghele@math.ohio-state.edu

Received: 27 December 2015 Revised: 20 December 2016

Algebraic & Geometric Topology 17 (2017) 2081–2124

A categorification of the Alexander polynomial in embedded contact homology

GILBERTO SPANO

Given a transverse knot K in a three-dimensional contact manifold (Y,α), Colin, Ghiggini, Honda and Hutchings defined a hat version $\widehat{\mathrm{ECK}}(K,Y,\alpha)$ of embedded contact homology for K and conjectured that it is isomorphic to the knot Floer homology $\widehat{\mathrm{HFK}}(K,Y)$.

We define here a full version $\mathrm{ECK}(K,Y,\alpha)$ and generalize the definitions to the case of links. We prove then that if $Y=S^3$, then ECK and $\widehat{\mathrm{ECK}}$ categorify the (multivariable) Alexander polynomial of knots and links, obtaining expressions analogous to that for knot and link Floer homologies in the minus and, respectively, hat versions.

57M27, 57R17, 57R58

Introduction

Given a 3–manifold Y, Ozsváth and Szabó [29] defined topological invariants of Y, indicated by $\mathrm{HF}^\infty(Y)$, $\mathrm{HF}^+(Y)$, $\mathrm{HF}^-(Y)$ and $\widehat{\mathrm{HF}}(Y)$. These groups are the *Heegaard Floer homologies of Y* in the respective versions.

Moreover, Ozsváth and Szabó [28] and Rasmussen [33] proved that any homologically trivial knot K in Y induces a "knot filtration" on the Heegaard Floer chain complexes. The first pages of the associated spectral sequences (in each version) are topological invariants of K: these are bigraded homology groups $\mathrm{HFK}^\infty(K,Y)$, $\mathrm{HFK}^+(K,Y)$, $\mathrm{HFK}^-(K,Y)$ and $\widehat{\mathrm{HFK}}(K,Y)$, called *Heegaard Floer knot homologies* (in the respective versions).

These homologies are powerful invariants for the couple (K,Y). For instance, in [28] and [33], it was proved that $\widehat{\mathrm{HFK}}(K,S^3)$ *categorifies the Alexander polynomial Δ_K of K*; ie

$$\chi(\widehat{\mathrm{HFK}}(K,S^3)) \doteq \Delta(K),$$

where $\doteq$ means that the two sides are equal up to sign change and multiplication by a monic monomial, and χ denotes the *graded Euler characteristic*.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2081

This was the first categorification of the Alexander polynomial; a second one (in Seiberg–Witten–Floer homology) was discovered later by Kronheimer and Mrowka [23].

Ozsváth and Szabó [31] developed a similar construction for any link L in S^3 and got invariants $\mathrm{HFL}^-(L, S^3)$ and $\widehat{\mathrm{HFL}}(L, S^3)$ for L, which they called *Heegaard Floer link homologies*. Now these homologies come with an additional $\mathbb{Z}^n$ degree, where n is the number of connected components of L. Ozsváth and Szabó proved moreover that $\mathrm{HFL}^-(L, S^3)$ *categorifies the multivariable Alexander polynomial of* L, which is a generalization of the classic Alexander polynomial. They found in particular that

$$\chi(\mathrm{HFL}^-(L, S^3)) \doteq \begin{cases} \Delta_L(t_1, \dots, t_n) & \text{if } n > 1, \\ \Delta_L(t)/(1-t) & \text{if } n = 1, \end{cases} \tag{1}$$

and

$$\chi(\widehat{\mathrm{HFL}}(L, S^3)) \doteq \begin{cases} \Delta_L \cdot \prod_{i=1}^n (t_i^{1/2} - t_i^{-1/2}) & \text{if } n > 1, \\ \Delta_L(t) & \text{if } n = 1. \end{cases} \tag{2}$$

In the series of papers [5; 6; 7; 8; 9], Colin, Ghiggini and Honda prove the equivalence between Heegaard Floer homology and *embedded contact homology* for three-manifolds. The last one is another Floer homology theory, first defined by Hutchings, which associates to a contact manifold (Y, α) two graded modules $\mathrm{ECH}(Y, \alpha)$ and $\widehat{\mathrm{ECH}}(Y, \alpha)$.

Theorem 0.1 (Colin, Ghiggini and Honda [5]–[9])

$$\mathrm{HF}^+(-Y) \cong \mathrm{ECH}(Y, \alpha),$$
$$\widehat{\mathrm{HF}}(-Y) \cong \widehat{\mathrm{ECH}}(Y, \alpha),$$

where $-Y$ is the manifold Y with the inverted orientation.

In light of Theorem 0.1, it is a natural problem to find an embedded contact counterpart of Heegaard Floer knot homology. In analogy with the *sutured Heegaard Floer theory* developed by Juhász [22], Colin, Ghiggini, Honda and Hutchings [10, Sections 6–7] define a sutured version of embedded contact homology. This can be thought of as a version of embedded contact homology for manifolds with boundary. In particular, given a knot K in a contact three-manifold (Y, ξ), using sutures they define a *hat version* $\widehat{\mathrm{ECK}}(K, Y, \alpha)$ *of embedded contact knot homology*.

Roughly speaking, this is the hat version of ECH for the contact manifold with boundary $(Y \setminus \mathcal{N}(K), \alpha)$, where $\mathcal{N}(K)$ is a suitable thin tubular neighborhood of K in Y and α is a contact form satisfying specific compatibility conditions with K. In [10, Conjecture 1.5], the following conjecture is stated:

Conjecture 0.2 $\quad \widehat{\mathrm{ECK}}(K, Y, \alpha) \cong \widehat{\mathrm{HFK}}(-K, -Y).$

In the present paper, we first define a *full version of embedded contact knot homology* $\mathrm{ECK}(K, Y, \alpha)$ for knots K in any contact three-manifold (Y, ξ) endowed with a (suitable) contact form α for ξ. Moreover, we generalize the definitions to the case of links L with more than one component to obtain homologies $\mathrm{ECK}(L, Y, \alpha)$ and $\widehat{\mathrm{ECK}}(L, Y, \alpha)$. We state then the following:

Conjecture 0.3 For any link L in Y, there exists a contact form α for which

$$\widehat{\mathrm{ECK}}(L, Y, \alpha) \cong \widehat{\mathrm{HFK}}(-L, -Y),$$
$$\mathrm{ECK}(L, Y, \alpha) \cong \mathrm{HFK}^{+}(-L, -Y).$$

We remark that $\mathrm{ECK}(L, Y, \alpha)$ (as well as $\widehat{\mathrm{ECK}}(L, Y, \alpha)$) is defined as the first page of a spectral sequence arising from a filtration induced by L on a suitable chain complex for $\mathrm{ECH}(Y)$. In light of the last conjecture, this fact is interesting because the analogous filtration for $\mathrm{HFK}^{-}(L, Y)$ is useful to study link surgery formulae in Heegaard Floer (see for example Ozsváth and Szabó [32] and Manolescu and Ozsváth [25]), and one can expect to find similar relations in ECH.

Next we compute the graded Euler characteristics of the ECK homologies for knots and links in homology three-spheres, and we prove the following:

Theorem 0.4 *Let L be an n–component link in a homology three-sphere Y. Then there exists a contact form α such that*

$$\chi(\mathrm{ECK}(L, Y, \alpha)) \doteq \mathrm{ALEX}(Y \setminus L).$$

Here $\mathrm{ALEX}(Y \setminus L)$ is the *Alexander quotient* of the complement of L in Y. The theorem is proved using Fried's dynamic reformulation of ALEX [14]. Classical relations between $\mathrm{ALEX}(S^3 \setminus L)$ and Δ_L imply the following result:

Theorem 0.5 *Let L be any n–component link in S^3. Then there exists a contact form α for which*

$$\chi(\mathrm{ECK}(L, S^3, \alpha)) \doteq \begin{cases} \Delta_L(t_1, \dots, t_n) & \text{if } n > 1, \\ \Delta_L(t)/(1-t) & \text{if } n = 1, \end{cases}$$

and

$$\chi(\widehat{\mathrm{ECK}}(L, S^3, \alpha)) \doteq \begin{cases} \Delta_L(t_1, \dots, t_n) \cdot \prod_{i=1}^{n}(1-t_i) & \text{if } n > 1, \\ \Delta_L(t) & \text{if } n = 1. \end{cases}$$

This implies that the homology ECK *is a categorification of the multivariable Alexander polynomial.* A straightforward consequence is:

Corollary 0.6 *In S^3, Conjectures 0.2 and 0.3 hold at the level of Euler characteristics.*

Acknowledgements I first thank my advisors Paolo Ghiggini and Vincent Colin for the patient and constant help and the trust they accorded to me during the years I spent in Nantes, where most of this work has been done. I also thank the referees for having reported my PhD thesis, containing essentially all the results presented here. I finally thank Vinicius Gripp, Thomas Guyard, Christine Lescop, Paolo Lisca, Margherita Sandon and Vera Vértesi for all the advice, support or stimulating conversations we had. This work was partially supported by University of Nantes, ERC Geodycon and ERC LTDBud.

1 Review of embedded contact homology

1.1 Preliminaries

This subsection is devoted to recall some basic notions about contact geometry, holomorphic curves, Morse–Bott theory and open books.

1.1.1 Contact geometry A (cooriented) *contact form* on a three-dimensional oriented manifold Y is a 1–form α on Y such that $\alpha \wedge d\alpha$ is a positive volume form. A *contact structure* is a smooth plane field ξ on Y such that there exists a contact form α for which $\xi = \ker \alpha$. The *Reeb vector field* of α is the (unique) vector field R_α determined by the equations $d\alpha(R_\alpha, \cdot) = 0$ and $\alpha(R_\alpha) = 1$. A *simple Reeb orbit* is a closed oriented orbit of $R = R_\alpha$; ie it is the image δ of an embedding $S^1 \hookrightarrow Y$ such that R_P is positively tangent to δ for any $P \in \delta$. A *Reeb orbit* is an m–fold cover of a simple Reeb orbit, with $m \geq 1$. The form α determines an *action* $\mathcal{A}$ on the set of its Reeb orbits defined by $\mathcal{A}(\gamma) = \int_\gamma \alpha$. By definition, $\mathcal{A}(\gamma) > 0$ for any nonempty orbit γ.

A basic result in contact geometry asserts that the flow of the Reeb vector field (abbreviated *Reeb flow*) $\phi = \phi_R$ preserves ξ, that is, $(\phi_t)_*(\xi_P) = \xi_{\phi_t(P)}$ for any $t \in \mathbb{R}$; see [15, Chapter 1]. Given a Reeb orbit δ, there exists $T \in \mathbb{R}^+$ such that $(\phi_T)_*(\xi_P) = \xi_P$ for any $P \in \delta$; if T is the smallest possible, the isomorphism $\mathfrak{L}_\delta := (\phi_T)_*\colon \xi_P \to \xi_P$ is called the (*symplectic*) *linearized first return map* of R in P.

The orbit δ is called *nondegenerate* if 1 is not an eigenvalue of $\mathfrak{L}_\delta$. There are two types of nondegenerate Reeb orbits, *elliptic* and *hyperbolic*: δ is elliptic if the eigenvalues of $\mathfrak{L}_\delta$ are on the unit circle and is hyperbolic if they are real. In the last case, we can make a further distinction: δ is called *positive* or *negative* hyperbolic if the eigenvalues are both positive or negative, respectively.

Definition 1.1 The *Lefschetz sign* of a nondegenerate Reeb orbit δ is

$$\epsilon(\delta) := \operatorname{sign}(\det(\mathbb{1} - \mathfrak{L}_\delta)) \in \{+1, -1\}.$$

Observation 1.2 It is easy to check that $\epsilon(\delta) = +1$ if δ is elliptic or negative hyperbolic and $\epsilon(\delta) = -1$ if δ is positive hyperbolic.

To any nondegenerate orbit δ and a trivialization τ of $\xi|_\delta$, we can associate also the *Conley–Zehnder (CZ) index* $\mu_\tau(\delta) \in \mathbb{Z}$ of δ with respect to τ; see for example [20, Section 3.2] or [12]. Even if $\mu_\tau(\delta)$ depends on τ, its parity depends only on δ, and

$$(-1)^{\mu_\tau(\delta)} = -\epsilon(\delta).$$

Definition 1.3 Given $X \subseteq Y$, we will indicate by $\mathcal{P}(X)$ the set of simple Reeb orbits of α contained in X. An *orbit set* (or *multiorbit*) in X is a formal finite product $\gamma = \prod_i \gamma_i^{k_i}$, where $\gamma_i \in \mathcal{P}(X)$ and $k_i \in \mathbb{N}$ is the *multiplicity* of γ_i in γ, with $k_i \in \{0, 1\}$ whenever γ_i is hyperbolic. The set of multiorbits in X will be denoted by $\mathcal{O}(X)$.

Note that the empty set is a legitimate orbit set, and it will be indicated by $\varnothing$. An orbit set $\gamma = \prod_i \gamma_i^{k_i}$ determines the homology class $[\gamma] = \sum_i k_i[\gamma_i] \in H_1(Y)$ (unless stated otherwise, all homology groups will be taken with integer coefficients). Moreover, the action of γ is defined by $\mathcal{A}(\gamma) = \sum_i k_i \int_{\gamma_i} \alpha$.

1.1.2 Holomorphic curves We recall here some definitions and properties about holomorphic curves in dimension 4. We refer the reader to [26] and [27] for the general theory and to [20] and [5; 7; 8; 9] for an approach which is more specialized to our context.

Let X be an oriented even-dimensional manifold. An *almost complex structure* on X is an isomorphism $J\colon TX \to TX$ such that $J(T_P X) = T_P X$ and $J^2 = -\mathrm{id}$. If (X_1, J_1) and (X_2, J_2) are two even-dimensional manifolds endowed with an almost complex structure, a map $u\colon (X_1, J_1) \to (X_2, J_2)$ is *pseudoholomorphic* if it satisfies the *Cauchy–Riemann equation*

$$du \circ J_1 = J_2 \circ du.$$

Definition 1.4 A *pseudoholomorphic curve* in a four-dimensional manifold (X, J) is a pseudoholomorphic map $u\colon (F, j) \to (X, J)$, where (F, j) is a (possibly disconnected) Riemann surface.

In this paper, we will be particularly interested in pseudoholomorphic curves (that sometimes we will simply call holomorphic curves) in "symplectizations" of contact three-manifolds. Given (Y, α), consider the four-manifold $\mathbb{R} \times Y$. Call s the $\mathbb{R}$–coordinate and let $R = R_\alpha$ be the Reeb vector field of α. The almost complex structure J on $\mathbb{R} \times Y$ is *adapted to* α if

(1) J is s–invariant;
(2) $J(\xi) = \xi$ and $J(\partial_s) = R$ at any point of $\mathbb{R} \times Y$;
(3) $J|_\xi$ is compatible with $d\alpha$; ie $d\alpha(\cdot\,, J\cdot)$ is a Riemannian metric on ξ.

For us, a holomorphic curve u in the symplectization of (Y, α) is a holomorphic curve $u\colon (\dot{F}, j) \to (\mathbb{R} \times Y,\ J)$, where

(i) J is adapted to α;
(ii) $(\dot{F}, j)$ is a Riemann surface obtained from a closed surface F by removing a finite number of points (called *punctures*);
(iii) for any puncture x there exists a neighborhood $U(x) \subset F$ such that $U(x) \setminus \{x\}$ is mapped by u asymptotically to a cover of a cylinder $\mathbb{R} \times \delta$ over an orbit δ of R in a way that $\lim_{y \to x} \pi_{\mathbb{R}}(u(y)) = \pm\infty$, where $\pi_{\mathbb{R}}$ is the projection on the $\mathbb{R}$–factor of $\mathbb{R} \times Y$.

We say that x is a *positive puncture* of u if in the last condition above the limit is $+\infty$: in this case the orbit δ is a *positive end* of u. If otherwise, the limit is $-\infty$, and we say x is a *negative puncture* and δ is a *negative end* of u.

By condition (iii) above, u near a puncture x determines a cover of the Reeb orbit δ corresponding to x: the number of sheets of this cover is the *local x–multiplicity of δ in u*. The sum of the x–multiplicities over all the punctures x associated to δ is the *(total) multiplicity* of δ in u.

If γ and γ' are the orbit sets determined by the sets of all positive and, respectively, negative ends of u counted with multiplicity, then we say that u is a *holomorphic curve from γ to γ'*.

Example 1.5 A *cylinder over an orbit set γ* of Y is the holomorphic curve $\mathbb{R} \times \gamma \subset \mathbb{R} \times Y$.

Observation 1.6 Note that if there exists a holomorphic curve u from γ to γ', then $[\gamma] = [\gamma'] \in H_1(Y, \mathbb{Z})$.

Theorem 1.7 [27, Lemma 2.4.1] *Let $u\colon (F, j) \to (\mathbb{R} \times Y, J)$ be a nonconstant holomorphic curve in (X, J). Then the critical points of $\pi_{\mathbb{R}} \circ u$ are isolated. In particular, if π_Y denotes the projection $\mathbb{R} \times Y \to Y$, then $\pi_Y \circ u$ is transverse to R_α away from a set of isolated points.*

From now on, if u is a map with image in $\mathbb{R} \times Y$, we will set $u_{\mathbb{R}} := \pi_{\mathbb{R}} \circ u$ and $u_Y := \pi_Y \circ u$.

Holomorphic curves also enjoy the following property, which will be essential for us; see for example Gromov [17] and, for the noncompact case, Siefring [34].

Theorem 1.8 (positivity of intersection) *Let u and v be two distinct holomorphic curves in a four-manifold (W,J). Then $\#(\mathrm{Im}(u)\cap\mathrm{Im}(v))<\infty$. Moreover, if P is an intersection point between $\mathrm{Im}(u)$ and $\mathrm{Im}(v)$, then its contribution m_P to the algebraic intersection number $\langle\mathrm{Im}(u),\mathrm{Im}(v)\rangle$ is strictly positive, and $m_P=1$ if and only if u and v are embeddings near P that intersect transversely in P.*

When the almost complex structure does not play an important role or is understood, it will be omitted from the notation.

1.1.3 Morse–Bott theory The Morse–Bott theory in contact geometry was first developed by Bourgeois [3]. We present in this subsection some basic notions and applications, mostly as presented in [5].

Definition 1.9 A *Morse–Bott (MB) torus* in a 3–dimensional contact manifold (Y,α) is an embedded torus T in Y foliated by a family γ_t, $t\in S^1$, of Reeb orbits, all in the same class in $H_1(T)$, that are nondegenerate in the Morse–Bott sense. Here this means the following: given any $P\in T$ and a positive basis (v_1,v_2) of ξ_P where $v_2\in T_P(T)$ (so that v_1 is transverse to $T_P(T)$), then the differential of the first return map of the Reeb flow on ξ_P is of the form

$$\begin{pmatrix}1 & 0\\ a & 1\end{pmatrix}$$

for some $a\neq 0$. T is a *positive* or *negative* MB torus if $a>0$ or $a<0$, respectively.

We say that α is a *Morse–Bott contact form* if all the Reeb orbits of α are either isolated and nondegenerate or come in S^1–families foliating MB tori.

As explained in [3] and [5, Section 4], it is possible to modify the Reeb vector field in a small neighborhood of a MB torus T preserving only two orbits, say e and h, of the S^1–family of Reeb orbits associated to T.

Moreover, for any fixed $L>0$, the perturbation can be done in such a way that e and h are the only orbits in a neighborhood of T with action less then L.

If T is a positive (respectively, negative) MB torus and τ is the trivialization of ξ along the orbits given pointwise by the basis (v_1,v_2) above, then one can make the MB perturbation in a way that h is positive hyperbolic with $\mu_\tau(h)=0$ and e is elliptic with $\mu_\tau(e)=1$ (respectively, $\mu_\tau(e)=-1$).

The orbits e and h can be seen as the only two critical points of a Morse function $f_T\colon S^1\to\mathbb{R}$ defined on the S^1–family of Reeb orbits foliating T and with maximum corresponding to the orbit with higher CZ index. Often MB tori will be implicitly given with such a function.

Observation 1.10 It is important to remark that, before the perturbation, T is foliated by Reeb orbits of α and so these are nonisolated. Moreover, the form of the differential of the first return map of the flow of ξ implies that these orbits are also degenerate. After the perturbation, T contains only two isolated and nondegenerate orbits, but other orbits are created in a neighborhood of T, and these orbits can be nonisolated and degenerate. See Figure 1 for a picture of a MB perturbation.

Proposition 1.11 [3, Section 3] *For any MB torus T and any $L \in \mathbb{R}$, there exists a MB perturbation of T such that, with the exception of e and h, all the periodic orbits in a neighborhood of T have action greater then L.*

A torus T foliated by Reeb orbits all in the same class of $H_1(T)$ (like, for example, a Morse–Bott torus) can be used to obtain constraints about the behavior of a holomorphic curve near T.

Following [5, Section 5], if γ is any of the Reeb orbits in T, we can define the *slope of T* as the equivalence class $s(T)$ of $[\gamma] \in H_1(T, \mathbb{R}) - \{0\}$ up to multiplication by positive real numbers.

Let $T \times [-\epsilon, \epsilon]$ be a neighborhood of $T = T \times \{0\}$ in Y with coordinates (ϑ, t, y) such that $(\partial_\vartheta, \partial_t)$ is a positive basis for $T(T)$ and ∂_y is directed as a positive normal vector to T.

Suppose that $u\colon (F, j) \to (\mathbb{R} \times Y,\ J)$ is a holomorphic curve in the symplectization of (Y, α); by Theorem 1.7, there exist at most finitely many points in $T \times [-\epsilon, \epsilon]$ where $u_Y(F)$ is not transverse to R_α. Then, if $T_y := T \times \{y\}$ and $u(F)$ intersects $\mathbb{R} \times T_y$, we can associate a slope $s_{T_y}(u)$ to $u_Y(F) \cap T_y$ for any $y \in [-\epsilon, \epsilon]$: this is defined exactly like $s(T)$, where $u_Y(F) \cap T_y$ is considered with the orientation induced by $\partial(u_Y(F) \cap (T \times [-\epsilon, y]))$.

Observation 1.12 If u has no ends in $T \times [y, y']$, then

$$\partial(u_Y(F) \cap T \times [y, y']) = u_Y(F) \cap T_{y'} - u_Y(F) \cap T_y,$$

and $s_{T_y}(u) = s_{T'_y}(u)$.

The following lemma is a consequence of the positivity of intersection in dimension four; see [5, Lemma 5.2.3].

Lemma 1.13 (blocking lemma) *Let T be linearly foliated by Reeb trajectories with slope $s = s(T)$ and u a holomorphic curve as above.*

(1) *If u is homotopic, by a compactly supported homotopy, to a map whose image is disjoint from $\mathbb{R} \times T$, then $u_Y(F) \cap T = \varnothing$.*

(2) *Let $T \times [-\epsilon, \epsilon]$ be a neighborhood of $T = T \times \{0\}$. Suppose that for some $y \in [-\epsilon, \epsilon] \setminus \{0\}$, u has no ends in $T \times (0, y]$ if $y \in (0, \epsilon]$ or in $T \times [y, 0)$ if $y \in [-\epsilon, 0)$. If $s_{T_y}(u) = \pm s(T)$, then u has an end which is a Reeb orbit in T.*

Let now x be a puncture of F whose associated end is an orbit γ in T; if there exists a neighborhood $U(x)$ of x in F such that $u_Y(U(x) \setminus \{x\}) \cap T = \varnothing$, then γ is a *one-sided end* of u in x. This is equivalent to requiring that $u_Y(U(x))$ is contained either in $T \times (-\epsilon, 0)$ or in $T \times (0, \epsilon)$.

Lemma 1.14 (trapping lemma [5, Lemma 5.3.2]) *If T is a positive (respectively, negative) MB torus and $\gamma \subset T$ is a one-sided end of u associated to the puncture x, then x is positive (respectively, negative).*

Definition 1.15 Let α be a Morse–Bott contact form on the three-manifold Y, and J a regular almost complex structure on $\mathbb{R} \times Y$. Suppose that any MB torus T in (Y, α) comes with a fixed a Morse function f_T. Let $\mathcal{P}(Y)$ be the set of simple Reeb orbits in Y minus the set of the orbits which correspond to some regular point of some f_T.

A *Morse–Bott building in* (Y, α) is a disjoint union of objects u of one of the following two types:

(1) u is the submanifold of a MB torus T corresponding to a gradient flow line of f_T: in this case, the positive (respectively, negative) end of u is the positive (respectively, negative) end of the flow line.

(2) u is a union of curves $\tilde{u} \cup u_1 \cup \cdots \cup u_n$ of the following kind: $\tilde{u}$ is a J–holomorphic curve in $\mathbb{R} \times Y$ with n ends $\{\delta_1, \dots, \delta_n\}$ corresponding to regular values of some $\{f_{T_1}, \dots, f_{T_n}\}$. Then, for each i, the curve $\tilde{u}$ is augmented by a gradient flow trajectory u_i of f_{T_i}: u_i goes from the maximum ϵ_i^+ of f_{T_i} to δ_i if δ_i is a positive end and goes from δ_i to the minimum ϵ_i^- of f_{T_i} if δ_i is a negative end. The ends of u are obtained from the ends of $\tilde{u}$ by substituting each δ_i with the respective ϵ_i^+ or ϵ_i^-.

A Morse–Bott building is *nice* if the $\tilde{u}$ above has at most one connected component which is not a cover of a trivial cylinder.

1.2 ECH for closed three-manifolds

We briefly review here Hutchings' original definitions of $\mathrm{ECH}(Y, \alpha)$ and $\widehat{\mathrm{ECH}}(Y, \alpha)$ for a closed contact three-manifold (Y, α).

Assume that α is nondegenerate (ie that any Reeb orbit of α is nondegenerate). For a fixed $\Gamma \in H_1(Y)$, define $\mathrm{ECC}(Y, \alpha, \Gamma)$ to be the free $\mathbb{Z}_2$–module generated by the orbit sets of Y in the homology class Γ, and set

$$\mathrm{ECC}(Y,\alpha) = \bigoplus_{\Gamma\in H_1(Y)} \mathrm{ECC}(Y,\alpha,\Gamma).$$

This is the ECH *chain group of* (Y,α). The ECH–*differential* ∂^{ECH} (called simply ∂ when no risk of confusion occurs) is defined in [19] in terms of holomorphic curves in the symplectization $(\mathbb{R}\times Y, d\alpha, J)$ of (Y,α) as follows.

Given $\gamma,\delta\in\mathcal{O}(Y)$, let $\mathcal{M}(\gamma,\delta)$ be the set of (possibly disconnected) holomorphic curves $u\colon (\dot F, j)\to(\mathbb{R}\times Y, J)$ from γ to δ, where $(\dot F, j)$ is a punctured compact Riemannian surface. It is clear that u determines a relative homology class $[\mathrm{Im}(u)]$ in $H_2(\mathbb{R}\times Y;\gamma,\delta)$ and that if such a curve exists, then $[\gamma]=[\delta]\in H_1(Y)$.

If $\xi=\ker(\alpha)$ and a trivialization τ of $\xi|_{\gamma\cup\delta}$ is given, then to any surface $C\subset\mathbb{R}\times Y$ with $\partial C=\gamma-\delta$, it is possible to associate an ECH–*index*

$$I(C) := c_\tau(C) + Q_\tau(C) + \mu_\tau^I(\gamma,\delta),$$

which depends only on the relative homology class of C. Here,

- $c_\tau(C) := c_1(\xi|_C,\tau)$ is the *first relative Chern class* of C;
- $Q_\tau(C)$ is the τ–*relative intersection paring* of $\mathbb{R}\times Y$ applied to C;
- $\mu_\tau^I(\gamma,\delta) := \sum_i\sum_{j=1}^{k_i}\mu_\tau(\gamma_i^j) - \sum_i\sum_{j=1}^{k_i}\mu_\tau(\delta_i^j)$, where μ_τ is the Conley–Zehnder index defined in Section 1.1.1.

We refer the reader to [20] for the details about these quantities. If u is a holomorphic curve from γ to δ, set $I(u)=I(\mathrm{Im}(u))$ (well-defined up to approximating $\mathrm{Im}(u)$ with a surface in the same homology class).

Define $\mathcal{M}_k(\gamma,\delta) := \{u\in\mathcal{M}(\gamma,\delta)\mid I(u)=k\}$. The ECH–differential is then defined on the generators of $\mathrm{ECC}(Y,\alpha)$ by

$$\partial^{\mathrm{ECH}}(\gamma) = \sum_{\delta\in\mathcal{O}(Y)} \#(\mathcal{M}_1(\gamma,\delta)/\mathbb{R})\cdot\delta, \tag{3}$$

where we quotient $\mathcal{M}_1(\gamma,\delta)$ by the $\mathbb{R}$–action on the curves given by the translation in the $\mathbb{R}$–direction in $\mathbb{R}\times Y$. In [20, Section 5], Hutchings proves that $\mathcal{M}_1(\gamma,\delta)/\mathbb{R}$ is a compact 0–dimensional manifold, so $\partial^{\mathrm{ECH}}(\gamma)$ is well-defined.

The *(full) embedded contact homology of* (Y,α) is

$$\mathrm{ECH}_*(Y,\alpha) := H_*(\mathrm{ECC}(Y,\alpha),\partial^{\mathrm{ECH}}).$$

It turns out that these groups do not depend on either the choices J in the symplectization or the contact form for ξ.

If $\gamma = \prod_i \gamma_i^{k_i}$ is a generator of $\mathrm{ECC}(Y, \alpha)$, set

$$\epsilon(\gamma) = \prod_i \epsilon(\gamma_i)^{k_i},$$

where $\epsilon(\gamma_i)$ is the Lefschetz sign of the simple orbit γ_i. Note that $\epsilon(\gamma)$ is given by the parity of the number of positive hyperbolic simple orbits in γ.

If u is a holomorphic curve from γ to δ, by simple computations it is possible to prove the *index parity formula* (see for example Section 3.4 in [20])

$$(-1)^{I(u)} = \epsilon(\gamma)\epsilon(\delta). \tag{4}$$

It follows then that the Lefschetz sign endows embedded contact homology with a well-defined absolute $\mathbb{Z}/2$–grading.

Fix now a generic point $(0, z) \in \mathbb{R} \times Y$. Given two orbit sets γ and δ, let

$$U_z\colon \mathrm{ECC}_*(Y, \alpha) \to \mathrm{ECC}_{*-2}(Y, \alpha)$$

be the map defined on the generators by

$$U_z(\gamma) = \sum_{\delta \in \mathcal{O}(Y)} \#\bigl\{u \in \mathcal{M}_2(\gamma, \delta) \mid (0, z) \in \mathrm{Im}(u)\bigr\} \cdot \delta.$$

Hutchings proves that U_z is a chain map that counts only a finite number of holomorphic curves and that this count does not depend on the choice of z. So it makes sense to define the map $U := U_z$ for any z as above. This is called the *U–map*.

The *hat version of embedded contact homology of* (Y, α) is defined as the homology $\widehat{\mathrm{ECH}}(Y, \alpha)$ of the mapping cone of the U–map. By this, we mean that $\widehat{\mathrm{ECH}}(Y, \alpha)$ is defined to be the homology of the chain complex

$$\mathrm{ECC}_{*-1}(Y, \alpha) \oplus \mathrm{ECC}_*(Y, \alpha)$$

with differential defined by the matrix

$$\begin{pmatrix} -\partial_{*-1} & 0 \\ U & \partial_* \end{pmatrix},$$

where the elements of the complex are thought as columns. Also, $\widehat{\mathrm{ECH}}(Y, \alpha)$ has the relative and the absolute gradings above.

We end this section by stating the following result; see for example [20, Section 1.4].

Theorem 1.16 *Let ξ be a contact structure on Y and α a contact form with* $\ker \alpha = \xi$. *Then the homology class $[\varnothing] \in \mathrm{ECH}(Y, \alpha)$ of the empty orbit set $\varnothing$ depends only on ξ, and it is called the* ECH–contact invariant of ξ.

1.3 ECH for manifolds with torus boundary

In order to define ECH for contact three-manifolds (N, α) with nonempty boundary, some compatibility between α and ∂N should be assumed. In this paper, we are particularly interested in three-manifolds whose boundary is a collection of disjoint tori.

In [5, Section 7], Colin, Ghiggini and Honda analyze this situation when ∂N is connected. If $\mathcal{T} = \partial N$ is homeomorphic to a torus, then they prove that the ECH–complex, and the differential can be defined almost as in the closed case, provided that $R = R_\alpha$ is tangent to $\mathcal{T}$ and that α is nondegenerate in $\operatorname{int}(N)$.

If the flow of $R|_{\mathcal{T}}$ is irrational, they define $\mathrm{ECH}(N, \alpha) = \mathrm{ECH}(\operatorname{int}(N), \alpha)$, and if it is rational, they consider the case of $\mathcal{T}$ Morse–Bott and do a MB perturbation of α near $\mathcal{T}$; this gives two Reeb orbits h and e on $\mathcal{T}$, and since α is now a MB contact form, the ECH–differential counts MB buildings.

1.3.1 Contact forms If Y is a closed 3–manifold and $K \subset Y$ is an oriented null-homologous knot, let $\mathcal{N}$ be a closed tubular neighborhood of K, and define N to be the closure of $Y \setminus \mathcal{N}$. Fix a neighborhood $[0, 2] \times T^2$ of $\partial\mathcal{N} = \{1\} \times T^2$ in Y with $[0, 1] \times T^2 \subset N$, and let $V \subset \mathcal{N}$ be the solid torus with core K and $\partial V = \{2\} \times T^2$. Obviously, $\mathcal{N} = ([1, 2] \times T^2) \cup V$.

Identify now $V \setminus K$ with $[2, 3) \times T^2$, and fix coordinates $(y, \vartheta, t) \in [0, 3) \times T^2 \cong \mathcal{N} \setminus K$ such that the natural projection $[0, 3) \times T^2 \twoheadrightarrow K$ sends (y, ϑ, t) to ϑ, and for any given y_0 and $t_0 \in [0, 1]/\langle 0 \sim 1\rangle$, the push off $\{(y_0, \vartheta, t_0) \mid \vartheta \in K\}$ of K has linking number 0 with K in Y (ie it gives the Seifert framing of K). Note in particular that each strip $\{t = t_0\}$ can be seen as the intersection between $\mathcal{N} \setminus K$ and some Seifert surface for K and with the inherited orientation, so that (y, ϑ, t_0) is a positive coordinate system and any $\{(y_0, \vartheta_0, t) \mid t \in [0, 1]/\langle 0 \sim 1\rangle$ is a positive meridian for K.

Definition 1.17 We say that the contact form α on Y is adapted to K if there exists a tubular neighborhood $\mathcal{N}$ of K as before such that

(1) α is a Morse–Bott contact form which is nondegenerate in $\operatorname{int}(N)$;

(2) the Reeb flow R_α is positively transverse to each strip $\{t = t_0\}$ in $\mathcal{N} \setminus K$;

(3) all the tori $y_0 \times T^2$ for $y_0 \in [1, 3)$ are linearly foliated by Reeb trajectories of α;

(4) $T_1 := \{1\} \times T^2$ and $T_2 := \{2\} \times T^2$ are respectively negative and positive MB tori foliated by Reeb orbits which are meridians of K;

(5) R_α is transverse to the disks of the form $\{\vartheta = \vartheta_0\} \cap \operatorname{int}(V)$;

(6) K is a Reeb orbit.

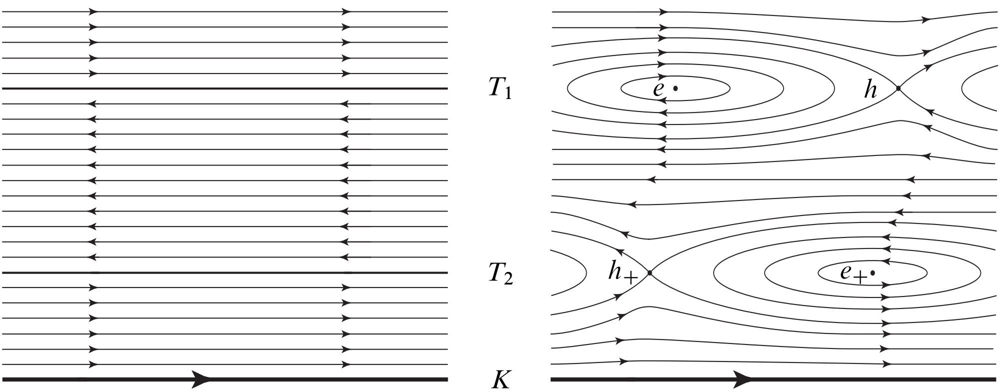


Figure 1: Reeb dynamic before and after a MB perturbation of the tori T_1 and T_2. Both pictures take place in a strip $\{t = t_0\} \subset \mathcal{N} \setminus K$. Each flow line represents an invariant subset of $\{t = t_0\}$ under the Reeb flow near K; the orientation gives the direction in which any point is mapped under the first return map of the flow.

The families of Reeb orbits in T_1 and T_2 can be perturbed into two pairs of Reeb orbits (e, h) and (e_+, h_+), where e and e_+ are elliptic with CZ index -1 and $+1$ respectively, and h and h_+ are positive hyperbolic, both with CZ index 0; see Figure 1.

Definition 1.18 A contact form α is adapted to a Seifert surface S for K if the R_α is positively transverse to $\mathrm{int}(S)$.

The proof of the following lemma is given in Sections 9.2 and 10.3 of [5]; compare also Section 7.2 of [10].

Lemma 1.19 [5] *Given a null-homologous knot K and a contact structure ξ on Y, there exists a contact form α for ξ and a genus minimizing Seifert surface S for K such that*

(1) *α is adapted to K;*

(2) *α is adapted to S.*

It is important to remark that the proof of (1) is obtained by locally modifying a given contact form near K using the Darboux–Weinstein neighborhood theorem; see for example [15]. Moreover, the (perturbed) contact form compatible with K obtained in [5] can be arranged to have all the orbits in $\mathcal{N} \setminus K$ that have arbitrarily large linking number with K, with the exception of the four relevant orbits e, h, e_+ and h_+.

Example 1.20 Let (K, S, ϕ) be an *open book decomposition* of Y, where S is the *page*, ϕ the *monodromy* and $K = \partial S$ the (not necessarily connected) *binding* of the open book. Let α be a *contact form adapted* to (K, S, ϕ) obtained via the Thurston–

Winkelnkemper construction [36]. Then α is compatible with S, and it can be easily adapted also to K; see for example [5, Section 9.3]. The strips $\{t = t_0\} \subset \mathcal{N} \setminus K$ can be obtained as intersections of the pages of the open book with $\mathcal{N} \setminus K$, and the flow depicted in Figure 1 is a dynamical representation of the restriction of ϕ to a strip.

1.3.2 The relative ECH With the notation above, if α is a contact form on Y which is compatible with K, in [5], the authors define relative versions $\mathrm{ECH}(N, \partial N, \alpha)$ and $\widehat{\mathrm{ECH}}(N, \partial N, \alpha)$ of embedded contact homology groups, and if α is also compatible with a Seifert surface S for K, they prove that

$$\mathrm{ECH}(N, \partial N, \alpha) \cong \mathrm{ECH}(Y, \alpha), \tag{5}$$

$$\widehat{\mathrm{ECH}}(N, \partial N, \alpha) \cong \widehat{\mathrm{ECH}}(Y, \alpha). \tag{6}$$

The notation suggests that these new homology groups are obtained by counting only orbits in N and quotienting by orbits on ∂N. Let us see the definition of these homologies in more detail.

In [5], the authors prove that it is possible to define the ECH–chain groups without taking into account the orbits in $\mathrm{int}(V)$ and in $T^2 \times (1,2)$, so that the only interesting orbits in $\mathcal{N}(K)$ are the four orbits above (plus, obviously, the empty orbit set). Moreover, the only curves counted by the (restriction of the) ECH–differential ∂ have projection on Y as depicted in Figure 2. These curves give the relations

$$\partial(e) = 0, \quad \partial(h) = 0, \quad \partial(h_+) = e + \varnothing, \quad \partial(e_+) = h. \tag{7}$$

Note that the two holomorphic curves from h to e, as well as the two from e_+ to h_+, cancel each other since we work with coefficients in $\mathbb{Z}/2$.

Observation 1.21 The compactification of the projection of the holomorphic curve that limits to the empty orbit set is topologically a disk with boundary h_+, which should be seen as a cylinder closing on some point of K. This curve contribute to the "$\varnothing$ part" of the third equation of (7), which gives $[e] = [\varnothing]$ in ECH–homology. In the rest of this manuscript, the fact that this disk is the only ECH index-1 connected holomorphic curve that crosses K will be essential.

Convention From now on, we will use the following notation. If (Y, α) is understood, given a submanifold $X \subset Y$ and a set of Reeb orbits $\{\gamma_1, \dots, \gamma_n\} \subset \mathcal{P}(Y \setminus X)$, we will denote by $\mathrm{ECC}^{\gamma_1,\dots,\gamma_n}(X, \alpha)$ the free $\mathbb{Z}/2$–module generated by orbit sets in $\mathcal{O}(X \sqcup \{\gamma_1, \dots, \gamma_n\})$.

Unless stated otherwise, the group $\mathrm{ECC}^{\gamma_1,\dots,\gamma_n}(X, \alpha)$ will come with the natural restriction of the ECH–differential of $\mathrm{ECC}(Y, \alpha)$, still denoted by ∂^{ECH}; if this restriction is

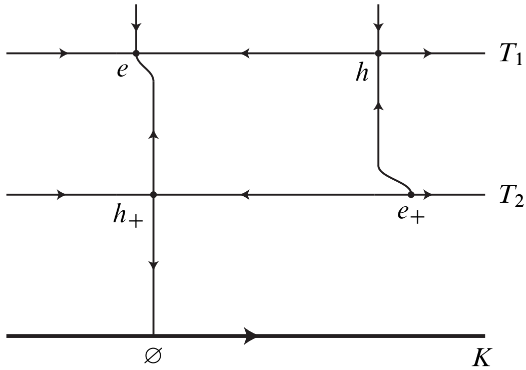


Figure 2: Orbits and holomorphic curves near K. Here the marked points denote the simple Reeb orbits, and the flow lines represent projections of the holomorphic curves counted by ∂^{ECH}. The two flow lines arriving from the top on e and h are depicted only to remember that, by the trapping lemma, holomorphic curves can only arrive at T_1.

still a differential, the associated homology is

$$\mathrm{ECH}^{\gamma_1,\dots,\gamma_n}(X,\alpha) := H_*(\mathrm{ECC}^{\gamma_1,\dots,\gamma_n}(X,\alpha), \partial^{\mathrm{ECH}}).$$

This notation is not used in [5], where the authors introduced a specific notation for each relevant ECH–group. In particular, with their notation,

$$\begin{aligned}\mathrm{ECC}^{\flat}(N,\alpha) &= \mathrm{ECC}^{e}(\mathrm{int}(N),\alpha),\\ \mathrm{ECC}^{\sharp}(N,\alpha) &= \mathrm{ECC}^{h}(\mathrm{int}(N),\alpha),\\ \mathrm{ECC}^{\natural}(N,\alpha) &= \mathrm{ECC}^{h_+}(N,\alpha).\end{aligned}$$

As mentioned before, even if there are other Reeb orbits in $\mathcal{N}$, it is possible to define chain complexes for the ECH homology of (Y,α) only taking into account the orbits $\{e,h,e_+,h_+\}$.

The blocking and trapping lemmas and the relations above imply that the restriction of the full ECH–differential of Y to the chain group $\mathrm{ECH}^{e_+,h_+}(N,\alpha)$ is given by

$$\partial(e_+^a h_+^b \gamma) = e_+^{a-1} h_+^b h\gamma + e_+^a h_+^{b-1}(1+e)\gamma + e_+^a h_+^b \partial\gamma, \tag{8}$$

where $\gamma \in \mathcal{O}(N)$ and where a term in the sum is meant to be zero if it contains some elliptic orbit with negative total multiplicity or a hyperbolic orbit with total multiplicity not in $\{0,1\}$; see [5, Section 9.5]. We remark that the blocking lemma implies also that $\partial\gamma \in \mathcal{O}(N)$.

The further restriction of the differential to $\mathrm{ECH}^{h_+}(N,\alpha)$ is then given by

$$\partial(h_+^b\gamma) = h_+^{b-1}(1+e)\gamma + h_+^b\partial\gamma. \tag{9}$$

Combining the computations of Sections 8 and 9 of [5] gives the following result.

Theorem 1.22 *Suppose that α is adapted to K and there exists a Seifert surface S for K such that α is adapted to S. Then*

$$\mathrm{ECH}(Y,\alpha) \cong \mathrm{ECH}^{e_+,h_+}(N,\alpha), \tag{10}$$

$$\widehat{\mathrm{ECH}}(Y,\alpha) \cong \mathrm{ECH}^{h_+}(N,\alpha). \tag{11}$$

Observation 1.23 It is important to remark that the empty orbit set is always taken into account as a generator of the groups above. This implies that if orbit sets with h_+ are considered, ∂^{ECH} counts also the holomorphic plane that contributes to the third of relations (7). Later we will give the definition of another differential, that we will call ∂^{ECK}, which is obtained from ∂^{ECH} by simply deleting that disk.

Define now the *relative embedded contact homology groups of* $(N,\partial N)$ by

$$\mathrm{ECH}(N,\partial N,\alpha) = \mathrm{ECH}^e(\mathrm{int}(N),\alpha)/\langle [e\gamma]\sim[\gamma]\rangle,$$
$$\widehat{\mathrm{ECH}}(N,\partial N,\alpha) = \mathrm{ECH}(N,\alpha)/\langle [e\gamma]\sim[\gamma]\rangle.$$

Since h_+ does not belong to the complexes $\mathrm{ECC}^e(\mathrm{int}(N),\alpha)$ and $\mathrm{ECC}(N,\alpha)$, the blocking lemma implies that the ECH–differentials count only holomorphic curves in N. This "lack" is balanced in the quotient by the equivalence relation

$$[e\gamma]\sim[\gamma]. \tag{12}$$

The reason behind this claim lies in the third of the relations (7). Indeed one can prove the following lemma; see Lemma 9.7.1 in [5].

Lemma 1.24 $\mathrm{ECH}^{e_+,h_+}(N,\alpha) \cong \mathrm{ECH}^{e_+}(N,\alpha)/\langle [e\gamma]\sim[\gamma]\rangle$.

Similarly, the fourth relation of (7) indicates why we can avoid considering h in the full $\mathrm{ECH}(Y,\alpha)$:

Lemma 1.25 [5, Lemma 9.9.1] $\mathrm{ECH}^{e_+}(N,\alpha) \cong \mathrm{ECH}^e(\mathrm{int}(N),\alpha)$.

Since $\partial(e\gamma) = e\partial(\gamma)$, the differential is compatible with the equivalence relation. So, instead of taking the quotient by $[e\gamma]\sim[\gamma]$ of the homology, we could take the homology of the quotient of the chain groups under the relation $e\gamma\sim\gamma$, and we would obtain the same homology groups. We will use this fact later. Note moreover that $[e^k] = [\varnothing]$ for every k. Equations (5) and (6) follow then from last two lemmas and Theorem 1.22.

1.4 ECH for knots

Let K be a homologically trivial knot in (Y, α). In this subsection, we recall the definition of a hat version of contact homology for the triple (K, Y, α). This was first defined by Colin, Ghiggini, Honda and Hutchings in [10, Section 7] as a particular case of *sutured* contact homology. On the other hand, following [5, Section 10], it is possible to proceed without dealing directly with sutures; we follow this approach here.

Let S be a Seifert surface for K. By standard arguments in homology, it is easy to compute that

$$H_1(Y \setminus K) \to H_1(Y) \times \mathbb{Z}, \quad [a] \mapsto \big(i_*[a], \langle a, S\rangle\big), \tag{13}$$

is an isomorphism. Here $i\colon Y \setminus K \to Y$ is the inclusion and $\langle a, S\rangle$ denotes the intersection number between a and S: this is a homological invariant of the pair (a, S) and is well-defined up to a slight perturbation of S (to make it transverse to a). Note that a preferred generator of $\mathbb{Z}$ is given by the homology class of a meridian for K, positively oriented with respect to the orientations of S and Y.

Example 1.26 If Y is a homology three-sphere, the number $\langle a, S\rangle$ depends only on a and K. This is the *linking number between a and K*, usually denoted by $\mathrm{lk}(a, K)$.

If $\gamma = \prod_i \gamma_i^{k_i}$ is a finite formal product of closed curves in $Y \setminus K$, then $\langle \gamma, S\rangle = \sum_i k_i \langle \gamma_i, S\rangle$.

Example 1.27 If (K, S, ϕ) is an open book decomposition of Y, α is an adapted contact form (in the sense of Thurston and Winkelnkemper) and $\gamma \in \mathcal{O}(Y \setminus K)$ is the orbit set $\prod_i \gamma_i^{k_i}$, then each γ_i is a periodic orbit of the diffeomorphism ϕ with *degree* d_i, and $\langle \gamma, S\rangle = \sum_i k_i d_i$.

Proposition 1.28 (see Proposition 7.1 in [10]) *Suppose that K is an orbit of R_α and let S be any Seifert surface for K. If γ and δ are two orbit sets in $Y \setminus K$ and $u\colon (F, j) \to (\mathbb{R} \times Y, J)$ is a holomorphic curve from γ to δ, then*

$$\langle \gamma, S\rangle \geq \langle \delta, S\rangle.$$

If α is a contact form adapted to K, a choice of (a homology class for) the Seifert surface S for K defines then a *knot filtration* on the chain complex $(\mathrm{ECC}^{h+}(N, \alpha), \partial^{\mathrm{ECH}})$

for $\widehat{\mathrm{ECH}}(Y,\alpha)$, where N is the complement of a neighborhood $\mathcal{N}(K)$ of K in which the only "interesting" orbits and holomorphic curves are those represented in Figure 2.

Let $\mathrm{ECC}^{h_+}_d(N,\alpha)$ be the free submodule of $\mathrm{ECC}^{h_+}(N,\alpha)$ generated by orbit sets γ in $\mathcal{O}(N \sqcup \{h_+\})$ such that $\langle \gamma, S\rangle = d$. Define moreover

$$\mathrm{ECC}^{h_+}_{\le d}(N,\alpha) := \bigoplus_{j\le d} \mathrm{ECC}^{h_+}_j(N,\alpha).$$

Definition 1.29 The *knot filtration* induced by K is the exhaustive filtration of the module $\mathrm{ECC}^{h_+}(N,\alpha)$ given by

$$\cdots \subseteq \mathrm{ECC}^{h_+}_{\le d-1}(N,\alpha) \subseteq \mathrm{ECC}^{h_+}_{\le d}(N,\alpha) \subseteq \mathrm{ECC}^{h_+}_{\le d+1}(N,\alpha) \subseteq \cdots.$$

The *filtration degree* of a generator γ of $\mathrm{ECC}^{h_+}_d(N,\alpha)$ is the integer d.

Corollary 1.30 *The* ECH*–differential respects the knot filtration.*

Proof The result follows by Proposition 1.28 applied to the MB buildings counted by ∂^{ECH}, which immediately implies that, for any $d \in \mathbb{Z}$,

$$\partial^{\mathrm{ECH}}(\mathrm{ECC}^{h_+}_{\le d}(N,\alpha)) \subseteq \mathrm{ECC}^{h_+}_{\le d}(N,\alpha). \qquad \square$$

If α is also adapted to S, in [5, Section 10.3], the authors prove that the filtration above induces a spectral sequence whose page ∞ is isomorphic to $\mathrm{ECH}^{h_+}(N,\alpha) \cong \widehat{\mathrm{ECH}}(Y,\alpha)$ and whose page 0 is the chain complex

$$\bigoplus_d \big(\mathrm{ECC}^{h_+}_d(N,\alpha), \partial^{\mathrm{ECK}}_d\big), \tag{14}$$

where $\mathrm{ECC}^{h_+}_d(N,\alpha) \cong \mathrm{ECC}^{h_+}_{\le d}(N,\alpha)/\mathrm{ECC}^{h_+}_{\le d-1}(N,\alpha)$ and $\partial^{\mathrm{ECK}}_d$ is the map on $\mathrm{ECC}^{h_+}_d(N,\alpha)$ induced by ∂^{ECH} on the quotient; ie it is the part of $\partial^{\mathrm{ECH}}|_{\mathrm{ECC}^{h_+}_d(N,\alpha)}$ that strictly preserves the filtration degree.

Observation 1.31 The proof of Proposition 1.28 implies that the holomorphic curves counted by ∂^{ECH} that strictly decrease the degree are exactly the curves that intersect K. So we can interpret ∂^{ECK} as the restriction of ∂^{ECH} (given by (8)) to the count of curves that do not cross a thin neighborhood of K. This is indeed the proper ECH–differential of the manifold $Y \setminus \mathrm{int}(V(K))$ (and not the restriction of the ECH–differential of Y to the orbit sets in $Y \setminus \mathrm{int}(V(K))$).

Note that, by definition of $\mathrm{ECC}^{h_+}(N,\alpha)$, all the holomorphic curves contained in $\mathbb{R}\times N$ strictly preserve the filtration degree. In fact the only holomorphic curve that contributes to $\partial^{\mathrm{ECH}}|_{\mathrm{ECC}^{h_+}(N,\alpha)}$ and decreases the degree (by 1) is the plane from h_+ to $\varnothing$. Then (9) gives

$$\partial(h_+^d\gamma) = h_+^{d-1}e\gamma + h_+^d\partial\gamma, \tag{15}$$

where $\gamma \in \mathcal{O}(N)$ and any term is meant to be zero if it contains some orbit with total multiplicity that is negative or, if the orbit is hyperbolic, not in $\{0,1\}$.

Definition 1.32 The *hat version of embedded contact (knot) homology* of the triple (K,Y,α) is

$$\widehat{\mathrm{ECK}}_*(K,Y,\alpha) := H_*(\mathrm{ECC}^{h_+}(N,\alpha), \partial^{\mathrm{ECK}}).$$

Observation 1.33 In order to define $\widehat{\mathrm{ECK}}(K,Y,\alpha)$, we supposed that α is compatible with S. This hypothesis is not present in the original definition (via sutures) in [10, Section 7.2]. Indeed, without this condition we can still apply all the arguments above and define the knot filtration on $\mathrm{ECC}^{h_+}(N,\alpha)$ exactly in the same way. Page 1 of the spectral sequence is again the well-defined homology in the definition above, and page ∞ is still isomorphic to $\mathrm{ECH}^{h_+}(N,\alpha)$. The only difference is that now we do not know that $\mathrm{ECH}^{h_+}(N,\alpha) \cong \widehat{\mathrm{ECH}}(Y,\alpha)$ since, in Theorem 1.22, the hypothesis that α is adapted to S is assumed.

This homology comes naturally with a further relative degree inherited by the filtered degree: if $\widehat{\mathrm{ECK}}_{*,d}(K,Y,\alpha) := H_*(\mathrm{ECC}_d^{h_+}(N,\alpha), \partial_d^{\mathrm{ECK}})$ then

$$\widehat{\mathrm{ECK}}_*(K,Y,\alpha) = \bigoplus_d \widehat{\mathrm{ECK}}_{*,d}(K,Y,\alpha).$$

Sometimes, in analogy with Heegaard Floer, we call this degree the *Alexander degree*.

Example 1.34 Suppose that (K,S,ϕ) is an open book decomposition of Y and that α is an adapted contact form. Since any nonempty Reeb orbit in $Y \setminus K$ has strictly positive intersection number with S,

$$\widehat{\mathrm{ECK}}_{*,0}(K,Y,\alpha) \cong \langle[\varnothing]\rangle_{\mathbb{Z}/2}.$$

This is the ECH–analogue of the fact that if K is fibered, then

$$\widehat{\mathrm{HFK}}_{*,-g}(K,Y) \cong \langle[c]\rangle_{\mathbb{Z}/2},$$

where g is the genus of K and c is the associated contact element; see Ozsváth and Szabó [30].

Observation 1.35 The Alexander degree can be considered as an absolute degree only once a relative homology class in $H_2(Y,K)$ for S has been fixed since the function $\langle\cdot, S\rangle$ defined on $H_1(Y\setminus K)$ changes if $[S]$ varies. On the other hand, if

$[\gamma] = [\delta] \in H_1(Y \setminus K)$ and $F \subset Y$ is a surface such that $\partial F = \gamma - \delta$, computations analogous to those in the proof of Proposition 1.28 imply that

$$\langle \gamma, S \rangle - \langle \delta, S \rangle = \langle F, K \rangle,$$

and the Alexander degree, considered as a relative degree, does not depend on the choice of a homology class for S. Obviously, if $H_2(Y) = 0$, the Alexander degree can be lifted to an absolute degree.

In [10] the authors conjectured that their sutured embedded contact homology is isomorphic to sutured Heegaard Floer homology. For knot complements, their conjecture becomes the following:

Conjecture 1.36 [10] For a homologically trivial knot K in Y,

$$\widehat{\mathrm{ECK}}(K, Y, \alpha) \cong \widehat{\mathrm{HFK}}(-K, -Y),$$

where α is a contact form on Y adapted to K.

2 Generalizations of $\widehat{\mathrm{ECK}}$

Let K be a homologically trivial knot in a contact three-manifold (Y, α). As recalled in Section 1.4, if α is adapted to K, a choice of a Seifert surface S for K induces a filtration on the chain complex $(\mathrm{ECC}^{h+}(N, \alpha), \partial^{\mathrm{ECH}})$, where $\mathrm{int}(N)$ is homeomorphic to $Y \setminus K$. Moreover, if α is also adapted to S, the homology of $(\mathrm{ECC}^{h+}(N, \alpha), \partial^{\mathrm{ECH}})$ is isomorphic to $\widehat{\mathrm{ECH}}(Y, \alpha)$, and the first page of the spectral sequence associated to the filtration is the hat version of embedded contact knot homology $\widehat{\mathrm{ECK}}(K, Y, \alpha)$. In this section, we generalize the knot filtration in two natural ways.

In Section 2.1, we extend to the chain complex $(\mathrm{ECC}^{h+,e+}(N, \alpha), \partial^{\mathrm{ECH}})$ the filtration induced by K. This filtration is defined in a way analogous to the hat case. We define the *full version of embedded contact knot homology* of (K, Y, α) to be the first page $\mathrm{ECK}(K, Y, \alpha)$ of the associated spectral sequence.

In Section 2.2, we generalize the knot filtration to n–component links L. The resulting homologies, defined in a similar way to the case of knots, are the *full and hat versions of embedded contact knot homologies of* (L, Y, α), which will be still denoted by $\mathrm{ECK}(L, Y, \alpha)$ and, respectively, $\widehat{\mathrm{ECK}}(L, Y, \alpha)$. Similarly to Heegaard Floer link homology, these homologies come endowed with an *Alexander (relative)* $\mathbb{Z}^n$*–degree*.

2.1 The full ECK

Let K be a homologically trivial knot in a contact three-manifold (Y, α) and suppose that α is adapted to K in the sense of Section 1.3. Recall in particular that there exist two concentric neighborhoods $V(K) \subset \mathcal{N}(K)$ of K whose boundaries are MB tori $T_1 = \partial\mathcal{N}(K)$ and $T_2 = \partial V(K)$ foliated by orbits of R_α in the homology class of meridians for K. These two families of orbits are modified into the two couples of orbits $\{e, h\}$ and, respectively, $\{e_+, h_+\}$.

Consider the chain complex $(\mathrm{ECC}^{e_+,h_+}(N,\alpha), \partial^{\mathrm{ECH}})$ where we recall that $N = Y \setminus \mathrm{int}(\mathcal{N}(K))$, $\mathrm{ECC}^{e_+,h_+}(N,\alpha) = \langle \mathcal{O}(N \sqcup \{h_+, e_+\})\rangle_{\mathbb{Z}/2}$ and ∂^{ECH} is the ECH–differential (obtained by restricting the differential on $\mathrm{ECC}(Y,\alpha)$) given by (8).

A Seifert surface S for K induces an *Alexander degree* $\langle \cdot, S\rangle$ on the generators of $\mathrm{ECC}^{h_+,e_+}(N,\alpha)$ exactly as in the case of $\mathrm{ECC}^{h_+}(N,\alpha)$. Let $\mathrm{ECC}^{h_+,e_+}_d(N,\alpha)$ be the submodule of $\mathrm{ECC}^{h_+,e_+}(N,\alpha)$ generated by the $\gamma \in \mathcal{O}(N) \sqcup \{h_+, e_+\}$ with $\langle \gamma, S\rangle = d$. If

$$\mathrm{ECC}^{h_+,e_+}_{\leq d}(N,\alpha) := \bigoplus_{j \leq d} \mathrm{ECC}^{h_+,e_+}_j(N,\alpha),$$

we have the exhaustive filtration

$$\cdots \subseteq \mathrm{ECC}^{h_+,e_+}_{\leq d-1}(N,\alpha) \subseteq \mathrm{ECC}^{h_+,e_+}_{\leq d}(N,\alpha) \subseteq \mathrm{ECC}^{h_+,e_+}_{\leq d+1}(N,\alpha) \subseteq \cdots$$

of $\mathrm{ECC}^{h_+,e_+}(N,\alpha)$. Proposition 1.28 again implies that ∂^{ECH} preserves the filtration. Let

$$\partial^{\mathrm{ECK}}_d \colon \mathrm{ECC}^{h_+,e_+}_d(N,\alpha) \to \mathrm{ECC}^{h_+,e_+}_d(N,\alpha)$$

be the part of ∂^{ECH} that strictly preserves the filtration degree d, that is, the differential induced by $\partial^{\mathrm{ECH}}|_{\mathrm{ECC}^{h_+,e_+}_{\leq d}(N,\alpha)}$ on the quotient

$$\mathrm{ECC}^{h_+,e_+}_{\leq d}(N,\alpha)/\mathrm{ECC}^{h_+,e_+}_{\leq d-1}(N,\alpha) = \mathrm{ECC}^{h_+,e_+}_d(N,\alpha).$$

Set

$$\partial^{\mathrm{ECK}} := \bigoplus_d \partial^{\mathrm{ECK}}_d \colon \mathrm{ECC}^{e_+,h_+}(N,\alpha) \to \mathrm{ECC}^{e_+,h_+}(N,\alpha).$$

Definition 2.1 We define the *full embedded contact knot homology of* (K, Y, α) by

$$\mathrm{ECK}(K, Y, \alpha) := H_*(\mathrm{ECC}^{e_+,h_+}(N,\alpha), \partial^{\mathrm{ECK}}).$$

Note that, as in the hat case, the only holomorphic curves counted by ∂^{ECH} that do not strictly respect the filtration degree are the curves that contain the plane from h_+ to $\varnothing$; see Observation 1.31. Recalling the expression of ∂^{ECH} given in (8), it follows that ∂^{ECK} is given by

$$\partial^{\mathrm{ECK}}(e_+^a h_+^b \gamma) = e_+^{a-1} h_+^b h\gamma + e_+^a h_+^{b-1} e\gamma + e_+^a h_+^b \partial\gamma, \tag{16}$$

where $\gamma \in \mathcal{O}(N)$ and any term is meant to be 0 if it contains an orbit with total multiplicity that is negative or, if the orbit is hyperbolic, not in $\{0, 1\}$.

Again the homology comes with an *Alexander degree*, which is well-defined once the homology class for S is fixed, and induces the natural splitting

$$\mathrm{ECK}_*(K,Y,\alpha) \cong \bigoplus_{d\in\mathbb{Z}} \mathrm{ECK}_{*,d}(K,Y,\alpha), \tag{17}$$

where

$$\mathrm{ECK}_{*,d}(K,Y,\alpha) := H_*(\mathrm{ECC}_d^{h+,e+}(N,\alpha), \partial_d^{\mathrm{ECK}}).$$

Lemma 2.2 *If $\mathcal{N}(K)$ is a neighborhood of K as above, then*

$$\mathrm{ECK}(K,Y,\alpha) \cong \mathrm{ECH}(Y \setminus \mathcal{N}(K), \alpha).$$

Proof Reasoning as in Lemma 1.24, it is easy to prove that

$$\begin{aligned} \mathrm{ECK}(K,Y,\alpha) &\cong H_*(\mathrm{ECC}^{e+,h+}(N,\alpha), \partial^{\mathrm{ECK}}) \\ &\cong H_*(\mathrm{ECC}^{e,h+}(\mathrm{int}(N),\alpha), \partial^{\mathrm{ECK}}) \\ &\cong H_*(\mathrm{ECC}(\mathrm{int}(N),\alpha), \partial^{\mathrm{ECK}}) \\ &\cong \mathrm{ECH}(\mathrm{int}(N),\alpha), \end{aligned}$$

which follows from the fact that $\partial^{\mathrm{ECK}}(\gamma) = \partial^{\mathrm{ECH}}(\gamma)$ for all $\gamma \in \mathcal{O}(N)$. □

Observation 2.3 Note that so far we only assumed that α is compatible with K, while we did not assume the condition

(♠) α is compatible with a Seifert surface S for K.

As remarked in Observation 1.33, we cannot prove Theorem 1.22 without (♠), and so we do not know if the spectral sequence whose 0–page is the ECK–chain complex limits to $\mathrm{ECH}(Y,\alpha)$. On the other hand, this spectral sequence is in any case well-defined, and so is $\mathrm{ECK}(K,Y,\alpha)$. Even if, in light of Lemma 1.19, we could assume (♠) here without restrictions on K, we prefer to avoid it in the general definition of $\mathrm{ECK}(K,Y,\alpha)$ in order to consider a wider class of contact forms.

In analogy with Conjecture 1.36 we state:

Conjecture 2.4 For any knot K in Y and any contact form α on Y adapted to K,

$$\mathrm{ECK}(K,Y,\alpha) \cong \mathrm{HFK}^+(-K,-Y).$$

2.2 The generalization to links

In this subsection, we extend the definitions of ECK and $\widehat{\mathrm{ECK}}$ to the case of homologically trivial links with more than one component. A (*strongly*) *homologically trivial*

n–link in Y is a disjoint union of n knots, each of which is homologically trivial in Y. Suppose that

$$L = K_1 \sqcup \cdots \sqcup K_n$$

is a homologically trivial n–link in Y. We say that a contact form α on Y is *adapted to* L if it is adapted to K_i for each i.

Lemma 2.5 *For any link L and contact structure ξ on Y there exists a contact form compatible with ξ which is adapted to L.*

Proof The proof of part (1) of Lemma 1.19 is local near the knot K and can then be applied recursively to each K_i. □

Fix $L = K_1 \sqcup \cdots \sqcup K_n$ homologically trivial and α an adapted contact form. Since α is adapted to each K_i, there exist pairwise disjoint tubular neighborhoods

$$V(K_i) \subset \mathcal{N}(K_i)$$

of K_i where α behaves exactly like in the neighborhoods $V(K) \subset \mathcal{N}(K)$ in Section 1.3. In particular, for each i, the tori $T_{i,1} := \partial\mathcal{N}(K_i)$ and $T_{i,2} := \partial V(K_i)$ are MB and foliated by families of orbits of R_α in the homology class of a meridian of K_i. We will consider these two families as perturbed into two pairs $\{e_i, h_i\}$ and $\{e_i^+, h_i^+\}$ in the usual way. Let

$$V(L) := \bigsqcup_i V(K_i) \quad \text{and} \quad \mathcal{N}(L) := \bigsqcup_i \mathcal{N}(K_i),$$

and set

$$N := Y \setminus \operatorname{int}(\mathcal{N}(L)).$$

Define moreover $\overline{e} := \bigsqcup_i e_i$, and let $\overline{h}$, $\overline{e}_+$ and $\overline{h}_+$ be similarly defined.

Consider $\mathrm{ECC}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ endowed with the restriction ∂^{ECH} of the ECH differential of (Y,α), and let $\mathrm{ECH}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ be the associated homology.

Lemma 2.6 $\mathrm{ECH}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ *is well-defined and the curves counted by ∂^{ECH} inside each $\mathcal{N}(K_i)$ are given by analogous expressions to those in (7).*

Proof The blocking and trapping lemmas can be applied locally near each component of ∂N and the proofs of Lemmas 7.1.1 and 7.1.2 in [5] work immediately in this context too. This imply that the homology of $(\mathrm{ECC}(N,\alpha), \partial^{\mathrm{ECH}})$ is well-defined.

Again the blocking and trapping lemmas, together with the local homological arguments in Lemmas 9.5.1 and 9.5.3 in [5], imply that the only holomorphic curves counted by ∂^{ECH} inside each $\mathcal{N}(K_i)$ are as required (see Figure 2), and so $\mathrm{ECH}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ is well-defined. □

An explicit formula for ∂^{ECH} can be obtained by generalizing (8) in the obvious way.

For each $i \in \{1, \dots, n\}$, fix a (homology class for a) Seifert surface S_i for K_i. These surfaces are not necessarily pairwise disjoint, and it is even possible that $S_i \cap K_j \neq \varnothing$ for some $i \neq j$. Consider then the *Alexander* $\mathbb{Z}^n$*–degree* on $\mathrm{ECH}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ given by the function

$$\mathrm{ECC}^{\overline{e}_+,\overline{h}_+}(N,\alpha) \to \mathbb{Z}^n, \quad \gamma \mapsto (\langle \gamma, S_1\rangle, \dots, \langle \gamma, S_n\rangle). \tag{18}$$

Define a partial ordering on $\mathbb{Z}^n$ by

$$(a_1, \dots, a_n) \le (b_1, \dots, b_n) \qquad \iff \qquad a_i \le b_i \quad \text{for all } i.$$

Proposition 1.28 applied to each K_i implies that if γ and δ are two orbit sets in $\mathcal{O}(N \sqcup \{\overline{e}_+, \overline{h}_+\})$, then for any k,

$$\mathcal{M}_k(\gamma,\delta)/\mathbb{R} \neq 0 \implies (\langle \delta, S_1\rangle, \dots, \langle \delta, S_n\rangle) \le (\langle \gamma, S_1\rangle, \dots, \langle \gamma, S_n\rangle).$$

This implies that ∂^{ECH} does not increase the Alexander degree, which induces a $\mathbb{Z}^n$–filtration on $(\mathrm{ECC}^{\overline{e}_+,\overline{h}_+}(N,\alpha), \partial^{\mathrm{ECH}})$. Like in the previous subsection, we are interested in the part of ∂^{ECH} that strictly respects the filtration degree. This can be defined again in terms of quotients as follows.

Let $d \in \mathbb{Z}^n$ and let $\mathrm{ECC}_d^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ be the submodule of $\mathrm{ECC}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ freely generated by orbit sets $\gamma \in \mathcal{O}(N \sqcup \{\overline{e}_+, \overline{h}_+\})$ such that

$$(\langle \gamma, S_1\rangle, \dots, \langle \gamma, S_n\rangle) = d.$$

Define

$$\mathrm{ECC}_{\le d}^{\overline{e}_+,\overline{h}_+}(N,\alpha) := \bigoplus_{j \le d} \mathrm{ECC}_d^{\overline{e}_+,\overline{h}_+}(N,\alpha),$$

and let $\mathrm{ECC}_{<d}^{\overline{e}_+,\overline{h}_+}(N,\alpha)$ be similarly defined.

Define the *full* ECK*–differential in degree* d to be the map

$$\partial_d^{\mathrm{ECK}} \colon \mathrm{ECC}_d^{\overline{e}_+,\overline{h}_+}(N,\alpha) \to \mathrm{ECC}_d^{\overline{e}_+,\overline{h}_+}(N,\alpha)$$

induced by $\partial^{\mathrm{ECH}}|_{\mathrm{ECC}_{\le d}^{\overline{e}_+,\overline{h}_+}(N,\alpha)}$ on the quotient

$$\mathrm{ECC}_{\le d}^{\overline{e}_+,\overline{h}_+}(N,\alpha)/\mathrm{ECC}_{<d}^{\overline{e}_+,\overline{h}_+}(N,\alpha) \cong \mathrm{ECC}_d^{\overline{e}_+,\overline{h}_+}(N,\alpha).$$

Define then the *full* ECK*–differential* by

$$\partial^{\mathrm{ECK}} := \bigoplus_d \partial_d^{\mathrm{ECK}} \colon \mathrm{ECC}^{\overline{e}_+,\overline{h}_+}(N,\alpha) \to \mathrm{ECC}^{\overline{e}_+,\overline{h}_+}(N,\alpha).$$

Observation 2.7 Observing the form of ∂^{ECH}, it is easy again to see that the only holomorphic curves that are counted by ∂^{ECH} and not by ∂^{ECK} are the ones containing a holomorphic plane from some h_i^+ to $\varnothing$.

Definition 2.8 The *full embedded contact knot homology of* (L, Y, α) is

$$\mathrm{ECK}(L, Y, \alpha) := H_*(\mathrm{ECC}^{\overline{e}+,\overline{h}+}(N, \alpha), \partial^{\mathrm{ECK}}).$$

The fact that $\mathrm{ECK}(L, Y, \alpha)$ is well-defined is a direct consequence of the good definition of $\mathrm{ECH}^{\overline{e}+,\overline{h}+}(N, \alpha)$ and the fact that ∂^{ECH} respects the Alexander filtration. Note that we have again a natural splitting

$$\mathrm{ECK}_*(L, Y, \alpha) = \bigoplus_{d \in \mathbb{Z}^n} \mathrm{ECK}_{*,d}(L, Y, \alpha), \tag{19}$$

where

$$\mathrm{ECK}_{*,d}(L, Y, \alpha) = H_*(\mathrm{ECC}_d^{\overline{e}+,\overline{h}+}(N, \alpha), \partial_d^{\mathrm{ECK}}).$$

The proof of the following lemma is the same of that of the analogous Lemma 2.2 for knots applied to each component of L.

Lemma 2.9 *If $\mathcal{N}(L)$ is a neighborhood of L as above, then*

$$\mathrm{ECK}(L, Y, \alpha) \cong \mathrm{ECH}(Y \setminus \mathcal{N}(L), \alpha).$$

Consider now the submodule $\mathrm{ECC}^{\overline{h}+}(N, \alpha)$ of $\mathrm{ECC}^{\overline{e}+,\overline{h}+}(N, \alpha)$ endowed with the restriction of ∂^{ECH}. Observe that its homology $\mathrm{ECH}^{\overline{h}+}(N, \alpha)$ is well-defined. Filtering $(\mathrm{ECC}^{\overline{h}+}(N, \alpha), \partial^{\mathrm{ECH}})$ by the Alexander degree, for any $d \in \mathbb{Z}^n$, we can define $\mathrm{ECC}_d^{\overline{h}+}(N, \alpha)$ with differential

$$\partial_d^{\mathrm{ECK}}\colon \mathrm{ECC}_d^{\overline{h}+}(N, \alpha) \to \mathrm{ECC}_d^{\overline{h}+}(N, \alpha).$$

Definition 2.10 The *hat version of embedded contact knot homology of* (L, Y, α) is

$$\widehat{\mathrm{ECK}}(L, Y, \alpha) := H_*(\mathrm{ECC}^{\overline{h}+}(N, \alpha), \partial^{\mathrm{ECK}}).$$

Observation 2.7 and a splitting like the one in (19) hold also for $\widehat{\mathrm{ECK}}(L, Y, \alpha)$. Moreover, it is easy to see that if L has only one connected component, we get the same theories of Sections 1.4 and 2.1.

Conjecture 2.11 If L is a link in Y and α is any contact form on Y adapted to L, then

$$\mathrm{ECK}(L, Y, \alpha) \cong \mathrm{HFK}^+(-L, -Y),$$
$$\widehat{\mathrm{ECK}}(L, Y, \alpha) \cong \widehat{\mathrm{HFK}}(L, Y).$$

Convention In order to simplify the notation in the rest of the paper, we will indicate the ECH chain groups for the knot embedded contact homology groups of links and knots by

$$\mathrm{ECC}(L, Y, \alpha) := \mathrm{ECC}^{\overline{e}+,\overline{h}+}(N, \alpha),$$
$$\widehat{\mathrm{ECC}}(L, Y, \alpha) := \mathrm{ECC}^{\overline{h}+}(N, \alpha).$$

These groups will implicitly come endowed with the differential ∂^{ECK}.

We end this section by saying some words about a further generalization of ECK to *weakly homologically trivial* links. We say that $L \subset Y$ is a weakly homologically trivial (or simply *weakly trivial*) n–component link if there exist surfaces with boundary $S_1, \dots, S_m \subset Y$ with $m \leq n$ and such that $\partial S_i \cap \partial S_j = \varnothing$ if $i \neq j$ and $\bigsqcup_{i=1}^m \partial S_i = L$. Also, here we do not require that S_i or even ∂S_i is disjoint from S_j for $j \neq i$. Clearly, L is a strongly trivial link if and only if it is weakly trivial with $m = n$.

If L is a weakly trivial link with $m \lneq n$, we cannot in general define a homology with a filtered n–degree. In fact, there exists $S \in \{S_1, \dots, S_m\}$ such that ∂S has more than one connected component. Suppose for instance that $\partial S = K_1 \sqcup K_2$. The arguments behind Proposition 1.28 imply that if $u\colon (F, j) \to (\mathbb{R} \times Y, J)$ is a holomorphic curve from γ to δ, then

$$\langle \gamma, S \rangle - \langle \delta, S \rangle = \langle \mathrm{Im}(u), \mathbb{R} \times (K_1 \sqcup K_2) \rangle \geq 0.$$

So in this case, we can still apply the arguments above and get well-defined ECH invariants for L. However, this time they will come only with a filtered (relative) $\mathbb{Z}^m$–degree on the generators γ of an ECH complex of Y, which is given by the m–tuple $(\langle \gamma, S_1 \rangle, \dots, \langle \gamma, S_m \rangle)$.

Example 2.12 Let (L, S, ϕ) be an open book decomposition of Y with possibly disconnected boundary. Using a (connected) page of (L, S, ϕ) to compute the Alexander degree, the generators of the chain complex for $\mathrm{ECK}_d(L, Y, \alpha)$ are d–periodic orbits of the diffeomorphism ϕ for any $d \in \mathbb{Z}$.

3 Euler characteristics

In this section, we compute the graded Euler characteristics of the embedded contact homology groups for knots and links in homology three-spheres Y with respect to suitable contact forms. The computations will be done in terms of the Lefschetz zeta function of the flow of the Reeb vector field.

Before proceeding, we briefly recall what the graded Euler characteristic is. Given a collection of chain complexes

$$(C, \partial) = \big\{(C_{*,(i_1,\dots,i_n)}, \partial_{(i_1,\dots,i_n)})\big\}_{(i_1,\dots,i_n) \in \mathbb{Z}^n},$$

where $*$ denotes a relative homological degree, its *graded Euler characteristic* is

$$\chi(C) = \sum_{i_1,\dots,i_n} \chi(C_{*,(i_1,\dots,i_n)}) t_1^{i_1} \cdots t_n^{i_n} \in \mathbb{Z}[t_1^{\pm 1}, \dots, t_n^{\pm 1}],$$

where $\chi(C_{*,(i_1,\dots,i_n)})$ is the standard Euler characteristic of $C_{*,(i_1,\dots,i_n)}$, and the t_j are formal variables. By definition, $\chi(C)$ is a Laurent polynomial, and the properties of the standard Euler characteristic imply

$$\chi(C) = \chi(H(C,\partial)).$$

In this case, the homology $H(C,\partial)$ is a *categorification* of $\chi(C)$.

The most important result of this section relates the Euler characteristic of ECK homologies of a link in S^3 with its multivariable Alexander polynomial.

Theorem 3.1 *Let L be any n–link in S^3. Then there exists a contact form α adapted to L such that*

$$\chi(\mathrm{ECK}(L,S^3,\alpha)) \doteq \begin{cases} \Delta_L(t_1,\dots,t_n) & \text{if } n>1,\\ \Delta_L(t)/(1-t) & \text{if } n=1,\end{cases} \tag{20}$$

and

$$\chi(\widehat{\mathrm{ECK}}(L,S^3,\alpha)) \doteq \begin{cases} \Delta_L\cdot\prod_{i=1}^n(1-t_i) & \text{if } n>1,\\ \Delta_L(t) & \text{if } n=1.\end{cases} \tag{21}$$

This theorem implies that ECK *categorifies the Alexander polynomial of knots and links in* S^3. This is the third known categorification of this kind, after the ones obtained in Heegaard Floer homology by Ozsváth and Szabó [28; 31] and Rasmussen [33] and in Seiberg–Witten–Floer homology by Kronheimer and Mrowka [24; 23].

An immediate consequence of Theorem 3.1 and Equations (1) and (2) is:

Corollary 3.2 *For any link L in S^3, there exists a contact form α such that*

$$\chi(\mathrm{ECK}(L,S^3,\alpha)) \doteq \chi(\mathrm{HFL}^+(-L,-S^3)),$$
$$\chi(\widehat{\mathrm{ECK}}(L,S^3,\alpha)) \doteq \chi(\widehat{\mathrm{HFL}}(-L,-S^3)).$$

This corollary implies that Conjecture 2.11 (which generalizes Conjectures 1.36 and 2.4) holds for links in S^3 at least at the level of Euler characteristic.

A key ingredient to prove Theorem 3.1 is the dynamical formulation of the Alexander quotient given by Fried [14].

3.1 A dynamical formulation of the Alexander polynomial

Given any link $L = K_1\sqcup\cdots\sqcup K_n$ in S^3, we can associate to it its *multivariable Alexander polynomial*

$$\Delta_L(t_1,\dots,t_n) \in \mathbb{Z}[t_1^{\pm1},\dots,t_n^{\pm1}]/\langle \pm t_1^{a_1}\cdots t_n^{a_n}\rangle$$

with $a_i \in \mathbb{Z}$. The quotient means that the Alexander polynomial is well-defined only up to multiplication by monomials of the form $\pm t_1^{a_1} \cdots t_n^{a_n}$.

A slightly simplified version is the *(classical) Alexander polynomial* $\Delta_L(t)$ defined by setting $t_1 = \cdots = t_n = t$, ie

$$\Delta_L(t) := \Delta_L(t, \dots, t).$$

If L is a knot, the two notions obviously coincide.

There are many possible definitions of the Alexander polynomial Δ_L. In this section, we give a formulation of Δ_L in terms of the dynamics of suitable vector fields in $S^3 \setminus L$.

The fact that the Alexander polynomial is related to dynamical properties of its complement in S^3 originates with the study of fibrations of S^3. For example, A'Campo [1] studied the *twisted Lefschetz zeta function* of the monodromy of an open book decomposition (S, ϕ) of S^3 associated to a Milnor fibration of a complex algebraic singularity. More generally, if (K, S, ϕ) is any open book decomposition of S^3, one can easily prove that

$$\Delta_K(t) \doteq \det(\mathbb{1} - t\phi_*^1),$$

where $\mathbb{1}$ and ϕ_*^1 are the identity map and, respectively, the application induced by ϕ, on $H_1(S, \mathbb{Z})$. The basic idea in this context is to express the right-hand side of the above equation in terms of traces of iterations of ϕ_*^1, then to apply the Lefschetz fixed point theorem to get expressions in terms of periodic points (ie periodic orbits) for the flow of some vector field in $S^3 \setminus K$ whose first return on a page is ϕ.

Suppose now that L is not a fibered link, so its complement is not globally fibered over S^1, and let R be a vector field in $S^3 \setminus L$. If one wants to apply arguments as above, it is necessary to decompose $S^3 \setminus L$ in "fibered-like" pieces with respect to R, in which it is possible to define at least a local first return map of the flow ϕ_R of R. Obviously, some condition on R is required. For example, Franks [13] considers *Smale vector fields*, ie vector fields with one-dimensional and hyperbolic chain recurrent set; see [35].

Here we are more interested in the approach used by Fried [14]. Consider a three-dimensional manifold X. Any *abelian cover* $\widetilde{X} \xrightarrow{\pi} X$ with deck-transformation group isomorphic to a fixed abelian group G is uniquely determined by the choice of a class $\rho = \rho(\pi) \in H^1(X, G) \cong \mathrm{Hom}(H_1(X, \mathbb{Z}), G)$. Here ρ is determined by the following property: for any $[\gamma] \in H_1(X)$, if $\widetilde{\gamma}\colon [0,1] \to \widetilde{X}$ is any lifting of the loop $\gamma\colon [0,1] \to X$, then $\rho([\gamma])$ is determined by $\rho([\gamma])(\widetilde{\gamma}(0)) = \widetilde{\gamma}(1)$.

Since the correspondence between abelian covers and cohomology classes is bijective, with abuse of notation sometimes we will refer to an abelian cover directly by identifying it with the corresponding ρ.

Example 3.3 The *universal abelian cover* of X is the abelian cover with deck-transformation group $G = H_1(X, \mathbb{Z})$ and corresponding to $\rho = \mathrm{id}$.

Example 3.4 Let $L = K_1 \sqcup \cdots \sqcup K_n$ be an n–component link in a three-manifold Y such that K_i is homologically trivial for any i, and fix a Seifert surface S_i for K_i. Let moreover μ_i be a positive meridian for K_i. If $i\colon Y \setminus L \hookrightarrow Y$ is the inclusion, the isomorphism

$$H_1(Y \setminus L) \to H_1(Y) \oplus \mathbb{Z}_{[\mu_1]} \oplus \cdots \oplus \mathbb{Z}_{[\mu_n]}, \quad [\gamma] \mapsto (i_*([\gamma]), \langle \gamma, S_1 \rangle, \ldots, \langle \gamma, S_n \rangle),$$

gives rise naturally to the abelian cover

$$\rho_L \in \mathrm{Hom}(H_1(Y \setminus L, \mathbb{Z}),\ \mathbb{Z}^n)$$

of $Y \setminus L$ defined by

$$\rho_L([\gamma]) = (\langle \gamma, S_1 \rangle, \ldots, \langle \gamma, S_n \rangle).$$

Setting $t_i = [\mu_i] \in H_1(Y \setminus L, \mathbb{Z})$, we can regard $\rho_L([\gamma])$ as a monomial in the t_i:

$$\rho_L([\gamma]) = t_1^{\langle \gamma, S_1 \rangle} \cdots t_n^{\langle \gamma, S_n \rangle}.$$

In the rest of the paper, we will often use this notation. Note finally that if Y is a homology three-sphere, ρ_L coincides with the universal abelian cover of $Y \setminus L$.

If R is a vector field on X satisfying some compatibility condition with ρ (and with ∂X if this is nonempty), Fried relates the Reidemeister–Franz torsion of $(X, \partial X)$ with the *twisted Lefschetz zeta function* of the flow ϕ_R.

3.1.1 Twisted Lefschetz zeta function of flows Let R be a vector field on X and γ a closed isolated orbit of ϕ_R. Pick any point $x \in \gamma$ and let D be a small disk transverse to γ such that $D \cap \gamma = \{x\}$. With this data it is possible to define the Lefschetz sign of γ exactly like we did in Section 1.1.1 for orbits of Reeb vector fields associated to a contact structure ξ, but using now $T_x D$ instead of ξ_x. Indeed, it is possible to prove that the Lefschetz sign of γ does not depend on the choice of x and D, and that it is an invariant $\epsilon(\gamma) \in \{-1, 1\}$ of ϕ_R near γ.

Definition 3.5 The *local Lefschetz zeta function* of ϕ_R near γ is the formal power series $\zeta_\gamma(t) \in \mathbb{Z}[\![t]\!]$ defined by

$$\zeta_\gamma(t) := \exp\biggl(\sum_{i \geq 1} \epsilon(\gamma^i) \frac{t^i}{i}\biggr).$$

Let now $\widetilde{X} \xrightarrow{\pi} X$ be an abelian cover with deck-transformation group G, and let $\rho = \rho(\pi) \in H^1(X, G)$. Suppose that all the periodic orbits of ϕ_R are isolated.

Definition 3.6 We define the *ρ–twisted Lefschetz zeta function of ϕ_R* by

$$\zeta_\rho(\phi_R) := \prod_\gamma \zeta_\gamma(\rho([\gamma])),$$

where the product is taken over the set of simple periodic orbits of ϕ_R.

When ρ is understood, we will write directly $\zeta(\phi_R)$ and we will call it the twisted Lefschetz zeta function of ϕ_R.

We remark that in [14], the author defines $\zeta_\rho(\phi_R)$ in a slightly different way, and then he proves in Theorem 2 that the two definitions coincide.

Convention Suppose that $\rho \in H^1(X, \mathbb{Z}^n)$ is an abelian cover of X and chose a generator $(t_1, \dots, t_n)$ of $\mathbb{Z}^n$. Then, with a similar notation to that of Example 3.4, we will often identify $\zeta_\rho(\phi_R)$ with an element of $\mathbb{Z}[\![t_1^{\pm 1}, \dots, t_n^{\pm 1}]\!]$.

3.1.2 Torsion and flows Fried [14] relates the Reidemeister torsion of an abelian cover ρ of a (not necessarily closed) three-manifold X with the twisted Lefschetz zeta function of certain flows. In particular, in Section 5, he considers a kind of torsion that he calls the *Alexander quotient* and denotes it by $\mathrm{ALEX}_\rho(X)$: the reason for the "quotient" comes from the fact that Fried uses a definition of Reidemeister torsion only up to the choice of a sign (this is the "refined Reidemeister torsion" of Turaev [38]), while $\mathrm{ALEX}_\rho(X)$ is defined up to an element in the abelian group of deck transformations of ρ.

In fact, one can check that $\mathrm{ALEX}_\rho(X)$ is exactly the Reidemeister–Franz torsion τ considered by Ozsváth and Szabó [31]. In particular, when X is the complement of an n–component link L in S^3 and ρ is the universal abelian cover of X, then

$$\mathrm{ALEX}(S^3 \setminus L) \doteq \begin{cases} \Delta_L(t_1, \dots, t_n) & \text{if } n > 1, \\ \Delta_L(t)/(1-t) & \text{if } n = 1, \end{cases} \tag{22}$$

where we removed $\rho = \mathrm{id}_{H_1(S^3 \setminus L, \mathbb{Z})}$ from the notation; see [14, Section 8] and [38].

Since the notation "τ" is ambiguous, we follow Fried [14] and we refer to the Reidemeister–Franz torsion as the Alexander quotient, indicated by $\mathrm{ALEX}_\rho(X)$.

In order to relate $\mathrm{ALEX}_\rho(X)$ to the twisted Lefschetz zeta function of the flow ϕ_R of a vector field R, Fried assumes some hypotheses on R. The first condition that R must satisfy is *circularity*.

Definition 3.7 A vector field R on X is *circular* if there exists a C^1 map $\theta\colon X \to S^1$ such that $d\theta(R) > 0$.

If $\partial X = \varnothing$, this is equivalent to say that R admits a global cross section. Intuitively, the circularity condition on R allows us to define a kind of first return map of ϕ_R. Suppose that R circular, and consider $S^1 \cong \mathbb{R}/\mathbb{Z}$ with $\mathbb{R}$–coordinate t. The cohomology class

$$u_\theta := \theta^*([dt]) \in H^1(X, \mathbb{Z})$$

is then well-defined.

Definition 3.8 Given an abelian cover $\widetilde{X} \xrightarrow{\pi} X$ with deck-transformation group G, let $\rho = \rho(\pi) \in H^1(X, G)$ be the corresponding cohomology class. A circular vector field R on X is *compatible* with ρ if there exists a homomorphism $v\colon G \to \mathbb{R}$ such that $v \circ \rho = u_\theta$, where θ and u_θ are as above.

Example 3.9 The universal abelian cover corresponds to $\rho = \mathrm{id}\colon H_1(X, \mathbb{Z}) \to H_1(X, \mathbb{Z})$, and it is compatible with any circular vector field on X.

The following theorem is not the most general result in [14], but it will be enough for our purposes.

Theorem 3.10 [14, Theorem 7] *Let X be a three-manifold and $\rho \in H^1(X, G)$ an abelian cover. Let R be a nonsingular, circular and nondegenerate vector field on X compatible with ρ. Suppose moreover that, if $\partial X \neq \varnothing$, then R is transverse to ∂X and pointing out of X. Then*

$$\mathrm{ALEX}_\rho(X) \doteq \zeta_\rho(\phi_R),$$

where the equivalence $\doteq$ is up to multiplication by $\pm g$ for any $g \in G$.

An immediate consequence is the following:

Corollary 3.11 *If L is an n–component link in S^3, let $\mathcal{N}(L)$ be a tubular neighborhood of L, and let $N = S^3 \setminus \mathcal{N}(L)$. Let R be a nonsingular circular vector field on N, transverse to ∂N and pointing out of N. Then*

$$\zeta(\phi_R) \doteq \begin{cases} \Delta_L(t_1, \dots, t_n) & \text{if } n > 1, \\ \Delta_L(t)/(1-t) & \text{if } n = 1. \end{cases} \tag{23}$$

3.2 Results

In the next subsections, we prove Theorem 3.1, which will be obtained as a consequence of the following more general result. Recall that an n–link $L \subset Y$ determines the abelian cover $\rho_L \in H^1(Y \setminus L, \mathbb{Z}^n)$ of $Y \setminus L$ given in Example 3.4. When Y is a homology three-sphere, we have

$$\rho_L \equiv \mathbb{1}\colon H_1(Y \setminus L) \to H_1(Y \setminus L) \cong \mathbb{Z}^n.$$

In order to simplify the notations, we remove ρ_L from the notation of the Alexander quotient and of the twisted Lefschetz zeta function:

$$\mathrm{ALEX}(Y \setminus L) := \mathrm{ALEX}_{\mathbb{1}}(Y \setminus L), \quad \zeta(\phi) := \zeta_{\mathbb{1}}(\phi).$$

Let $(t_1, \dots, t_n)$ be a basis for $H_1(Y \setminus L)$, where $[\mu_i] = t_i$ for μ_i a positively oriented meridian of K_i.

Theorem 3.12 *Let L be an n–link in a homology three-sphere Y. Then there exists a contact form α such that*

$$\chi(\mathrm{ECK}(L, Y, \alpha)) \doteq \mathrm{ALEX}(Y \setminus L).$$

The proofs of Theorems 3.1 and 3.12 will be carried out in two main steps: in Section 3.3, we will prove the theorems in the case of fibered links, while the general case will be treated in Section 3.4.

3.3 Fibered links

In this subsection, we prove Theorems 3.1 and 3.12 for fibered links. Let (L, S, ϕ) be an open book decomposition of a homology three-sphere Y, and let α be an adapted contact form on Y. In particular, with our definition, α is also adapted to L.

In order to prove the theorems above, we want to express the Euler characteristic $\chi(\mathrm{ECK}(L, Y, \alpha))$ in terms of the twisted Lefschetz zeta function of the Reeb flow ϕ_R of $R = R_\alpha$ and then apply Theorem 3.10. The first thing that one should do is then to check if ϕ_R and ρ_L satisfy the hypotheses of that theorem. Unfortunately, this is not the case. The needed properties are, in fact, that R is

(1) nonsingular and circular;

(2) compatible with ρ_L;

(3) nondegenerate;

(4) transverse to $\partial V(L)$ and pointing out of $Y \setminus \mathring{V}(L)$, where $\mathring{V}(L) = \mathrm{int}(V(L))$.

In our situation, only properties (1) and (2) are satisfied. Indeed, by the definition of open book decomposition, there is a natural fibration $\theta\colon Y \setminus \mathring{V}(L) \to S^1 \cong \mathbb{R}/\mathbb{Z}$ such that the surfaces $\theta^{-1}(t)$ are the pages of the open book. The fact that α is adapted to (L, S, ϕ) implies that R is always positively transverse to the pages. This evidently implies that $d\theta(R) > 0$, so R is circular. The fact that R is compatible with ρ_L (that coincides with the universal abelian cover of $Y \setminus \mathring{V}(L)$) comes from Example 3.9.

On the other hand, properties (3) and (4) above are not satisfied. Indeed, after the MB perturbation of T_2, the vector field R is tangent to $\partial V(L)$ on $\overline{e}_+$ and $\overline{h}_+$. Moreover, as observed in Section 1.1.3, the MB perturbations near the two tori T_1 and T_2 may create degenerate orbits. We will then perturb R to get a new vector field R'. This vector field will be defined in $Y \setminus V'(L)$, where $V'(L) \subset \mathring{V}(L)$ is an open tubular neighborhood of L defined by $V'(L) = V'(K_1) \sqcup \cdots \sqcup V'(K_n)$, where, using the coordinates of Section 1.3.1, $\partial(V'(K_i)) = \{y = 2.5\}$.

Lemma 3.13 *There exists a (noncontact) vector field R' such that*

(i) *R' coincides with R outside a neighborhood of $\mathcal{N}(L)$;*

(ii) *R' satisfies properties* (1)–(4) *above with $V(L)$ replaced by $V'(L)$;*

(iii) *the only periodic orbits of R' in $\mathcal{N}(V) \setminus V'(L)$ are the four sets of nondegenerate orbits $\overline{e}, \overline{h}, \overline{e}_+, \overline{h}_+$.*

Observe that property (i) implies that the twisted Lefschetz zeta functions of the restrictions of the flows ϕ_R and $\phi_{R'}$ to $Y \setminus \mathcal{N}(K)$ coincide, while property (ii) allows us to apply Theorem 3.10 to $\phi_{R'}$.

Proof A perturbation of R into an R' satisfying the conditions (i)–(iii) can be obtained in more than one way. An example is pictured in Figure 3; see also Figure 1. We briefly explain how it is obtained. Since the modification of R is nontrivial only inside disjoint neighborhoods of each K_i, we will describe it only for a fixed component K of L. The characterization of the a perturbation will be presented in terms of perturbation of the lines in a page S of (L, S, ϕ) that are invariant under the first return map ϕ of ϕ_R: we will refer to these curves as *ϕ–invariant lines on S*. Note that these curves are naturally oriented by the flow.

Outside a neighborhood of $\partial V'$, one can see this perturbation in terms of a perturbation of ϕ into another monodromy ϕ', and R' is the vector field ∂_t in $Y \setminus V'(L) \cong S \times [0,1]/\langle (x,1) \sim (\phi'(x), 0)\rangle$, where t is the coordinate of $[0,1]$.

Observe first that the only periodic orbit in the (singular) ϕ–invariant line a_1 containing h (in correspondence to the singularity) is exactly h. Similarly, the only periodic orbit in the ϕ–invariant singular flow line a_2 containing h_+ is precisely h_+. Denote by $A_i \subset Y$ the mapping torus of $(a_i, \phi|_{a_i})$, $i = 1, 2$. We modify R separately inside the regions of $(Y \setminus V'(K)) \setminus (A_1 \sqcup A_2)$ as follows.

In the region containing e (and with boundary A_1), the set of ϕ–invariant lines (the elliptic lines in Figure 3 (left)) is perturbed in a set of ϕ'–invariant spiral-like lines (Figure 3 (right)), each of which is negatively asymptotic to a_1 and positively

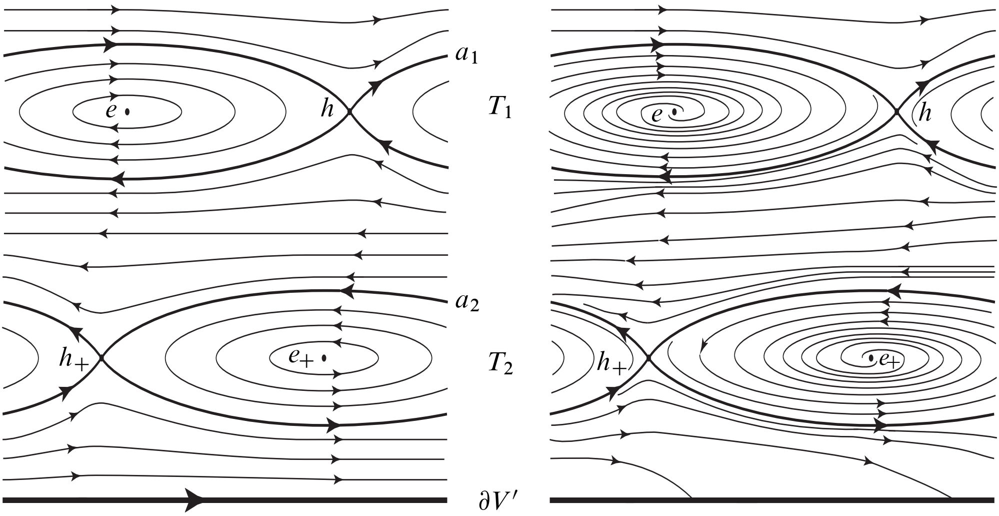


Figure 3: The dynamics of the vector fields R and R' near $\mathcal{N}(V) \setminus V'(L)$. Each oriented line represents an invariant subset of a page of (L, S, ϕ) under the first return map ϕ (left) and ϕ' (right); the invariant lines a_1 and a_2 are stressed. The situation at the left is the same depicted in Figure 1.

asymptotic to e. It is easy to see that after the perturbation, the only periodic orbit in the interior of this region is e. Moreover, we can arrange the perturbation such in a way that the differential $\mathfrak{L}_e^{R'}$ of the first return map on S of $\phi_{R'}$ along e coincides, up to a positive factor smaller then 1, with $\mathfrak{L}_e^{R}$, so that the Lefschetz sign $\epsilon(e)$ of e is still $+1$.

A similar perturbation is done in the region of $(Y \setminus V'(K)) \setminus (A_1 \sqcup A_2)$ containing e_+ in such a way that e_+ is the only periodic orbit of the perturbed vector field R', with still $\epsilon(e_+) = +1$.

The perturbation in the region between A_1 and A_2 is done by slightly pushing the monodromy in the positive y–direction in such a way that the set of ϕ–invariant lines is perturbed into a set of ϕ'–invariant lines, each of which is negatively asymptotic to a_1 and positively asymptotic to a_2 (in particular, there can not exist periodic orbits in this region).

A similar perturbation is done also inside the region between A_2 and $\partial V'(K)$, but in this case each ϕ'–invariant line is negatively asymptotic to a_2 and intersects $\partial V'(K)$ pointing out of the three-manifold.

Finally, we leave $R' = R$ in the rest of the manifold, where R was supposed having only isolated and nondegenerate periodic orbits.

Note that the two bases of eigenvectors of $\mathfrak{L}_h^R$ and $\mathfrak{L}_{h_+}^R$ are contained in the tangent spaces of the curves a_1 and a_2, and since $\phi_R = \phi_{R'}$ on these curves, the Lefschetz signs of the two orbits are preserved by the perturbation.

It is easy to convince ourselves that R' satisfies the properties (i)–(iii) above. □

Set $\zeta = \zeta_{\mathbb{1}}$. Since the Lefschetz zeta function of a flow depends only on its periodic orbits and their signs, we have the following:

Corollary 3.14 *If R' is obtained from R as above, then*

$$\zeta(\phi_{R'}) = \zeta(\phi_{R'}|_{(Y\setminus\mathcal{N}(K)\sqcup\{\bar{e},\bar{h},\bar{e}_+,\bar{h}_+\})}) = \zeta(\phi_R|_{(Y\setminus\mathcal{N}(K))})\cdot\prod_{\gamma\in\{\bar{e},\bar{h},\bar{e}_+,\bar{h}_+\}}\zeta_\gamma([\gamma]),$$

where $[\gamma]$ is the homology class of γ in $H_1(Y\setminus\mathcal{N}(K))$.

Now we want to compute more explicitly the twisted Lefschetz zeta function $\zeta(\phi_{R'})$. Let us begin with the local Lefschetz zeta function of the simple orbits; see Definition 3.5.

Lemma 3.15 *Let γ be an orbit of R or R'. Then*

$$\zeta_\gamma(t) = \begin{cases} (1-t)^{-1} = 1+t+t^2+\cdots & \text{if } \gamma \text{ elliptic},\\ 1-t & \text{if } \gamma \text{ positive hyperbolic},\\ 1+t & \text{if } \gamma \text{ negative hyperbolic}.\end{cases} \tag{24}$$

Proof This is just matter of replacing the Lefschetz signs given in Observation 1.2. For example, if γ is positive hyperbolic, then all the iterates are also positive hyperbolic, $\epsilon(\gamma^i) = -1$ for every $i>0$, and

$$\zeta_\gamma(t) = \exp\left(\sum_{i\geq 1} -\frac{t^i}{i}\right) = \exp(\log(1-t)) = 1-t.$$ □

Observation 3.16 Note that the equations above are exactly the generating functions given by Hutchings in [20, Section 2].

Let μ_i be a positive meridian of K_i for $i\in\{1,\dots,n\}$, and set $t_i = [\mu_i]\in H_1(Y\setminus K)$; fix moreover a Seifert surface S_i for each K_i. Recall that, for a given $X\subset Y$, we denote by $\mathcal{P}(X)$ the set of simple Reeb orbits contained in X.

Corollary 3.17 *The twisted Lefschetz zeta function of $\phi_R|_{(Y\setminus\mathcal{N}(L))}$ is*

$$\zeta(\phi_R|_{(Y\setminus\mathcal{N}(L))}) = \prod_{\gamma\in\mathcal{P}(Y\setminus\mathcal{N}(L))}\zeta_\gamma([\gamma]),$$

where $\zeta_\gamma([\gamma])$ is determined as follows:

$$\zeta_\gamma(\rho_L(\gamma)) = \left(1 - \prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle}\right)^{-1} = \sum_{l=0}^\infty \left(\prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle}\right)^l \quad (\gamma \text{ elliptic}),$$

$$\zeta_\gamma(\rho_L(\gamma)) = 1 - \prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle} \quad (\gamma \text{ positive hyperbolic}),$$

$$\zeta_\gamma(\rho_L(\gamma)) = 1 + \prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle} \quad (\gamma \text{ negative hyperbolic}).$$

Proof of Theorem 3.12 for fibered links To finish the proof, it remains essentially to prove that

$$\chi(\mathrm{ECC}(L, Y, \alpha)) = \zeta(\phi_R|_{(Y\setminus\mathcal{N}(L))}) \cdot \prod_{\gamma\in\{\bar{e},\bar{h},\bar{e}_+,\bar{h}_+\}} \zeta_\gamma([\gamma]). \tag{25}$$

This is easy to verify recursively on the set of simple orbits. Suppose $\delta = \prod_j \delta_j^{k_j}$ is an orbit set and let γ be an orbit such that $\gamma \neq \delta_j$ for any j. Then the set of all multiorbits that we can build using δ and γ can be expressed via the product formulae

$$\begin{array}{ll} \delta\cdot\{\varnothing, \gamma, \gamma^2, \dots\} & \text{if } \gamma \text{ is elliptic,} \\ \delta\cdot\{\varnothing, \gamma\} & \text{if } \gamma \text{ is hyperbolic.} \end{array} \tag{26}$$

As remarked in Section 1.2, the index parity formula (4) implies that the Lefschetz sign endows the ECH–chain complex with an absolute degree, and it coincides with the parity of the ECH–index. Then the contribution to the graded Euler characteristic of $\delta\cdot\gamma^l$, for any l ($l \in \mathbb{N}$ if γ is elliptic and $l \in \{0, 1\}$ if γ is hyperbolic), is

$$\epsilon(\delta)\prod_{i=1}^n t_i^{\langle \delta, S_i\rangle} \cdot \left(\epsilon(\gamma)\prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle}\right)^l.$$

Substituting the last formula in (26), the total contribution of the product formulae to the Euler characteristic are

- $\epsilon(\delta)\prod_{i=1}^n t_i^{\langle \delta, S_i\rangle} \cdot \sum_{l=0}^\infty \left(\prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle}\right)^l$ if γ is elliptic,
- $\epsilon(\delta)\prod_{i=1}^n t_i^{\langle \delta, S_i\rangle} \cdot \left(1 - \prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle}\right)$ if γ is positive hyperbolic,
- $\epsilon(\delta)\prod_{i=1}^n t_i^{\langle \delta, S_i\rangle} \cdot \left(1 + \prod_{i=1}^n t_i^{\langle \gamma, S_i\rangle}\right)$ if γ is negative hyperbolic,

that is,

$$\epsilon(\delta)\prod_{i=1}^n t_i^{\langle \delta, S_i\rangle} \cdot \zeta_\gamma([\gamma]).$$

Starting from $\delta = \varnothing$, (25) follows by induction on the set of the simple Reeb orbits in $(Y \setminus \mathcal{N}(L)) \sqcup \{\overline{e}, \overline{h}, \overline{e}_+, \overline{h}_+\}$. The theorem follows then by applying Corollary 3.14 and Theorem 3.10 to the flow of R'. □

Proof of Theorem 3.1 for fibered links Theorem 3.12 and (22) imply (20) immediately. To prove the result in the hat version, we reason again at the level of chain complexes. Recall that if $N := Y \setminus \mathring{\mathcal{N}}(L)$, by the definition of the ECK–chain complexes,

$$\begin{aligned} \mathrm{ECC}(L, Y, \alpha) &= \mathrm{ECC}^{\overline{e}_+, \overline{h}_+}(N, \alpha) \\ &= \mathrm{ECC}^{\overline{h}_+}(N, \alpha) \otimes \bigotimes_{i=1}^{n} \langle \varnothing, e_i^+, (e_i^+)^2, \dots \rangle \\ &= \widehat{\mathrm{ECC}}(L, Y, \alpha) \otimes \bigotimes_{i=1}^{n} \langle \varnothing, e_i^+, (e_i^+)^2, \dots \rangle, \end{aligned}$$

where the second line comes from the product formula (26) and the fact that e_i^+ is elliptic for any i. Taking the graded Euler characteristics as above, we have

$$\begin{aligned} \chi(\mathrm{ECC}(L, Y, \alpha)) &= \chi(\widehat{\mathrm{ECC}}(L, Y, \alpha)) \cdot \prod_{i=1}^{n} \zeta_{e_i^+}([e_i^+]) \\ &= \chi(\widehat{\mathrm{ECC}}(L, Y, \alpha)) \cdot \prod_{i=1}^{n} \frac{1}{1 - t_i}, \end{aligned}$$

where the last equality comes from the fact that $[e_i^+] = [\mu_i] = t_i \in H_1(Y \setminus L)$. If $Y = S^3$, then the last equation and (20) evidently imply (21). □

Observation 3.18 (symplectic Floer homology) If (L, S, ϕ) is an open book decomposition of Y, one can think of $\mathrm{ECK}(L, Y, \alpha)$ and $\widehat{\mathrm{ECK}}(L, Y, \alpha)$ as invariants of the pair (S, ϕ) and the adapted α. It is interesting to note that the Euler characteristic of $\mathrm{ECK}_1(L, Y, \alpha)$ with respect to the surface S (see Example 2.12) coincides with the sum of the Lefschetz signs of the Reeb orbits of period 1 in the interior of S, ie the Lefschetz number $\Lambda(\phi)$ of ϕ.

In fact, given Y (not necessarily an homology three-sphere) we can say even more about this fact by relating $\mathrm{ECK}_1(L, Y, \alpha)$ to the *symplectic Floer homology* $\mathrm{SH}(S, \phi)$ of (S, ϕ), whose Euler characteristic is precisely $\Lambda(\phi)$. Here we are considering the version of $\mathrm{SH}(S, \phi)$ for surfaces with boundary that is slightly rotated by ϕ in the positive direction, with respect to the orientation induced by S on ∂S; see for example Cotton-Clay [11].

Combining the definition of ECK, the relation between the *periodic Floer homology* PFH and ECH for a mapping torus (see Theorem 3.6.1 in [7]) and between PFH and SH (see for example Hutchings and Sullivan [21]) one can easily prove that

$$\mathrm{ECK}_1(L, Y, \alpha) \cong \mathrm{SH}(S, \phi), \tag{27}$$

where the degree of $\mathrm{ECK}(L, Y, \alpha)$ is computed with respect to the Alexander degree induced by S. We remark that an analogous result for HFK is currently unknown.

3.4 The general case

The first approach that one could use to attempt to apply Theorem 3.10 to a general link $L \subset Y$ is to look for a contact form on Y that is compatible with L and whose Reeb vector field is circular outside a neighborhood of L. Unfortunately we will not be able to find such a contact form. The basic idea to solve the problem consists of two steps:

Step 1 Find a contact form α on Y which is compatible with L and for which there exists a finite decomposition $Y \setminus L = \bigsqcup_i X_i$ for which $R = R_\alpha$ is circular in each X_i.

Step 2 Apply repeatedly the *Torres formula* for links to get the result.

As we will see, the Torres formula is a classical result which explains how to compute the Alexander polynomial of any sublink of a given link L starting from the Alexander polynomial of L.

3.4.1 Preliminary The key ingredient for Step 1 of our strategy is the following result; see Baker, Etnyre and Van Horn-Morris [2] and Guyard [18] for slightly different proofs.

Proposition 3.19 *Let $L = K_1 \sqcup \cdots \sqcup K_n \subset Y$ be an n–component link and let ξ be any fixed contact structure on Y. Then there exists an m–component link $L' \subset Y$ with $m \geq n$ and such that*

(1) $L' = L \sqcup K_{n+1} \sqcup \cdots \sqcup K_m$;

(2) *L' is fibered and the associated open book supports ξ.*

Proof The proof makes a deep use of the proof of the Giroux correspondence between open book decompositions and contact structures; see Giroux [16] and Colin [4]. Given a contact structure ξ on Y, Giroux explicitly constructs an open book decomposition of Y that supports a contact form α such that $\ker(\alpha) = \xi$. Such decomposition is built starting from a cellular decomposition $\mathcal{D}$ of Y that is compatible (in a specific sense) with ξ: for us it is important that, up to taking a refinement, any cellular decomposition of Y can be made compatible with ξ by an isotopy.

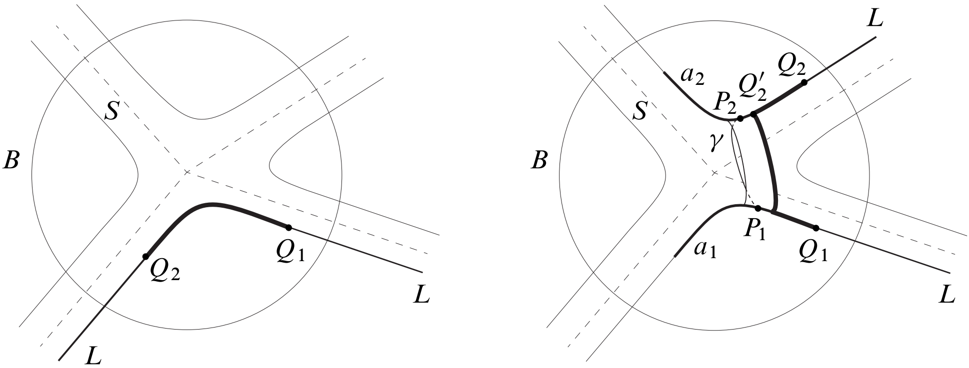


Figure 4: Making L contained in ∂S in $\mathcal{N}(\mathcal{D}^0)$: easy case (left) and general (right). The dotted lines are 1–simplexes in $\mathcal{D}^1$, while the bold segments from Q_1 to Q_2 represent the push-offs of L in $\mathcal{N}(\mathcal{D}^0)$.

Using the simplicial approximation theorem, it is possible to choose a triangulation $\mathcal{D}$ of Y in such a way that, up to isotopy, L is contained in $\mathcal{D}^1$, where $\mathcal{D}^i$ denotes the i–skeleton of $\mathcal{D}$. Up to taking a refinement, we can then suppose that $\mathcal{D}$ is adapted to ξ.

Let $\mathcal{N}(\mathcal{D}^1)$ be a tubular neighborhood of $\mathcal{D}^1$. Suppose that $\mathcal{N}(\mathcal{D}^0) \subset \mathcal{N}(\mathcal{D}^1)$ is a tubular neighborhood of $\mathcal{D}^0$ such that $\mathcal{N}(\mathcal{D}^1) \setminus \mathcal{N}(\mathcal{D}^0)$ is homeomorphic to a tubular neighborhood of $\mathcal{D}^1 \setminus \mathcal{N}(\mathcal{D}^0)$. The 0–page S of the associated open book built via the proof of Giroux satisfies then the following properties:

(1) $S \subset \mathcal{N}(\mathcal{D}^1)$, $L' := \partial S \subset \partial\mathcal{N}(\mathcal{D}^1)$ and $\mathcal{D}^1 \subset \operatorname{int}(S)$;

(2) $S \cap (\mathcal{N}(\mathcal{D}^1) \setminus \mathcal{N}(\mathcal{D}^0))$ is a disjoint union of strips which are diffeomorphic to $(\mathcal{D}^1 \setminus \mathcal{N}(\mathcal{D}^0)) \times [-1, 1]$, with $\mathcal{D}^1 \setminus \mathcal{N}(\mathcal{D}^0)$ corresponding to $(\mathcal{D}^1 \setminus \mathcal{N}(\mathcal{D}^0)) \times \{0\}$.

These properties imply that $L \subset \operatorname{int}(S)$ and that it is possible to push $L \setminus \mathcal{N}(\mathcal{D}^0)$ inside S to make it contained in ∂S. Note that in each strip composing $S \setminus \mathcal{N}(\mathcal{D}^0)$, we have only one possible choice for the direction in which to push $L \setminus \mathcal{N}(\mathcal{D}^0)$ to ∂S in such a way that the orientation of L coincides with that of ∂S.

We would like to extend this isotopy also to $L \cap \mathcal{N}(\mathcal{D}^0)$ to make the whole of L contained in ∂S. Suppose that B is a connected component (homeomorphic to a ball) of $\mathcal{N}(\mathcal{D}^0)$. In particular, we suppose that $B \cap S$ is connected. Then $L \cap \partial B$ consists of two points Q_1 and Q_2. The extension is done differently in the following two cases (see Figure 4):

Easy case This is when Q_1 and Q_2 belong to the same connected component of $\partial S \cap B$. The isotopy is then extended to B by pushing $L \cap B$ to $\partial S \cap B$ inside $S \cap B$; see Figure 4 (left).

General case If Q_1 and Q_2 belong to (the boundary of) different connected components a_1 and a_2 of $\partial S \cap B$, we proceed as follows: Let P_i be a point in the interior of a_i for $i = 1, 2$. Let γ be a simple arc in $S \cap B$ from P_1 to P_2 (there exists only one choice for γ up to isotopy). Let S' be obtained by positive Giroux stabilization of S along γ; see Figure 4 (right). Now we can connect Q_1 with a_2 by an arc in $\partial S'$ crossing the belt sphere of the 1–handle of the stabilization once; let Q_2' be the end point of this arc. Since a Giroux stabilization is compatible with the orientation of ∂S, the points Q_2' and Q_2 are in the same connected component of $a \setminus \{P_2\}$, so we can connect them inside $\partial S \cap B$, and we are done.

Pushing L to ∂S (and changing L and S as before where necessary) gives a link $\bar{L}$ that is contained in ∂S. To see that $\bar{L}$ is isotopic to L, we have to prove that, for any B as before, the two kinds of push-offs we use do not change the isotopy class of L.

Clearly, the isotopy class of L is preserved in the easy case. For the general case, it suffices to show that substituting the arc $L \cap S \cap B$ from Q_1 to Q_2 with an arc crossing the belt sphere of the handle once does not change the isotopy class of L. This is equivalent to proving that if γ is the path of the Giroux stabilization and $\bar{\gamma} = \gamma \cup c$, where c is the core curve of the handle, then $\bar{\gamma}$ bounds a disk in $Y \setminus L$. This can be proved for example by using the particular kind of Heegaard diagrams used in [7]. Observe that if b is the cocore of the handle, then $\bar{\gamma}$ is isotopic in S to $b \cup \phi'(b)$, where ϕ' is the monodromy on S' given by the Giroux stabilization. We finish by observing that $b \cup \phi'(b)$ is isotopic, up to a small perturbation near ∂S, to an attaching curve of a Heegaard diagram of Y. □

We now recall the Torres formula that we will use in the second step of our proof of Theorem 3.12. Since we need to consider the Alexander quotient as a polynomial, when we need to highlight its variables $t_1, \dots, t_k$, we will indicate them as subscripts and write $\mathrm{ALEX}_{t_1,\dots,t_k}$ instead of ALEX.

Theorem 3.20 (Torres formula) *Let $L = K_1 \sqcup \dots \sqcup K_n$ be an n–link in a homology three-sphere Y, K_{n+1} a knot in $Y \setminus L$ and $L' = L \sqcup K_{n+1}$. Let S_i be a Seifert surface for K_i for $i \in \{1, \dots, n+1\}$. Then*

$$\mathrm{ALEX}_{t_1,\dots,t_n,1}(Y \setminus L') \doteq \mathrm{ALEX}_{t_1,\dots,t_n}(Y \setminus L) \cdot \left(1 - \prod_{i=1}^{n} t_i^{\langle K_{n+1}, S_i \rangle}\right),$$

where $\mathrm{ALEX}_{t_1,\dots,t_n,1}(Y \setminus L')$ indicates the polynomial $\mathrm{ALEX}_{t_1,\dots,t_{n+1}}(Y \setminus L')$ evaluated at $t_{n+1} = 1$.

We refer the reader to Torres [37, Theorem 3] for the original proof. See also Franks [13, Theorem 6.4] for a proof making use of techniques of dynamics and Turaev [38, Section 1.4] for a generalization of the formula to links in any three-manifold.

Observation 3.21 One can see the condition $t_{n+1}=1$ from a purely topological point of view. Imagine taking the manifold $Y \setminus L'$ and then gluing back K_{n+1}. The effect on $H_1(Y \setminus L')$ is that the generator $[\mu_{n+1}]$ is killed, and now the homology class of a loop $\gamma \subset Y \setminus L'$ is determined only by the numbers $\langle \gamma, S_i \rangle$ for $S_i \in \{1,\dots,n\}$ (ie by $\rho_L(\gamma)$).

3.4.2 Proof of the result in the general case

Proof of Theorem 3.12 Let $L = K_1 \sqcup \cdots \sqcup K_n$ be a given link in Y. Proposition 3.19 implies that there exists an open book decomposition (L', S, ϕ) of Y with binding

$$L' = L \sqcup K_{n+1} \sqcup \cdots \sqcup K_m$$

for some $m \geq n$. Let α be a contact form on Y adapted to (L', S, ϕ). Let $R = R_\alpha$ be its Reeb vector field. As remarked in Section 3.3, and using the same notation, R is circular in $Y \setminus \mathring{V}'(L')$, where we recall that $V'(L)$ is an union of tubular neighborhoods $V'(K_i) \subsetneq V(K_i)$, for $i \in \{1,\dots,m\}$, of L.

Since α is also adapted to L', each $\mathring{V}(K_i)$ is, by definition, foliated by concentric tori, which in turn are linearly foliated by Reeb orbits that intersect positively a meridian disk for K_i in $V(K_i)$. Now, we can choose α in such a way that for each $i \in \{n+1,\dots,m\}$, the tori contained in $V'(K_i)$ are foliated by orbits of R with fixed irrational slope. This condition can be achieved by applying the Darboux–Weinstein theorem in $V(K_i)$ to make $\alpha|_{V'(K_i)}$ like in Example 6.2.3 of [5]. It follows that for each $i \in \{n+1,\dots,m\}$, the only closed orbit of R in $V'(K_i)$ is K_i. Define $U(L') = \bigsqcup_{i=1}^m U(K_i)$, where

$$U(K_i) = \begin{cases} V(K_i) & \text{if } i \in \{1,\dots,n\}, \\ V'(K_i) & \text{if } i \in \{n+1,\dots,m\}. \end{cases}$$

We have

$$\begin{aligned}
\chi(\mathrm{ECC}(L,Y,\alpha)) &= \zeta_{\rho_L}(\phi_R|_{Y\setminus V(L)}) \\
&= \zeta_{\rho_L}(\phi_R|_{Y\setminus U(L')}) \cdot \prod_{i=n+1}^{m} \prod_{\gamma \in \mathcal{P}(V'(K_i))} \zeta_\gamma(\rho_L([\gamma])) \\
&= \zeta_{\rho_L}(\phi_R|_{Y\setminus U(L')}) \cdot \prod_{i=n+1}^{m} \zeta_{K_i}(\rho_L([K_i])) \\
&= \zeta_{\rho_{L'}}(\phi_R|_{Y\setminus U(L')})|_{t_1,\dots,t_n,1,\dots,1} \cdot \prod_{i=n+1}^{m} \zeta_{K_i}(\rho_L([K_i])) \\
&\doteq \mathrm{ALEX}_{t_1,\dots,t_n,1,\dots,1}(Y \setminus L') \cdot \prod_{i=n+1}^{m} \zeta_{K_i}(\rho_L([K_i]))
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{ALEX}_{t_1,\dots,t_n,1,\dots,1}(Y \setminus L') \cdot \prod_{i=n+1}^{m} \Bigl(1 - \prod_{j=1}^{n} t_j^{\langle K_i, S_j \rangle}\Bigr)^{-1} \\
&= \mathrm{ALEX}_{t_1,\dots,t_n}(Y \setminus L),
\end{aligned}
$$

where

- line 2 follows from reasoning as in the proof of (25);
- line 3 holds since K_i, for $i \in \{n+1,\dots,m\}$, is the only Reeb orbit of α in $V'(K_i)$;
- line 4 comes from the idea in Observation 3.21: ρ_L and $\rho_{L'}$ coincide on the generators t_i of $H_1(Y \setminus L)$ for $i \in \{1,\dots,n\}$, and $t_i = [\mu_i] = 1 \in H_1(Y \setminus L)$ for $i \in \{n+1,\dots,m\}$;
- line 5 holds since, up to a slight perturbation of R near each $\partial U(K_i)$ to make it nondegenerate and transverse to the boundary like in the proof in Section 3.3, $\rho_{L'}$ and $R|_{Y \setminus U(L')}$ satisfy the hypothesis of Theorem 3.10;
- line 6 is due to the fact that the K_i are elliptic;
- line 7 is obtained by applying repeatedly the Torres formula on the components $K_{n+1},\dots,K_m$. □

The proof of Theorem 3.1 works then exactly as in the fibered case.

References

[1] **N A'Campo**, *La fonction zêta d'une monodromie*, Comment. Math. Helv. 50 (1975) 233–248 MR

[2] **K L Baker**, **J B Etnyre**, **J Van Horn-Morris**, *Cabling, contact structures and mapping class monoids*, J. Differential Geom. 90 (2012) 1–80 MR

[3] **F Bourgeois**, *A Morse–Bott approach to contact homology*, from "Symplectic and contact topology: interactions and perspectives" (Y Eliashberg, B Khesin, F Lalonde, editors), Fields Inst. Commun. 35, Amer. Math. Soc., Providence, RI (2003) 55–77 MR

[4] **V Colin**, *Livres ouverts en géométrie de contact (d'après Emmanuel Giroux)*, from "Séminaire Bourbaki, 2006/2007", Astérisque 317, Soc. Math. France, Paris (2008) exposé 969, 91–117 MR

[5] **V Colin**, **P Ghiggini**, **K Honda**, *Embedded contact homology and open book decompositions*, preprint (2010) arXiv

[6] **V Colin**, **P Ghiggini**, **K Honda**, *Equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions*, Proc. Natl. Acad. Sci. USA 108 (2011) 8100–8105 MR

[7] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions, I*, preprint (2012) arXiv

[8] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions, II*, preprint (2012) arXiv

[9] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology, III: From hat to plus*, preprint (2012) arXiv

[10] **V Colin**, **P Ghiggini**, **K Honda**, **M Hutchings**, *Sutures and contact homology, I*, Geom. Topol. 15 (2011) 1749–1842 MR

[11] **A Cotton-Clay**, *Symplectic Floer homology of area-preserving surface diffeomorphisms*, Geom. Topol. 13 (2009) 2619–2674 MR

[12] **Y Eliashberg**, **L Traynor** (editors), *Symplectic geometry and topology*, IAS/Park City Mathematics Series 7, Amer. Math. Soc., Providence, RI (1999) MR

[13] **J M Franks**, *Knots, links and symbolic dynamics*, Ann. of Math. 113 (1981) 529–552 MR

[14] **D Fried**, *Homological identities for closed orbits*, Invent. Math. 71 (1983) 419–442 MR

[15] **H Geiges**, *An introduction to contact topology*, Cambridge Studies in Advanced Mathematics 109, Cambridge Univ. Press (2008) MR

[16] **E Giroux**, *Géométrie de contact: de la dimension trois vers les dimensions supérieures*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed. Press, Beijing (2002) 405–414 MR

[17] **M Gromov**, *Pseudo holomorphic curves in symplectic manifolds*, Invent. Math. 82 (1985) 307–347 MR

[18] **T Guyard**, *Sur le calcul d'invariants et l'engendrement des noeuds transverses dans les variétés de contact de dimension trois*, PhD thesis, Université de Nantes (2015) Available at `https://tel.archives-ouvertes.fr/tel-01387461/`

[19] **M Hutchings**, *The embedded contact homology index revisited*, from "New perspectives and challenges in symplectic field theory" (M Abreu, F Lalonde, L Polterovich, editors), CRM Proc. Lecture Notes 49, Amer. Math. Soc., Providence, RI (2009) 263–297 MR

[20] **M Hutchings**, *Lecture notes on embedded contact homology*, from "Contact and symplectic topology" (F Bourgeois, V Colin, A Stipsicz, editors), Bolyai Soc. Math. Stud. 26, János Bolyai Math. Soc., Budapest (2014) 389–484 MR

[21] **M Hutchings**, **M Sullivan**, *The periodic Floer homology of a Dehn twist*, Algebr. Geom. Topol. 5 (2005) 301–354 MR

[22] **A Juhász**, *Holomorphic discs and sutured manifolds*, Algebr. Geom. Topol. 6 (2006) 1429–1457 MR

[23] **P Kronheimer**, **T Mrowka**, *Instanton Floer homology and the Alexander polynomial*, Algebr. Geom. Topol. 10 (2010) 1715–1738 MR

[24] **P Kronheimer**, **T Mrowka**, *Knots, sutures, and excision*, J. Differential Geom. 84 (2010) 301–364 MR

[25] **C Manolescu**, **P Ozsváth**, *Heegaard Floer homology and integer surgeries on links*, preprint (2010) arXiv

[26] **D McDuff**, *Singularities and positivity of intersections of J–holomorphic curves*, from "Holomorphic curves in symplectic geometry" (M Audin, J Lafontaine, editors), Progr. Math. 117, Birkhäuser, Basel (1994) 191–215 MR

[27] **D McDuff**, **D Salamon**, *J–holomorphic curves and symplectic topology*, Amer. Math. Soc. Colloq. Pub. 52, Amer. Math. Soc., Providence, RI (2004) MR

[28] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and knot invariants*, Adv. Math. 186 (2004) 58–116 MR

[29] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. 159 (2004) 1027–1158 MR

[30] **P Ozsváth**, **Z Szabó**, *Heegaard Floer homology and contact structures*, Duke Math. J. 129 (2005) 39–61 MR

[31] **P Ozsváth**, **Z Szabó**, *Holomorphic disks, link invariants and the multi-variable Alexander polynomial*, Algebr. Geom. Topol. 8 (2008) 615–692 MR

[32] **P S Ozsváth**, **Z Szabó**, *Knot Floer homology and integer surgeries*, Algebr. Geom. Topol. 8 (2008) 101–153 MR

[33] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) MR Available at `http://search.proquest.com/docview/305332635`

[34] **R Siefring**, *Intersection theory of punctured pseudoholomorphic curves*, Geom. Topol. 15 (2011) 2351–2457 MR

[35] **S Smale**, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. 73 (1967) 747–817 MR

[36] **W P Thurston**, **H E Winkelnkemper**, *On the existence of contact forms*, Proc. Amer. Math. Soc. 52 (1975) 345–347 MR

[37] **G Torres**, *On the Alexander polynomial*, Ann. of Math. 57 (1953) 57–89 MR

[38] **V G Turaev**, *Reidemeister torsion in knot theory*, Uspekhi Mat. Nauk 41 (1986) 97–147 MR In Russian; translated in Russian Math. Surveys 41 (1986) 119–182

Institut Fourier, Université Grenoble Alpes
38000 Grenoble, France

gilbertospano.math@gmail.com

Received: 9 February 2016 Revised: 5 December 2016

Algebraic & Geometric Topology 17 (2017) 2125–2144

On mod p A_p–spaces

RUIZHI HUANG
JIE WU

We prove a necessary condition for the existence of an A_p–structure on mod p spaces, and also derive a simple proof for the finiteness of the number of mod p A_p–spaces of given rank. As a direct application, we compute a list of possible types of rank 3 mod p homotopy associative H–spaces.

55P45, 55S25; 55N15, 55P15, 55S05

1 Introduction

A longstanding problem in algebraic topology is to classify finite H–spaces. However, this problem is rather complicated, and has only been solved in few cases. There is Zabrodsky's localization and mixing theorem [27] yielding that a simply connected finite complex is an H–space if and only if each of its p–localizations is an H–space. One would also like to know for which primes p the localization at p fails to be an H–space, so it is natural to consider the p–local version of H–spaces.

Let X be a CW-complex whose cohomology is an exterior algebra generated by r elements of odd dimension; we call r the rank of X. For $r = 1$, J F Adams [1; 2] has determined that S^1, S^3, S^7 are the only H–spaces localized at 2 by solving the famous Hopf invariant one problem, and all odd spheres are H–spaces localized at any odd prime p. For $r = 2$, the case $p = 2$ (then the integral case) has been solved in a series of papers: see Adams [3], Hubbuck [15], Zabrodsky [28; 29], Douglas and Sigrist [7], Mimura, Nishida and Toda [19], as well as the case $p > 3$ by N Hagelgans [9]. The remaining case $p = 3$ is challenging and has been an open question for decades; recent progress on it can be found in Grbić, Harper, Mimura, Theriault and Wu [8].

The phenomenon that the H–structures are largely controlled by the prime $p = 2$ appears similarly when we consider higher homotopy associative structures. Namely, if we consider A_p–spaces in the sense of J Stasheff [21; 22], the A_p–structure is controlled by that of the localization at p, where a connected A_2–space is just an H–space. In general, for any A_n–space X, Stasheff suggests an n–projective space $P_n(X)$ over X, which is analogous to Milnor's classifying space for topological groups.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2125

(See Definition 3.5 and the paragraph before that for the explicit definition of A_n–spaces and related comments.)

Let $n = p$. It is well-known that there exists some nontrivial p^{th} power in the cohomology of p–stage projective space $P_p(X)$ which exactly detects the A_p–structure. Furthermore, Hemmi [12] has defined a modified projective space $R_n(X)$ for a special family of A_n–spaces, which is our main concern in this paper. Based on these ideas and constructions we prove the following theorem, which generalizes the result of Wilkerson [25] for local spheres:

Theorem 1.1 *Fix an odd prime $p \geq 3$ and let X be a connected p–local A_p–space with cohomology ring $H^*(X, \mathbb{Z}/p\mathbb{Z}) \cong \bigwedge(x_{2m_1-1}, \dots, x_{2m_r-1})$, where $m_1 \leq m_j$ for all j. Define*

$$m = \gcd\{m_i \mid m_i \leq pm_1\}.$$

Then $m \mid p-1$.

For the converse of the theorem, we recall that Stasheff [23] has constructed a realization for polynomial algebras $\mathbb{Z}/p\mathbb{Z}[x_{2m}, x_{4m}, \dots, x_{2km}]$ with $m \mid p-1$ using a theorem of Quillen. Here, our proof of this theorem is based on a generalization of a method of Adams and Atiyah [4]; (see also Section 2), using which we also derive a simple proof of a finiteness theorem of Hubbuck and Mimura [16] (also see Theorem 3.7) which claims that there are only finitely many possible homotopy types of spaces with fixed rank r which are A_p–spaces.

For the special case when $p = 3$, a mod 3 A_3–space is a usual 3–local homotopy associative H–space. The only simply connected homotopy associative H–space at 3 of rank 1 is S^3. If we define the increasing sequence $(m_1, \dots, m_r)$ to be the type of X in Theorem 1.1, then the complete list of types for rank 2 3–local simply connected homotopy associative H–spaces are $(2,3)$, $(2,4)$, $(2,6)$ and $(6,8)$; see Wilkerson [24, Theorem 5.1]. It is clear that

$$S^3 \times S^5 \overset{3}{\simeq} \mathrm{SU}(3)$$

provides an example for $(2,3)$, $\mathrm{Sp}(2)$ for $(2,4)$, and G_2 for $(2,6)$. Harper [10] gives a decomposition

$$F_4 \overset{3}{\simeq} K \times B_5(3),$$

where $B_5(3)$ is the S^{11} bundle over S^{15} classified by α_1, and, further, Zabrodsky [30] shows that $B_5(3)$ is a loop space, which provides an example for $(6,8)$. In this paper, we consider the case of rank 3. With the help of the method of Adams and Atiyah, and some results of Wilkerson (see [24] or Theorem 4.2), we prove the following theorem by careful analysis of the effect of both Steenrod operations and Adams' ψ–operations.

Theorem 1.2 *Let X be an indecomposable 3–local homotopy associative H–space with cohomology ring $H^*(X,\mathbb{Z}/3\mathbb{Z}) \cong \bigwedge(x_{2r-1}, x_{2n-1}, x_{2m-1})$, where $\deg(x_k)=k$ and $1<r<n<m$. Then the type of X (r,n,m) can only be one of*

$$(2,4,6),\ (2,6,8),\ (3,5,7),\ (3,6,8),\ (6,8,10),\ (6,8,12).$$

In this list, the only known example is $\mathrm{Sp}(3)$, which is of type $(2,4,6)$. Here are a few things we know about potential examples of rank 3 3–local A_3–spaces of the remaining five types. For $(2,6,8)$, we can form a space X as the total space of a G_2–principal fibration over S^{15}, which is classified by the generator of

$$\pi_{15}(BG_2) \cong \pi_{14}(G_2) \overset{3}{\cong} \pi_{14}(S^3) \overset{3}{\cong} \mathbb{Z}/3\mathbb{Z}.$$

Then the classifying map factors as $S^{15} \xrightarrow{f} BS^3 \to BG_2$, and we get $X \overset{3}{\simeq} (G_2 \times Y)/S^3$, where Y is the total space of the fibration classified by f and also an H–space by Theorem 7.1 of Grbić, Harper, Mimura, Theriault and Wu [8]. However, we still do not know whether X is an H–space or not. For the case $(3,5,7)$ we have Nishida's $B_2^3(3)$, which is a 3–component of $\mathrm{SU}(7)$ (see Mimura, Nishida and Toda [20]). Still, we do not know whether $B_2^3(3)$ is homotopy associative. If X is of type $(3,6,8)$, then X has a generating complex of the form $S^5 \vee A$ by the knowledge of the homotopy groups of spheres, where A is of type $(6,8)$. For $(6,8,10)$, Harper and Zabrodsky [11] have proved that if the exterior algebra of rank p generated by $\{x_{2n-1}, \mathscr{P}^1 x_{2n-1}, \dots, \mathscr{P}^{p-1} x_{2n-1}\}$ can be realized by an H–space, then $p \mid n$, and the converse is still open for $n>p$. For the last possible case of type $(6,8,12)$, we have $\mathscr{P}^1(x_{11}) = x_{15}$ and $\mathscr{P}^3(x_{11}) = x_{23}$.

The article is organized as follows. In Section 2 we will introduce a refined version of Adams and Atiyah's method from [4]. In Section 3 we use number theory to prove Theorem 1.1 and the finiteness theorem of Hubbuck and Mimura. Section 4 is devoted to the proof of Theorem 1.2.

2 A method of Adams and Atiyah

In [4], Adams and Atiyah develop a method to detect the p^{th} power of cohomology elements using Adams' ψ–operations. For our purpose, we need to modify it slightly.

Given a connected CW-complex X with no p–torsion in $H^*(X,\mathbb{Z})$, suppose there exists a subalgebra $\bar{\mathcal{H}}$ of $H^*(X;\mathbb{Z}/p\mathbb{Z})$ such that

$$\bar{\mathcal{H}} \cong \bar{A} \oplus \bar{B}$$

as rings, where $\bar{A}$ contains $\bar{\mathcal{H}}^0$, $\bar{B}$ is an ideal and also $\bar{\mathcal{H}}$ and $\bar{B}$ are closed under the action of the mod p Steenrod algebra $\mathscr{A}_p$. Then by the Atiyah–Hirzebruch–Whitehead spectral sequence and [5, Theorem 6.5], we have the corresponding filtered subalgebra $\mathcal{H}$ of $K(X)\otimes\mathbb{Z}_{(p)}$ such that

$$\mathcal{H}\cong A\oplus B,$$

as filtered rings, and also $\mathcal{H}$ and B are closed under ψ^p–action. Write the Chern character of an element $x\in K(X)\otimes\mathbb{Z}_{(p)}$ as

$$\mathrm{ch}(x)=a_0+\sum\nolimits_i a_{2i}+\sum\nolimits_j b_{2j},$$

with $a_0\in\mathbb{Q}$, $a_{2i}\in\bar{A}^{>0}\otimes\mathbb{Q}$ and $b_{2j}\in\bar{B}^{>0}\otimes\mathbb{Q}$ (the subscripts refer to the degree). Then we have

$$\mathrm{ch}(\psi^k(x))=a_0+\sum\nolimits_i k^i a_{2i}+\sum\nolimits_j k^j b_{2j}.$$

Hence ψ^k is indeed a semisimple linear transformation if we use the Chern character to identify $K(X)\otimes\mathbb{Q}$ with $H^{\mathrm{even}}(X;\mathbb{Q})$, and the eigenspace decomposition of $\widetilde{K}(X)\otimes\mathbb{Q}$ is independent of the choice of ψ^k. In particular, $\mathcal{H}\otimes\mathbb{Q}$ and $B\otimes\mathbb{Q}$ are invariant under ψ^k for any k, as they are invariant under ψ^p, and then $\mathcal{H}$ and B are also invariant under each ψ^k. Then, as in [4], we get a (partial) eigenspace decomposition

$$\widetilde{\mathcal{H}}\cong\bigoplus_{i=1}^{r}V_i\oplus W,\qquad B^{>0}\otimes\mathbb{Q}\cong W,$$

where $\widetilde{\mathcal{H}}=\mathcal{H}^{>0}\otimes\mathbb{Q}$, $\deg(V_i)=2m_i$ (which means the degree of its elements) and V_i is allowed to be the 0 vector space. For each ψ^k, V_i is the eigenspace corresponding to the eigenvalue k^{m_i}. We also notice that $A^{>0}\otimes\mathbb{Q}\cong\bigoplus_{i=1}^r V_i$ but only as vector spaces. Now define a linear transformation on $\widetilde{K}(X)\otimes\mathbb{Q}$ by

$$\pi_i=\prod_{\substack{1\le j\le r\\ j\ne i}}\frac{\psi^{k_j}-k_j^{m_j}}{k_j^{m_i}-k_j^{m_j}},$$

and a number

$$d_i(m_1,\dots,m_r)=\gcd\Big\{\prod_{\substack{1\le j\le r\\ j\ne i}}(k_j^{m_i}-k_j^{m_j})\;\Big|\;k_j\in\mathbb{N}^+\text{ for }1\le j\le r,\ j\ne i\Big\}.$$

Notice that π_i induces a linear transformation $\bar{\pi}_i$ on $\bigoplus_{i=1}^r V_i$ which is the natural projection onto the i^{th} component V_i. For any $x\in\widetilde{\mathcal{H}}$, we have

$$\pi_i(x)\cdot\prod_{\substack{1\le j\le r\\ j\ne i}}(k_j^{m_i}-k_j^{m_j})=\prod_{\substack{1\le j\le r\\ j\ne i}}(\psi^{k_j}-k_j^{m_j})(x)\in\widetilde{\mathcal{H}}.$$

Accordingly,

$$\pi_i(x)d_i(m_1,\ldots,m_r)\in\widetilde{\mathcal{H}}.$$

If we write $x=\sum_i\bar{\pi}_i(x-v)+v$ for some $v\in B$, then we also have

$$\bar{\pi}_i(x-v)d_i(m_1,\ldots,m_r)\in\widetilde{\mathcal{H}}.$$

Now we make a crucial assumption that for each i

$$p^{m_i}\nmid d_i(m_1,\ldots,m_r). \tag{2-1}$$

Since B is a $\{\psi^p\}$–module, we have

$$\begin{aligned}\psi^p(x)&=\sum_i\psi^p(\bar{\pi}_i(x-v))+\psi^p(v)\\&=\sum_i p^{m_i}\frac{\bar{\pi}_i(x-v)d_i(m_1,\ldots,m_r)}{d_i(m_1,\ldots,m_r)}+\psi^p(v)\\&=py+\psi^p(v)\in p\widetilde{\mathcal{H}}+B,\end{aligned}$$

ie $x^p\equiv\psi^p(x)\equiv 0 \bmod (p,B)$. Again, as in [4], $\bar{x}^p\equiv 0 \bmod (\bar{B})$ on the cohomology level, where $\bar{x}$ denotes the corresponding element of x in $\bar{\mathcal{H}}\subset H^*(X,\mathbb{Z}/p\mathbb{Z})$.

Remark 2.1 Notice that when $\bar{\mathcal{H}}=H^*(X,\mathbb{Z}/p\mathbb{Z})$ and $\bar{B}=0$, the above result is exactly [4, Corollary].

3 Proof of Theorem 1.1 and the finiteness theorem

3.1 Proof of Theorem 1.1

We prove the theorem by contradiction. The main task is to prove the condition (2-1) holds. We have to do some number theory first.

Definition 3.1 Let n be a positive integer.

(1) Define $e(n)=f$ if $n=p^f\cdot x$ and $p\nmid x$.

(2) Define ν by

$$\nu(n)=\begin{cases}f+1 & \text{if } n=p^f(p-1)x \text{ and } p\nmid x,\\ 0 & \text{if } p{-}1\nmid n.\end{cases}$$

Suppose k is a primitive root modulo p^2. Then k is also a primitive root modulo p^f for all $f \in \mathbb{N}^+$. Then for any positive integer n, we have

$$k^n \equiv 1 \bmod p^f \iff n \equiv 0 \bmod p^{f-1}(p-1). \tag{3-1}$$

So $\nu(n)$ is the exact exponent of p in the prime factorization of $k^n - 1$ if $p{-}1 \mid n$.

The following lemma is well known and basic in number theory:

Lemma 3.2 (Legendre 1808) *We have*

$$e(n!) = \sum_{k=1}^{\infty} \left\lfloor \frac{n}{p^k} \right\rfloor = \frac{n - s_p(n)}{p-1},$$

where $s_p(n) = a_k + a_{k-1} + \cdots + a_1 + a_0$ is the sum of all the digits in the expansion of n in base p.

From above, we easily get:

Corollary 3.3 (1) $e(a!) + e(b!) \le e((a+b)!)$*;*

(2) $e((ab)!) \le a + e(a!)$ *if* $b \le p$*.*

Now we are ready to prove our main lemma, which is a generalization of [4, Lemma 3.5]:

Lemma 3.4 *Let p be an odd prime, k be a primitive root modulo p^2, $m, t \in \mathbb{N}^+$ be such that $m \nmid p{-}1$, and set*

$$\Pi := \prod_{\substack{j \le t \le tp \\ j \ne i}} (k^{mi} - k^{mj}).$$

Then we have

$$e(\Pi) < mt. \tag{3-2}$$

Proof We set $\gcd(m, p-1) = h$, $m = ah$, and $p - 1 = bh$. Then $a > 1$ since $m \nmid p{-}1$. Then we have

$$\prod_{\substack{j \le t \le tp \\ j \ne i}} (k^{mi} - k^{mj}) = \prod_{t \le j < i} k^{mj}(k^{m(i-j)} - 1) \cdot \prod_{i < j \le tp} k^{mi}(1 - k^{m(j-i)}).$$

By (3-1), we only need to consider values of j satisfying $p{-}1 \mid m(i-j)$, ie $b \mid i{-}j$. Then we have

$$\begin{aligned}
e(\Pi) &= \prod_{t\le j<i} e(k^{m(i-j)}-1)\cdot \prod_{i<j\le tp} e(1-k^{m(j-i)})\\
&= \prod_{1\le \frac{i-j}{b}\le \lfloor \frac{i-t}{b}\rfloor} e(k^{mb\frac{i-j}{b}}-1)\cdot \prod_{1\le \frac{j-i}{b}\le \lfloor \frac{tp-i}{b}\rfloor} e(k^{mb\frac{j-i}{b}}-1)\\
&= \prod_{1\le j\le \lfloor \frac{i-t}{b}\rfloor} e(k^{mbj}-1)\cdot \prod_{1\le l\le \lfloor \frac{tp-i}{b}\rfloor} e(k^{mbl}-1)\\
&= \sum_{1\le j\le \lfloor \frac{i-t}{b}\rfloor} \nu(mbj)+\sum_{1\le l\le \lfloor \frac{tp-i}{b}\rfloor} \nu(mbl)\\
&= (e(m)+1)\Big(\Big\lfloor \frac{i-t}{b}\Big\rfloor+\Big\lfloor \frac{tp-i}{b}\Big\rfloor\Big)+e\Big(\Big\lfloor \frac{i-t}{b}\Big\rfloor !\Big)+e\Big(\Big\lfloor \frac{tp-i}{b}\Big\rfloor !\Big)\\
&\le (e(m)+1)\frac{tp-t}{b}+e\Big(\Big(\Big\lfloor \frac{i-t}{b}\Big\rfloor+\Big\lfloor \frac{tp-i}{b}\Big\rfloor\Big)!\Big)\\
&\le (e(m)+1)th+e((th)!).
\end{aligned}$$

Now if $h=1$, then

$$\begin{aligned}
e(\Pi) &\le (e(m)+1)t+e(t!)\\
&= (e(m)+1)t+\frac{t-s_p(t)}{p-1}\\
&< t\Big(e(m)+1+\frac{1}{p-1}\Big).
\end{aligned}$$

If $h\ge 2$, then

$$\begin{aligned}
e(\Pi) &\le (e(m)+1)th+t+e(t!)\\
&= (e(m)+1)th+t+\frac{t-s_p(t)}{p-1}\\
&< t\Big((e(m)+1)h+1+\frac{1}{p-1}\Big).
\end{aligned}$$

On the other hand, the inequality $a-e(m)-1\ge 1$ always holds, for otherwise $e(a)+1=e(m)+1=a$ implies $a=1$ (we use $p\ge 3$ here). Now combining all above, it is easy to see $e(\Pi)<mt$ in both cases. □

Now we are going to prove Theorem 1.1. First we recall some background on A_n–spaces, for which Stasheff's original papers [21; 22] are the standard reference. Stasheff's A_n–spaces can be defined inductively with the help of Stasheff polytopes, which are also called associahedra. Explicitly, an associahedron K_n is an $(n-2)$–dimensional convex polytope whose vertices are in one to one correspondence with the parenthesizings of the word $x_1x_2\dots x_n$ and whose edges correspond to single application of the associativity rule. In particular, K_2 is a point, K_3 is a interval and K_4 is the convex hull of a pentagon. There are canonical maps between the K_n. Indeed, the family $\mathcal{K} = \{K_n\}$ can be endowed with an operadic structure such that any $\mathcal{K}$–space is the so-called A_∞–space ($\mathcal{K}$ is called A_∞–operad). Then an A_n–space is just an space with the action of $\mathcal{K}$ only up to the n–stage (the corresponding operad is called the A_n–operad). Stasheff also gave another equivalent description of A_n–spaces, which he used as definition:

Definition 3.5 [21, Definition 1] An A_n–structure on a space X consists of an n–tuple of maps

$$\begin{tikzcd} X \arrow[r, equal] & E_1 \arrow[r, hook] \arrow[d, "p_1"] & E_2 \arrow[r, hook] \arrow[d, "p_2"] & \cdots \arrow[r, hook] & E_n \arrow[d, "p_n"] \\ * \arrow[r, equal] & B_1 \arrow[r, hook] & B_2 \arrow[r, hook] & \cdots \arrow[r, hook] & B_n \end{tikzcd}$$

such that each p_i is a quasifibration and there is a contracting homotopy $h\colon CE_{n-1} \to E_n$ such that $h(CE_{i-1}) \subset E_i$.

Note that if A_n–structure is given by the operadic action, the above diagram can be constructed such that B_i is the i^{th} "projective space" $P_n(X)$ over X (as in Milnor's construction). The reverse process was done by Stasheff. The projective space is crucial for there are nontrivial n^{th} powers in its cohomology ring.

Here, the key construction for our proof of Theorem 1.1 is the so-called modified projective space of Hemmi [13] which is an analogy of Stasheff's n–projective space [21]. Since we will not use the explicit construction of this concept, we only recall some properties stated in the following lemma.

Lemma 3.6 (see [13, Theorem 1.1]) *Let $n \geq 3$ and let X be a finite A_n–space with cohomology ring*

$$H^*(X, \mathbb{Z}/p\mathbb{Z}) \cong \bigwedge(x_{2m_1-1}, \dots, x_{2m_r-1}), \quad \deg(x_{2m_i-1}) = 2m_i - 1.$$

Then there exists a modified projective space $R_n(X)$ with a map $\varepsilon\colon \Sigma X \to R_n(X)$ such that

$$\overline{\mathcal{H}} \cong \overline{A} \oplus \overline{B} = \mathbb{Z}/p\mathbb{Z}[y_{2m_1}, \dots, y_{2m_r}]/(\text{height } n+1) \oplus \overline{B}$$

as rings, for some subalgebra $\overline{\mathcal{H}}$ of $H^*(R_n(X), \mathbb{Z}/p\mathbb{Z})$ and $\varepsilon^*(y_{2m_i}) = \sigma^*(x_{2m_i-1})$, where the ideal under quotient in the first factor is generated by monomials of length greater than or equal to $n+1$. Further, $\overline{\mathcal{H}}$ and $\overline{B}$ are closed under the action of the mod p *Steenrod algebra $\mathscr{A}_p$.*

Now we are ready to prove Theorem 1.1.

Proof of Theorem 1.1 We prove the theorem by contradiction, and assume $m \nmid p-1$. By Lemma 3.6, $H^*(R_p(X))$ contains a truncated polynomial algebra

$$\mathbb{Z}/p\mathbb{Z}[y_{2m_1}, \dots, y_{2m_r}]/(\text{height } p+1) \hookrightarrow H^*(R_p(X)).$$

Let us define $Y(X) = R_p^{2pm_1+1}(X)$ to be the $(2pm_1+1)$–skeleton of $R_p(X)$. We then have a ring decomposition

$$i^*(\bar{\mathcal{H}}) \cong i^*(\bar{A}) \oplus i^*(\bar{B}),$$

where $i\colon Y(X) \hookrightarrow R_p(X)$ is the canonical inclusion. Then $y_{2m_1}^p \not\equiv 0 \bmod (i^*(\bar{B}))$. We then set $m_i = ms_i$, and apply Lemma 3.4 for $t = s_1$ and $m = m$ since $m \nmid p-1$ by assumption. Then we get $e(\Pi) < ms_1 = m_1$, which implies the condition (2-1) holds for $Y(X)$ since m_1 is the lowest degree. Further, $i^*(\bar{\mathcal{H}})$ and $i^*(\bar{B})$ are closed under the action of $\mathscr{A}_p$, hence by the argument in Section 2, $\bar{x}^p \equiv 0 \bmod (i^*(\bar{B}))$ for any $\bar{x} \in i^*(\bar{\mathcal{H}})$, which contradicts the fact that $y_{2m_1}^p \not\equiv 0 \bmod (i^*(\bar{B}))$. The proof of Theorem 1.1 is completed. □

3.2 The finiteness theorem for finite A_p–spaces

As another application, we prove the following theorem of Hubbuck and Mimura:

Theorem 3.7 [16] *Let X be a connected finite* mod p *A_p–space of rank r. Then there are only finitely many possible homotopy types for the space X.*

Proof Suppose X has the type $(m_1, m_2, \dots, m_r)$ with $m_1 \le m_2 \le \dots \le m_r$, and form the space

$$Y(X) = \frac{R_p^{2pm_r+1}(X)}{R_p^{2m_r-1}(X)},$$

which is the $(2pm_r+1)$–skeleton of $R_p(X)$ with the $(2m_r-1)$–skeleton pinched to a point. As in the proof of Theorem 1.1, we can get a ring decomposition

$$p^{*-1}i^*(\bar{\mathcal{H}}) \cong p^{*-1}i^*(\bar{A}) \oplus p^{*-1}i^*(\bar{B})$$

using the canonical inclusion and projection, such that $p^{*-1}i^*(\bar{\mathcal{H}})$ and $p^{*-1}i^*(\bar{B})$ are closed under the action of $\mathscr{A}_p$, and $y_{2m_r}^p$ is the nontrivial module $p^{*-1}i^*(\bar{B})$. We may also fix a number $N(p,r)$ only depending on p and r such that $N(p,r) \ge \dim\ p^{*-1}i^*(\bar{\mathcal{H}})$, and notice that the largest difference of the degrees of any two elements in $p^{*-1}i^*(\bar{\mathcal{H}})$ is bounded by $2(p-1)m_r$. Suppose the even part of $p^{*-1}i^*(\bar{\mathcal{H}})$

concentrates in dimension $2t_1, 2t_2, \dots$. Then for sufficiently large m_r we have

$$\begin{aligned} e\Big(\prod_{j\neq i}(k^{t_i}-k^{t_j})\Big) &\le \sum_{j\neq i}(e(t_i-t_j)+1) \\ &\le N(p,r)\lfloor \log_p(2(p-1)m_r)\rfloor + N(p,r) \\ &< m_r \end{aligned}$$

for any i, ie the condition (2-1) holds, which contradicts the existence of the nontrivial p^{th} power in $p^{*-1}i^*(\bar{\mathcal{H}})$. Accordingly the largest dimension of the generators is bounded and there are only finitely many possible types for X. Also by [6, Corollary 4.2], there are only finitely many homotopy types for each certain type. Then in all there are finitely many homotopy types for fixed rank. □

4 Rank 3 mod 3 homotopy associative H–spaces

For rank 3 mod 3 homotopy associative H–spaces, we will consider Stasheff's 3–projective space instead of Hemmi's modified projective space used in the proof of Theorem 1.1. The key lemma analogous to Lemma 3.6 for projective spaces is the following well-known result.

Lemma 4.1 (see eg [17]) *Let $n \ge 3$ and X be a finite A_n–space with cohomology ring*

$$H^*(X,\mathbb{Z}/p\mathbb{Z}) \cong \bigwedge(x_{2m_1-1},\dots,x_{2m_r-1}), \quad \deg(x_{2m_i-1}) = 2m_i-1,$$

such that each x_{2m_i-1} is A_n–primitive, ie x_{2m_i-1} lies in the image of a series of natural morphisms

$$H^*(P_n(X)) \to H^*(P_{n-1}(X)) \to \cdots \to H^*(P_1(X)=\Sigma X) \xleftarrow{\cong} H^{*-1}(X).$$

Then we have ring isomorphism

$$H^*(P_n(X),\mathbb{Z}/p\mathbb{Z}) \cong A \oplus B = \mathbb{Z}/p\mathbb{Z}[y_{2m_1},\dots,y_{2m_r}]/(\text{height } n+1) \oplus B$$

as $\mathscr{A}_p$–modules and $A^+\cdot B = 0$, where $\deg(y_{2m_i}) = 2m_i$.

Notice that the corresponding result in the context of K–theory can be easily deduced, and for rank 3 mod 3 homotopy associative H–spaces, the primitivity assumption is automatically satisfied. To prove Theorem 1.2, we will also use the following theorem of Wilkerson.

Theorem 4.2 [24, Theorems 6.1 and 6.2] *Let X be a finite* mod p *A_p–space with cohomology ring $H^*(X,\mathbb{Z}/p\mathbb{Z}) \cong \bigwedge(x_{2m_1-1},\dots,x_{2m_r-1})$, with $m_1 \le m_2 \le \cdots \le m_r$* and *$m_r > p$. Then:*

(1) *There is an x_{2m_k-1} with $m_r - m_k = s(p-1)$ for some $1 \le s \le e(m_r)+1$.*

(2) *If $p \nmid m_i$ for some i, there is an x_{2m_j-1} such that $m_j = k_j m_i - p + 1$ for some $1 \le k_j \le p$.*

Combining Theorem 1.1 and Theorem 4.2, we are left to consider the following four cases for the possible types of the mod 3 A_3–space X in Theorem 1.2:

Case 1 $3 \mid m$, $3 \mid n$ and $m-n=2s$ with $1 \le s \le e(m)+1$,

Case 2 $3 \mid m$, $3 \nmid n$ and $m-n=2s$ with $1 \le s \le e(m)+1$,

Case 3 $3 \nmid m$ and $m-n=2s$ with $1 \le s \le e(m)+1$,

Case 4 $m-r=2t$ with $1 \le t \le e(m)+1$, and $m-n \ne 2s$ for any s such that $1 \le s \le e(m)+1$.

For Case 1, we need the following lemma:

Lemma 4.3 *Under the condition of Theorem 1.2 and Case 1, we have:*

(1) *If $r=2$, $m>n>6$ and $e(m) \ge e(n)+2$, then*
$$8e(n)+23 \ge n.$$

(2) *If $r=2$, $m>n>6$ and $e(m)=e(n)+1$, then*
$$8\max\{e(3n-m),e(3n-2m)\}+15 \ge n.$$

(3) *If $m \le 3r$, $e(m) \ge e(n)+2$, then*
$$7e(n)+\lfloor \log_3(m-r)\rfloor+24 \ge m \quad \text{or} \quad 8\lfloor \log_3(m-r)\rfloor+24 \ge 3r.$$

(4) *If $m \le 3r$, $e(m)=e(n)+1$, then*
$$7\max\{e(3n-m),e(3n-2m)\}+\lfloor \log_3(m-r)\rfloor+17 \ge m$$
or
$$8\lfloor \log_3(m-r)\rfloor+24 \ge 3r.$$

Proof By the condition, we have a $\{\psi^k\}$–module $K=\mathbb{Z}_{(3)}[x_r,x_n,x_m]/(\text{height }4)$, where the subscripts refer to the filtration degree. For (1) and (2) we have $r=2$, and we only need to consider $K'=K-\{x_r^i\mid i=1,2,3\}$. We can set

$$S=\bigl\{2i+jn+km\mid (i,j,k)\neq(i,0,0),\ 0\le|j|,|k|\le 3,\ 0\le|i|\le 2\bigr\},$$

and define $\Phi(i,j,k)=|2i+jn+km|$. For (1) we have $e(\Phi(0,j,k))\le e(n)+1$ and $e(\Phi(i,j,k))=0$ if $|i|=1$, or 2. And we notice that there are nine elements of the form $x_n^*x_m^*$, five elements of the form $x_r^1x_n^*x_m^*$, and two elements of the form $x_r^2x_n^*x_m^*$ in K'. Then

$$\begin{aligned}e\Bigl(\prod_{(\tilde{i},\tilde{j},\tilde{k})\neq(0,j,k)}(2^{jn+km}-2^{2\tilde{i}+\tilde{j}n+\tilde{k}m})\Bigr)&\le\sum e\bigl(\Phi(-\tilde{i},j-\tilde{j},k-\tilde{k})\bigr)+15\\&\le 8(e(n)+1)+15\\&=8e(n)+23.\end{aligned}$$

Similarly, we have

$$e\Bigl(\prod_{(\tilde{i},\tilde{j},\tilde{k})\neq(1,j,k)}\Bigr)\le 4e(n)+19\quad\text{and}\quad e\Bigl(\prod_{(\tilde{i},\tilde{j},\tilde{k})\neq(2,j,k)}\Bigr)\le e(n)+16.$$

Since condition (2-1) should fail for X, we must have $8e(n)+23\ge n$.

The remaining three claims can be proved similarly, and notice that for (3) and (4), we work with $K'=K-\{x_r,x_n\}$ if $m\le 2r$ and with $K'=K-\{x_r,x_n,x_r^2\}$ if $m>2r$. □

Now we are ready to deal with Case 1:

Proposition 4.4 *Under the condition of Theorem 1.2 and Case 1, the only possible types of X are*

$$(2,3,9),\ (2,12,18),\ (2,21,27),\ (2,30,36),\ (2,39,45),$$
$$(7,12,18),\ (10,12,18),\ (16,30,36),\ (19,30,36).$$

Proof By Theorem 1.1, we have $\gcd(r,n,m)\le 2$, so $3\nmid r$. Hence by Theorem 4.2, we have $x=\lambda r-2$ with $\lambda\in\{1,2,3\}$ and $x\in\{r,n,m\}$. Then $r=2$ or $n=2r-2$ or $m=2r-2$.

We prove the proposition under the condition $e(m)>e(n)$ first:

(1) If $r=2$, $n>6$ and $e(m)\geq e(n)+2$, by Lemma 4.3 we have $8e(n)+23\geq n$. Then

$$\begin{aligned} 3^{e(m)}\cdot f &= m = n+2s \\ &\leq 8e(n)+23+2(e(n)+1) \\ &= 10e(n)+25 \\ &\leq 10e(m)+5. \end{aligned}$$

Since $e(m)\geq 3$, we have $m=27$ and $e(m)=3$. Then $e(n)=e(s)=1$ and n is odd. Now it is not hard to check that $(2,21,27)$ is the only possible type satisfying all the conditions.

(2) If $r=2$, $n>6$ and $e(m)=e(n)+1$, by Lemma 4.3,

$$8\max\{e(3n-m),e(3n-2m)\}+15\geq n.$$

If $8e(3n-m)+15\geq n$, then

$$8e(n-s)+12\geq 8e(n-s)+15-s\geq n-s$$

for

$$e(n-s)=e(2(n-s)=3n-m)\geq e(m)\geq 2 \quad\text{and}\quad s\geq 3.$$

Then it is easy to show that $n-s=9,18$ or 27. In any case, $s\leq e(m)+1\leq 4$, which implies $s=3$. And then $m-n=6$ and $n=12,21$ or 30. But since $e(m)=e(n)+1=2$, only $(2,12,18)$ or $(2,30,36)$ is possible for our X.

If $8e(3n-2m)+15\geq n$, then

$$8e(n-4s)+3\geq 8e(n-4s)+15-4s\geq n-4s$$

for

$$n-4s=3n-2m \quad\text{and}\quad s\geq 3.$$

Then we get $n-4s=9,18$ or 27. Again since $e(n-4s)\geq e(m)\geq 2$ and $s\leq e(m)+1$, we have $s=3$. Then $m-n=6$ and $n=21,30$ or 39 and only $(2,30,36)$ and $(2,39,45)$ survive.

(3) If $m\leq 3r$, $e(m)\geq e(n)+2$, by Lemma 4.3 we have

$$7e(n)+\lfloor\log_3(m-r)\rfloor+24\geq m \quad\text{or}\quad 8\lfloor\log_3(m-r)\rfloor+24\geq 3r.$$

We also notice that $r \neq 2$, which by our earlier discussion implies $n = 2r-2$ or $n = 2m-2$. If the first inequality and $n = 2r-2$ hold, then

$$\begin{aligned} 2r-2 = n < m &\leq 7e(n) + \lfloor \log_3(m-r) \rfloor + 24 \\ &\leq 7e(r-1) + \lfloor \log_3(2r) \rfloor + 24 \\ &\leq 8\lfloor \log_3 r \rfloor + 25, \end{aligned}$$

which implies $r \leq 21$. Then $m \leq 3r \leq 63$ implies $m = 27$ or 54 for $e(m) \geq 3$. So $e(m) = 3$ and $e(n) = 1$. Since $3 \mid s$ and $s \leq e(m)+1$, we have $s = 3$ and $m-n = 6$. Then we see $m = 27$ is impossible for n is even, while $m = 54$ leads to $r = 25$, which contradicts our previous calculation. Similar arguments can be applied to the other three cases, which will show there are no types left.

(4) If $m \leq 3r$, $e(m) = e(n)+1$, by Lemma 4.3 and similar calculations as in part (3), we get $(r,n,m) = (7,12,18)$, $(10,12,18)$, $(16,30,36)$ or $(19,30,36)$.

(5) By Theorem 1.1, the only remaining case under condition $e(m) > e(n)$ is $n \leq 3r$ but $m > 3r$. If $r = 2$, then $n = 3$ or 6, which gives $(r,n,m) = (2,3,9)$. When $n = 2r-2$, we have $\frac{1}{3}m + 2 < m-n = 2s \leq 2e(m)+2$, which is impossible. Further, $m = 2r-2$ can not hold by our assumption.

We have proved the proposition when $e(m) > e(n)$. If $e(n) \geq e(m)$, then $e(s) = e(m-n) \geq e(m) \geq s-1 \geq 0$, which implies $s = 1$ and $m-n = 2$. However, since $3 \mid m$ and $3 \mid n$, this is impossible. □

For the remaining cases, we will also use a theorem of Hemmi:

Theorem 4.5 ([12, Theorem 1.2]; also see [13, Section 8]) *Let X be a homotopy H–space with $H^*(X;\mathbb{Z}/3\mathbb{Z})$ being finite. Then for any $n \in \mathbb{Z}$ with $n \not\equiv 0 \bmod 3$ and $n > 3$, if*

$$QH^{2(3^a \cdot 2t)-1}(X,\mathbb{Z}/3\mathbb{Z}) = 0 \quad \text{for } t \geq n-1, \tag{4-1}$$

then

$$\mathscr{P}^{3^a}\colon QH^{2(3^a(n-2))-1}(X,\mathbb{Z}/3\mathbb{Z}) \to QH^{2(3^a n)-1}(X,\mathbb{Z}/3\mathbb{Z}) \tag{4-2}$$

is an epimorphism, where $QH^*(X,\mathbb{Z}/3\mathbb{Z}) = H^*(X,\mathbb{Z}/3\mathbb{Z})/DH^*(X,\mathbb{Z}/3\mathbb{Z})$ and $DH^*(X,\mathbb{Z}/3\mathbb{Z})$ is the submodule consisting of decomposable elements.

Proposition 4.6 *Under the conditions of Theorem 1.2 and Case 2, the only possible types of X are*

$$(2,4,6),\ (3,4,6),\ (3,5,9),\ (6,8,12).$$

Proof Since $3 \nmid n$, by Theorem 4.2, we have $x = \lambda n - 2$ with x and λ as before. Then either $r = n-2$ or $m = 2n-2$.

(1) $r = n-2$. If $m > 3r$, then $m - n > m - \left(\frac{1}{3}m + 2\right) = \frac{2}{3}m - 2$. So we have $\frac{2}{3}m - 2 = 2s < 2e(m) + 2$, which implies $m = 9$. Then $(r,n,m) = (2,4,9)$ contradicts the fact that $m-n$ is even.

If $2n-2 = 2r+2 \le m \le 3r$, then $\frac{1}{2}m - 1 \le m - n = 2s \le 2e(m)+2$, which implies $(r,n,m) = (2,4,6)$ or $(3,5,9)$.

If $m < 2n-2$ and $n = 3k+2$ for some k, then in the $\mathscr{A}_p$–module

$$\bar{K} = \mathbb{Z}/3\mathbb{Z}[x_r, x_n, x_m]/(\text{height } 4),$$

$\mathscr{P}^1(x_r) = c x_n$ with $c \not\equiv 0 \bmod 3$ by Theorem 4.5. By the Adem relation

$$\mathscr{P}^1 \mathscr{P}^3 \mathscr{P}^{3k-1} = \epsilon \mathscr{P}^1 \mathscr{P}^{3k+2} + 2\mathscr{P}^{3k+2} \mathscr{P}^1, \tag{4-3}$$

we have $\mathscr{P}^{3k-1}(x_r) \neq 0$, which implies $9k-2 = 3n-8$ has to be the degree of some monomial in K. Then by direct computation, we get $n = 8$ and $r = 6$, which implies $m < 14$. Since $3 \mid m$, we have $m = 9$ or 12. When $m = 9$, $m - n = 1$ is odd, which is impossible. So we have $(r,n,m) = (6,8,12)$.

If $m < 2n-2$ and $n = 3k+1$, then $r = 3k-1$ which by Theorem 4.2 implies $x = \lambda r - 2$ with $x \in \{r,n,m\}$ and $\lambda \in \{1,2,3\}$. Then we have $r = 2$ or $n = 2r-2$, both of which are impossible.

(2) $m = 2n-2$. We have $\frac{1}{2}m - 1 = m - n = 2s \le 2e(m)+2$, which implies $(r,n,m) = (2,4,6)$ or $(3,4,6)$. □

Proposition 4.7 *Under the conditions of Theorem 1.2 and Case 3, the only possible types of X are*

$$(2,3,5),\ (2,6,8),\ (3,5,7),\ (3,6,8),\ (4,6,8),\ (5,6,8), (6,8,10),$$
$$(8,12,14),\ (12,18,20),\ (18,24,26),\ (21,27,29),\ (30,36,38).$$

Proof Since $3 \nmid m$, we have $m - n = 2$. Then by Theorem 4.5, we have $\mathscr{P}^1(x_n) \neq 0$.

(1) If $m = 3k+1$, we have $n = 3k-1$, which by Theorem 4.2 implies $x = \lambda n - 2$ as before. Then $r = n-2$, or $m = 2n-2$, or $m = 3n-2$; the latter two cases are easy to check and are impossible. For $r = n-2$, we apply Theorem 4.5 to get $\mathscr{P}^1(x_r) \neq 0$, and again by Adem relation (4-3), we get $\mathscr{P}^{r-1}(x_r) \neq 0$, which implies $(r,n,m) = (3,5,7)$ or $(6,8,10)$.

(2) If $m = 3k+2$, again by Adem relation (4-3) we have $\mathscr{P}^{n-1}(x_n) \neq 0$. By comparing the degree and applying Theorem 4.2, we get a list of possible types:

$(2,3,5)$, $(2,6,8)$, $(3,6,8)$, $(4,6,8)$, $(5,6,8)$, $(8,12,14)$ and also a special type $(r, r+6, r+8)$ with $3 \mid r$. For this remaining case, if $r = 3l$ with $l \not\equiv 1 \bmod 3$, Theorem 4.5 implies $\mathscr{P}^3(x_r) \neq 0$. By the Adem relation

$$\mathscr{P}^9\mathscr{P}^{3l-1} = \epsilon_1\mathscr{P}^{3l+8} + \epsilon_2\mathscr{P}^{3l+7}\mathscr{P}^1 + \epsilon_3\mathscr{P}^{3l+6}\mathscr{P}^2 + \mathscr{P}^{3l+5}\mathscr{P}^3, \tag{4-4}$$

we have $\mathscr{P}^{3l-1}(x_r) \neq 0$, which gives $(r,n,m) = (18,24,26)$.

For $l \equiv 1 \bmod 3$, we argue similarly as in Lemma 4.3 to get the condition $m \leq 44$. Then the possible types are $(12,18,20)$, $(21,27,29)$ and $(30,36,38)$. □

Proposition 4.8 *Under the condition of Theorem 1.2 and Case 4, the only possible types of X are*

$$(2,3,4),\ (2,3,6).$$

Proof If $m > 3r$, then $2t = m - r > 2r$, ie $r < t$. Then we have

$$m = r + 2t < 3t \leq 3e(m) + 3,$$

which is impossible. So we have $m \leq 3r$.

If $3 \nmid m$, then $m - r = 2$ and $(r,n,m) = (r, r+1, r+2)$. Further, if $3 \mid r$, then $3 \nmid n$, which implies $x = \lambda n - 2$ as usual. However, it is easy to check the latter is impossible. Then we get $3 \nmid r$, which implies $x = \lambda r - 2$. In this case, the only possible type is $(r,n,m) = (2,3,4)$.

Now suppose $3 \mid m$. If $3 \nmid r$, we have $r = 2$, $n = 2r-2$, $n = 3r-2$ or $m = 2r-2$ by Theorem 4.2. When $r = 2$, we get $(r,n,m) = (2,3,6)$, while $(2,5,6)$ is impossible since $\lambda 5 - 2 \in \{3, 8, 13\}$. When $n = 2r-2$, we have $r = \frac{1}{2}n + 1 < \frac{1}{2}m + 1$. Then $\frac{1}{2}m - 1 < m - r = 2t \leq 2e(m) + 2$, which implies $m = 6$ or 9. When n is even, $n = 4$ when $m = 6$, which implies $r = 3$. But $3 \nmid r$, so $m = 6$ is impossible. If $m = 9$, then we have $(r,n,m) = (4,6,9)$ or $(5,8,9)$, both of which are impossible since $9 - 4 \neq 2t$ and $\lambda 8 - 2 \in \{6, 14, 22\}$. The other two cases can be treated similarly and lead to no possible types.

If $3 \mid r$, then $3 \nmid n$, which implies $r = n-2$ or $m = 2n-2$. When $r = n-2$, we argue exactly as in the proof of the first case in Proposition 4.6 and get no possible types in this case. When $m = 2n-2$, we see $r < n = \frac{1}{2}m + 1$, which implies $\frac{1}{2}m - 1 < m - r = 2t \leq 2e(m) + 2$. Again, no types survive. □

We recall the following theorem of Wilkerson and Zabrodsky [26], which was also reproved by McCleary [18], and later strengthened by Hemmi in [14] where the assumption of the primitivity of the generators was removed:

Theorem 4.9 *Let X be a simply connected* mod p *H–space with cohomology ring $H^*(X,\mathbb{Z}/p\mathbb{Z}) = \bigwedge(x_{2m_1-1},\dots,x_{2m_r-1})$, with $m_1 \le m_2 \le \cdots \le m_r$. If $m_r - m_1 < 2(p-1)$, then X is p–quasiregular, ie X is p–equivalent to a product of odd spheres and copies of $B_n(p)$, where $B_n(p)$ is the S^{2n+1}–fibration over $S^{2n+1+2(p-1)}$ characterized by α_p.*

Now we are ready to prove Theorem 1.2.

Proof of Theorem 1.2 We collect all the types obtained from Propositions 4.4, 4.6, 4.7 and 4.8, and prove the theorem case by case.

First, we notice that $(2,3,4)$, $(2,3,5)$, $(3,4,6)$ and $(5,6,8)$ are quasiregular by Theorem 4.9.

If $(r,n,m) = (4,6,8)$, we already know $\mathscr{P}^1(x_n) = x_m$ in $\bar{K} = \mathbb{Z}/3\mathbb{Z}[x_r,x_n,x_m]/(\text{height } 4)$. Then for degree reasons we have

$$\mathscr{P}^4(x_r) = \mathscr{P}^1\mathscr{P}^3(x_r) = \mathscr{P}^1(\lambda x_r x_n) = \lambda\mathscr{P}^1(x_r)x_n + \lambda x_r x_m,$$

which contradicts that $\mathscr{P}^4(x_r) = x_r^3$. So $(4,6,8)$ cannot be the type of X.

If $(r,n,m) = (3,5,9)$, we still have $\mathscr{P}^1(x_r) = x_n$ by Theorem 4.5. Then by Adem relation (4-3), we have $\mathscr{P}^2(x_r) \ne 0$, which is impossible since $K_7 = 0$.

If $(r,n,m) = (8,12,14)$, we know $\mathscr{P}^1(x_n) = x_m$ in $\bar{K}$. Then for degree reasons we have

$$2\mathscr{P}^8(x_r) = \mathscr{P}^1\mathscr{P}^1\mathscr{P}^6(x_r) = \mathscr{P}^1\mathscr{P}^1(\lambda x_r x_n) = \lambda x_r \mathscr{P}^1(x_m),$$

which implies $\mathscr{P}^1(x_m) = \mu x_r^2$ with $3 \nmid \mu$. On the other hand, we have $\mathscr{P}^{11}(x_n) \ne 0$ from the proof of Proposition 4.7, which implies that $\mathscr{P}^1\colon \bar{K}_{30} = \mathbb{Z}/p\mathbb{Z}(x_r^2 x_m) \to \bar{K}_{32}$ is not the zero map. But $\mathscr{P}^1(x_r^2 x_m) = x_r^2\mathscr{P}^1(x_m) = 0$ and then $(r,n,m) = (8,12,14)$ is impossible.

If $(r,n,m) = (10,12,18)$, we have $\mathscr{P}^1\mathscr{P}^9(x_r) = \mathscr{P}^{10}(x_r) = x_r^3$, which implies $\mathscr{P}^1(x_r x_m) = x_r\mathscr{P}^1(x_m) + \mathscr{P}^1(x_r)x_m = \lambda x_r^3$ with $3 \nmid \lambda$. Then we have $\mathscr{P}^1(x_r) = 0$ and $\mathscr{P}^1(x_m) = \lambda x_r^2$. Then by the Adem relation

$$\mathscr{P}^3\mathscr{P}^7 = -\mathscr{P}^{10} + \mathscr{P}^9\mathscr{P}^1, \tag{4-5}$$

we have $\mathscr{P}^3(x_n^2) = \mu x_r^3$ with $3 \nmid \mu$. However, $\mathscr{P}^3(x_n^2) = 2x_n\mathscr{P}^3(x_n)$ is not equal to μx_r^3, so $(10,12,18)$ cannot be the type of X.

If $(r,n,m) = (12,18,20)$, we have $\mathscr{P}^1(x_n) = x_m$. Again, by Adem relation (4-3), we have $\mathscr{P}^{17}(x_n) \ne 0$ and $\mathscr{P}^3\colon \bar{K}_{52} = \mathbb{Z}/p\mathbb{Z}(x_r x_m^2) \to \bar{K}_{58} = \mathbb{Z}/p\mathbb{Z}(x_n x_m^2)$ is not the zero map, which implies $\mathscr{P}^3(x_r) = x_n$. However, the Adem relation

$$\mathscr{P}^3\mathscr{P}^9 = \mathscr{P}^{12} + \mathscr{P}^{11}\mathscr{P}^1 \tag{4-6}$$

implies $\mathscr{P}^3(x_r x_n) = \pm x_r^3$, which contradicts the equality

$$\mathscr{P}^3(x_r x_n) = x_r \mathscr{P}^3(x_n) + \mathscr{P}^3(x_r) x_n = x_r \mathscr{P}^3(x_n) + x_n^2.$$

So $(r,n,m) = (12,18,20)$ is impossible.

For $(r,n,m) = (2,12,18)$, or $(7,12,18)$, we first prove the following lemma:

Lemma 4.10 *Let X be a p–local A_p–space with cohomology ring $H^*(X,\mathbb{Z}/p\mathbb{Z}) \cong \bigwedge(x_{2m_1-1},\dots,x_{2m_r-1})$, such that each x_{2m_i-1} is A_p–primitive, $m_1 \le m_j$ for all j, and $p < m_r$. Then there is an x_{2m_k-1} such that $\mathscr{P}^i(x_{2m_k-1}) = x_{2m_r-1}$ for some suitable nonzero i.*

Proof This is essentially [24, Lemma 4.4], which claims that in the $\{\psi^p\}$–submodule $K = \mathbb{Z}_{(p)}[x_{m_1},\dots,x_{m_r}]/(\text{height } p+1)$ of $K(P_p(X)) \otimes \mathbb{Z}_{(p)}$, there is an x_{m_k} such that

$$\psi^p(x_{m_k}) = \lambda x_{m_r} + \text{other terms}$$

with $\lambda \neq 0$, for in [5, Theorem 6.5], Atiyah has shown that if $\psi^p(x_q) = \sum_i p^{q-i} x_i$, then $\mathscr{P}^i(\bar{x}_q) = \bar{x}_i$ holds on the cohomology level. □

Now we return to the proof Theorem 1.2. Using Lemma 4.10, we see $\mathscr{P}^3(x_{12}) = x_{18}$ holds in $\bar{K} \subset H^*(P_3(X))$ for both mentioned cases. Then we apply Adem relation (4-6) to x_{12}. Since in both cases $\mathscr{P}^{11}\mathscr{P}^1(x_{12}) = 0$, we have $\mathscr{P}^3\mathscr{P}^9(x_{12}) = \pm x_{12}^3$. However, $\bar{K}_{30} = \mathbb{Z}/p\mathbb{Z}(x_{12}x_{18})$, and since $\bar{K}$ is truncated,

$$\mathscr{P}^3(x_{12}x_{18}) = x_{12}\mathscr{P}^3(x_{18}) + \mathscr{P}^3(x_{12})x_{18} = x_{12}\mathscr{P}^3(x_{18}) + x_{18}^2,$$

which is not equal to $\pm x_{12}^3$. Accordingly, neither case can be the type of X.

We notice that $(r,n,m) = (2,3,6)$ is impossible directly by the above lemma.

For the remaining cases which do not appear in the final list, we can check whether the condition (2-1) fails or not in an appropriate $\{\psi^k\}$–module K' constructed from K (with the help of a computer), and find that (2-1) holds when (r,n,m) is one of $(2,3,9)$, $(2,21,27)$, $(2,30,36)$, $(2,39,45)$, $(18,24,26)$, $(16,30,36)$, $(19,30,36)$, $(21,27,29)$ or $(30,36,38)$, which implies X cannot be a mod 3 A_3–space. □

Acknowledgements

The authors would like to thank Professor Stephen D Theriault and Professor Mamoru Mimura for helpful discussions and comments, and are also indebted to Professor John R Harper for suggesting the reference [12] and valuable knowledge about H–spaces of rank p and higher associativity. We wish to thank the referee most warmly for suggestions and comments on using the modified projective space of Hemmi [13] which has greatly improved the article, and also the careful reading of our manuscript. We are also indebted to Professor Jérôme Scherer and Professor Fred Cohen for careful reading of the manuscript and many valuable suggestions which have improved the paper.

The authors are partially supported by the Singapore Ministry of Education research grant (AcRF Tier 1 WBS No. R-146-000-222-112). The second author is also supported by a grant (No. 11329101) of NSFC of China.

References

[1] **J F Adams**, *On the non-existence of elements of Hopf invariant one*, Ann. of Math. 72 (1960) 20–104 MR

[2] **J F Adams**, *The sphere, considered as an H–space* mod p, Quart. J. Math. Oxford. Ser. 12 (1961) 52–60 MR

[3] **J F Adams**, *H–spaces with few cells*, Topology 1 (1962) 67–72 MR

[4] **J F Adams**, **M F Atiyah**, *K–theory and the Hopf invariant*, Quart. J. Math. Oxford Ser. 17 (1966) 31–38 MR

[5] **M F Atiyah**, *Power operations in K–theory*, Quart. J. Math. Oxford Ser. 17 (1966) 165–193 MR

[6] **R A Body**, **R R Douglas**, *Homotopy types within a rational homotopy type*, Topology 13 (1974) 209–214 MR

[7] **R R Douglas**, **F Sigrist**, *Sphere bundles over spheres and H–spaces*, Topology 8 (1969) 115–118 MR

[8] **J Grbić**, **J Harper**, **M Mimura**, **S Theriault**, **J Wu**, *Rank $p-1$ mod-p H–spaces*, Israel J. Math. 194 (2013) 641–688 MR

[9] **N L Hagelgans**, *Local spaces with three cells as H–spaces*, Canad. J. Math. 31 (1979) 1293–1306 MR

[10] **J R Harper**, *The* mod 3 *homotopy type of* F_4, from "Localization in group theory and homotopy theory, and related topics" (P Hilton, editor), Lecture Notes in Math. 418, Springer (1974) 58–67 MR

[11] **J Harper**, **A Zabrodsky**, *Evaluating a p^{th} order cohomology operation*, Publ. Mat. 32 (1988) 61–78 MR

[12] **Y Hemmi**, *Homotopy associative finite H–spaces and the* mod 3 *reduced power operations*, Publ. Res. Inst. Math. Sci. 23 (1987) 1071–1084 MR

[13] **Y Hemmi**, *On exterior A_n–spaces and modified projective spaces*, Hiroshima Math. J. 24 (1994) 583–605 MR

[14] **Y Hemmi**, *Mod p decompositions of mod p finite H–spaces*, Mem. Fac. Sci. Kochi Univ. Ser. A Math. 22 (2001) 59–65 MR

[15] **J R Hubbuck**, *Generalized cohomology operations and H–spaces of low rank*, Trans. Amer. Math. Soc. 141 (1969) 335–360 MR

[16] **J R Hubbuck**, **M Mimura**, *The number of mod p $A(p)$–spaces*, Illinois J. Math. 33 (1989) 162–169 MR

[17] **N Iwase**, *On the K–ring structure of X–projective n–space*, Mem. Fac. Sci. Kyushu Univ. Ser. A 38 (1984) 285–297 MR

[18] **J McCleary**, *Mod p decompositions of H–spaces; another approach*, Pacific J. Math. 87 (1980) 373–388 MR

[19] **M Mimura**, **G Nishida**, **H Toda**, *On the classification of H–spaces of rank 2*, J. Math. Kyoto Univ. 13 (1973) 611–627 MR

[20] **M Mimura**, **G Nishida**, **H Toda**, *Mod p decomposition of compact Lie groups*, Publ. Res. Inst. Math. Sci. 13 (1977/78) 627–680 MR

[21] **J D Stasheff**, *Homotopy associativity of H–spaces, I*, Trans. Amer. Math. Soc. 108 (1963) 275–292 MR

[22] **J D Stasheff**, *Homotopy associativity of H–spaces, II*, Trans. Amer. Math. Soc. 108 (1963) 293–312 MR

[23] **J D Stasheff**, *The mod p decomposition of Lie groups*, from "Localization in group theory and homotopy theory, and related topics" (P Hilton, editor), Lecture Notes in Math. 418, Springer (1974) 142–149 MR

[24] **C Wilkerson**, *K–theory operations in mod p loop spaces*, Math. Z. 132 (1973) 29–44 MR

[25] **C Wilkerson**, *Spheres which are loop spaces mod p*, Proc. Amer. Math. Soc. 39 (1973) 616–618 MR

[26] **C Wilkerson**, *Mod p decomposition of mod p H–spaces*, from "Algebraic and geometrical methods in topology" (L F McAuley, editor), Lecture Notes in Math. 428, Springer (1974) 52–57 MR

[27] **A Zabrodsky**, *Homotopy associativity and finite CW complexes*, Topology 9 (1970) 121–128 MR

[28] **A Zabrodsky**, *The classification of simply connected H–spaces with three cells, I*, Math. Scand. 30 (1972) 193–210 MR

[29] **A Zabrodsky**, *The classification of simply connected H–spaces with three cells, II*, Math. Scand. 30 (1972) 211–222 MR

[30] **A Zabrodsky**, *On the realization of invariant subgroups of $\pi_*(X)$*, Trans. Amer. Math. Soc. 285 (1984) 467–496 MR

Department of Mathematics, National University of Singapore
10 Lower Kent Ridge Road, Singapore 119076, Singapore

a0123769@u.nus.edu, matwuj@nus.edu.sg

http://www.math.nus.edu.sg/~matwujie

Received: 11 February 2016 Revised: 9 December 2016

Algebraic & Geometric Topology 17 (2017) 2145–2176

Acylindrical group actions on quasi-trees

SAHANA H BALASUBRAMANYA

A group G is acylindrically hyperbolic if it admits a non-elementary acylindrical action on a hyperbolic space. We prove that every acylindrically hyperbolic group G has a generating set X such that the corresponding Cayley graph Γ is a (non-elementary) quasi-tree and the action of G on Γ is acylindrical. Our proof utilizes the notions of hyperbolically embedded subgroups and projection complexes. As an application, we obtain some new results about hyperbolically embedded subgroups and quasi-convex subgroups of acylindrically hyperbolic groups.

20F67; 20F65, 20E08

1 Introduction

Definition 1.1 An isometric action of a group G on a metric space (S, d) is *acylindrical* if for every $\epsilon > 0$ there exist $R, N > 0$ such that for every two points x, y with $\mathrm{d}(x, y) \geq R$, there are at most N elements $g \in G$ satisfying

$$\mathrm{d}(x, gx) \leq \epsilon \quad \text{and} \quad \mathrm{d}(y, gy) \leq \epsilon.$$

Obvious examples are provided by geometric (ie proper and cobounded) actions; note, however, that acylindricity is a much weaker condition.

In order to define an acylindrically hyperbolic group, we must define non-elementary actions, for which we will need the following definition and theorem.

Definition 1.2 Let G be a group acting on a hyperbolic metric space S. An element $g \in G$ is called *loxodromic* if the map $\mathbb{Z} \to S$ given by

$$n \mapsto g^n s$$

is a quasi-isometric embedding for some (equivalently any) $s \in S$. Every loxodromic element has exactly two limit points $\{g^{\pm\infty}\}$ on the Gromov boundary ∂S. Two loxodromic elements g, h are said to be *independent* if the sets $\{g^{\pm\infty}\}$ and $\{h^{\pm\infty}\}$ are disjoint.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2145

Theorem 1.3 (Osin [12, Theorem 1.1]) *Let G be a group acting acylindrically on a hyperbolic space S. Then exactly one of the following holds:*

(a) *G has bounded orbits.*

(b) *G is virtually cyclic and contains a loxodromic element.*

(c) *G has infinitely many independent loxodromic elements.*

Definition 1.4 An acylindrical action of a group G is said to be *elementary* in cases (a) and (b) above, and *non-elementary* is case (c). Equivalently, a non-elementary acylindrical action of a group G on a hyperbolic space is an action with unbounded orbits, and where G is not virtually cyclic.

Definition 1.5 A group G is called *acylindrically hyperbolic* if it admits a non-elementary acylindrical action on a hyperbolic space.

Over the last few years, the class of acylindrically hyperbolic groups has received considerable attention. It is broad enough to include many examples of interest, eg non-elementary hyperbolic and relatively hyperbolic groups, all but finitely many mapping class groups of punctured closed surfaces, $\mathrm{Out}(F_n)$ for $n \geq 2$, most 3–manifold groups, and finitely presented groups of deficiency at least 2. On the other hand, the existence of a non-elementary acylindrical action on a hyperbolic space is a rather strong assumption, which allows one to prove non-trivial results. In particular, acylindrically hyperbolic groups share many interesting properties with non-elementary hyperbolic and relatively hyperbolic groups. For details we refer to Dahmani, Guirardel and Osin [5], Minasyan and Osin [10], Osin [12; 11] and references therein.

The main goal of this paper is to answer the following.

Question 1.6 Which groups admit non-elementary cobounded acylindrical actions on quasi-trees?

By a quasi-tree we mean a connected graph which is quasi-isometric to a tree. Quasi-trees form a very particular subclass of the class of all hyperbolic spaces. From the asymptotic point of view, quasi-trees are exactly "1–dimensional hyperbolic spaces".

The motivation behind our question comes from the following observation. If instead of cobounded acylindrical actions we consider cobounded proper (ie geometric) ones, then there is a crucial difference between the groups acting on hyperbolic spaces and quasi-trees. Indeed, a group G acts geometrically on a hyperbolic space if and only if G is a hyperbolic group. On the other hand, Stallings' theorem on groups with infinitely many ends and Dunwoody's accessibility theorem implies that groups admitting geometric actions on quasi-trees are exactly virtually free groups. Yet another related observation is that acylindrical actions on unbounded locally finite graphs are necessarily proper. Thus if we restrict to quasi-trees of bounded valence in Question 1.6, we again obtain the class of virtually free groups. Other known examples of groups having non-elementary, acylindrical and cobounded actions on quasi-trees include groups associated with special cube complexes and right-angled Artin groups (see Behrstock, Hagen and Sisto [1], Hagen [6] and Kim and Koberda [8]).

Thus one could expect that the answer to Question 1.6 would produce a proper subclass of the class of all acylindrically hyperbolic groups, which generalizes virtually free groups in the same sense as acylindrically hyperbolic groups generalize hyperbolic groups. Our main result shows that this does not happen.

Theorem 1.7 *Every acylindrically hyperbolic group admits a non-elementary cobounded acylindrical action on a quasi-tree.*

In other words, being acylindrically hyperbolic is equivalent to admitting a non-elementary acylindrical action on a quasi-tree. Although this result does not produce any new class of groups, it can be useful in the study of acylindrically hyperbolic groups and their subgroups. In this paper we concentrate on proving Theorem 1.7 and leave applications for future papers to explore (for some applications, see [10]).

It was known before that every acylindrically hyperbolic group admits a non-elementary cobounded action on a quasi-tree satisfying the so-called *weak proper discontinuity* property, which is weaker than acylindricity. Such a quasi-tree can be produced by using projection complexes introduced by Bestvina, Bromberg and Fujiwara [2]. To the best of our knowledge, whether the corresponding action is acylindrical is an open question. The main idea of the proof of Theorem 1.7 is to combine the Bestvina–Bromberg–Fujiwara approach with an "acylindrification" construction from Osin [12], in order to make the action acylindrical. An essential role in this process is played by the notion of a hyperbolically embedded subgroup, introduced by Dahmani, Guirardel and Osin [5]. This fact is of independent interest since it provides a new setting for the application of the Bestvina–Bromberg–Fujiwara construction.

The above-mentioned construction has been applied in the setting of geometrically separated subgroups (see [5, Section 4.5]). However, not every hyperbolically embedded subgroup $H \leq G$ arises from an action of G on a hyperbolic space in which H is geometrically separated. Nevertheless, it is possible to employ hyperbolically embedded subgroups in this construction, possibly with interesting applications. If fact, we prove much stronger results in terms of hyperbolically embedded subgroups (see Theorem 3.1) of which Theorem 1.7 is an easy consequence, and derive an application which is stated below (see Corollary 3.24).

Corollary 1.8 *Let G be a group. If $H \leq K \leq G$, H is countable and H is hyperbolically embedded in G, then H is hyperbolically embedded in K.*

This result continues to hold even when we have a finite collection $\{H_1, H_2, \dots, H_n\}$ of hyperbolically embedded subgroups in G such that $H_i \leq K$ for $i = 1, 2, \dots, n$. Interestingly, A Sisto obtains a similar result in [14, Corollary 6.10]. His result does not require H to be countable, but under the assumption that $H \cap K$ is a virtual retract of K, it states that $H \cap K \hookrightarrow_h K$. Although similar, these two theorems are independent in the sense that neither follows from the other.

Another application of Theorem 3.1 is to the case of finitely generated subgroups, as stated below (see Corollary 3.27).

Corollary 1.9 *Let H be a finitely generated subgroup of an acylindrically hyperbolic group G. Then there exists a subset $X \subset G$ such that*

(a) *$\Gamma(G, X)$ is hyperbolic, and the action of G on $\Gamma(G, X)$ is non-elementary and acylindrical, and*

(b) *H is quasi-convex in $\Gamma(G, X)$.*

This result indicates that in order to develop a theory of quasi-convex subgroups in acylindrically hyperbolic groups, the notion of quasi-convexity is not sufficient, ie a stronger set of conditions is necessary in order to prove results similar to those known for quasi-convex subgroups in hyperbolic groups. For example, using Rips' construction [13] and Corollary 1.9, one can easily construct an example of an infinite, infinite-index, normal subgroup in an acylindrically hyperbolic group, which is quasi-convex with respect to some non-elementary acylindrical action.

Acknowledgements My heartfelt gratitude to my advisor Denis Osin for his guidance and support, and to Jason Behrstock and Yago Antolin Pichel for their remarks. I am also grateful to the referee for remarks. My sincere thanks to Bryan Jacobson for his thorough proofreading and comments on this paper.

2 Preliminaries

We recall some definitions and theorems which we will need to refer to.

2.1 Relative metrics on subgroups

Definition 2.1 (relative metric) Let G be a group and $\{H_\lambda\}_{\lambda\in\Lambda}$ a fixed collection of subgroups of G. Let $X \subset G$ such that G is generated by X along with the union of all $\{H_\lambda\}_{\lambda\in\Lambda}$. Let $\mathcal{H} = \bigsqcup_{\lambda\in\Lambda} H_\lambda$. We denote the corresponding Cayley graph of G (whose edges are labeled by elements of $X \sqcup \mathcal{H}$) by $\Gamma(G, X \sqcup \mathcal{H})$.

Remark 2.2 It is important that the union in the definition above is disjoint. This disjoint union leads to the following observation: for every $h \in H_i \cap H_j$, the alphabet $\mathcal{H}$ will have two letters representing h in G, one from H_i and another from H_j. It may also be the case that a letter from $\mathcal{H}$ and a letter from X represent the same element of the group G. In this situation, the corresponding Cayley graph $\Gamma(G, X \sqcup \mathcal{H})$ has bigons (or multiple edges in general) between the identity and the element, one corresponding to each of these letters.

We think of $\Gamma(H_\lambda, H_\lambda)$ as a complete subgraph in $\Gamma(G, X \sqcup \mathcal{H})$. A path p in $\Gamma(G, X \sqcup \mathcal{H})$ is said to be *λ–admissible* if it contains no edges of the subgraph $\Gamma(H_\lambda, H_\lambda)$. In other words, the path p does not travel through H_λ in the Cayley graph. Using this notion, we can define a metric $\hat{d}_\lambda\colon H_\lambda \times H_\lambda \to [0, \infty]$, known as the relative metric, by setting $\hat{d}_\lambda(h, k)$ for $h, k \in H_\lambda$ to be the length of the shortest admissible path in $\Gamma(G, X \sqcup \mathcal{H})$ that connects h to k. If no such path exists, we define $\hat{d}_\lambda(h, k) = \infty$. It is easy to check that $\hat{d}_\lambda$ is a metric.

Definition 2.3 Let q be a path in the Cayley graph of $\Gamma(G, X \sqcup \mathcal{H})$. A non-trivial subpath p of q is said to be an *H_λ–subpath* if the label of p (denoted $\mathrm{Lab}(p)$) is a word in the alphabet H_λ. Such a subpath is further called an *H_λ–component* if it is not contained in a longer H_λ–subpath of q. If q is a loop, we must also have that p is not contained in a longer H_λ–subpath of any cyclic shift of q.

We refer to an H_λ–component of q (for some $\lambda \in \Lambda$) simply by calling it a *component* of q. We note that, on a geodesic, H_λ–components must be single H_λ–edges. In general, however, the subpath p of q may consist of more than one edge.

Let p_1, p_2 be two H_λ–components of a path q for some $\lambda \in \Lambda$. These components are said to be *connected* if there exists a path p in $\Gamma(G, X \sqcup \mathcal{H})$ such that $\mathrm{Lab}(p)$ is a word consisting only of letters from H_λ, and p connects some vertex of p_1 to some vertex of p_2. In algebraic terms, this means that all vertices of p_1 and p_2 belong to the same (left) coset of H_λ. We refer to a component of a path q as *isolated* if it is not connected to any other component of q.

If p is a path, we denote its initial point by p_- and its terminating point by p_+.

Lemma 2.4 [5, Proposition 4.13] *Let G be a group and $\{H_\lambda\}_{\lambda\in\Lambda}$ a fixed collection of subgroups in G. Let $X \subset G$ such that G is generated by X together with the union of all $\{H_\lambda\}_{\lambda\in\Lambda}$. Then there exists a constant $C > 0$ such that for any n–gon p with geodesic sides in $\Gamma(G, X \sqcup \mathcal{H})$, any $\lambda \in \Lambda$ and any isolated H_λ–component a of p, $\widehat{d}_\lambda(a_-, a_+) \le Cn$.*

2.2 Hyperbolically embedded subgroups

Hyperbolically embedded subgroups will be our main tool in constructing the quasi-tree. The notion has been taken from Dahmani, Guirardel and Osin [5], where it was introduced. We recall the definition here.

Definition 2.5 (hyperbolically embedded subgroups) Let G be a group. Let X be a (not necessarily finite) subset of G and let $\{H_\lambda\}_{\lambda\in\Lambda}$ be a collection of subgroups of G. We say that $\{H_\lambda\}_{\lambda\in\Lambda}$ is *hyperbolically embedded* in G with respect to X (denoted by $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h (G, X)$) if the following conditions hold:

(a) The group G is generated by X together with the union of all $\{H_\lambda\}_{\lambda\in\Lambda}$.

(b) The Cayley graph $\Gamma(G, X \sqcup \mathcal{H})$ is hyperbolic, where $\mathcal{H} = \bigsqcup_{\lambda\in\Lambda} H_\lambda$.

(c) For every $\lambda \in \Lambda$, the metric space $(H_\lambda, \widehat{d}_\lambda)$ is proper, ie every ball of finite radius has finite cardinality.

Furthermore, we say that $\{H_\lambda\}_{\lambda\in\Lambda}$ is hyperbolically embedded in G (denoted by $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h G$) if $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h (G, X)$ for some $X \subseteq G$. The set X is called a *relative generating set*.

Since the notion of a hyperbolically embedded subgroup plays a crucial role in this paper, we include two examples borrowed from [5].

Example 2.6 Let $G = H \times \mathbb{Z}$ and $\mathbb{Z} = \langle x\rangle$. Let $X = \{x\}$. Then $\Gamma(G, X \sqcup H)$ is quasi-isometric to a line and is hence hyperbolic. The corresponding relative metric satisfies the inequality $\widehat{d}(h_1, h_2) \le 3$ for every $h_1, h_2 \in H$, which is easy to see from the Cayley graph (see Figure 1, left). Indeed, if Γ_H denotes the Cayley graph $\Gamma(H, H)$, then in its shifted copy $x\Gamma_H$, there is an edge e connecting xh_1 to xh_2 (labeled by $h_1^{-1}h_2 \in H$). There is thus an admissible path of length 3 connecting h_1 to h_2. We conclude that if H is infinite, then H is not hyperbolically embedded in (G, X), since the relative metric will not be proper. In this example, one can also note that the admissible path from h_1 to h_2 contains an H–subpath, namely the edge e, which is also an H–component of this path.

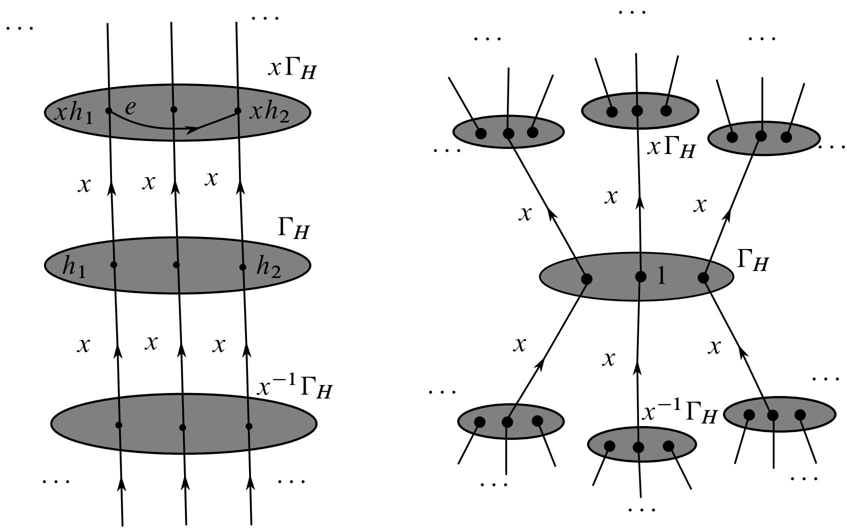


Figure 1: $H \times \mathbb{Z}$ (left) and $H * \mathbb{Z}$ (right)

Example 2.7 Let $G = H * \mathbb{Z}$ and $\mathbb{Z} = \langle x \rangle$. As in the previous example, let $X = \{x\}$. In this case $\Gamma(G, X \sqcup H)$ is quasi-isometric to a tree (see Figure 1, right) and it is easy to see that $\widehat{d}(h_1, h_2) = \infty$ unless $h_1 = h_2$. This means that every ball of finite radius in the relative metric has cardinality 1. We can thus conclude that $H \hookrightarrow_h (G, X)$.

2.3 A slight modification to the relative metric

The aim of this section is to modify the relative metric on countable subgroups that are hyperbolically embedded, so that the resulting metric takes values only in $\mathbb{R}$, ie is finite-valued. This will be of importance in Section 3. The main result of this section is the following.

Theorem 2.8 *Let G be a group. Let $H < G$ be countable and such that $H \hookrightarrow_h G$. Then there exists a left-invariant metric $\widetilde{d}\colon H \times H \to \mathbb{R}$ such that*

(a) $\widetilde{d} \le \widehat{d}$*, and*

(b) $\widetilde{d}$ *is proper, ie every ball of finite radius has finitely many elements.*

Proof There exists a collection of finite, symmetric (closed under inverses) subsets $\{F_i\}$ of H such that $H = \bigcup_{i=1}^{\infty} F_i$ and $1 \subseteq F_1 \subseteq F_2 \subseteq \cdots$.

Let $\widehat{d}$ be the relative metric on H. Let $H_0 = \{h \in H \mid \widehat{d}(1, h) < \infty\}$.

Define a function $w\colon H \to \mathbb{N}$ by

$$w(h) = \begin{cases} \widehat{d}(1, h) & \text{if } h \in H_0, \\ \min\{i \mid h \in F_i\} & \text{otherwise.} \end{cases}$$

Since the F_i are symmetric, $w(h) = w(h^{-1})$ for all $h \in H$. Define a function l on H as follows: for every word $u = x_1 x_2 \cdots x_k$ in the elements of H, set

$$l(u) = \sum_{i=1}^{k} w(x_i).$$

Define a length function on H by

$$|g|_w = \min\{l(u) \mid u \text{ is a word in the elements of } H \text{ that represents } g\},$$

for each g in H. We can now define a metric $d_w\colon H \times H \to \mathbb{N}$ by

$$d_w(g,h) = |g^{-1}h|_w.$$

It is easy to check that d_w is a (finite-valued) well defined metric. Since

$$d_w(ag,ah) = |(ag)^{-1}ah|_w = |g^{-1}a^{-1}ah|_w = |g^{-1}h|_w = d_w(g,h)$$

for all $a, g, h \in G$, the metric d_w is left-invariant. Further, it is easy to see that for all $h \in H$,

$$d_w(1,h) \le w(h).$$

It remains to show that d_w is proper. Let $N \in \mathbb{N}$. Suppose $h \in H$ such that $w(h) \le N$. If $h \in H_0$, then $\hat{d}(1,h) \le N$, which implies that there are finitely many choices for h, since $\hat{d}$ is proper. If $h \notin H_0$, then $h \in F_i$ for some minimal i. But each F_i is a finite set, so there are finitely many choices for h. Thus $|\{h \in H \mid w(h) \le N\}| < \infty$ for all $N \in \mathbb{N}$. This implies d_w is proper.

Indeed, if $y \neq 1$ is such that $|y|_w \le n$, then there exists a word u, written without the identity element (which has weight zero), representing y in the alphabet H such that $u = x_1 x_2 \cdots x_r$ and $\sum_{i=1}^{r} w(x_i) \le n$. Since $w(x_i) \ge 1$ for every $x_i \neq 1$, we have $r \le n$. Further, $w(x_i) \le n$ for all i. Thus $x_i \in \{x \in H \mid w(x) \le n\}$ for all i. So there are only finitely many choices for each x_i, which implies there are finitely many choices for y. By definition, $d_w \le \hat{d}$. So we can set $\tilde{d} = d_w$. □

2.4 Acylindrically hyperbolic groups

In the following theorem, ∂ represents the Gromov boundary.

Theorem 2.9 *For any group G, the following are equivalent:*

(AH_1) *There exists a generating set X of G such that the corresponding Cayley graph $\Gamma(G,X)$ is hyperbolic, $|\partial\Gamma(G,X)| \ge 2$ and the natural action of G on $\Gamma(G,X)$ is acylindrical.*

(AH_2) *G admits a non-elementary acylindrical action on a hyperbolic space.*

(AH_3) *G contains a proper infinite hyperbolically embedded subgroup.*

It follows from the definitions that $(\mathrm{AH}_1) \Longrightarrow (\mathrm{AH}_2)$. The implication $(\mathrm{AH}_2) \Longrightarrow (\mathrm{AH}_3)$ is non-trivial and was proved by Dahmani, Guirardel and Osin [5]. The implication $(\mathrm{AH}_3) \Longrightarrow (\mathrm{AH}_1)$ was proved by Osin [12].

Definition 2.10 We call a group G *acylindrically hyperbolic* if it satisfies any of the equivalent conditions (AH_1)–(AH_3) from Theorem 2.9.

Lemma 2.11 [5, Corollary 4.27] *Let G be a group, $\{H_\lambda\}_{\lambda\in\Lambda}$ a collection of subgroups of G, and X_1 and X_2 be relative generating sets. Suppose that $|X_1 \Delta X_2| < \infty$. Then $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h (G, X_1)$ if and only if $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h (G, X_2)$.*

Theorem 2.12 [12, Theorem 5.4] *Let G be a group, $\{H_\lambda\}_{\lambda\in\Lambda}$ a finite collection of subgroups of G, and X a subset of G. Suppose that $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h (G, X)$. Then there exists $Y \subset G$ such that the following conditions hold:*

(a) $X \subset Y$.

(b) $\{H_\lambda\}_{\lambda\in\Lambda} \hookrightarrow_h (G, Y)$. *In particular, the Cayley graph $\Gamma(G, Y \sqcup \mathcal{H})$ is hyperbolic.*

(c) *The action of G on $\Gamma(G, Y \sqcup \mathcal{H})$ is acylindrical.*

Definition 2.13 Let (X, d_X) and (Y, d_Y) be two metric spaces. A map $\phi\colon X \to Y$ is said to be a (λ, C)*–quasi-isometry* if there exist constants $\lambda > 1$, $C > 0$ such that

(a) $\frac{1}{\lambda} d_X(a,b) - C \le d_Y(\phi(a), \phi(b)) \le \lambda d_X(a,b) + C$, for all $a, b \in X$, and

(b) Y is contained in the C–neighborhood of $\phi(X)$.

The spaces X and Y are said to be *quasi-isometric* if such a map $\phi\colon X \to Y$ exists. It is easy to check that being quasi-isometric is an equivalence relation. If the map ϕ satisfies only condition (a), then it is said to be a (λ, C)*–quasi-isometric embedding*.

Definition 2.14 A graph Γ with the combinatorial metric d_Γ is said to be a *quasi-tree* if it is quasi-isometric to a tree T.

Definition 2.15 A *quasi-geodesic* is a quasi-isometric embedding of an interval $I \subseteq \mathbb{R}$ (bounded or unbounded) into a metric space X. Note that geodesics are $(1,0)$–quasi-geodesics. By slight abuse of notation, we may identify the map that defines a quasi-geodesic with its image in the space.

Theorem 2.16 [9, Theorem 4.6, bottleneck property] *Let Y be a geodesic metric space. The following are equivalent:*

(a) *Y is quasi-isometric to some simplicial tree Γ.*

(b) *There is some $\mu > 0$ such that for all $x, y \in Y$, there is a midpoint $m = m(x, y)$ with $d(x, m) = d(y, m) = \frac{1}{2}d(x, y)$ and the property that any path from x to y must pass within less than μ of the point m.*

We remark that if m is replaced with any point p on a geodesic between x and y, then the property that any path from x to y passes within less than μ of the point p still follows from (a), as proved below in Lemma 2.18. We will need the following lemma.

Lemma 2.17 [4, Proposition 3.1] *For all $\lambda \geq 1$, $C \geq 0$, $\delta \geq 0$, there exists an $R = R(\delta, \lambda, C)$ such that if X is a δ–hyperbolic space, γ is a (λ, C)–quasi-geodesic in X, and γ' is a geodesic segment with the same endpoints, then γ' and γ are Hausdorff distance less than R from each other.*

Lemma 2.18 *If Y is a quasi-tree, then there exists $\mu > 0$ such that for any point z on a geodesic connecting two points, any other path between the same endpoints passes within μ of z.*

Proof Let T be a tree and $q\colon Y \to T$ be the (λ, C)–quasi-isometry. Let d_Y and d_T denote the metrics in the spaces Y and T, respectively. Note that since T is 0–hyperbolic, Y is δ–hyperbolic for some δ.

Let x, y be two points in Y, joined by a geodesic γ. Let z be any point of γ and let α be another path from x to y. Let V denote the vertex set of α, ordered according to the geodesic γ. Take its image $q(V)$ and connect consecutive points by geodesics (of length at most $\lambda + C$) to get a path β in T from $q(x)$ to $q(y)$. Then the unique geodesic σ in T must be a subset of β. Since $q(V) \subset q \circ \alpha$, we get that any point of σ is at most $\lambda + C$ from $q \circ \alpha$. Also, $q \circ \gamma$ is a (λ, C)–quasi-isometric embedding of an interval, and hence a (λ, C)–quasi-geodesic. Thus, by Lemma 2.17 the distance from $q(z)$ to σ is less than $R = R(0, \lambda, C)$.

Let p be the point on σ closest to $q(z)$. There is a point $w \in Y$ on α such that $d(q(w), p) \leq \lambda + C$. Since $d(p, q(z)) < R$, we have $d(q(w), q(z)) \leq \lambda + C + R$. Thus

$$d(z, w) \leq \lambda^2 + 2\lambda C + R\lambda.$$

Thus α must pass within $\mu = \lambda^2 + 2\lambda C + R\lambda$ of the point z. □

2.5 A modified version of Bowditch's lemma

In this section, $\mathcal{N}_k(X)$ denotes the closed k–neighborhood of a set X in a metric space (S, d_S), ie

$$\mathcal{N}_k(X) = \{s \in S \mid \exists x \in X \text{ such that } d_S(s, x) \leq k\}.$$

In particular, $\mathcal{N}_k(x)$ denotes the closed k–neighborhood of a point x in a metric space.

The following theorem will be used in Section 3. Part (a) is a simplified form of a result taken from [7], which is in fact derived from a hyperbolicity criterion developed by Bowditch [3].

Theorem 2.19 *Let Σ be a hyperbolic graph and Δ be a graph obtained from Σ by adding edges.*

(a) [3] *Suppose there exists $M > 0$ such that for all vertices $x, y \in \Sigma$ joined by an edge in Δ and for all geodesics p in Σ between x and y, all vertices of p lie in an M–neighborhood of x, ie $p \subseteq \mathcal{N}_M(x)$ in Δ. Then Δ is also hyperbolic, and there exists a constant k such that for all vertices $x, y \in \Sigma$, every geodesic q between x and y in Σ lies in a k–neighborhood in Δ of every geodesic in Δ between x and y.*

(b) *If, under the assumptions of* (a), *we additionally assume that Σ is a quasi-tree, then Δ is also a quasi-tree.*

Lemma 2.20 *Let p, q be two paths in a metric space S between points x and y, such that p is a geodesic and $q \subseteq \mathcal{N}_k(p)$. Then $p \subseteq \mathcal{N}_{2k}(q)$.*

Proof Let z be any point on p. Let p_1, p_2 denote the segments of the geodesic p with endpoints x, z and z, y, respectively.

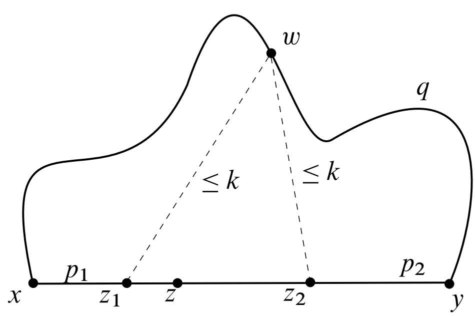


Figure 2: Lemma 2.20

Define a function $f\colon q \to \mathbb{R}$ by $f(s) = d(s, p_1) - d(s, p_2)$. Then f is a continuous function. Further, $f(x) < 0$ and $f(y) > 0$. By the intermediate value theorem, there exists a point w on q such that $f(w) = 0$. Thus $d(w, p_1) = d(w, p_2)$ (see Figure 2). For $i = 1, 2$, let z_i be a point of p_i such that $d(p_i, w) = d(z_i, w)$. Then $d(z_1, w) = d(z_2, w)$. By the hypothesis, $d(w, p) = \min\{d(w, p_1), d(w, p_2)\} \leq k$. So we get that $d(w, p_1) = d(w, p_2) \leq k$. Thus $d(z_1, z_2) \leq 2k$, which implies $d(z, w) \leq 2k$. □

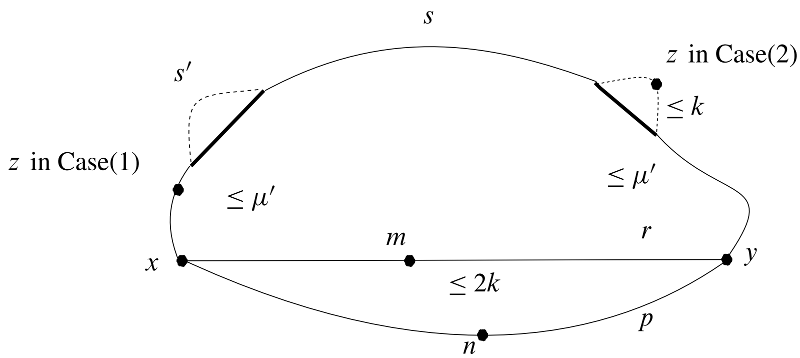


Figure 3: Theorem 2.19

Proof of Theorem 2.19 We proceed with the proof of part (b).

We prove that Δ is a quasi-tree by verifying the bottleneck property from Theorem 2.16. Let d_Σ and d_Δ denote the distances in the graphs Σ and Δ, respectively. Note that the vertex sets of the two graphs are equal.

Let x, y be two vertices. Let m be the midpoint of a geodesic r in Δ connecting them. Let s be any path from x to y in Δ. The path s consists of edges of two types:

(i) edges of the graph Σ;

(ii) edges added in transforming Σ to Δ (marked as bold edges on Figure 3).

Let p be a geodesic in Σ between x and y. By part (a), there exists k such that p is in the k–neighborhood of r in Δ. Applying Lemma 2.20 , we get a point n on p such that

$$d_\Delta(m,n) \leq 2k.$$

Let s' be the path in Σ between x and y, obtained from s by replacing every edge e of type (ii) by a geodesic path $t(e)$ in Σ between its endpoints (marked by dotted lines in Figure 3). Since Σ is a quasi-tree, by Lemma 2.18, there exists $\mu' > 0$ and a point z on s' such that

$$d_\Sigma(z,n) \leq \mu'.$$

Case 1 If z lies on an edge of s of type (i), then

$$d_\Delta(z,m) \leq d_\Delta(z,n) + d_\Delta(n,m) \leq d_\Sigma(z,n) + d_\Delta(n,m) \leq \mu' + 2k.$$

Case 2 If z lies on a path $t(e)$ that replaced an edge e of type (ii), then by part (a),

$$d_\Delta(e_-,m) \leq d_\Delta(e_-,z) + d_\Delta(z,n) + d_\Delta(n,m) \leq k + \mu' + 2k = \mu' + 3k.$$

Thus the bottleneck property holds for $\mu = \mu' + 3k > 0$. □

3 Proof of the main result

Our main result is the following theorem, from which Theorem 1.7 and other corollaries stated in the introduction can be easily derived (see Section 3.5).

Theorem 3.1 *Let $\{H_1, H_2, \dots, H_n\}$ be a finite collection of countable subgroups of a group G such that $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (G, Z)$ for some $Z \subset G$. Let K be a subgroup of G such that $H_i \le K$ for all i. Then there exists a subset $Y \subset K$ such that:*

(a) $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (K, Y)$.

(b) $\Gamma(K, Y \sqcup \mathcal{H})$ *is a quasi-tree, where* $\mathcal{H} = \bigsqcup_{i=1}^n H_i$.

(c) *The action of K on $\Gamma(K, Y \sqcup \mathcal{H})$ is acylindrical.*

(d) $Z \cap K \subset Y$.

3.1 Outline of the proof

Step 1 In order to prove Theorem 3.1, we first prove the following proposition. It is distinct from Theorem 3.1 since it does not require the action of K on the Cayley graph $\Gamma(K, X \sqcup \mathcal{H})$ to be acylindrical.

Proposition 3.2 *Let $\{H_1, H_2, \dots, H_n\}$ be a finite collection of countable subgroups of a group G such that $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h G$ with respect to a relative generating set Z. Let K be a subgroup of G such that $H_i \le K$ for all i. Then there exists $X \subset K$ such that:*

(a) $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (K, X)$.

(b) $\Gamma(K, X \sqcup \mathcal{H})$ *is a quasi-tree, where* $\mathcal{H} = \bigsqcup_{1=1}^n H_i$.

(c) $Z \cap K \subset X$.

Step 2 Once we have proved Proposition 3.2, we will utilize an "acylindrification" construction from [12] to make the action acylindrical, which will prove Theorem 3.1. The details of this step are as follows.

Proof By Proposition 3.2, there exists $X \subseteq K$ such that:

(a) $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (K, X)$.

(b) $\Gamma(K, X \sqcup \mathcal{H})$ is a quasi-tree.

(c) $Z \cap K \subset X$.

By applying Theorem 2.12 to the above, we get that there exists $Y \subset K$ such that:

(a) $X \subseteq Y$.

(b) $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (K, Y)$. In particular, the Cayley graph $\Gamma(K, Y \sqcup \mathcal{H})$ is hyperbolic.

(c) The action of K on $\Gamma(K, Y \sqcup \mathcal{H})$ is acylindrical.

From the proof of Theorem 2.12 (see [12, Lemma 5.6] in particular), it is easy to see that the Cayley graph $\Gamma(G, Y \sqcup \mathcal{H})$ is obtained from $\Gamma(G, X \sqcup \mathcal{H})$ in a manner that satisfies the assumptions of Theorem 2.19, with $M = 1$ (see [12, Lemma 5.6]). Thus by Theorem 2.19, $\Gamma(K, Y \sqcup \mathcal{H})$ is also a quasi-tree. Further,

$$K \cap Z \subset X \subset Y.$$

Thus Y is the required relative generating set. □

We will thus now focus on proving Proposition 3.2. To do so, we will use a construction introduced by Bestvina, Bromberg and Fujiwara in [2]. We describe the construction below and will retain the same terminology as introduced by the authors in [2].

3.2 The projection complex

Definition 3.3 Let $\mathbb{Y}$ be a set and $\xi > 0$ be a constant. Suppose that for each $Y \in \mathbb{Y}$ we have a function

$$d_Y^\pi\colon (\mathbb{Y}\backslash\{Y\} \times \mathbb{Y}\backslash\{Y\}) \to [0, \infty)$$

that satisfies the following axioms:

(A1) $d_Y^\pi(A, B) = d_Y^\pi(B, A)$ for all $Y \in \mathbb{Y}$ and all $A, B \in \mathbb{Y}\backslash\{Y\}$.

(A2) $d_Y^\pi(A, B) + d_Y^\pi(B, C) \geq d_Y^\pi(A, C)$ for all $Y \in \mathbb{Y}$ and all $A, B, C \in \mathbb{Y}\backslash\{Y\}$.

(A3) $\min\{d_Y^\pi(A, B), d_B^\pi(A, Y)\} < \xi$ for all distinct $Y, A, B \in \mathbb{Y}$.

(A4) $\#\{Y \mid d_Y^\pi(A, B) \geq \xi\}$ is finite for all $A, B \in \mathbb{Y}$.

Let J be a positive constant. Then associated to this data we have the *projection complex* $P_J(\mathbb{Y})$, which is a graph constructed in the following manner: the set of vertices of $P_J(\mathbb{Y})$ is the set $\mathbb{Y}$. To specify the set of edges, one first defines a new function $d_Y\colon (\mathbb{Y}\backslash\{Y\} \times \mathbb{Y}\backslash\{Y\}) \to [0, \infty)$, which can be thought of as a small perturbation of d_Y^π. The exact definition of d_Y can be found in [2]. An essential property of the new function is the following inequality, which is an immediate corollary of [2, Proposition 3.2].

For every $Y \in \mathbb{Y}$ and every $A, B \in \mathbb{Y}\backslash\{Y\}$, we have

(1) $$|d_Y^\pi(A, B) - d_Y(A, B)| \leq 2\xi.$$

The set of edges of the graph $P_J(\mathbb{Y})$ can now be described as follows: two vertices $A, B \in \mathbb{Y}$ are connected by an edge if and only if $d_Y(A,B) \le J$ for every $Y \in \mathbb{Y} \setminus \{A, B\}$. This construction strongly depends on the constant J. Complexes corresponding to different J are not isometric in general.

We would like to mention that if $\mathbb{Y}$ is endowed with an action of a group G that preserves projections (ie $d^{\pi}_{g(Y)}(g(A), g(B)) = d^{\pi}_{Y}(A, B)$), then the action of G can be extended to an action on $P_J(\mathbb{Y})$. We also mention the following proposition, which has been proved under the assumptions of Definition 3.3.

Proposition 3.4 [2, Theorem 3.16] *For a sufficiently large $J > 0$, $P_J(\mathbb{Y})$ is connected and quasi-isometric to a tree.*

Definition 3.5 (nearest point projection) In a metric space (S, d), given a set Y and a point $a \in S$, we define the nearest point projection as

$$\mathrm{proj}_Y(a) = \{y \in Y \mid d(Y, a) = d(y, a)\}.$$

If A, Y are two sets in S, then

$$\mathrm{proj}_Y(A) = \bigcup_{a \in A} \mathrm{proj}_Y(a).$$

We note that in our case, since elements of $\mathbb{Y}$ will come from a Cayley graph, which is a combinatorial graph, the nearest point projection will exist. This is because distances on a combinatorial graph take discrete values in $\mathbb{N} \cup \{0\}$. Since this set is bounded below, we cannot have an infinite strictly decreasing sequence of distances.

We make all geometric considerations in the Cayley graph $\Gamma(G, Z \sqcup \mathcal{H})$. Let $\mathrm{d}_{Z \sqcup \mathcal{H}}$ denote the metric on this graph. Since $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h G$ under the assumptions of Proposition 3.2, from Remark 4.26 of [5] it follows that $H_i \hookrightarrow_h G$ for $i = 1, 2, \dots, n$. By Theorem 2.8, we can define a finite-valued, proper metric $\widetilde{d}_i$ on H_i, for $i = 1, 2, \dots, n$, satisfying

$$\widetilde{d}_i(x, y) \le \widehat{d}_i(x, y) \tag{2}$$

for all $x, y \in H_i$.

We can extend both $\widehat{d}_i$ and $\widetilde{d}_i$ to all cosets gH_i of H_i by setting $\widetilde{d}_i(gx, gy) = \widetilde{d}_i(x, y)$ and $\widehat{d}_i(gx, gy) = \widehat{d}_i(x, y)$ for all $x, y \in H_i$. Let $\widehat{\mathrm{diam}}$ (resp. $\widetilde{\mathrm{diam}}$) denote the diameter of a subset of H_i or a coset of H_i with respect to the $\widehat{d}_i$ (resp. $\widetilde{d}_i$) metric.

Let

$$\mathbb{Y} = \{kH_i \mid k \in K, i = 1, 2, \dots, n\}$$

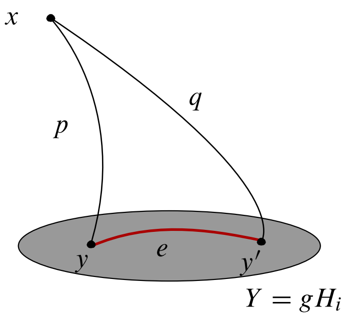


Figure 4: The bold red edge e denotes a single edge labeled by an element of $\mathcal{H}$

be the set of cosets of all H_i in K. We think of cosets of H_i as a subset of vertices of $\Gamma(G, Z \sqcup \mathcal{H})$.

For each $Y \in \mathbb{Y}$ and $A, B \in \mathbb{Y} \backslash \{Y\}$, define

$$d_Y^{\pi}(A, B) = \widetilde{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B)), \tag{3}$$

where $\operatorname{proj}_Y(A)$ is as in Definition 3.5. The fact that (3) is well-defined will follow from Lemma 3.6 and Lemma 3.8, which are proved below. We will also proceed to verify the axioms (A1)–(A4) of the Bestvina–Bromberg–Fujiwara construction in the above setting.

Lemma 3.6 *For any $Y \in \mathbb{Y}$ and any $x \in G$, $\widetilde{\operatorname{diam}}(\operatorname{proj}_Y(x)) \le 3C$, where C is the constant as in Lemma 2.4. As a consequence, $\widetilde{\operatorname{diam}}(\operatorname{proj}_Y(x))$ is bounded.*

Proof By (2), it suffices to prove that $\widehat{\operatorname{diam}}(\operatorname{proj}_Y(x))$ is bounded. Let $y, y' \in \operatorname{proj}_Y(x)$. Then $\mathrm{d}_{Z\sqcup\mathcal{H}}(x, y) = \mathrm{d}_{Z\sqcup\mathcal{H}}(x, y') = \mathrm{d}_{Z\sqcup\mathcal{H}}(x, Y)$. Without loss of generality, $x \notin Y$, or else the diameter is zero.

Let $Y = gH_i$. Let e denote the edge connecting y and y', which is labeled by an element of H_i. Let p and q denote geodesics between x and y and between x and y', respectively (see Figure 4).

Consider the geodesic triangle T with sides e, p, q. Since p and q are geodesics between the point x and Y, e is an isolated component in T, ie e cannot be connected to either p or q. Indeed, if e is connected to, say, a component of p, then since e_+ and e_- are in Y, it would imply that the geodesic p passes through a point of Y before y. But then y is not the nearest point from Y to x, which is a contradiction. By Lemma 2.4, $\widehat{d}_i(y, y') \le 3C$. Hence

$$\widehat{\operatorname{diam}}(\operatorname{proj}_Y(x)) \le 3C. \qquad \square$$

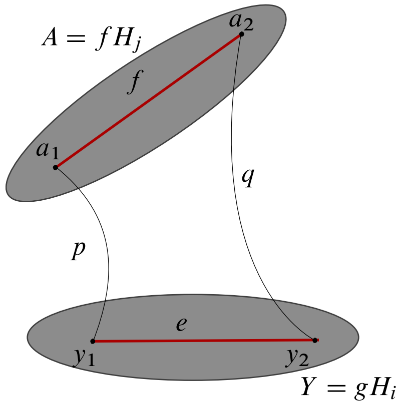


Figure 5: Lemma 3.8

Remark 3.7 Observe that in the previous lemma, we proved the following fact: If x is a point in G and $y \in \mathrm{proj}_Y(x)$, then every geodesic path p between x and y satisfies the property that no vertex of p, except for y, can belong to the coset Y. We will use this fact repeatedly in the following lemmas.

Lemma 3.8 *For every pair of distinct elements $A, Y \in \mathbb{Y}$, $\widehat{\mathrm{diam}}(\mathrm{proj}_Y(A)) \le 4C$, where C is the constant as in Lemma 2.4. As a consequence, $\widetilde{\mathrm{diam}}(\mathrm{proj}_Y(A))$ is bounded.*

Proof Let $Y = gH_i$ and $A = fH_j$. Let $y_1, y_2 \in \mathrm{proj}_Y(A)$. Then there exist $a_1, a_2 \in A$ such that $\mathrm{d}_{Z \sqcup \mathcal{H}}(a_1, y_1) = \mathrm{d}_{Z \sqcup \mathcal{H}}(a_1, Y)$ and $\mathrm{d}_{Z \sqcup \mathcal{H}}(a_2, y_2) = \mathrm{d}_{Z \sqcup \mathcal{H}}(a_2, Y)$. Now y_1 and y_2 are connected by a single edge e, labeled by an element of H_i, and similarly, a_1 and a_2 are connected by an edge f, labeled by an element of H_j (see Figure 5). Let p and q be geodesics that connect y_1 to a_1 and y_2 to a_2, respectively. We note that p and/or q may be trivial paths (consisting of a single point), but this does not alter the proof.

Consider e in the quadrilateral Q with sides p, f, q, e. By Remark 3.7, e cannot be connected to a component of p or q.

If $i = j$, then e cannot be connected to f since $A \ne Y$. If $i \ne j$, then obviously e and f cannot be connected. Thus e is isolated in this quadrilateral Q. By Lemma 2.4, $\widehat{d}_i(y_1, y_2) \le 4C$. Thus

$$\widehat{\mathrm{diam}}(\mathrm{proj}_Y(A)) \le 4C. \qquad \square$$

Corollary 3.9 *The function d_Y^{π} defined by (3) is well-defined.*

Proof Since the $\widetilde{d}_i$ metric takes finite values for $i = 1, 2, \dots, n$, using Lemma 3.8, we have that d_Y^π also takes only finite values. □

Lemma 3.10 *The function d_Y^π defined by (3) satisfies conditions* (A1) *and* (A2) *in Definition 3.3.*

Proof (A1) is obviously satisfied. For any $Y \in \mathbb{Y}$ and any $A, B, C \in \mathbb{Y} \backslash \{Y\}$, by the triangle inequality, we have that

$$\begin{aligned} d_Y^\pi(A, C) &= \widetilde{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(C)) \\ &\le \widetilde{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B)) + \widetilde{\operatorname{diam}}(\operatorname{proj}_Y(B) \cup \operatorname{proj}_Y(C)) \\ &= d_Y^\pi(A, B) + d_Y^\pi(B, C). \end{aligned}$$

Thus (A2) also holds. □

Lemma 3.11 *The function d_Y^π from (3) satisfies condition* (A3) *in Definition 3.3 for any $\xi > 14C$, where C is the constant from Lemma 2.4*

Proof By (2), it suffices to prove that

$$\min\bigl\{\widehat{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B)), \widehat{\operatorname{diam}}(\operatorname{proj}_B(A) \cup \operatorname{proj}_B(Y))\bigr\} < \xi.$$

Let $A, B \in \mathbb{Y} \backslash \{Y\}$ be distinct. Let $Y = gH_i$, $A = fH_j$ and $B = tH_k$. If

$$\widehat{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B)) \le 14C,$$

then we are done. So suppose that

$$\widehat{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B))) > 14C. \tag{4}$$

Choose $a \in A$, $b \in B$ and $x, y \in Y$ such that $\mathrm{d}_{Z \sqcup \mathcal{H}}(A, Y) = \mathrm{d}_{Z \sqcup \mathcal{H}}(a, x)$ and $\mathrm{d}_{Z \sqcup \mathcal{H}}(B, Y) = \mathrm{d}_{Z \sqcup \mathcal{H}}(b, y)$. In particular,

$$x \in \operatorname{proj}_Y(A), \qquad y \in \operatorname{proj}_Y(B) \tag{5}$$

and $b \in \operatorname{proj}_B(Y)$. Let p, q denote geodesics connecting a to x and b to y, respectively. Let h_1 denote the edge connecting x and y, which is labeled by an element of H_i.

By (5), we have that

$$\widehat{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B)) \le \widehat{\operatorname{diam}}(\operatorname{proj}_Y(A)) + \widehat{\operatorname{diam}}(\operatorname{proj}_Y(B)) + \widehat{d}_i(x, y).$$

Combining this with (4) and Lemma 3.8, we get

$$\begin{aligned} \widehat{d}_i(x, y) &\ge \widehat{\operatorname{diam}}(\operatorname{proj}_Y(A) \cup \operatorname{proj}_Y(B)) - \widehat{\operatorname{diam}}(\operatorname{proj}_Y(A)) - \widehat{\operatorname{diam}}(\operatorname{proj}_Y(B)) \\ &> 14C - 8C = 6C. \end{aligned}$$

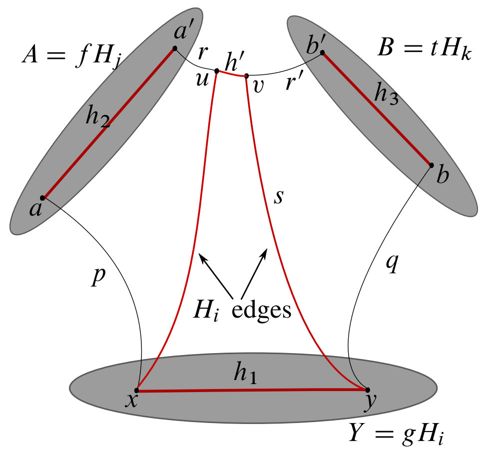


Figure 6: Condition (A3)

Choose any $a' \in A$ and $b' \in \mathrm{proj}_B(a')$. Then $\mathrm{d}_{Z \sqcup \mathcal{H}}(a', B) = \mathrm{d}_{Z \sqcup \mathcal{H}}(a', b')$ (see Figure 6). (Note that if $a' = a$, the following arguments still hold.) Let h_2 and h_3 denote the edges connecting a to a' and b to b', which are labeled by elements of H_j and H_k, respectively. Let r denote a geodesic connecting a' and b'. Consider the geodesic hexagon W with sides p, h_1, q, h_3, r, h_2. Then h_1 is not isolated in W; otherwise, by Lemma 2.4, $\widehat{d}_i(x,y) \le 6C$, a contradiction.

Thus h_1 is connected to another H_i–component in W. By Remark 3.7, h_1 cannot be connected to a component of p or q. Since A, B, Y are all distinct, h_1 cannot be connected to h_2 or h_3. So h_1 must be connected to an H_i–component on the geodesic r. Let this edge be h', with endpoints u and v, as shown in Figure 6. Let s denote the edge (labeled by an element of H_i) that connects y and v. Let r' denote the segment of r that connects v to b'. Then r' is also a geodesic.

Consider the quadrilateral Q with sides r', h_3, q, s. By using arguments similar to those in the previous paragraph, h_3 cannot be connected to r', q or s. Thus h_3 is isolated in Q. By Lemma 2.4,

$$\widehat{d}_k(b, b') \le 4C.$$

Since the above argument holds for any $a' \in A$ and for $b' \in \mathrm{proj}_B(A)$, we have that $\widehat{d}_k(b, b') \le 4C$. Using Lemma 3.8 (see Figure 7), we get that

$$\widehat{\mathrm{diam}}(\mathrm{proj}_B(Y) \cup \mathrm{proj}_B(A)) \le 4C + 4C = 8C < \xi. \qquad \square$$

Lemma 3.12 *The function d^{π}_{Y} defined by (3) satisfies condition* (A4) *in Definition 3.3, for $\xi > 14C$, where C is the constant from Lemma 2.4*

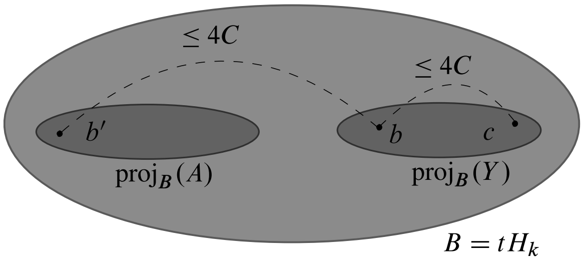


Figure 7: Estimating the distance between arbitrary points b' and c of $\mathrm{proj}_B(A)$ and $\mathrm{proj}_B(Y)$

Proof If $d_Y^{\pi}(A,B) \ge \xi$, then by (2),

$$\widehat{\mathrm{diam}}(\mathrm{proj}_Y(A) \cup \mathrm{proj}_Y(B)) \ge d_Y^{\pi}(A,B) \ge \xi.$$

Thus it suffices to prove that the number of elements $Y \in \mathbb{Y}$ satisfying

$$\widehat{\mathrm{diam}}(\mathrm{proj}_Y(A) \cup \mathrm{proj}_Y(B)) \ge \xi \tag{6}$$

is finite. Let $A, B \in \mathbb{Y}$, $A = fH_j$ and $B = tH_k$. Let $Y \in \mathbb{Y} \backslash \{A, B\}$, $Y = gH_i$. Let $a' \in A$, $b' \in \mathrm{proj}_B(a')$. By repeating the computations in Lemma 3.11, we can show that if Y is such that $\widehat{\mathrm{diam}}(\mathrm{proj}_Y(A) \cup \mathrm{proj}_Y(B)) \ge \xi$, then for any $a \in A$, $b \in B$, $x \in \mathrm{proj}_Y(a)$, $y \in \mathrm{proj}_Y(b)$, we have that $\widehat{d}_i(x,y) > 6C$.

Let h_1 denote the edge connecting x, y, which is labeled by an element of H_i (see Figure 8). Let h_2 denote the edge connecting a, a', which is labeled by an element of H_j, and let h_3 denote the edge connecting b, b', which is labeled by an element of H_k. Let p be a geodesic between a, x, let q be a geodesic between b, y and let r be a geodesic between a', b'. As argued in Lemma 3.11, we can show that h_1 cannot be isolated in the hexagon W with sides p, h_1, q, h_2, r, h_3 and must be connected to an H_i–component of r, say the edge h'.

We claim that the edge h' uniquely identifies Y. Indeed, let Y' be a member of $\mathbb{Y}$, with elements x', y' connected by an edge e (labeled by an element of the corresponding subgroup). Suppose that e is connected to h'. Then we must have that Y' is also a coset of H_i. But cosets of a subgroup are either disjoint or equal, so $Y = Y'$. Thus, the number of $Y \in \mathbb{Y}$ satisfying (6) is bounded by the number of distinct H_i–components of r, which is finite. □

3.3 Choosing a relative generating set

We now have the necessary details to choose a relative generating set X which will satisfy conditions (a) and (b) of Proposition 3.2. This set will later be altered slightly

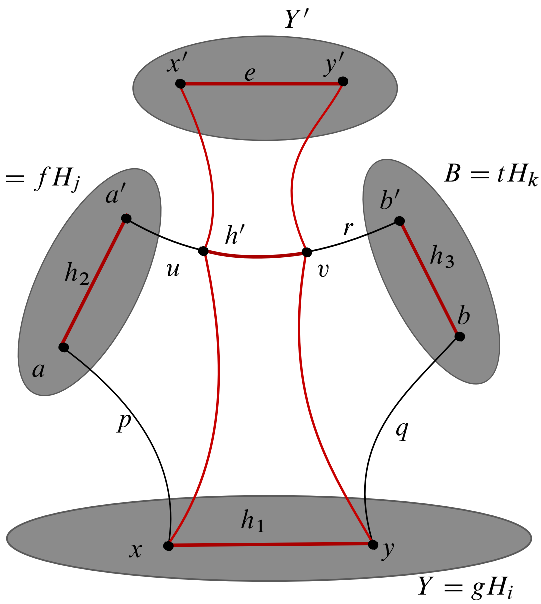


Figure 8: Condition (A4)

to obtain another relative generating set which will satisfy all three conditions of Proposition 3.2. We will repeat arguments similar to those made by Dahmani, Guirardel and Osin in [5, pages 60–63].

Recall that $\mathcal{H} = \bigsqcup_{i=1}^{n} H_i$ and that Z is the relative generating set for this collection of subgroups such that $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (G, Z)$. Let $P_J(\mathbb{Y})$ be the projection complex corresponding to the vertex set $\mathbb{Y}$ as specified in Section 3.2, where the constant J is as in Proposition 3.4, ie $P_J(\mathbb{Y})$ is connected and a quasi-tree. Let d_P denote the combinatorial metric on $P_J(\mathbb{Y})$. Our definition of projections is K–equivariant and hence the action of K on $\mathbb{Y}$ extends to a cobounded action of K on $P_J(\mathbb{Y})$.

In what follows, by considering H_i to be vertices of the projection complex $P_J(\mathbb{Y})$, we denote by $\operatorname{star}(H_i)$ the set

$$\{e \text{ is an edge in } P_J(\mathbb{Y}) \mid e \text{ connects } H_i \text{ to } kH_j \text{ for some } k \in K \text{ and } 1 \le j \le n\}.$$

We choose the set X in the following manner. For all $i = 1, 2, \dots, n$ and each edge e in $\operatorname{star}(H_i)$ in $P_J(\mathbb{Y})$ that connects H_i to kH_j, choose all elements $x_e \in H_i k H_j$ such that

$$\mathrm{d}_{Z \sqcup \mathcal{H}}(1, x_e) = \mathrm{d}_{Z \sqcup \mathcal{H}}(1, H_i k H_j).$$

We say that all such x_e have *type* (i, j). Since $H_i \le K$ for all i, $x_e \in K$. We observe the following:

(a) For each x_e of type (i,j) as above, there is an edge in $P_J(\mathbb{Y})$ connecting H_i and $x_e H_j$. Indeed if $x_e = h_1 k h_2$ for $h_1 \in H_i$ and $h_2 \in H_j$, then

$$\begin{aligned} d_P(H_i, x_e H_j) &= d_P(H_i, h_1 k h_2 H_j) = d_P(H_i, h_1 k H_j) \\ &= d_P(h_1^{-1} H_i, k H_j) = d_P(H_i, k H_j) \\ &= 1. \end{aligned}$$

(b) For each edge e connecting H_i and kH_j, there is a dual edge f connecting H_j and $k^{-1}H_i$. Thus for every element x_e of type (i,j), there is an element $x_f = (x_e)^{-1}$ of type (j,i). In particular, the set given by

$$X = \{x_e \neq 1 \mid e \in \operatorname{star}(H_i),\ i = 1, 2, \dots, n\} \tag{7}$$

is symmetric, ie closed under taking inverses. Obviously, $X \subset K$.

(c) If $x_e \in X$ is of type (i,j), then x_e is not an element of H_i or H_j. Indeed, if $x_e = h_1 k h_2$ for some $h_1 \in H_i$ and some $h_2 \in H_j$ and x_e is an element of H_i or H_j, then $k = hf$ for some $h \in H_i$ and some $f \in H_j$. Consequently

$$\mathrm{d}_{Z \sqcup \mathcal{H}}(1, H_i k H_j) = \mathrm{d}_{Z \sqcup \mathcal{H}}(1, H_i H_j) = 0 = \mathrm{d}_{Z \sqcup \mathcal{H}}(1, x_e),$$

which implies $x_e = 1$, which is a contradiction to (7).

Lemma 3.13 (cf [5, Lemma 4.49]) *The subgroup K is generated by X together with the union of all the H_i. Further, the Cayley graph $\Gamma(K, X \sqcup \mathcal{H})$ is quasi-isometric to $P_J(\mathbb{Y})$, and hence a quasi-tree.*

Proof Let $\Sigma = \{H_1, H_2, \dots, H_n\} \subseteq \mathbb{Y}$. Let $\operatorname{diam}(\Sigma)$ denote the diameter of the set Σ in the combinatorial metric d_P. Since Σ is a finite set, $\operatorname{diam}(\Sigma)$ is finite. Define

$$\phi\colon K \to \mathbb{Y}, \quad \phi(k) = kH_1.$$

By property (a) above, if $x_e \in X$ is of type (i,j),

$$\begin{aligned} d_P(x_e H_1, H_1) &\le d_P(x_e H_1, x_e H_j) + d_P(x_e H_j, H_i) + d_P(H_i, H_1) \\ &= d_P(H_1, H_j) + 1 + d_P(H_i, H_1) \\ &\le 2\operatorname{diam}(\Sigma) + 1. \end{aligned}$$

Further, for $h \in H_i$,

$$\begin{aligned} d_P(hH_1, H_1) &\le d_P(hH_1, hH_i) + d_P(hH_i, H_1) \\ &= d_P(H_1, H_i) + d_P(H_i, H_1) \\ &\le 2\operatorname{diam}(\Sigma). \end{aligned}$$

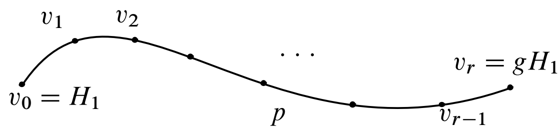


Figure 9: The geodesic p

Thus, for all $g \in \langle X \cup H_1 \cup H_2 \cup \cdots \cup H_n \rangle$, we have

$$d_P(\phi(1), \phi(g)) \le (2\,\mathrm{diam}(\Sigma) + 1)|g|_{X \sqcup \mathcal{H}}, \tag{8}$$

where $|g|_{X \sqcup \mathcal{H}}$ denotes the length of g in the generating set $X \cup H_1 \cup H_2 \cup \cdots \cup H_n$. (We use this notation for the sake of uniformity).

Now let $g \in K$ and suppose $d_P(\phi(1), \phi(g)) = r$, ie $d_P(H_1, gH_1) = r$. If $r = 0$, then $H_1 = gH_1$, thus $g \in H_1$ and $|g|_{X \sqcup \mathcal{H}} \le 1$. If $r > 0$, consider a geodesic p in $P_J(\mathbb{Y})$ connecting H_1 and gH_1. Let

$$\begin{aligned} v_0 &= H_1 = g_0 H_1 \quad (g_0 = 1), \\ v_1 &= g_1 H_{\lambda_1}, \\ v_2 &= g_2 H_{\lambda_2}, \\ &\ \ \vdots \\ v_{r-1} &= g_{r-1} H_{\lambda_{r-1}}, \\ v_r &= gH_1 \qquad\qquad (g_r = g) \end{aligned}$$

be the sequence of vertices of p, for some $\lambda_j \in \{1, 2, \dots, n\}$ and some $g_i \in K$ (see Figure 9).

Now $g_i H_{\lambda_i}$ is connected by a single edge to $g_{i+1} H_{\lambda_{i+1}}$. Thus

$$d_P(g_i H_{\lambda_i}, g_{i+1} H_{\lambda_{i+1}}) = 1,$$

which implies

$$d_P(H_{\lambda_i}, g_i^{-1} g_{i+1} H_{\lambda_{i+1}}) = 1.$$

Then there exists $x \in X$ such that

$$x \in H_{\lambda_i} g_i^{-1} g_{i+1} H_{\lambda_{i+1}} \quad \text{and} \quad \mathrm{d}_{Z \sqcup \mathcal{H}}(1, x) = \mathrm{d}_{Z \sqcup \mathcal{H}}(1, H_{\lambda_i} g_i^{-1} g_{i+1} H_{\lambda_{i+1}}).$$

Thus $x = h g_i^{-1} g_{i+1} k$ for some $h \in H_{\lambda_i}$ and some $k \in H_{\lambda_{i+1}}$ which implies $g_i^{-1} g_{i+1} = h^{-1} x k^{-1}$. So $|g_i^{-1} g_{i+1}|_{X \sqcup \mathcal{H}} \le 3$, which implies

$$|g|_{X \sqcup \mathcal{H}} = \left| \prod_{i=1}^{r} g_{i-1}^{-1} g_i \right|_{X \sqcup \mathcal{H}} \le \sum_{i=1}^{r} |g_{i-1}^{-1} g_i|_{X \sqcup \mathcal{H}} \le 3r = 3 d_P(\phi(1), \phi(g)). \tag{9}$$

The above argument also provides a representation for every element $g \in K$ as a product of elements from $X \cup H_1 \cup H_2 \cup \dots \cup H_n$. Thus K is generated by the union of X and all the H_i. By (8) and (9), ϕ is a quasi-isometric embedding of $(K, |\cdot|_{X \sqcup \mathcal{H}})$ into $(P_J(\mathbb{Y}), d_P)$ satisfying

$$\tfrac{1}{3}|g|_{X \sqcup \mathcal{H}} \leq d_P(\phi(1), \phi(g)) \leq (2\,\mathrm{diam}(\Sigma) + 1)|g|_{X \sqcup \mathcal{H}}.$$

Since $\mathbb{Y}$ is contained in the closed diam(Σ)–neighborhood of $\phi(K)$, ϕ is a quasi-isometry. This implies that $\Gamma(K, X \sqcup \mathcal{H})$ is a quasi-tree. □

Let $\widetilde{d}_i$ denote the modified relative metric on H_i associated with the Cayley graph $\Gamma(G, Z \sqcup \mathcal{H})$ from Theorem 2.8. Let $\widehat{d_i^X}$ denote the relative metric on H_i associated with the Cayley graph $\Gamma(K, X \sqcup \mathcal{H})$. We will now show that $\widehat{d_i^X}$ is proper for all $i = 1, 2, \dots, n$. We will use the fact that $\widetilde{d}_i$ is proper and derive a relation between $\widetilde{d}_i$ and $\widehat{d_i^X}$.

Lemma 3.14 (cf [5, Lemma 4.50]) *There exists a constant α such that for any $Y \in \mathbb{Y}$ and any $x \in X \sqcup \mathcal{H}$, if*

$$\widetilde{\mathrm{diam}}(\mathrm{proj}_Y\{1, x\}) > \alpha,$$

then $x \in H_j$ and $Y = H_j$ for some j.

Proof We prove the result for

$$\alpha = \max\{J + 2\xi, 6C\}.$$

Suppose that $\widetilde{\mathrm{diam}}(\mathrm{proj}_Y\{1, x\}) > \alpha$ and $x \in X$ has type (k, l), ie there exists an edge connecting H_k and gH_l in $P_J(\mathbb{Y})$, where $g \in K$. We consider three possible cases and arrive at a contradiction in each case.

Case 1 ($H_k \neq Y \neq xH_l$) Then

$$\widetilde{\mathrm{diam}}(\mathrm{proj}_Y\{1, x\}) \leq d_Y^{\pi}(H_k, xH_l) \leq d_Y(H_k, xH_l) + 2\xi \leq J + 2\xi \leq \alpha,$$

using (1) and the fact that H_k and xH_l are connected by an edge in $P_J(\mathbb{Y})$, which is a contradiction.

Case 2 ($H_k = Y$) Since $x \notin H_k$, let $y \in \mathrm{proj}_Y(x)$, ie

$$\mathrm{d}_{Z \sqcup \mathcal{H}}(x, y) = \mathrm{d}_{Z \sqcup \mathcal{H}}(x, H_k) = \mathrm{d}_{Z \sqcup \mathcal{H}}(x, Y).$$

By Lemma 3.6, if $\widehat{d}_k(1,y) \le 3C$, then

$$\begin{aligned}\widehat{\operatorname{diam}}(\operatorname{proj}_Y\{1,x\}) &\le \widehat{\operatorname{diam}}(\operatorname{proj}_Y(1)) + \widehat{\operatorname{diam}}(\operatorname{proj}_Y(x)) + \widehat{d}_k(\operatorname{proj}_Y(1),\operatorname{proj}_Y(x)) \\ &\le 0 + 3C + \widehat{d}_k(1,y) \\ &\le 6C \le \alpha.\end{aligned}$$

Then by (2), we have

$$\widetilde{\operatorname{diam}}(\operatorname{proj}_Y\{1,x\}) \le \alpha,$$

which is a contradiction. Thus $\widehat{d}_k(1,y) > 3C$. This implies that $1 \notin \operatorname{proj}_Y(x)$ (see Figure 10). By definition of the nearest point projection, $\mathrm{d}_{Z\sqcup\mathcal{H}}(1,x) > \mathrm{d}_{Z\sqcup\mathcal{H}}(y,x)$, which implies $\mathrm{d}_{Z\sqcup\mathcal{H}}(1,x) > \mathrm{d}_{Z\sqcup\mathcal{H}}(1,y^{-1}x)$. Since $y^{-1}x \in H_k g H_l$, we obtain $\mathrm{d}_{Z\sqcup\mathcal{H}}(1,x) > \mathrm{d}_{Z\sqcup\mathcal{H}}(1,H_k g H_l)$, which is a contradiction to the choice of x.

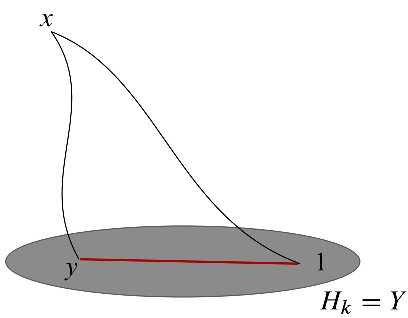


Figure 10: Case 2

Case 3 $(Y = xH_l,\ H_k \ne Y)$ This case reduces to Case 2, since we can translate everything by x^{-1}.

Thus we must have $x \in H_j$ for some j. Suppose that $H_j \ne Y$. But then

$$\widetilde{\operatorname{diam}}(\operatorname{proj}_Y\{1,x\}) \le \widetilde{\operatorname{diam}}(\operatorname{proj}_Y(H_j)) \le 4C \le \alpha,$$

by Lemma 3.8, which is a contradiction. □

Lemma 3.15 (cf [5, Lemma 4.45]) *If $H_i = fH_j$, then $H_i = H_j$ and $f \in H_i$. Consequently, if $gH_i = fH_j$, then $H_i = H_j$ and $g^{-1}f \in H_i$.*

Proof If $H_i = fH_j$, then $1 = fk$ for some $k \in H_j$. Then $f = k^{-1} \in H_j$, which implies $H_i = H_j$. □

Lemma 3.16 (cf [5, Theorem 4.42]) *For $i = 1, 2, \dots, n$ and any $h \in H_i$, we have*

$$\alpha \widehat{d}_i^X(1,h) \ge \widetilde{d}_i(1,h),$$

where α is the constant from Lemma 3.14. Thus $\widehat{d}_i^X$ is proper.

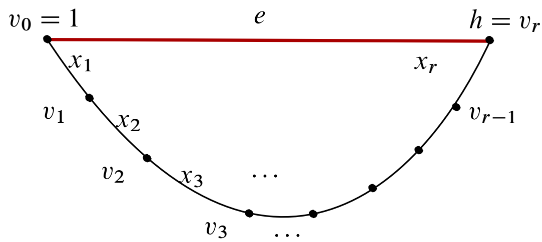


Figure 11: The cycle ep

Proof Let $h \in H_i$ be such that $\widehat{d}_i^X(1,h) = r$. Let e denote the H_i–edge in the Cayley graph $\Gamma(K, X \sqcup \mathcal{H})$ connecting h to 1, labeled by h^{-1}. Let p be an admissible (see Definition 2.1) path of length r in $\Gamma(K, X \sqcup \mathcal{H})$ connecting 1 and h. Then ep forms a cycle (see Figure 11). Since p is admissible, e is isolated in this cycle.

Let $\mathrm{Lab}(p) = x_1 x_2 \cdots x_r$ for some $x_1, x_2, \dots, x_r \in X \sqcup \mathcal{H}$. Let

$$v_0 = 1, \quad v_1 = x_1, \quad v_2 = x_1 x_2, \quad \dots, \quad v_r = x_1 x_2 \cdots x_r = h.$$

Since these are also elements of G, for $k = 1, 2, \dots, r$ we have

$$\begin{aligned}\widetilde{\mathrm{diam}}(\mathrm{proj}_{H_i}\{v_{k-1}, v_k\}) &= \widetilde{\mathrm{diam}}(\mathrm{proj}_{H_i}\{x_1 x_2 \cdots x_{k-1}, x_1 x_2 \cdots x_{k-1} x_k\}) \\ &= \widetilde{\mathrm{diam}}(\mathrm{proj}_Y\{1, x_k\}),\end{aligned}$$

where $Y = (x_1 x_2 \cdots x_{k-1})^{-1} H_i$.

If $\widetilde{\mathrm{diam}}(\mathrm{proj}_Y\{1, x_k\}) > \alpha$ for some k, then by Lemma 3.14, $x_k \in H_j$ and $Y = H_j$ for some j. By Lemma 3.15, $H_i = H_j$ and $x_1 x_2 \cdots x_{k-1} \in H_j$. But then e is not isolated in the cycle ep, which is a contradiction.

Hence

$$\widetilde{\mathrm{diam}}(\mathrm{proj}_{H_i}\{v_{k-1}, v_k\}) \le \alpha$$

for all $k = 1, 2, \dots, r$, which implies

$$\begin{aligned}\widetilde{d}_i(1,h) &\le \widetilde{\mathrm{diam}}(\mathrm{proj}_{H_i}\{v_0, v_r\}) \\ &\le \sum_{j=1}^{r} \widetilde{\mathrm{diam}}(\mathrm{proj}_{H_i}\{v_{j-1}, v_j\}) \\ &\le r\alpha = \alpha \widehat{d}_i^X(1,h).\end{aligned}$$ □

3.4 Proof of Proposition 3.2

The goal of this section is to alter our relative generating set X from Section 3.3, so that we obtain another relative generating set that satisfies all the conditions of

Proposition 3.2. To do so, we need to establish a relation between the set X and the set Z. We will need the following obvious lemma.

Lemma 3.17 *Let X and Y be generating sets of G, and suppose that*

$$\sup_{x\in X} |x|_Y < \infty \quad \text{and} \quad \sup_{y\in Y} |y|_X < \infty.$$

Then $\Gamma(G,X)$ is quasi-isometric to $\Gamma(G,Y)$. In particular, $\Gamma(G,X)$ is a quasi-tree if and only if $\Gamma(G,Y)$ is a quasi-tree.

Remark 3.18 This lemma implies that if we change a generating set by adding finitely many elements, then the property that the Cayley graph is a quasi-tree still holds.

We also need to note that from (1) in Definition 3.3, it easily follows that for each $Y\in\mathbb{Y}$ and every $A,B\in\mathbb{Y}\backslash\{Y\}$, we have

$$d_Y(A,B)\le d_Y^{\pi}(A,B)+2\xi. \tag{10}$$

Lemma 3.19 *For a large enough J, the set X constructed in Section 3.3 satisfies the following property: if $z\in Z\cap K$ does not represent an element of H_i for any $i=1,2,\dots,n$, then $z\in X$.*

Proof Recall that $\mathrm{d}_{Z\sqcup\mathcal{H}}$ denotes the combinatorial metric on $\Gamma(G,Z\sqcup\mathcal{H})$. Let $z\in Z\cap K$ be as in the statement of the lemma. Then $z\in H_izH_i$ for all i and $1\notin H_izH_i$. Thus

$$\mathrm{d}_{Z\sqcup\mathcal{H}}(1,H_izH_i)\ge 1=\mathrm{d}_{Z\sqcup\mathcal{H}}(1,z)\ge \mathrm{d}_{Z\sqcup\mathcal{H}}(1,H_izH_i),$$

which implies

$$\mathrm{d}_{Z\sqcup\mathcal{H}}(1,H_izH_i)=\mathrm{d}_{Z\sqcup\mathcal{H}}(1,z) \quad \text{for all } i.$$

In order to prove $z\in X$, we must show that H_i and zH_i are connected by an edge in $P_J(\mathbb{Y})$. By Definition 3.3, this is true if

$$d_Y(H_i,zH_i)\le J \quad \text{for all } Y\ne H_i,zH_i.$$

In view of (10), we will estimate $d_Y^{\pi}(H_i,zH_i)$.

Let $\mathrm{d}_{Z\sqcup\mathcal{H}}(h,x)=\mathrm{d}_{Z\sqcup\mathcal{H}}(H_i,Y)$ and $\mathrm{d}_{Z\sqcup\mathcal{H}}(f,y)=\mathrm{d}_{Z\sqcup\mathcal{H}}(zH_i,Y)$ for some $h\in H_i$, $f\in zH_i$ and for some $x,y\in Y=gH_j$. Let p be a geodesic connecting h and x, and let q be a geodesic connecting y and f. Let h_2 denote the edge connecting x and y, labeled by an element of H_j. Similarly, let s,t denote the edges connecting h to 1 and z to f respectively, which are labeled by elements of H_i. Let e denote the

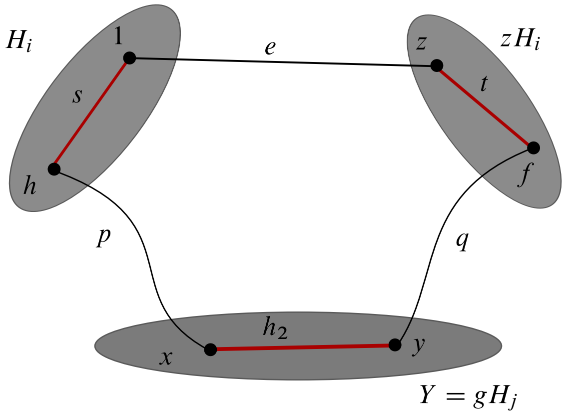


Figure 12: Dealing with elements of $Z \cap K$ that represent elements of $\mathcal{H}$

edge connecting 1 and z, labeled by z. Consider the geodesic hexagon W with sides p, h_2, q, t, e, s (see Figure 12).

By using Remark 3.7 and the fact that $Y \neq H_i, zH_i$, we can show that h_2 cannot be connected to q, p, s or t. Since z does not represent an element of H_i for any i, h_2 cannot be connected to e. Thus, h_2 is isolated in W. By Lemma 2.4, $\widehat{d}_j(x, y) \leq 6C$. By Lemma 3.8,

$$d_Y(H_i, zH_i) \leq d_Y^{\pi}(H_i, zH_i) + 2\xi \leq 14C + 2\xi.$$

We conclude that by taking the constant J to be sufficiently large so that Proposition 3.4 holds and J exceeds $14C + 2\xi$, we can ensure that $z \in X$ and the arguments of the previous section still hold. □

Lemma 3.20 *There are only finitely many elements of $Z \cap K$ that can represent an element of H_i for some $i \in \{1, 2, \ldots, n\}$.*

Proof Let $z \in Z \cap K$ represent an element of H_i for some $i = 1, 2, \ldots, n$. Then in the Cayley graph $\Gamma(G, Z \sqcup \mathcal{H})$, we have a bigon between the elements 1 and h, where one edge is labeled by z and the other edge is labeled by an element of H_i, say h_1 (see Remark 2.2 and Figure 13).

This implies that $\widehat{d}_i(1, z) \leq 1$, so $\widetilde{d}_i(1, z) \leq 1$. But then $z \in \widetilde{B}_i(1, 1)$, ie the ball of radius 1 in the subgroup H_i in the relative metric, centered at the identity. But this is a finite ball. Take

$$\rho = \left| \bigcup_{i=1}^{n} \widetilde{B}_i(1, 1) \right|.$$

Then z has at most ρ choices, which is finite. □

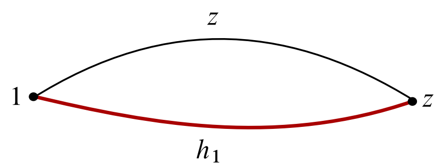


Figure 13: Bigons in the Cayley graph

By Lemma 3.20 and by selecting the constant J as specified in Lemma 3.19, we conclude that the set X from Section 3.3 does not contain at most finitely many elements of $Z \cap K$. By adding these finitely many remaining elements of $Z \cap K$ to X, we obtain a new relative generating set X' such that $|X' \Delta X| < \infty$. By Lemma 2.11, $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h (K, X')$ and $Z \cap K \subset X'$. By Remark 3.18, $\Gamma(K, X' \sqcup \mathcal{H})$ is also a quasi-tree. Thus X' is the required set in the statement of Proposition 3.2, which completes the proof.

3.5 Applications of Theorem 3.1

In order to prove Theorem 1.7, we first need to recall the following definitions.

Definition 3.21 (loxodromic element) Let G be a group acting on a hyperbolic space S. An element $g \in G$ is called *loxodromic* if the map $\mathbb{Z} \to S$ defined by $n \mapsto g^n s$ is a quasi-isometric embedding for some (equivalently, any) $s \in S$.

Definition 3.22 (elementary subgroup [5, Lemma 6.5]) Let G be a group acting acylindrically on a hyperbolic space S and $g \in G$ a loxodromic element. Then g is contained in a unique maximal *elementary subgroup* $E(g)$ of G given by

$$E(g) = \{h \in G \mid d_{\mathrm{Hau}}(l, h(l)) < \infty\},$$

where d_{Hau} denotes the Hausdorff distance and l is a quasi-geodesic axis of g in S.

Corollary 3.23 *A group G is acylindrically hyperbolic if and only if G has an acylindrical and non-elementary action on a quasi-tree.*

Proof If G has an acylindrical and non-elementary action on a quasi-tree, Theorem 2.9 implies that G is acylindrically hyperbolic. Conversely, let G be acylindrically hyperbolic, with an acylindrical non-elementary action on a hyperbolic space X. Let g be a loxodromic element for this action. By Lemma 6.5 of [5] the elementary subgroup $E(g)$ is virtually cyclic and thus countable. By Theorem 6.8 of [5], $E(g)$ is hyperbolically embedded in G. Taking $K = G$ and $E(g)$ to be the hyperbolically embedded subgroup in the statement of Theorem 3.1 now gives us the result. Since $E(g)$ is non-degenerate, by Lemma 5.12 of [12], the resulting action of G on the associated Cayley graph $\Gamma(G, X \sqcup E(g))$ is also non-elementary. $\square$

The following corollary is an immediate consequence of Theorem 3.1.

Corollary 3.24 *Let $\{H_1, H_2, \dots, H_n\}$ be a finite collection of countable subgroups of a group G such that $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h G$. Let K be a subgroup of G. If $H_i \leq K$ for all $i = 1, 2, \dots, n$, then $\{H_1, H_2, \dots, H_n\} \hookrightarrow_h K$.*

Definition 3.25 Let (M, d) be a geodesic metric space, and $\epsilon > 0$ a fixed constant. A subset $S \subset M$ is said to be *ϵ–coarsely connected* if for any two points x, y in S, there exist points $x_0 = x, x_1, x_2, \dots, x_{n-1}, x_n = y$ in S such that for all $i = 0, \dots, n-1$,

$$d(x_i, x_{i+1}) \leq \epsilon.$$

We say that S is *coarsely connected* if it is ϵ–coarsely connected for some $\epsilon > 0$.

Recall that we denote the closed σ–neighborhood of S by $S^{+\sigma}$.

Definition 3.26 Let (M, d) be a geodesic metric space, and $\sigma > 0$ a fixed constant. A subset $S \subset M$ is said to be *σ–quasi-convex* if for any two points x, y in S, any geodesic connecting x and y is contained in $S^{+\sigma}$. Further, we say that S is *quasi-convex* if it is σ–quasi-convex for some $\sigma > 0$.

Corollary 3.27 *Let H be a finitely generated subgroup of an acylindrically hyperbolic group G. Then there exists a subset $X \subset G$ such that*

(a) *$\Gamma(G, X)$ is hyperbolic, and the action of G on $\Gamma(G, X)$ is non-elementary and acylindrical, and*

(b) *H is quasi-convex in $\Gamma(G, X)$.*

To prove this corollary, we need the following two lemmas.

Lemma 3.28 *Let T be a tree, and let $Q \subset T$ be ϵ–coarsely connected. Then Q is ϵ–quasi-convex.*

Proof Let $\epsilon > 0$ be the constant from Definition 3.25. Let x, y be two points in Q and p be any geodesic between them. Then there exist points $x_0 = x, x_1, x_2, \dots, x_{n-1}, x_n = y$ in Q such that $d(x_i, x_{i+1}) \leq \epsilon$ for all $i = 0, \dots, n-1$. Let p_i denote the geodesic segments between x_i and x_{i+1} for $i = 0, 1, \dots, n-1$. Since T is a tree, we must have that

$$p \subseteq \bigcup_{i=0}^{n-1} p_i.$$

By definition, $p_i \subseteq B(x_i, \epsilon)$, the ball of radius ϵ centered at x_i, for $i = 0, 1, \dots, n-1$. Since $x_i \in Q$ for $i = 0, 1, \dots, n-1$, we obtain

$$p_i \subseteq Q^{+\epsilon}.$$

This implies $p \subseteq Q^{+\epsilon}$. □

Lemma 3.29 *Let Γ be a quasi-tree, and $S \subset \Gamma$ be coarsely connected. Then S is quasi-convex.*

Proof Let T be a tree such that Γ is quasi-isometric to T. Let d_Γ and d_T denote distances in Γ and T, respectively. Let $\delta > 0$ be the hyperbolicity constant of Γ. Let $q\colon T \to \Gamma$ be a (λ, C)–quasi-isometry, ie

$$-C + \frac{1}{\lambda} d_T(a,b) \le d_\Gamma(q(a), q(b)) \le \lambda d_T(a,b) + C.$$

Let $\epsilon > 0$ be the constant from Definition 3.25 for S. Set $Q = q^{-1}(S)$. Then $Q \subset T$. It is easy to check that Q is ρ–coarsely connected with constant $\rho = \lambda(\epsilon + C)$. By Lemma 3.28, Q is ρ–quasi-convex.

Let x, y be two points in S and p be a geodesic between them. Choose points a, b in Q such that $q(a) = x$ and $q(b) = y$. Let r denote the (unique) geodesic in T between a and b. Since Q is ρ–quasi-convex, we have

$$r \subseteq Q^{+\rho}.$$

Set $\sigma = \lambda\rho + C$. Then

$$q(r) \subseteq S^{+\sigma}.$$

Further, $q \circ r$ is a quasi-geodesic between x and y. By Lemma 2.17, there exists a constant $R = R(\lambda, C, \delta)$ such that $q(r)$ and p are Hausdorff distance less than R from each other. This implies that $p \subseteq S^{+(R+\sigma)}$. Thus S is quasi-convex. □

Proof of Corollary 3.27 By Corollary 3.23, there exists a generating set X of G such that $\Gamma(G, X)$ is a quasi-tree (hence hyperbolic) and the action of G on $\Gamma(G, X)$ is acylindrical and non-elementary. Let d_X denote the metric on $\Gamma(G, X)$ induced by the generating set X. Let $H = \langle x_1, x_2, \dots, x_n \rangle$. Set

$$\epsilon = \max\{d_X(1, x_i^{\pm 1}) \mid i = 1, 2, \dots, n\}.$$

We claim that H is coarsely connected with constant ϵ. Indeed if u, v are elements of H, then $u^{-1}v = \prod_{j=1}^{k} w_j$, where $w_j \in \{x_1^{\pm 1}, \dots, x_n^{\pm 1}\}$. Set

$$z_0 = u, \quad z_1 = uw_1, \quad \dots, \quad z_{k-1} = uw_1w_2\cdots w_{k-1}, \quad z_k = v.$$

Clearly $z_i \in H$ for all $i = 0, 1, \ldots, k-1$. Further,

$$d_X(z_i, z_{i+1}) = d_X(1, w_{i+1}) \leq \epsilon$$

for all $i = 0, 1, \ldots, k-1$. By Lemma 3.29, H is quasi-convex in $\Gamma(G, X)$. □

References

[1] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, preprint (2015) arXiv

[2] **M Bestvina**, **K Bromberg**, **K Fujiwara**, *Constructing group actions on quasi-trees and applications to mapping class groups*, Publ. Math. Inst. Hautes Études Sci. 122 (2015) 1–64 MR

[3] **B H Bowditch**, *Intersection numbers and the hyperbolicity of the curve complex*, J. Reine Angew. Math. 598 (2006) 105–129 MR

[4] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR

[5] **F Dahmani**, **V Guirardel**, **D Osin**, *Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces*, Mem. Amer. Math. Soc. 1156, Amer. Math. Soc., Providence, RI (2017) MR

[6] **M F Hagen**, *Weak hyperbolicity of cube complexes and quasi-arboreal groups*, J. Topol. 7 (2014) 385–418 MR

[7] **I Kapovich**, **K Rafi**, *On hyperbolicity of free splitting and free factor complexes*, Groups Geom. Dyn. 8 (2014) 391–414 MR

[8] **S-H Kim**, **T Koberda**, *The geometry of the curve graph of a right-angled Artin group*, Internat. J. Algebra Comput. 24 (2014) 121–169 MR

[9] **J F Manning**, *Geometry of pseudocharacters*, Geom. Topol. 9 (2005) 1147–1185 MR

[10] **A Minasyan**, **D Osin**, *Small subgroups of acylindrically hyperbolic groups*, preprint (2016)

[11] **D Osin**, *On acylindrical hyperbolicity of groups with positive first ℓ^2–Betti number*, Bull. Lond. Math. Soc. 47 (2015) 725–730 MR

[12] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368 (2016) 851–888 MR

[13] **E Rips**, *Subgroups of small cancellation groups*, Bull. London Math. Soc. 14 (1982) 45–47 MR

[14] **A Sisto**, *On metric relative hyperbolicity*, preprint (2012) arXiv

Department of Mathematics, Vanderbilt University
1326 Stevenson Center Ln, Nashville, TN, 37240, United States

sahana.balasubramanya@vanderbilt.edu

Received: 4 March 2016 Revised: 19 October 2016

Algebraic & Geometric Topology 17 (2017) 2177–2237

Translation surfaces and the curve graph in genus two

DUC-MANH NGUYEN

Let S be a (topological) compact closed surface of genus two. We associate to each translation surface $(X,\omega) \in \Omega\mathcal{M}_2 = \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$ a subgraph $\hat{\mathcal{C}}_{\rm cyl}$ of the curve graph of S. The vertices of this subgraph are free homotopy classes of curves which can be represented either by a simple closed geodesic or by a concatenation of two parallel saddle connections (satisfying some additional properties) on X. The subgraph $\hat{\mathcal{C}}_{\rm cyl}$ is by definition $\mathrm{GL}^+(2,\mathbb{R})$–invariant. Hence it may be seen as the image of the corresponding Teichmüller disk in the curve graph. We will show that $\hat{\mathcal{C}}_{\rm cyl}$ is always connected and has infinite diameter. The group $\mathrm{Aff}^+(X,\omega)$ of affine automorphisms of (X,ω) preserves naturally $\hat{\mathcal{C}}_{\rm cyl}$, we show that $\mathrm{Aff}^+(X,\omega)$ is precisely the stabilizer of $\hat{\mathcal{C}}_{\rm cyl}$ in $\mathrm{Mod}(S)$. We also prove that $\hat{\mathcal{C}}_{\rm cyl}$ is Gromov-hyperbolic if (X,ω) is completely periodic in the sense of Calta.

It turns out that the quotient of $\hat{\mathcal{C}}_{\rm cyl}$ by $\mathrm{Aff}^+(X,\omega)$ is closely related to McMullen's prototypes in the case that (X,ω) is a Veech surface in $\mathcal{H}(2)$. We finally show that this quotient graph has finitely many vertices if and only if (X,ω) is a Veech surface for (X,ω) in both strata $\mathcal{H}(2)$ and $\mathcal{H}(1,1)$.

51H20; 54H15

1 Introduction

1.1 The curve complex

Let S be a compact surface. The *curve complex* of S is a simplicial complex whose vertices are free homotopy classes of essential simple closed curves on S, and k–simplices are defined to be the sets of (homotopy classes of) $k+1$ curves that can be realized pairwise disjointly on S. This complex was introduced by Harvey [20] in order to use its combinatorial structure to encode the asymptotic geometry of the Teichmüller space. It turns out that its geometry is intimately related to the geometry and topology of Teichmüller space; see eg Rafi [43]. The curve complex has now become a central subject in Teichmüller theory, low-dimensional topology, and geometric group theory. Note that this complex is quasi-isometric to its 1–skeleton, which is referred to as the *curve graph* of S. In this paper we will denote the curve graph by $\mathcal{C}(S)$.

The mapping class group $\mathrm{Mod}(S)$ naturally acts on the curve complex by isomorphisms. In most cases, all automorphisms of the curve complex are induced by elements of $\mathrm{Mod}(S)$; see Ivanov [24] and Luo [31]. Based on this relation, topological and

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2177

combinatorial properties of the curve complex have been used to study the mapping class group, for example in Harer [19] and Bestvina and Fujiwara [2]. Masur and Minsky [33] showed that the curve graph (and the curve complex) is Gromov-hyperbolic; see also Bowditch [6]. A stronger result, that the hyperbolicity constant is independent of the surface S, has recently been proved simultaneously by several people: Aougab [1], Bowditch [7], Clay, Rafi and Schleimer [12], and Hensel, Przytycki and Webb [21]. Its boundary at infinity has been studied by Klarreich [27] and Hamenstädt [16]. Those results have led to numerous applications and a fast growing literature on the subject. In particular, the hyperbolicity of the curve graph has been exploited in the resolution of the ending lamination conjecture by Brock, Canary and Minsky [8]. For a nice survey on the curve complex and its applications we refer to Bowditch [5].

1.2 Teichmüller disks and translation surfaces

Another important notion in Teichmüller theory are the Teichmüller disks. These are isometric embeddings of the hyperbolic disk $\mathbb{H}$ in the Teichmüller space. Such a disk can be viewed as a complex geodesic generated by a quadratic differential q on a Riemann surface X. This quadratic differential defines a flat metric structure on X with conical singularities such that the holonomy of any closed curve on X belongs to the subgroup $\{\pm\mathrm{Id}\}\times\mathbb{R}^2$ of $\mathrm{Isom}(\mathbb{R}^2)$. If this quadratic differential is the square of a holomorphic 1–form ω on X, then the holonomy of any closed curve is a translation of $\mathbb{R}^2$, and we have a translation surface (X,ω).

Using the flat metric viewpoint, one can easily define a natural action of $\mathrm{GL}^+(2,\mathbb{R})$ on the space of translation surfaces as follows: given a matrix $A\in\mathrm{GL}^+(2,\mathbb{R})$ and an atlas $\{\phi_i \mid i\in I\}$ defining a translation surface structure, we get an atlas for another translation surface structure defined by $\{A\circ\phi_i \mid i\in I\}$. The Teichmüller disk generated by a holomorphic 1–form (X,ω) is precisely the projection into the Teichmüller space of its $\mathrm{GL}^+(2,\mathbb{R})$–orbit. Translations surfaces and their $\mathrm{GL}^+(2,\mathbb{R})$–orbit also arise in different contexts such as dynamics of billiards in rational polygons, interval exchange transformations and pseudo-Anosov homeomorphisms.

The importance of the $\mathrm{GL}^+(2,\mathbb{R})$–action is related to the fact that the $\mathrm{GL}^+(2,\mathbb{R})$–orbit closure of a translation surface encodes information on its geometric and dynamical properties. A remarkable illustration of this phenomenon is the famous Veech dichotomy, which states that if the stabilizer of (X,ω) for the action of $\mathrm{GL}^+(2,\mathbb{R})$ is a lattice in $\mathrm{SL}(2,\mathbb{R})$, then the linear flow in any direction on X is either periodic or uniquely ergodic. By a result of Smillie (see [49; 46]) the stabilizer of (X,ω), denoted by $\mathrm{SL}(X,\omega)$, is a lattice in $\mathrm{SL}(2,\mathbb{R})$ if and only if the $\mathrm{GL}^+(2,\mathbb{R})$–orbit of (X,ω) is a closed subset of the moduli space. For more details on translation surfaces and related problems we refer to the excellent surveys by Masur and Tabachnikov [35] and Zorich [53].

The group $\mathrm{SL}(X,\omega)$ is closely related to the subgroup of the mapping class group that stabilizes the Teichmüller disk generated by (X,ω). This subgroup consists of elements of $\mathrm{Mod}(S)$ that are realized by homeomorphisms of X preserving the set of singularities (for the flat metric), and given by affine maps in local charts of the flat metric structure. This subgroup is denoted by $\mathrm{Aff}^+(X,\omega)$. There is a natural homomorphism from $\mathrm{Aff}^+(X,\omega)$ to $\mathrm{SL}(X,\omega)$ which associates to each element of $\mathrm{Aff}^+(X,\omega)$ its derivative. It is not difficult to see that this homomorphism is surjective and has finite kernel. The study of $\mathrm{Aff}^+(X,\omega)$ and $\mathrm{SL}(X,\omega)$ is a recurrent theme in the theory of dynamics in Teichmüller space; see eg McMullen [36], Hubert and Schmidt [23], Hubert and Lanneau [22], Möller [40] and Lehnert [30].

1.3 The flat metric and curve complex

Consider now the flat metric defined by a holomorphic 1–form ω on a (compact) Riemann surface X. By compactness, there exists a curve of minimal length in the free homotopy class of any essential simple closed curve. In general this curve of minimal length may not be a geodesic as it may contain some singularity in its interior. Nevertheless, following a result by Masur [32], we know that there are infinitely many curves that can be realized as simple closed geodesics for ω. Thus (X,ω) specifies a subset of vertices of $\mathcal{C}(S)$. Note that unlike the situation of hyperbolic surfaces, closed geodesics of minimal length are not unique in their homotopy class. They actually arise in family, that is, simple closed geodesics in the same homotopy class fill out a subset of X which is isometric to $(\mathbb{R}/c\mathbb{Z})\times(0,h)$. We will call such a subset a *geometric cylinder*, and the corresponding simple closed geodesics its *core curves*.

Mimicking the construction of the curve graph, we can add an edge between two vertices representing two cylinders if there exist two curves, one in each homotopy class, that can be realized disjointly (this condition is equivalent to requiring that the corresponding geodesics for the flat metric are disjoint). Thus, for each translation surface, we have a subgraph $\mathcal{C}_{\mathrm{cyl}}$ of the curve graph.

Let A be a matrix in $\mathrm{GL}^+(2,\mathbb{R})$, and consider the surface $(X',\omega') := A\cdot(X,\omega)$. Since the action of A preserves the affine structure, a geodesic on X corresponds to a geodesic on X' and vice-versa. Therefore, the subgraphs associated to (X',ω') and to (X,ω) are the same. This subgraph is actually associated to the Teichmüller disk generated by (X,ω). As $\mathcal{C}(S)$ can be viewed as the combinatorial model for the Teichmüller space, $\mathcal{C}_{\mathrm{cyl}}$ can be viewed as the counterpart of a Teichmüller disk in this setting. By definition, elements of $\mathrm{Aff}^+(X,\omega)$ preserve $\mathcal{C}_{\mathrm{cyl}}$ and act on $\mathcal{C}_{\mathrm{cyl}}$

by isomorphisms. As properties of the mapping class group can be studied via its action on the curve complex, one can expect the knowledge about the combinatorial and geometric structure of $\mathcal{C}_{\mathrm{cyl}}$ to be useful for the study of $\mathrm{Aff}^+(X,\omega)$.

1.4 Statement of results

The main purpose of this paper is to investigate $\mathcal{C}_{\mathrm{cyl}}$ when X is a surface of genus two. The reason for this restriction is the technical difficulties for the general cases. Hopefully, the results and techniques used in this situation may inspire further results in higher genera.

Recall that the moduli space of translation surfaces is naturally stratified by the zero orders of the 1–form ω (or equivalently, the cone angles at the singularities). In genus two, we have two strata: $\mathcal{H}(2)$ which contains pairs (X,ω) such that ω has a unique double zero, and $\mathcal{H}(1,1)$ which contains pairs (X,ω) such that ω has two simple zeros. Our first result shows that the geometry of $\mathcal{C}_{\mathrm{cyl}}$ does depend on the stratum of (X,ω).

Theorem A (Theorem 2.6) *If $(X,\omega)\in\mathcal{H}(2)$ then $\mathcal{C}_{\mathrm{cyl}}$ contains no triangles, but if $(X,\omega)\in\mathcal{H}(1,1)$ then $\mathcal{C}_{\mathrm{cyl}}$ always contains triangles.*

Note that a triangle in $\mathcal{C}_{\mathrm{cyl}}$ is a triple of simple closed pairwise disjoint curves that are simultaneously realized as core curves of three cylinders in (X,ω).

From its definition, the geometric structure of the subgraph $\mathcal{C}_{\mathrm{cyl}}$ depends very much on the flat metric of (X,ω). It is not difficult to see that $\mathcal{C}_{\mathrm{cyl}}$ is not connected in general; see Section 3. To get a nicer subgraph of $\mathcal{C}(S)$, we enlarge $\mathcal{C}_{\mathrm{cyl}}$ by adjoining to it the vertices of $\mathcal{C}(S)$ representing *degenerate cylinders*. Roughly speaking, a degenerate cylinder on X is a union of two saddle connections in the same direction such that there are deformations of (X,ω) on which this union is freely homotopic to the core curves of a geometric cylinder. We refer to Section 3 for a more precise definition. In particular, any degenerate cylinder is freely homotopic to a simple closed curve. Thus it corresponds to a vertex of $\mathcal{C}(S)$.

We define $\widehat{\mathcal{C}}^{(0)}_{\mathrm{cyl}}$ to be the set of vertices of $\mathcal{C}(S)$ that correspond to geometric cylinders and degenerate cylinders in (X,ω). We then define $\widehat{\mathcal{C}}^{(1)}_{\mathrm{cyl}}$ to be the set of the edges of $\mathcal{C}(S)$ both of whose endpoints belong to $\widehat{\mathcal{C}}^{(0)}_{\mathrm{cyl}}$. We thus get a subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ of $\mathcal{C}(S)$. By a slight abuse of notation, we will also call $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ the *cylinder graph* of (X,ω). Subsequently, this subgraph will be the main object of our investigation. We summarize the results concerning $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ in the following theorem:

Theorem B *For any $(X,\omega)\in\mathcal{H}(1,1)\sqcup\mathcal{H}(2)$, the subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is connected and has infinite diameter. The subgroup of* $\mathrm{Mod}(S)$ *that stabilizes $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is precisely* $\mathrm{Aff}^+(X,\omega)$. *Moreover, if (X,ω) is completely periodic in the sense of Calta, then $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is Gromov-hyperbolic.*

Theorem B actually comprises several statements, which are proved in Corollary 4.2, Propositions 5.1 and 6.1 and Theorem 7.1. The contexts and precise statements will be given in the corresponding sections.

We finally consider the quotient of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ by the action of $\mathrm{Aff}^+(X,\omega)$ in the case that (X,ω) is a Veech surface, that is, $\mathrm{SL}(X,\omega)$ is a lattice of $\mathrm{SL}(2,\mathbb{R})$.

Theorem C *Let $\mathcal{G}$ be the quotient of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ by the group of affine automorphisms. Then $(X,\omega)\in\mathcal{H}(2)\sqcup\mathcal{H}(1,1)$ is a Veech surface if and only if $\mathcal{G}$ has finitely many vertices. For any Veech surface in $\mathcal{H}(2)$ the set of edges of $\mathcal{G}$ is also finite. There exist Veech surfaces in $\mathcal{H}(1,1)$ such that $\mathcal{G}$ has infinitely many edges.*

The statements of Theorem C are proved in Theorem 8.1 and Proposition 8.2.

1.5 Outline

In Section 2 we recall standard notions concerning translation surfaces. We show some geometric and topological features of translation surfaces of genus two. We end this section with the proof of Theorem A.

In Section 3, we introduce the notion of degenerate cylinders and define the cylinder graphs $\mathcal{C}_{\mathrm{cyl}}$ and $\widehat{\mathcal{C}}_{\mathrm{cyl}}$. We show that $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is connected and has infinite diameter in Sections 4 and 5. These results follow from Theorem 4.1, which gives a bound on the distance in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ using the intersection number.

Section 6 is devoted to the proof of the fact that the stabilizer subgroup of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ in $\mathrm{Mod}(S)$ is precisely the group of affine automorphisms.

In Section 7 we show that if (X,ω) is completely periodic in the sense of Calta, then $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is Gromov-hyperbolic. Our proof follows a strategy of Bowditch and uses a hyperbolicity criterion by Masur and Schleimer.

We give the proof of Theorem C in Section 8. Finally, in Section 9, we give the connection between the quotient graph $\mathcal{G}=\widehat{\mathcal{C}}_{\mathrm{cyl}}/\mathrm{Aff}^+$ and the set of prototypes for Veech surfaces in $\mathcal{H}(2)$, which were introduced by McMullen [37].

Acknowledgements The author warmly thanks Arnaud Hilion for very helpful and stimulating discussions.

2 Preliminaries

In this section we will prove some topological properties of saddle connections and cylinders on translation surfaces in genus two. The main result of this section is Theorem 2.6.

Let (X,ω) be a translation surface. A saddle connection on X is a geodesic segment whose endpoints are singularities, but which contains no singularities in its interior. A (geometric) cylinder of X is a subset C isometric to $(\mathbb{R}/c\mathbb{Z})\times(0,h)$, with $c,h\in\mathbb{R}_{>0}$, which is not properly contained in another subset with the same property. The parameter c is called the *circumference* and h the *width* or *height* of this cylinder.

The isometry from $(\mathbb{R}/c\mathbb{Z})\times(0,h)$ to C can be extended by continuity to a map from $(\mathbb{R}/c\mathbb{Z})\times[0,h]$ to X. We will call the images of $(\mathbb{R}/c\mathbb{Z})\times\{0\}$ and $(\mathbb{R}/c\mathbb{Z})\times\{h\}$ the boundary components of C. Each boundary component is a concatenation of some saddle connections. It may happen that the two boundary components coincide as subsets of X. We say that C is a *simple cylinder* if each of its boundary components is a single saddle connection. It is worth noticing that on a translation surface of genus two, every cylinder is invariant by the hyperelliptic involution. Therefore, the two boundary components of any cylinder contain the same number of saddle connections.

Throughout this paper, for any cycle $c\in H_1(X,\{\text{zeros of }\omega\};\mathbb{Z})$, we will use the notation $\omega(c):=\int_c\omega$, and for any saddle connection s, its euclidean length will be denoted by $|s|$. Let us start by the following elementary lemma.

Lemma 2.1 *Let (X,ω) be a translation surface in one of the hyperelliptic components $\mathcal{H}^{\mathrm{hyp}}(2g-2)$ or $\mathcal{H}^{\mathrm{hyp}}(g-1,g-1)$, and s be a saddle connection invariant by the hyperelliptic involution τ of X. We assume that s is not vertical. Then there exist a parallelogram $\boldsymbol{P}=(P_1P_2P_3P_4)$ in $\mathbb{R}^2$ and a locally isometric mapping $\varphi\colon \boldsymbol{P}\to X$ such that the following hold:*

(a) *The vertical lines through the vertices P_3 and P_4 intersect the diagonal $\overline{P_1P_2}$.*

(b) *The vertices of $\boldsymbol{P}$ are mapped to the singularities of X, and $\overline{P_1P_2}$ is mapped isometrically to s.*

(c) *The restriction of φ into $\mathrm{int}(\boldsymbol{P})$ is an embedding.*

(d) *Let $\eta>0$ be the length of the vertical segment from P_3 or P_4 to a point in $\overline{P_1P_2}$. Then for any vertical segment u in X from a singular point to a point in s, we have $|u|\geq\eta$, where $|u|$ is the euclidean length of u.*

We will call $\boldsymbol{P}$ the embedded parallelogram *associated to s.*

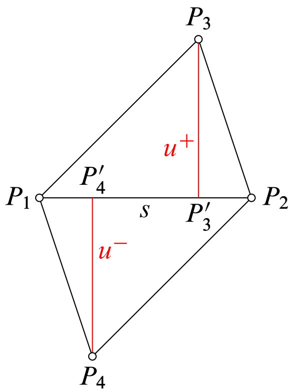


Figure 1: Here $s = \varphi(\overline{P_1 P_2})$, $u^+ = \varphi(\overline{P_3 P_3'})$, $u^- = \varphi(\overline{P_4 P_4'})$, and $\boldsymbol{P} = (P_1 P_2 P_3 P_4)$ is the embedded parallelogram associated to s

Remark 2.2

- Since s in invariant by τ, we must have $\tau(\varphi(\boldsymbol{P})) = \varphi(\boldsymbol{P})$.
- The sides of $\boldsymbol{P}$ are mapped to saddle connections on X. Even though the restriction of φ into $\mathrm{int}(\boldsymbol{P})$ is one-to-one, it may happen that φ maps the opposite sides of $\boldsymbol{P}$ to the same saddle connection.
- This lemma is also valid for translation surfaces in $\mathcal{H}(0)$ and $\mathcal{H}(0,0)$.

Proof of Lemma 2.1 We will only give the proof for the case $(X,\omega) \in \mathcal{H}^{\mathrm{hyp}}(2g-2)$, as the proof for $\mathcal{H}^{\mathrm{hyp}}(g-1,g-1)$ is the same. Using

$$U_- = \left\{ \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\},$$

we can assume that s is horizontal. Let Ψ_t be the vertical flow on X generated by the vertical vector field $(0,1)$; this flow moves regular points of X vertically, upward if $t > 0$.

Consider the vertical geodesic rays emanating from the unique zero x_0 of ω in direction $(0,-1)$. We claim that one of the rays in this direction must meet s. Indeed, if this is not the case, then $\Psi_t(s)$ does not contain x_0 for any $t \in \mathbb{R}_{>0}$, and it follows that one can embed a rectangle of infinite area into X. Let u^+ be a vertical geodesic segment of minimal length from x_0 to a point in s which is included in a ray in direction $(0,-1)$. Since s is invariant by τ, the segment $u^- := \tau(u^+)$ is vertical of minimal length from x_0 to a point in s which is included in a ray in direction $(0,1)$. Using the developing map, we can realize s as a horizontal segment $\overline{P_1 P_2} \subset \mathbb{R}^2$, u^+ (resp. u^-) as a vertical segment $\overline{P_3 P_3'}$ (resp. $\overline{P_4 P_4'}$) where $P_3', P_4' \in \overline{P_1 P_2}$; see Figure 1. We remark that the central symmetry fixing the midpoint of $\overline{P_1 P_2}$ exchanges $\overline{P_3 P_3'}$ and $\overline{P_4 P_4'}$.

Let $\boldsymbol{P}$ denote the parallelogram $(P_1P_3P_2P_4)$. We define a map $\varphi\colon \boldsymbol{P} \to X$ as follows: for any point $M \in \boldsymbol{P}$, let M' be the orthogonal projection of M in $\overline{P_1P_2}$, and t be the length of $\overline{MM'}$. Let $\widehat{M}'$ be the point in s corresponding to M' by the identification between $\overline{P_1P_2}$ and s. We then define $\varphi(M) := \Psi_t(\widehat{M}')$ if M is above $\overline{P_1P_2}$, and $\varphi(M) = \Psi_{-t}(\widehat{M}')$ if M is below $\overline{P_1P_2}$. By definition, φ is a local isometry and maps the vertices of $\boldsymbol{P}$ to x_0.

Note that we have $|\overline{MM'}| \leq |\overline{P_3P_3'}| = |\overline{P_4P_4'}|$, and the equality only occurs when $M = P_3$ or $M = P_4$. Thus, for all $M \in \boldsymbol{P} \setminus \{P_1, P_2, P_3, P_4\}$, $\varphi(M)$ is a regular point in X; otherwise we would have a vertical segment from P_0 to a point in s of length smaller than $|u^+|$.

We now claim that $\varphi|_{\mathrm{int}(\boldsymbol{P})}$ is an embedding. Assume that there exist two points $M_1, M_2 \in \mathrm{int}(\boldsymbol{P})$ such that $\varphi(M_1) = \varphi(M_2)$. Set $\vec{v} := \overrightarrow{M_1M_2}$; then for any $M, M' \in \boldsymbol{P}$ such that $\overrightarrow{MM'} = \vec{v}$, we have $\varphi(M) = \varphi(M')$. Since $\boldsymbol{P}$ is a parallelogram, there exists a vertex $P_i \in \{P_1, P_2, P_3, P_4\}$ and a point $M' \in \boldsymbol{P} \setminus \{P_1, P_2, P_3, P_4\}$ such that $\overrightarrow{P_iM'} = \vec{v}$, which implies that $\varphi(M') = x_0$, and we have a contradiction to the observation above.

It is now straightforward to verify that $\boldsymbol{P}$ and φ satisfy all the required properties. □

In what follows, by a *slit torus* we will mean a triple (X, ω, s) where X is an elliptic curve, ω a nonzero holomorphic 1–form and s an embedded geodesic segment (with respect to the flat metric defined by ω) on X. We consider the endpoints of s as marked points on X. Note that there is a unique involution of X that preserves s and permutes its endpoints. The following lemma is useful for us in the sequel.

Lemma 2.3 *Let (X, ω, s) be a slit torus and x_1, x_2 be the endpoints of s. Assume that the segment (slit) s is not vertical, that is, $|\mathrm{Re}\,\omega(s)| > 0$. Then there exists a pair of parallel simple closed geodesics c_1, c_2 with c_i passing through x_i such that $c_i \cap \mathrm{int}(s) = \varnothing$, and $|\mathrm{Re}\,\omega(c_i)| \leq |s|$. In particular, the geodesics c_1, c_2 cut X into two cylinders, one of which contains $\mathrm{int}(s)$. Moreover, any leaf of the vertical foliation intersecting c_i must intersect s, and if every leaf of the vertical foliation meets s, then we have $|\mathrm{Re}\,\omega(c_i)| > 0$.*

Proof We remark that a slit torus can be considered as hyperelliptic translation surface with the hyperelliptic involution being the unique one that preserves s and exchanges its endpoints. Let $\boldsymbol{P} = (P_1P_2P_3P_4)$ be the parallelogram associated to s, and $\varphi\colon \boldsymbol{P} \to X$ the corresponding embedding defined as in Lemma 2.1. Since we have $\varphi(P_3) \in \{x_1, x_2\}$, either $\varphi(P_3) = \varphi(P_1)$ or $\varphi(P_3) = \varphi(P_2)$. It follows that one pair of opposite sides of $\boldsymbol{P}$ are mapped to a pair of parallel simple closed geodesics c_1, c_2 of X with c_i passing through x_i. The other pair of opposite sides of $\boldsymbol{P}$ are mapped

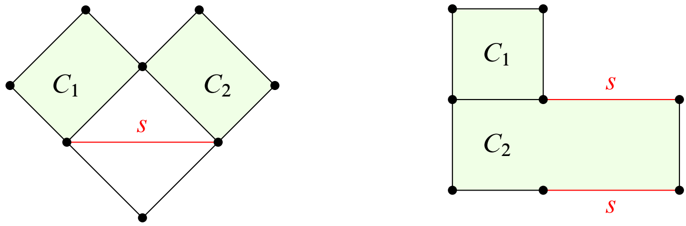


Figure 2: Configurations of C_1, C_2 with respect to s: none of C_1, C_2 contains s in its boundary (left) and s is contained in the boundary of C_2 (right).

to the same geodesic segment joining x_1 and x_2. Thus $\varphi(\boldsymbol{P})$ is a cylinder in X that contains s. Since X is a torus, the complement of $\varphi(\boldsymbol{P})$ is also a cylinder. It is straightforward to check that the pair $\{c_1, c_2\}$ satisfy all the required properties. □

We now turn to translation surfaces in genus two. Let (X, ω) be a translation surface in $\mathcal{H}(2) \sqcup \mathcal{H}(1,1)$. We denote by τ the hyperelliptic involution of X.

Lemma 2.4 *Let s_1, s_2 be a pair of saddle connections in X which are permuted by τ. If $(X,\omega) \in \mathcal{H}(2)$, then s_1 and s_2 bound a simple cylinder. If $(X,\omega) \in \mathcal{H}(1,1)$ then we have two cases:*

- *If s_i joins a zero of ω to itself, then s_1 and s_2 bound a simple cylinder.*
- *If s_i joins two different zeros of ω, then $s_1 \cup s_2$ decomposes X as a connected sum of two slit tori.*

Proof Since τ acts by $-\mathrm{Id}$ on $H_1(X, \mathbb{Z})$, s_1 and s_2 must be homologous. This lemma follows from an inspection on the configurations of rays originating from the zero(s) of ω in the same direction. □

Lemma 2.5 *Let (X,ω) be a surface in $\mathcal{H}(2)$ and s be a saddle connection in X invariant by the hyperelliptic involution τ. Then there exist two disjoint cylinders C_1, C_2 that do not intersect s, that is, $C_1 \cap C_2 = \varnothing$, and the core curves of C_1 and C_2 do not meet s. We remark that s may be contained in the boundary of C_1 or C_2. The possible configurations of C_1 and C_2 with respect to s are shown in Figure 2.*

Proof Without loss of generality, we can assume that s is horizontal. Let $\boldsymbol{P} = (P_1 P_3 P_2 P_4)$ be the embedded parallelogram associated to s, and $\varphi\colon \boldsymbol{P} \to X$ be the embedding map such that $s = \varphi(\overline{P_1 P_2})$; see Lemma 2.1. We choose the labeling of the vertices of $\boldsymbol{P}$ such that P_3 is the highest vertex, and P_4 is the lowest one. Throughout the proof, we will refer to Figure 3.

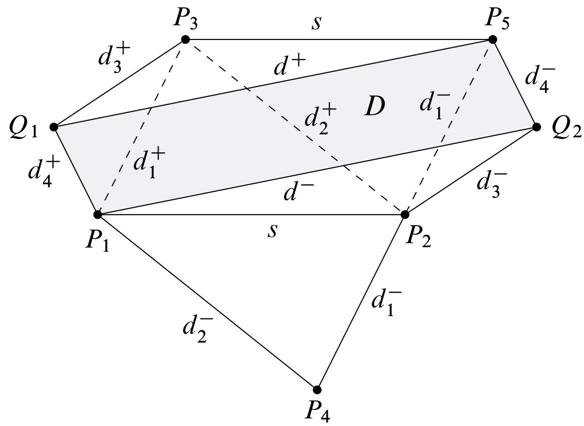


Figure 3: Finding a cylinder disjoint from s

Let $d_1^+ = \varphi(\overline{P_3P_1}), d_2^+ = \varphi(\overline{P_3P_2}), d_1^- = \varphi(\overline{P_4P_2}), d_2^- = \varphi(\overline{P_4P_1})$. We have $d_i^- = \tau(d_i^+)$. By Lemma 2.4, either $d_i^+ = d_i^-$ as subsets of X or the pair $d_i^\pm$ bound a simple cylinder. We remark that d_1^+ and d_2^+ cannot both be invariant by τ, otherwise we would have $X = \varphi(\boldsymbol{P})$, and X must be a torus. Thus we must only consider two cases:

(i) Both pairs $d_1^\pm$ and $d_2^\pm$ are respectively boundaries of two simple cylinders C_1, C_2 in X. In this case, it is not difficult to see that both C_1 and C_2 are disjoint from $\varphi(\boldsymbol{P})$, and $C_1 \cap C_2 = \varnothing$. We then get the configuration Figure 2 (left).

(ii) One of the pairs $d_1^\pm, d_2^\pm$ bound a simple cylinder, the other consist of a single saddle connection invariant by τ. In this case, $\varphi(\boldsymbol{P})$ is actually a simple cylinder. Without loss of generality, we can assume that the pair $d_1^\pm$ bound the cylinder $C = \varphi(\boldsymbol{P})$, and $d_2^+ = d_2^-$.

Let P_5 be the point in $\mathbb{R}^2$ such that the triangle $(P_3P_5P_2)$ is the image of $(P_1P_2P_4)$ by the translation by $\overrightarrow{P_1P_3}$. Using the assumption that $d_2^+ = d_2^-$, that is, $\varphi(\overline{P_3P_2}) = \varphi(\overline{P_1P_4})$, we see that φ extends to a local isometric map from $\boldsymbol{P}' = (P_1P_2P_5P_3)$ to X such that $\varphi(\boldsymbol{P}') = C$ and $\varphi|_{\mathrm{int}(\boldsymbol{P}')}$ is an embedding; see Figure 3.

Consider the horizontal rays emanating from the unique zero x_0 of ω to the outside of C. By the same argument as in Lemma 2.1, we see that one of the rays in direction $(1,0)$ reaches $d_1^+ = \varphi(\overline{P_3P_1})$ from the outside of C. It follows that we can then extend φ to a convex hexagon $\boldsymbol{H} := (P_1P_2Q_2P_5P_3Q_1)$, which is the union of $\boldsymbol{P}'$ and two triangles $(P_2Q_2P_5)$ and $(P_3Q_1P_1)$. Note that $(P_2Q_2P_5)$ and $(P_3Q_1P_1)$ are exchanged by the central symmetry fixing the midpoint of $\overline{P_2P_3}$, and all the vertices of $\boldsymbol{H}$ are mapped to x_0.

Let $d_3^+ = \varphi(\overline{P_3 Q_1})$, $d_4^+ = \varphi(\overline{Q_1 P_1})$, $d_3^- = \varphi(\overline{P_2 Q_2})$ and $d_4^- = \varphi(\overline{Q_2 P_5})$. Again, for $i = 3, 4$, we have either $d_i^+ = d_i^-$ or the pair $d_i^\pm$ bound a simple cylinder. If $d_i^+ = d_i^-$ for both $i = 3, 4$, then $X = \varphi(\boldsymbol{H})$ and X must be a flat torus, so we have a contradiction. If both pairs $d_3^\pm, d_4^\pm$ are the boundaries of simple cylinders, then these cylinders are disjoint, and also disjoint from $\varphi(\boldsymbol{H})$. It follows that the total angle at x_0 is at least 8π (the total angle of $\boldsymbol{H}$ plus 4π), thus we have again a contradiction. We can then conclude that one of the pairs $d_3^\pm, d_4^\pm$ consists of a single saddle connection, and the other pair bounds a simple cylinder. Without loss of generality, we can assume that $d_3^\pm$ bounds a simple cylinder C_3, and $d_4^+ = d_4^- = d_4$. Note that C_3 must be disjoint from $\varphi(\boldsymbol{H})$, and in particular it is disjoint from s.

Let $d^+ = \varphi(\overline{Q_1 P_5})$ and $d^- = \varphi(\overline{P_1 Q_2})$; then the pair $d^\pm$ is the boundary of a cylinder D whose core curves cross $d_4^\pm$. If $\boldsymbol{H}$ is strictly convex then D is a simple cylinder, but if $\overline{P_2 Q_2}$ is parallel to $\overline{P_1 P_2}$ then D is not simple (in this case we actually have $\overline{D} = \varphi(\boldsymbol{H})$). Nevertheless, in both cases the core curves of D do not intersect s. Since D is contained in $\varphi(\boldsymbol{H})$, we have $C_3 \cap D = \varnothing$. Since both C_3 and D are disjoint from s, the lemma is proved. □

We are now ready to show the following theorem:

Theorem 2.6 (a) *On any $(X, \omega) \in \mathcal{H}(2)$, there always exist two disjoint simple cylinders. There cannot exist a triple of pairwise disjoint cylinders in X.*

(b) *On any $(X, \omega) \in \mathcal{H}(1,1)$, there always exists a triple of cylinders which are pairwise disjoint.*

Remark 2.7

- The cylinders in Theorem 2.6 are not necessarily parallel.
- There cannot exist more than three simple pairwise disjoint closed curves on S. Statement (b) means that given any holomorphic 1–form in $\mathcal{H}(1,1)$, there always exists a family of three disjoint (simple closed) curves, realized simultaneously as simple closed geodesics for the flat metric induced by this 1–form.
- The statement (a) of the theorem is a direct consequence of [42, Proposition A.1].

Proof of Theorem 2.6, case $\mathcal{H}(2)$ Lemma 2.5 almost proves the statement for $\mathcal{H}(2)$ except that it does not guarantee that both cylinders are simple. We will give here a proof by using [41, Lemma 2.1]. Let s be a saddle connection that is invariant by the hyperelliptic involution τ (one can find such a saddle connection by picking a geodesic segment of minimal length $\hat{s}$ joining a regular Weierstrass point of X to the unique zero of ω, then taking $s = \hat{s} \cup \tau(\hat{s})$). By [41, Lemma 2.1], there exists a simple cylinder C_1 that contains s. Cut off C_1 from X then identify the two geodesic segments on the boundary of the resulting surface, we obtain a flat torus (X', ω') with a marked geodesic segment s'.

We can consider (X', ω', s') as a slit torus. By Lemma 2.3, there exists a cylinder C' in X' that contains s' whose complement in X' is another cylinder C_2 disjoint from s'. By construction C_2 is a simple cylinder in X and disjoint from C_1, hence the first assertion follows.

For the second assertion, we observe that any triple of pairwise disjoint simple closed curves disconnect X into two three-holed spheres. If all the curves in this triple are simple closed geodesics (core curves of cylinders), then we get two flat surfaces with geodesic boundary. Since X has only one singularity, one of the surfaces has no singularities in its interior. But the Euler characteristic of a three-holed sphere is -1, thus we have a contradiction to the Gauss–Bonnet formula. We can then conclude that X can not contain three disjoint cylinders. □

Proof of Theorem 2.6, case $\mathcal{H}(1,1)$ By [41, Lemma 2.1], we know that there exists a simple cylinder C_0 on (X, ω) that is invariant by τ. Cut off C_0 and glue the two boundary components of the resulting surface; we obtain a surface $(\widehat{X}, \widehat{\omega}) \in \mathcal{H}(2)$ with a marked saddle connection $\hat{s}$. Note that $\hat{s}$ is invariant by the hyperelliptic involution of $\widehat{X}$. By Lemma 2.5, we know that there exist two cylinders C_1 and C_2 on $\widehat{X}$ disjoint from $\hat{s}$ such that $C_1 \cap C_2 = \varnothing$. It follows immediately that C_1 and C_2 are actually cylinders in X and disjoint from C_0, from which we get the desired conclusion. □

3 Degenerate cylinders and the cylinder graph

3.1 Cylinders and the curve graph

Each cylinder in a translation surface is filled by simple closed geodesics in the same free homotopy class. The following elementary lemma shows that two (freely) homotopic closed geodesics must belong to the same cylinder.

Lemma 3.1 *Let c_1 and c_2 be two simple closed geodesics in (X, ω) which are freely homotopic. Then c_1 and c_2 are contained in the same cylinder.*

Proof Since c_1, c_2 are freely homotopic, they are homologous, hence $\omega(c_1) = \omega(c_2)$. It follows that c_1 and c_2 are parallel, thus must be disjoint. The pair c_1, c_2 cut X into two components, one of which must be an annulus denoted by A; see Proposition A.11 of [9]. We have a flat metric on A induced by the flat metric of X. Let $\theta_1, \dots, \theta_k$ be the cone angles at the singularities in A. Since the boundary of A is geodesic, the Gauss–Bonnet formula gives

$$\sum_{1 \le i \le k} (2\pi - \theta_i) = 2\pi \chi(A) = 0.$$

Since any singularity on a translation surface has cone angle at least 4π, the equation above actually shows that A contains no singularities. Thus A is a flat annulus, which must be contained in a cylinder of X. Therefore, c_1 and c_2 are contained in the same cylinder. □

Let S be a fixed topological compact closed surface of genus two. Let $\mathcal{C}(S)$ denote the curve graph of S. Let $\Omega\mathcal{T}_2$ be the abelian differential bundle over the Teichmüller space $\mathcal{T}_2$. Elements of $\Omega\mathcal{T}_2$ are equivalence classes of triples (X,ω,f), where X is a Riemann surface of genus two, ω is a holomorphic 1–form on X, and f is a homeomorphism from S to X; two triples (X,ω,f) and (X',ω',f') are identified if there exists an isomorphism $\varphi\colon X\to X'$ such that $\varphi^*\omega'=\omega$ and $f'^{-1}\circ\varphi\circ f\colon S\to S$ is isotopic to Id_S. The equivalence class of (X,ω,f) will be denoted by $[X,\omega,f]$.

Each element $[X,\omega,f]$ of $\Omega\mathcal{T}_2$ defines naturally a subgraph $\mathcal{C}_{\mathrm{cyl}}(X,\omega,f)$ of $\mathcal{C}(S)$. The vertices of this subgraph are free homotopy classes of the core curves of all cylinders on the translation surface (X,ω). The set $\mathcal{C}^{(1)}_{\mathrm{cyl}}(X,\omega,f)$ consists of the edges in $\mathcal{C}^{(1)}(S)$ both of whose endpoints belong to $\mathcal{C}^{(0)}_{\mathrm{cyl}}(X,\omega,f)$.

3.2 Degenerate cylinders

If C is a cylinder in X that fills X (ie $\overline{C}=X$), then C represents an isolated vertex in $\mathcal{C}_{\mathrm{cyl}}(X,\omega,f)$. This is because the core curve of any other cylinder in X must cross C. So in general $\mathcal{C}_{\mathrm{cyl}}(X,\omega,f)$ is not a connected graph. To fix this issue we introduce the notion of *degenerate cylinders*. Roughly speaking, a degenerate cylinder in X is a union of parallel saddle connections such that there exist deformations of (X,ω) where this union is freely homotopic to the core curves of a simple cylinder.

To be more precise, let x_0 be a singularity on a translation surface (X,ω). For any pair (r_1,r_2) of geodesic rays emanating from x_0, we will denote the counterclockwise angle from r_1 to r_2 by $\vartheta(r_1,r_2)$. If s is an oriented saddle connection from a singularity x_1 to a singularity x_2, then we denote by s^+ (resp. s^-) the intersection of s with a neighborhood of x_1 (resp. a neighborhood of x_2). This definition also makes sense when $x_1=x_2$, in which case the orientation of s is to start in s^+ and end in s^-.

Definition 3.2 (degenerate cylinder) We will call the union of two saddle connections s_1,s_2 in $(X,\omega)\in\mathcal{H}(2)\sqcup\mathcal{H}(1,1)$ a *degenerate cylinder* if they are both invariant by the hyperelliptic involution, and up to an appropriate choice for the orientations of s_1 and s_2, we have

$$\vartheta(s_1^-,s_2^+)=\vartheta(s_1^+,s_2^-)=\pi.$$

In Figure 4, we represent the configurations of a degenerate cylinder at the singularities.

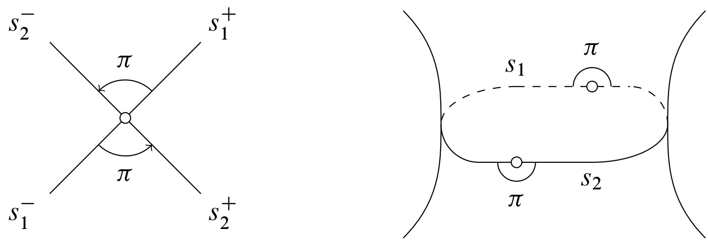


Figure 4: Configuration of a degenerate cylinder at the singularities for $\mathcal{H}(2)$ (left) and $\mathcal{H}(1,1)$ (right)

Remark 3.3

- If (X,ω) is in $\mathcal{H}(2)$, then a degenerate cylinder is not a simple curve: the zero of ω is its unique double point.
- If (X,ω) is in $\mathcal{H}(1,1)$, then the hyperelliptic involution τ of X permutes the zeros of ω, thus a saddle connection invariant by τ must connect the two zeros of ω. Therefore a degenerate cylinder must be a simple closed curve.

Examples Assume that $(X,\omega) \in \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$ is horizontally periodic, and has a unique (geometric) horizontal cylinder C. If $(X,\omega) \in \mathcal{H}(2)$ then it has 3 horizontal saddle connections s_1, s_2, s_3, which are contained in the boundary of C; see Figure 5. Note that all of them are invariant by the hyperelliptic involution. By definition $s_1 \cup s_2$, $s_2 \cup s_3$ and $s_3 \cup s_1$ are three degenerate cylinders. Similarly, if $(X,\omega) \in \mathcal{H}(1,1)$, then we have 4 horizontal saddle connections denoted by $s_1, \dots, s_4$ (see Figure 5) such that $s_i \cup s_{i+1}$ is a degenerate cylinder for $i = 1, \dots, 4$, with the convention $s_5 = s_1$.

We will now prove some key properties of degenerate cylinders.

Lemma 3.4 *Let $s_1 \cup s_2$ be a horizontal degenerate cylinder in $(X,\omega) \in \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$. Then there exists in a neighborhood of (X,ω) a continuous family of translation surfaces $\{(X_t,\omega_t) \mid t \in [0,\epsilon)\}$ in the same stratum as (X,ω), with $\epsilon \in \mathbb{R}_{>0}$, such that*

- *$(X_0,\omega_0) = (X,\omega)$;*
- *for any $t \in (0,\epsilon)$, (X_t,ω_t) contains two saddle connections $s_{1,t}$ and $s_{2,t}$ corresponding to s_1 and s_2 and satisfying the following property: $s_{1,t} \cup s_{2,t}$ is freely homotopic to the core curves of a simple cylinder C_t in X_t;*
- *as $t \to 0$, the width of C_t decreases to zero.*

Moreover, for all $t \in (0,\epsilon)$, any vertical saddle connection (resp. regular geodesic) in (X,ω) corresponds to a vertical saddle connection (resp. regular geodesic) in (X_t,ω_t).

Proof Let us define a *half cylinder* to be the quotient $(\mathbb{R}\times[0,h])/\Gamma$, where $\Gamma \simeq \mathbb{Z}_2 \ltimes \mathbb{Z}$ is generated by $t\colon (x,y) \mapsto (x+\ell, y)$ and $s\colon (x,y) \mapsto (-x, h-y)$. We will call h

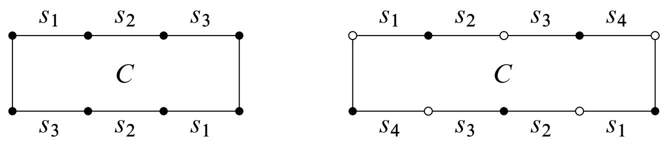


Figure 5: Degenerate cylinders on a horizontally periodic surface with a unique geometric horizontal cylinder for $\omega \in \mathcal{H}(2)$ (left) and $\omega \in \mathcal{H}(1,1)$ (right)

and ℓ the *width* and *circumference* of the half cylinder, respectively. We will refer to the projection of $(0,0)$ as the marked point on its boundary. Equivalently, a half cylinder is a closed disc equipped with a flat metric structure with geodesic boundary and two singularities of angle π in the interior.

Recall that all Riemann surfaces of genus two are hyperelliptic. Let $p\colon X \to \mathbb{CP}^1$ be the hyperelliptic double cover of X. There exists a meromorphic quadratic differential η on $\mathbb{CP}^1$ with at most simple poles such that $\omega^2 = p^*\eta$. Note that η has one zero and k poles, where $k = 5$ if $\omega \in \mathcal{H}(2)$, and $k = 6$ if $\omega \in \mathcal{H}(1,1)$. Let P_0 denote the unique zero of η, and $P_1, \dots, P_k$ its simple poles. Let Y be the flat surface defined by η on $\mathbb{CP}^1$. Observe that the cone angle of Y at P_0 is 3π if $\omega \in \mathcal{H}(2)$, and 4π if $\omega \in \mathcal{H}(1,1)$. The cone angle at P_i is π for $1, \dots, k$.

Since s_i, $i = 1, 2$, is invariant by τ, its projection in Y is a geodesic segment s_i' joining P_0 to a pole of η. By the definition of degenerate cylinder, one of the angles at P_0 specified by s_1' and s_2' is π. Let $\widehat{Y}$ be the flat surface obtained by slitting open Y along s_1' and s_2'. By construction, $\widehat{Y}$ is a flat disc with $k-2$ singularities (of cone angle π) in its interior, and whose boundary is a geodesic loop c based at P_0. Note that P_0 is also a singular point of $\widehat{Y}$.

Let c denote the boundary of $\widehat{Y}$, and ℓ be the length of c. Fix an $\epsilon > 0$. For any $t \in (0, \epsilon)$, let $\widehat{C}_t$ be the half cylinder of circumference ℓ and width t. We can glue $\widehat{C}_t$ to $\widehat{Y}$ such that the marked point in the boundary of $\widehat{C}_t$ is identified with P_0. Let Y_t' denote the resulting flat surface. Observe that Y_t' corresponds to a meromorphic differential η_t' on $\mathbb{CP}^1$ which has a unique zero at P_0 and the same number of simple poles as η. It follows that the orienting double cover of $(\mathbb{CP}^1, \eta_t')$ is an abelian differential (X_t, ω_t) in the same stratum as (X, ω). We also remark that the double cover of $\widehat{C}_t$ is a simple cylinder of width to t. We define (X_0, ω_0) to be (X, ω). It is now straightforward to check that the family $\{(X_t, \omega_t) \mid t \in [0, \epsilon)\}$ satisfies the properties in the statement of the lemma. □

As a byproduct of Lemma 3.4, we also have the following:

Lemma 3.5 *Let $s := s_1 \cup s_2$ be a degenerate horizontal cylinder in the surface $(X, \omega) \in \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$.*

(i) *If $(X,\omega) \in \mathcal{H}(2)$, then there exist a pair of homologous saddle connections $r^{\pm}$ that cut out a slit torus containing s satisfying the following condition: any vertical leaf crossing $r^{\pm}$ must intersect s.*

(ii) *If $(X,\omega) \in \mathcal{H}(1,1)$, then either*
  (a) *there exist a pair of homologous saddle connections $r^{\pm}$ that cut out a slit torus containing s such that any vertical leaf crossing $r^{\pm}$ must intersect s, or*
  (b) *there are two simple cylinders C_1, C_2 disjoint from s such that any vertical leaf crossing C_1 or C_2 must intersect s.*

Proof Let us use the same notation as in the proof of Lemma 3.4. Recall that by slitting open Y along the projections of s_1 and s_2, we obtain a flat surface $\widehat{Y}$ whose boundary is a geodesic loop c based at P_0. One can construct a new flat surface homeomorphic to the sphere $\mathbb{CP}^1$ by "sewing up" c. This operation produces an extra singular point of angle π at the midpoint of c.

Let Y' denote the resulting surface. On Y', we have $k-1$ singularities of cone angles π and a singularity at P_0 of cone angle 2π if $\omega \in \mathcal{H}(2)$, or 3π if $\omega \in \mathcal{H}(1,1)$. The loop c corresponds to a segment c' on Y' joining P_0 to a singularity of angle π. Let (X', ω') be the orienting double cover of Y'. Then either $(X',\omega') \in \mathcal{H}(0,0)$ or $(X',\omega') \in \mathcal{H}(2)$. In both cases, c' gives rise to a saddle connection s' invariant by the hyperelliptic involution of X'. Note that by construction, we can identify $X' \setminus s'$ with $X \setminus s$.

Let $\varphi\colon \boldsymbol{P} \to X'$ be the embedded parallelogram associated to s' introduced in Lemma 2.1. By construction, φ maps the sides of $\boldsymbol{P}$ to saddle connections on X' which do not intersect s' in their interior. Thus those saddle connections correspond to some saddle connections on X. It follows that $\varphi(\boldsymbol{P}) \subset X'$ corresponds to a subsurface of X containing s. The conclusions of the lemma then follow from a careful inspection on the boundary of $\varphi(\boldsymbol{P})$. □

3.3 The cylinder graph

We now define a new subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ of $\mathcal{C}(S)$ as follows: the vertices of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ are free homotopy classes of core curves of cylinders, or free homotopy classes of degenerate cylinders in X. Elements of $\widehat{\mathcal{C}}^{(1)}_{\mathrm{cyl}}(X,\omega,f)$ are the edges of $\mathcal{C}(S)$ both of whose endpoints are in $\widehat{\mathcal{C}}^{(0)}_{\mathrm{cyl}}(X,\omega,f)$.

Let $\boldsymbol{d}^{\mathcal{C}}$ denote the distance in $\mathcal{C}(S)$. Recall that by definition each edge of $\mathcal{C}(S)$ has length equal to one. Let a, b be two simple closed curves on S, and $[a]$, $[b]$ be their free homotopy classes, considered as vertices of $\mathcal{C}(S)$. We have

$$\boldsymbol{d}^{\mathcal{C}}([a],[b]) = \min\bigl\{\mathrm{leng}(\gamma) \mid \gamma \text{ is a path in } \mathcal{C}(S) \text{ from } [a] \text{ to } [b]\bigr\}.$$

We define a distance $\boldsymbol{d}$ in $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ in the same manner, that is, every edge has length equal to one, and given $[a],[b] \in \widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$,

$$\boldsymbol{d}([a],[b]) = \min\bigl\{\mathrm{leng}(\gamma) \mid \gamma \text{ is a path in } \widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f) \text{ from } [a] \text{ to } [b]\bigr\}.$$

By convention, if there are no paths in $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ from $[a]$ to $[b]$, then we define $\boldsymbol{d}([a],[b]) = \infty$. The subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$, called the *cylinder graph*, will be the main subject of our investigation in the remainder of this paper. To lighten notation, when (X,ω) and a marking mapping $f\colon S \to X$ are fixed, we will write $\mathcal{C}_{\mathrm{cyl}}$ and $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ instead of $\mathcal{C}_{\mathrm{cyl}}(X,\omega,f)$ and $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$.

Convention In the sequel, a "cylinder" could mean a usual geometric cylinder or a degenerate one. We will refer to usual geometric cylinders as *nondegenerate* cylinders. The term *core curve* will have the usual meaning for nondegenerate cylinder, for a degenerate one it just means the cylinder itself.

3.4 Intersection numbers

Let $\iota(\cdot,\cdot)$ denote the geometric intersection form on the set of free homotopy classes of simple closed curves on S. Let a, b be two simple closed curves in S, and $[a]$, $[b]$ their free homotopy classes, respectively. Recall that $[a]$ and $[b]$ are connected by an edge in $\mathcal{C}(S)$ if and only if $\iota([a],[b]) = 0$.

Assume now that a and b are simple closed geodesics in (X,ω). If a and b are parallel, then they do not have intersection, hence $\iota([a],[b]) = 0$. If they are not parallel, then they intersect transversally at every intersection point. By using the bigon criterion (see [14, Section 1.2.4]), it is not difficult to show that $\iota([a],[b]) = \#\{a \cap b\}$. However, if a or b is a degenerate cylinder then we must be a little more careful since in this case a or b may be not a simple curve (ie in $\mathcal{H}(2)$), and their intersections are not always transversal.

To deal with this complication, if a and b are core curves of two cylinders in X (possibly degenerate), we will fix some parametrizations $\alpha\colon \mathbb{S}^1 \to X$ for a, and $\beta\colon \mathbb{S}^1 \to X$ for b such that α and β are local homeomorphisms onto their images, and the restriction of α (resp. of β) to $\mathbb{S}^1 \setminus \alpha^{-1}(\{\text{singularities of } X\})$ (resp. to $\mathbb{S}^1 \setminus \beta^{-1}(\{\text{singularities of } X\})$) is one-to-one.

By an *intersection* of a and b, we will mean a pair $(t,t') \in \mathbb{S}^1 \times \mathbb{S}^1$ such that $\alpha(t) = \beta(t')$. This intersection is said to be *transversal* if there exist $\epsilon, \epsilon' > 0$ such that $a_1 := \alpha((t-\epsilon, t+\epsilon))$ and $b_1 := \beta((t'-\epsilon', t'+\epsilon'))$ are two simple arcs in X, a_1 intersects b_1 transversally at $p = \alpha(t) = \beta(t')$, and a_1 and b_1 have no other intersections. We denote by $a \cap b$ the set of intersections of a and b, and by $a \mathbin{\widehat{\cap}} b$ the subset of transversal intersections.

Lemma 3.6 *Let C and D be two cylinders on (X,ω) (both possibly degenerate) that are not parallel. Let c and d be respectively a core curve of C and a core curve of D. We denote by $[c]$ and $[d]$ the free homotopy classes of c and d, respectively. Let $c \mathbin{\widehat{\cap}} d$ denote the set of transversal intersections of c and d. Then we have*

$$\iota([c],[d]) = \#\{c \mathbin{\widehat{\cap}} d\}.$$

Since a nontransversal intersection of c and d can only occur at a singularity, it follows in particular that $\iota([c],[d]) = \#\{c \cap d\}$ if one of c and d is a regular geodesic.

Proof Let $\pi\colon \Delta = \{z \in \mathbb{C} : |z| < 1\} \to X$ denote the universal cover of X. The pull-back $\pi^*\omega$ of ω is a holomorphic 1–form, which defines a flat metric with cone singularities on Δ.

Fix a base point x for c and a base point y for d, which are not the singularities of X. Through any point in $\pi^{-1}(\{x\})$ (resp. any point in $\pi^{-1}(\{y\})$), there is a unique lift of c (resp. a unique lift d). Since c and d are not necessarily simple curves, a priori each lift of c and d may not be a simple arc. But this actually does not happen.

Claim 3.7 (i) *Each lift of c or of d is a simple arc in Δ.*

(ii) *Two lifts of c or of d can meet at at most one point (which is a nontransversal intersection).*

(iii) *A lift of c and a lift of d can meet at at most one point.*

Proof of the claim Since the argument for the three assertions are the same, we only give the proof of (iii). Let $\tilde{c}_0$ and $\tilde{d}_0$ be a lift of c and a lift of d in Δ, respectively. Let us assume that $\tilde{c}_0$ and $\tilde{d}_0$ intersect at two points. There exists then a disc $B \subset \Delta$ bounded by a subarc $c_0 \subset \tilde{c}_0$ and a subarc $d_0 \subset \tilde{d}_0$. Let p, q be the common endpoints of c_0 and d_0, and α and β be respectively the interior angles of B at p and q. Since c_0 and d_0 are geodesic segments for the flat metric on Δ, we have $\alpha > 0$ and $\beta > 0$ ($\alpha = 0$ or $\beta = 0$ means that c and d are parallel).

Let $p_1, \dots, p_r$ be the points in ∂B that correspond to the zeros of $\pi^*\omega$ and which are different from p, q. Let θ_i be the interior angle of B at p_i. By the definition of cylinders, we have $\theta_i \geq \pi$ for all $i = 1, \dots, r$. Let $x_1, \dots, x_s$ be the zeros of $\pi^*\omega$ in

$\mathrm{int}(B)$, and $\hat{\theta}_i$ be the angles at x_i. The Gauss–Bonnet formula gives (see, for instance, [48, Proposition 1])

$$\sum_{i=1}^{s}(2\pi-\hat{\theta}_i)+\sum_{i=1}^{r}(\pi-\theta_i)+2\pi-(\alpha+\beta)=2\pi\chi(B)=2\pi.$$

Since $\alpha+\beta>0$, $\pi-\theta_i\leq 0$ and $2\pi-\hat{\theta}_i<0$, we see that the equality above cannot be realized. Therefore, B cannot exist, which means that $\tilde{c}_0$ and $\tilde{d}_0$ can only meet at at most one point. □

Since nontransversal intersections of c and d can only occur at the singularities of X (zeros of ω), we can deform c and d slightly in a neighborhood of each zero of ω to get simple closed curves c' and d' in the same free homotopy classes as c and d, respectively, such that $\#\{c\mathbin{\widehat{\cap}}d\}=\#\{c'\cap d'\}$. Claim 3.7 then implies that any lift of c' in Δ intersects a lift of d' at at most one point and all the intersections are transversal. It follows from the bigon criterion (see eg [14, Proposition 1.7]) that

$$\iota([c],[d])=\#\{c'\cap d'\}=\#\{c\mathbin{\widehat{\cap}}d\}.$$

The lemma is then proved. □

Remark 3.8

- If C and D are not parallel, we can assume that C is horizontal and D is vertical. In the case both C and D are degenerate, to compute their intersection number, one can use Lemma 3.4 to get a deformation (X_t,ω_t) of (X,ω) in which C corresponds to a simple (horizontal) cylinder C_t. In X_t, D corresponds to a vertical cylinder D_t. Consequently, c is freely homotopic to a regular horizontal geodesic c_t in X_t, while d is freely homotopic to a core curve d_t of D_t. It follows from Lemma 3.6 that $\iota([c],[d])=\iota([c_t],[d_t])=\#\{c_t\cap d_t\}$.
- It may happen that two degenerate cylinders in the same direction have a positive intersection number.

4 Reducing numbers of intersection

In what follows, given two cylinders C, D in X, by $\iota(C,D)$ we will mean the geometric intersection number $\iota([c],[d])$, where c and d are some core curves of C and D, respectively. Our first goal is to estimate the distance in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ using intersection numbers.

Theorem 4.1 *There exist two positive constants K_1, K_2 such that for any $[X,\omega,f]$ in $\Omega\mathcal{T}_2$, and any cylinders C and D in X (both possibly degenerate) considered as vertices of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$, we have*

$$\boldsymbol{d}(C,D)\leq K_1\iota(C,D)+K_2. \tag{1}$$

As a direct consequence of inequality (1), we get the following:

Corollary 4.2 *The subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ is connected.*

4.1 Reducing to simple cylinders

In what follows, we will fix a point $[X,\omega,f] \in \Omega\mathcal{T}_2$, and use the term "cylinder" to refer to both degenerate and nondegenerate cylinders. Our first step is to reduce the problem to the case where C and D are simple cylinders.

Lemma 4.3 *Let C be horizontal cylinder that does not fill X, ie $\overline{C} \neq X$, and D be a vertical cylinder. Assume that $\iota(C,D) > 0$. Then there exists a simple cylinder C' such that $\boldsymbol{d}(C,C') \leq 1$ and $\iota(C',D) \leq \iota(C,D)$.*

Proof We first consider the case that C is nondegenerate. Let c be a core curve of C and d a core curve of D. Since c is a regular simple closed geodesic, by Lemma 3.6, we have $\iota(C,D) = \#\{c \cap d\}$. Obviously, we only need to consider the case that C is not simple.

If $(X,\omega) \in \mathcal{H}(2)$, then the complement of $\overline{C}$ is a simple cylinder C' whose boundary is a pair of homologous saddle connections contained in the boundary of C. In particular, C' is also horizontal, and we have $\iota(C,C') = 0$, hence $\boldsymbol{d}(C,C') = 1$. Any time d crosses C', it must cross C before returning to C'. Therefore, we have $\iota(C',D) \leq \iota(C,D)$.

If $(X,\omega) \in \mathcal{H}(1,1)$ then the complement of $\overline{C}$ is either (a) a horizontal simple cylinder, (b) two disjoint horizontal simple cylinders, or (c) a torus with a horizontal slit. In case (a) and case (b), the boundaries of the horizontal cylinders in the complement are contained in the boundary of C. Therefore, it suffices to choose one of them to be C'. In case (c), let (X',ω',s') be the slit torus which is the complement of $\overline{C}$. Note that the slit s' corresponds to a pair of homologous saddle connections in the boundary of C. By Lemma 2.3 we know that X' contains a simple cylinder C' disjoint from the slit s' such that any vertical line crossing C' must cross s'. Since C' is disjoint from C we have $\boldsymbol{d}(C,C') = 1$. Any time d crosses C', it must cross the slit s' and hence C. Therefore, we also have $\iota(C',D) \leq \iota(C,D)$.

We now turn to the case that C is degenerate. If $(X,\omega) \in \mathcal{H}(2)$, from Lemma 3.5, we know that C is contained in a slit torus cut out by a pair of homologous saddle connections $r^{\pm}$ such that every vertical leaf crossing $r^{\pm}$ intersects C. Since $(X,\omega) \in \mathcal{H}(2)$, the complement of the slit torus is a simple cylinder C' bounded by $r^{\pm}$. Clearly, we have $\boldsymbol{d}(C,C') = 1$. If the core curves of D are regular geodesics (that is, D is nondegenerate), then we can immediately conclude that $\iota(C',D) \leq \iota(C,D)$. When D is degenerate, we consider the deformations $\{(X_t,\omega_t) \mid t \in [0,\epsilon)\}$ of (X,ω) given by

Lemma 3.4. For $t \in (0, \epsilon)$, in (X_t, ω_t), D becomes a simple cylinder D_t, while the cylinders C and C' persist and have the same properties. Since $\iota(C', D) = \iota(C', D_t)$ and $\iota(C, D) = \iota(C, D_t)$, we also get $\iota(C', D) \leq \iota(C, D)$.

The case $(X, \omega) \in \mathcal{H}(1,1)$ also follows from similar arguments. □

Lemma 4.4 *Assume that C is a horizontal cylinder that fills X, and D is a vertical cylinder. Then there exists a simple cylinder C' such that*

$$\boldsymbol{d}(C', C) = 2, \quad \iota(C', D) \leq \iota(C, D).$$

Proof Let c be a core curve of C. If $(X, \omega) \in \mathcal{H}(2)$ then the complement of C is the union of three horizontal saddle connections s_1, s_2, s_3, all invariant by the hyperelliptic involution. We remark that the union of any two of these saddle connections is a degenerate cylinder. One can easily find a transverse simple cylinder C' containing s_1, disjoint from the union $s_2 \cup s_3$, whose core curves cross c once. Furthermore, we can choose C' such that the horizontal component of its core curves has length smaller than the length of c. Clearly, we have $\boldsymbol{d}(C, C') = 2$. Since any vertical geodesic crossing C' crosses also C, we have $\iota(C', D) \leq \iota(C, D)$. Thus the lemma is proved for this case.

The case $(X, \omega) \in \mathcal{H}(1,1)$ follows from the same arguments. □

In what follows, a geodesic line on X that does not contain any singularity is called *regular*.

Lemma 4.5 *Let C be a horizontal cylinder and D be a vertical cylinder in X. If there exists a regular vertical leaf which does not cross C, then $\boldsymbol{d}(C, D) \leq 2$.*

Proof Obviously we only need to consider the case that $\iota(C, D) > 0$. Assume that there is a regular vertical closed geodesic that does not intersect C. Then there exists another vertical cylinder D' which is disjoint from both C and D. Consequently, we have $\boldsymbol{d}(C, D) = 2$.

Assume now that there is an infinite regular vertical leaf that does not intersect C. The closure of this leaf is a subsurface X' of X bounded by some vertical saddle connections. Let s be a saddle connection in the boundary of X'. Note that s and $\tau(s)$ are homologous. Thus they decompose X into two subsurfaces X_1 and X_2 both invariant by τ. Since C is invariant by τ, it must be contained in one of the subsurfaces, say X_1. Since s and $\tau(s)$ are vertical, the core curves of D cannot cross s and $\tau(s)$, which means that D is also contained in one subsurface. Since we have assumed that $\iota(C, D) > 0$, D must be contained in X_1.

The subsurface X_2 must be either a slit torus or a surface in $\mathcal{H}(2)$ with a marked saddle connection. However, the latter case does not occur because it would imply that X_1 is a vertical simple cylinder containing both C and D, which is impossible. Now, by Lemma 2.3, one can find in the torus X_2 a simple cylinder C' that does not meet the slit. Since C' corresponds to a simple cylinder of X which is disjoint from both C and D, and we have $\boldsymbol{d}(C, D) = 2$. The lemma is then proved. □

From Lemmas 4.3, 4.4, we know that if C is not simple then there exists a simple cylinder C' such that $\boldsymbol{d}(C, C') \le 2$ and $\iota(C', D) \le \iota(C, D)$. Consequently, we can find simple cylinders C', D' such that

$$\boldsymbol{d}(D, D') \le 2, \quad \boldsymbol{d}(C, C') \le 2, \quad \iota(C', D') \le \iota(C, D).$$

It follows in particular that $\boldsymbol{d}(C, D) \le \boldsymbol{d}(C', D') + 4$. Therefore, in order to prove Theorem 4.1, we only need to prove (1) for the case that C and D are simple cylinders. Moreover, by Lemma 4.5, we can further assume that all the leaves of the foliation in the direction of D intersect $\overline{C}$. Thus, Theorem 4.1 is a consequence of the following:

Proposition 4.6 *Let C and D be two simple cylinders such that all the leaves of the foliation in the direction of D intersect $\overline{C}$. Then there always exists a simple cylinder C' such that*

$$\boldsymbol{d}(C', C) \le 3, \quad \iota(C', D) < \iota(C, D). \tag{2}$$

To prove this proposition we will make use of the representation of translation surfaces as polygons in $\mathbb{R}^2$. In Appendix A, we give a uniform construction from symmetric polygons of translation surfaces in genus two satisfying the hypothesis of Proposition 4.6.

4.2 Proof of Proposition 4.6, case $\mathcal{H}(2)$

By using $\mathrm{GL}^+(2, \mathbb{R})$, we can assume that C is a horizontal cylinder, and D is vertical. From Proposition A.1(i), we can construct (X, ω) from a symmetric polygon $\boldsymbol{P} := (P_0 \cdots P_3 Q_0 \cdots Q_3)$ in $\mathbb{R}^2$. Note that by construction, the hyperelliptic involution of X lifts to the central symmetry fixing the midpoint of $\overline{P_0 Q_0}$.

Let X_1, X_2 and Y be respectively the vertical projections of P_1, P_2 and Q_0 on $\overline{P_0 P_3}$. Let x_1, x_2, x_3, y be respectively the lengths of $\overline{P_0 X_1}$, $\overline{P_0 X_2}$, $\overline{P_0 P_3}$, $\overline{P_0 Y}$. Clearly, we have $0 \le x_1 \le x_2 \le x_3$ and $0 \le y \le x_3$. By cutting and regluing, we see that the cases $y = 0$ ($Y \equiv P_0$) and $y = x_3$ ($Y \equiv P_3$) are equivalent. Therefore we can always suppose $0 < y \le x_3$.

By symmetry, we can assume that $|\overline{P_1 X_1}| \ge |\overline{P_2 X_2}|$; see Figure 6. Observe that the union of the projections of $(P_0 P_1 P_2)$ and $(Q_0 Q_1 Q_2)$ in X is a cylinder E which is disjoint from C. Similarly, the union of the projections of $(P_2 P_3 Q_0)$ and $(Q_2 Q_3 P_0)$

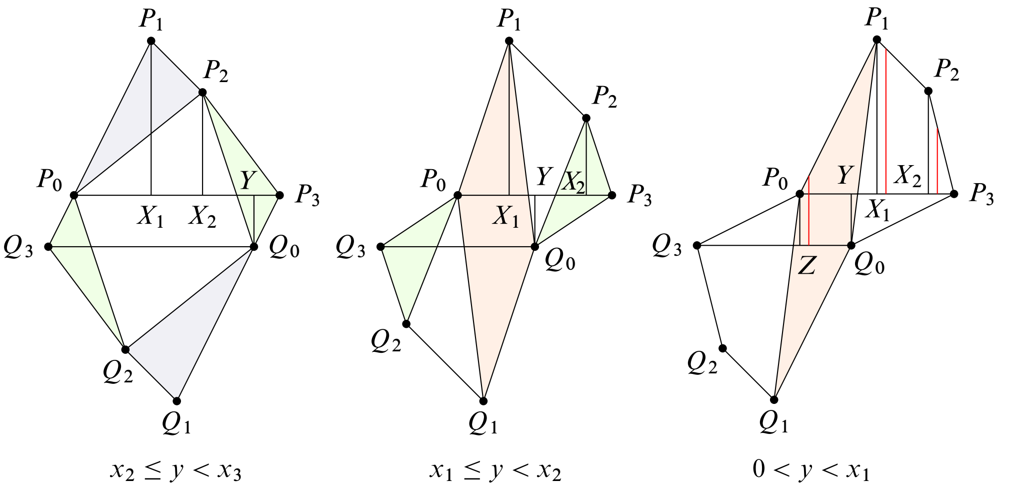


Figure 6: Finding simple cylinders having fewer intersections with D than C in the case $(X,\omega) \in \mathcal{H}(2)$. C is represented by the parallelogram $(P_0P_3Q_0Q_3)$; D is supposed to be vertical.

is also a cylinder F in X, which is disjoint from E. Observe that by assumption, E is always a simple cylinder, but F can be a degenerate one (that is, when both $\overline{P_2P_3}$ and $\overline{P_3Q_0}$ are vertical). Note that we have $\boldsymbol{d}(C,E) = 1$ and $\boldsymbol{d}(C,F) = 2$.

Let d be a core curve of D and $\widehat{d}$ be the preimage of d in $\boldsymbol{P}$. We remark that $\widehat{d}$ is a (finite) union of vertical segments with endpoints in the boundary of $\boldsymbol{P}$ and none of the vertices of $\boldsymbol{P}$ is contained in $\widehat{d}$. We first consider the generic case, where none of the sides of $\boldsymbol{P}$ is vertical. By assumption, we have

$$0 < x_1 < x_2 < x_3 \quad \text{and} \quad 0 < y < x_3.$$

We have three possibilities:

(a) $\boldsymbol{x_2 \leq y < x_3}$ We observe that if a vertical line intersects $\overline{P_0P_2}$ or $\overline{P_2Q_0}$ then it must intersect $\overline{P_0X_2}$ or $\overline{X_2Y}$, respectively. Thus, we have

$$\#\{\widehat{d} \cap \overline{P_0P_3}\} \geq \#\{\widehat{d} \cap \overline{P_0P_2}\} + \#\{\widehat{d} \cap \overline{P_2Q_0}\}.$$

It follows that at least one of the following inequalities is true:

$$\#\{\widehat{d} \cap \overline{P_0P_2}\} < \#\{\widehat{d} \cap \overline{P_0P_3}\} \quad \Rightarrow \quad \iota(E,D) < \iota(C,D),$$
$$\#\{\widehat{d} \cap \overline{P_2Q_0}\} < \#\{\widehat{d} \cap \overline{P_0P_3}\} \quad \Rightarrow \quad \iota(F,D) < \iota(C,D).$$

Therefore, in this case, we can choose C' to be either E or F.

(b) $\boldsymbol{x_1 \leq y < x_2}$ In this case, the parallelogram $(P_0P_1Q_0Q_1)$ is contained in $\boldsymbol{P}$, thus it projects to a simple cylinder G in X, which is disjoint from F. In particular, we have $\boldsymbol{d}(G,C)\le 3$. We now observe that

$$\#\{\widehat{d}\cap\overline{X_1X_2}\}=\#\{\widehat{d}\cap\overline{P_1Q_0}\}+\#\{\widehat{d}\cap\overline{P_2Q_0}\}\le\#\{\widehat{d}\cap\overline{P_0P_3}\}.$$

Therefore, at least one of the following inequalities is true $\iota(F,D)<\iota(C,D)$ or $\iota(G,D)<\iota(C,D)$. Hence we can choose C' to be either F or G.

(c) $\boldsymbol{0<y<x_1}$ We will show that in this case $\iota(G,D)<\iota(C,D)$. Let Z be the vertical projection of P_0 to $\overline{Q_0Q_3}$. We choose a core curve d of D which is contained in the ϵ–neighborhood of the left boundary of D, with $\epsilon>0$ small. The left boundary of D is a vertical saddle connection, thus it contains (the projection of) one of the following segments: $\overline{P_0Z}$, $\overline{P_1X_1}$, $\overline{P_2X_2}$. It follows that $\widehat{d}$ contains a vertical segment $\widehat{d}_0$ which is ϵ–close to one of $\overline{P_0Z}$, $\overline{P_1X_1}$, $\overline{P_2X_2}$ from the right. Observe that $\widehat{d}_0$ always intersects $\overline{P_0P_3}$, but when ϵ is chosen to be small enough, $\widehat{d}_0$ does not intersect $\overline{P_1Q_0}$. Since any vertical segment in $\boldsymbol{P}$ intersecting $\overline{P_1Q_0}$ must intersect $\overline{YX_1}\subset\overline{P_0P_3}$, it follows that $\iota(G,D)<\iota(C,D)$, and we can choose C' to be G.

It remains to show that the same arguments work in the degenerating situations, that is, when one of the sides of $\boldsymbol{P}$ is vertical. First, let us suppose that $\overline{P_2P_3}$ is vertical, ie $x_2=x_3$.

- If $y=x_3$, then F becomes a degenerate cylinder. Clearly F and D are disjoint since they are both vertical. Therefore $\boldsymbol{d}(C,D)\le\boldsymbol{d}(C,F)+1\le 3$, hence we can choose C' to be D.
- If $0<y<x_3$, then case (a) and case (b) follow from the same arguments. For case (c), we observe that the left boundary of D is not invariant by the hyperelliptic involution, and $\overline{P_2P_3}$ projects to an invariant saddle connection. Therefore $\widehat{d}_0$ is either ϵ–close to $\overline{P_0Z}$ or $\overline{P_1X_1}$. Hence we can use the same argument to conclude that $\iota(G,D)<\iota(C,D)$ and we can choose C' to be G.

Other degenerations are easy to deal with in similar manner; details are left for the reader. □

4.3 Proof of Proposition 4.6, case $\mathcal{H}(1,1)$

Using the notation in Proposition A.1(ii), we know that (X,ω) is obtained from a decagon $\boldsymbol{P}:=(P_0\cdots P_4Q_0\cdots Q_4)\subset\mathbb{R}^2$. Our arguments depend on the properties of this decagon. We have three different models for $\boldsymbol{P}$ (see Figure 7): (I) both $\mathrm{int}(\overline{P_0P_2})$ and $\mathrm{int}(\overline{P_2P_4})$ are contained in $\mathrm{int}(\boldsymbol{P})$, (II) only one of $\mathrm{int}(\overline{P_0P_2})$ and $\mathrm{int}(\overline{P_2P_4})$ is contained in $\mathrm{int}(\boldsymbol{P})$, and (III) none of $\mathrm{int}(\overline{P_0P_2})$ and $\mathrm{int}(\overline{P_2P_4})$ is contained in $\mathrm{int}(\boldsymbol{P})$.

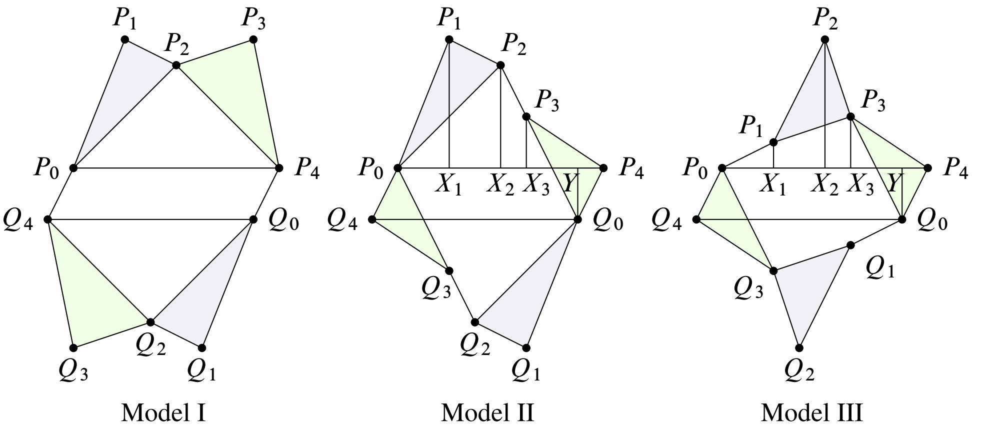


Figure 7: Finding a simple cylinder with fewer intersections with D than C in the case $(X,\omega)\in\mathcal{H}(1,1)$. C is represented by the parallelogram $(P_0P_4Q_0Q_4)$, D is supposed to be vertical.

Let X_1, X_2, X_3 and Y be respectively the vertical projections of P_1, P_2, P_3 and Q_0 on $\overline{P_0P_4}$. The lengths of $\overline{P_0X_i}$, $\overline{P_0P_4}$ and $\overline{P_0Y}$ are denoted by x_i, x_4 and y, respectively. As in the previous case, we have $0\le x_i\le x_{i+1}$, $i=1,2,3$, and $0<y\le x_4$. Let d be a core curve of D, and $\hat{d}$ its preimage in $\boldsymbol{P}$.

4.3.1 Model I In this model, the sets $(P_0P_1P_2)\cup(Q_0Q_1Q_2)$ and $(P_2P_3P_4)\cup(Q_2Q_3Q_4)$ project to two disjoint simple cylinders in X which will be denoted by E and F, respectively. Note that $\boldsymbol{d}(C,E)=\boldsymbol{d}(C,F)=1$. Clearly, we have

$$\#\{\hat{d}\cap\overline{P_0P_4}\}=\#\{\hat{d}\cap\overline{P_0P_2}\}+\#\{\hat{d}\cap\overline{P_2P_4}\}\quad\Rightarrow\quad \iota(C,D)=\iota(E,D)+\iota(F,D).$$

Therefore, we can pick C' to be E or F.

4.3.2 Model II By symmetry, we only need to consider the case that $\operatorname{int}(\overline{P_0P_2})\subset\operatorname{int}(\boldsymbol{P})$, and $\operatorname{int}(\overline{P_2P_4})\not\subset\operatorname{int}(\boldsymbol{P})$. Let E be the simple cylinder on X which is the projection of $(P_0P_1P_2)\cup(Q_0Q_1Q_2)$. Let F be the cylinder which is the projection of $(P_3P_4Q_0)\cup(Q_3Q_4P_0)$. We have $\boldsymbol{d}(C,E)=1$ and $\boldsymbol{d}(C,F)=2$.

We first consider the generic situation, that is, $0<x_i<x_{i+1}$, $i=1,2,3$, and $0<y<x_4$. Note that in this situation F is a simple cylinder. We have three cases: (a) $x_2\le y<x_4$, (b) $x_1\le y<x_2$ and (c) $0<y<x_1$. In all of these cases, one can find a simple cylinder having the desired property by the same arguments as the case that $(X,\omega)\in\mathcal{H}(2)$.

Consider now the degenerating situations: (1) $\overline{P_0P_1}$ is vertical, equivalently $x_1=0$; (2) $\overline{P_1P_2}$ is vertical, equivalently $x_1=x_2$; (3) $\overline{P_2P_3}$ is vertical, equivalently $x_2=x_3$; (4) $\overline{P_3P_4}$ is vertical, equivalently $x_3=x_4$; (5) $Y\equiv P_4$, equivalently $y=x_4$. If (4) or (5) does not occur then F is always a simple cylinder, hence the arguments above apply. If (4) and (5) hold then F is a vertical degenerate cylinder. Since F must be disjoint from D, we have $\boldsymbol{d}(C, D) \leq 3$. Therefore, we can choose C' to be D.

4.3.3 Model III In this case P_2 must be the highest point of $\boldsymbol{P}$, and $\overline{P_1 P_3}$ must be contained in $\boldsymbol{P}$. Consequently, the union $(P_1 P_2 P_3) \cup (Q_1 Q_2 Q_3)$ projects to a simple cylinder E in X. Let F denote the cylinder in X which is the projection of $(P_3 P_4 Q_0) \cup (Q_3 Q_4 P_0)$. We remark that $\boldsymbol{d}(C, E) = 1$ and $\boldsymbol{d}(C, F) = 2$. It is not difficult to see that the same arguments as the previous cases also allow us to get the desired conclusion. □

4.4 Proof of Theorem 4.1

By Lemmas 4.3 and 4.4, we know that there exist two simple cylinders C' and D' such that

$$\iota(C', D') \leq \iota(C, D) \quad \text{and} \quad \boldsymbol{d}(C, D) \leq \boldsymbol{d}(C', D') + 4.$$

It follows from Lemma 4.5 and Proposition 4.6 that $\boldsymbol{d}(C', D') \leq 3\iota(C', D') + 2$. Therefore

$$\boldsymbol{d}(C, D) \leq 3\iota(C, D) + 6. \qquad \square$$

5 Infinite diameter

In this section we prove the following proposition.

Proposition 5.1 *For any $(X, \omega) \in \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$, the diameter of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X, \omega, f)$ is infinite.*

The geometry of the curve complex is closely related to the Teichmüller space $\mathcal{T}(S)$. Recall that given a simple closed curve γ on S, for any $x \in \mathcal{T}(S)$ the extremal length $\mathrm{Ext}_x(\gamma)$ of γ is defined to be

$$\mathrm{Ext}_x(\gamma) = \sup_h |\gamma^*|_h^2,$$

where h ranges over the set of Riemannian metrics of area one in the conformal class of x, and $|\gamma^*|_h$ is the length of the shortest curve (with respect to h) in the homotopy class of γ. Alternatively, one can define $\mathrm{Ext}_x(\gamma)$ to be the inverse of the largest modulus of an annulus homotopic to γ on S. There is a natural coarse mapping Φ from $\mathcal{T}(S)$ to $\mathcal{C}(S)$ defined as follows: we assign to each $x \in \mathcal{T}(S)$ a curve of minimal x–extremal length on S. It is a well-known fact (see [33, Lemma 2.4]) that there is a universal constant c depending only the topology of S, such that the diameter of the subset of $\mathcal{C}(S)$ consisting of simple curves having minimal x–extremal length is at most c for any $x \in \mathcal{T}(S)$.

Teichmüller geodesics in $\mathcal{T}(S)$ through x are the projections of the lines $a_t \cdot q$, where q is a holomorphic quadratic differential on S equipped with the conformal structure x, and

$$a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \quad t \in \mathbb{R}.$$

It is proven in [33] that if $L_q\colon \mathbb{R} \to \mathcal{T}(S)$ is a Teichmüller geodesic, then $\Phi(L_q(\mathbb{R}))$ is an unparametrized quasigeodesic in $\mathcal{C}(S)$. It may happen that this quasigeodesic has finite diameter.

The curve graph $\mathcal{C}(S)$ has infinite diameter; see [33]. Klarreich [27] shows that the boundary at infinity $\partial_\infty \mathcal{C}(S)$ of $\mathcal{C}(S)$ can be identified with the space of topological minimal foliations $\mathcal{F}_{\min}(S)$ on S. Recall that a foliation on S is minimal if it has no leaf which is a simple closed curve, here we consider foliations up to isotopy and Whitehead moves. A characterization of sequences of curves converging to a foliation in $\partial_\infty \mathcal{C}(S)$ is given by Hamenstädt [16]. It follows from this result that if the vertical foliation of q are minimal then $\Phi \circ L_q([0,\infty))$ is a quasigeodesic of infinite diameter in $\mathcal{C}(S)$; see [17; 18].

Recall that a geometric (nondegenerate) cylinder on a translation surface is modeled by $\mathbb{R} \times (0,h)/((x,y) \sim (x+c,y))$, where $c > 0$ is its circumference and h is its width. Vorobets [50], developing Smillie's ideas in [45], showed the following:

Theorem 5.2 (Smillie and Vorobets) *Given any stratum $\mathcal{H}(\kappa)$ of translation surface, there exists a constant $K > 0$ depending on κ such that, on every translation surface of area one in $\mathcal{H}(\kappa)$, one can find a geometric cylinder of width bounded below by K.*

Proposition 5.1 follows easily from this and the results of Klarreich and Hamenstädt.

Proof of Proposition 5.1 Using the action of $\mathrm{GL}^+(2,\mathbb{R})$, we can always assume that $\mathrm{Area}(X,\omega) = 1$ and the vertical foliation of (X,ω) is minimal. Let $L\colon \mathbb{R} \to \mathcal{T}(S)$ be the Teichmüller geodesic defined by $q = \omega^2$. By the results of Klarreich and Hamenstädt, the quasigeodesic $\Phi \circ L(\mathbb{R}_{>0})$ has infinite diameter.

Denote by $\boldsymbol{d}^{\mathcal{C}}$ the distance in $\mathcal{C}(S)$, and by $\boldsymbol{d}$ the distance in $\hat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$. For any pair (α,β) in $\hat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$, we have $\boldsymbol{d}^{\mathcal{C}}(\alpha,\beta) \le \boldsymbol{d}(\alpha,\beta)$.

For each $t \in \mathbb{R}$, let $(X_t,\omega_t) := a_t \cdot (X,\omega)$. For any $R \in \mathbb{R}_{>0}$ there exist $t_1, t_2 \in (0,+\infty)$ such that $\boldsymbol{d}^{\mathcal{C}}(\Phi \circ L(t_1), \Phi \circ L(t_2)) \ge R$. Let $\alpha_i := \Phi \circ L(t_i)$. By Theorem 5.2 we know that there is a geometric cylinder C_i of width bounded below by K in (X_{t_i},ω_{t_i}). Let β_i be a core curve of C_i.

The extremal length of α_i in X_i is bounded by a universal constant $e_0(S)$; see eg [39, Lemma 2.1]. Thus by definition, the length of the shortest curve α_i^* in the homotopy class of α_i with respect to ω_{t_i} is bounded by $e_0(S)$. Since the width of C_i is at least K, we have $\#\{\alpha_i^* \cap \beta_i\} \leq e_0(S)/K$, which implies that $\iota([\alpha_i],[\beta_i]) \leq e_0(S)/K$. It is well known that the distance in $\mathcal{C}(S)$ is bounded by a linear function of the intersection number; see eg [33, Lemma 2.1] or [6, Lemma 1.1]. Thus there is a constant M depending only on S such that $\boldsymbol{d}^{\mathcal{C}}([\alpha_i],[\beta_i]) \leq M$. Therefore, we have

$$\boldsymbol{d}^{\mathcal{C}}([\beta_1],[\beta_2]) \geq \boldsymbol{d}^{\mathcal{C}}([\alpha_1],[\alpha_2]) - \boldsymbol{d}^{\mathcal{C}}([\alpha_1],[\beta_1]) - \boldsymbol{d}^{\mathcal{C}}([\alpha_2],[\beta_2]) \geq R - 2M.$$

Since $\boldsymbol{d}(C_1,C_2) = \boldsymbol{d}([\beta_1],[\beta_2]) \geq \boldsymbol{d}^{\mathcal{C}}([\beta_1],[\beta_2])$, the proposition follows. □

6 Automorphisms of the cylinder graph

Let $\mathrm{Aff}^+(X,\omega)$ denote the group of affine automorphisms of (X,ω). Recall that elements of $\mathrm{Aff}^+(X,\omega)$ are orientation-preserving homeomorphisms of X that preserve the zero set of ω, and are given by affine maps in local charts of the flat metric out side of this set; see [25; 35]. Note that the derivative of such a map (in local charts associated to the flat metric) is a constant matrix in $\mathrm{SL}(2,\mathbb{R})$. Thus we have a group homomorphism $D\colon \mathrm{Aff}^+(X,\omega) \to \mathrm{SL}(2,\mathbb{R})$ which associates to each element of $\mathrm{Aff}^+(X,\omega)$ its derivative. The image of D in $\mathrm{SL}(2,\mathbb{R})$ is called the *Veech group* of (X,ω) and usually denoted by $\mathrm{SL}(X,\omega)$. Note that the kernel of D is contained in the group $\mathrm{Aut}(X)$ of automorphisms of X, thus must be finite. The group $\mathrm{SL}(X,\omega)$ can also be viewed as the stabilizer of (X,ω) for the action of $\mathrm{SL}(2,\mathbb{R})$.

Given a point $[X,\omega,f] \in \Omega\mathcal{T}_2$, via the marking $f\colon S \to X$, one can identify $\mathrm{Aff}^+(X,\omega)$ with a subgroup of the mapping class group $\mathrm{Mod}(S)$ of S; see [35, Section 5]. An element of $\mathrm{Mod}(S)$ induces naturally an automorphism of the curve graph $\mathcal{C}(S)$. It is a well-known fact that every automorphism of $\mathcal{C}(S)$ arises from an element of $\mathrm{Mod}(S)$; see [24; 31]. Since an affine homeomorphism maps cylinders into cylinders, and saddle connections into saddle connections, it is clear that any element of $\mathrm{Aff}^+(X,\omega)$ induces an automorphism of the subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$. The aim of this section is to show the following.

Proposition 6.1 *Let ϕ be an element of $\mathrm{Mod}(S)$ which preserves the subgraph $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$, that is, $\phi(\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)) \subset \widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$. Then ϕ is induced by an affine automorphism in $\mathrm{Aff}^+(X,\omega)$. In particular, ϕ realizes an automorphism of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$.*

Remark 6.2 Proposition 6.1 is equivalent to the following statement: if $\psi\colon X \to X$ is a homeomorphism such that for any regular simple closed geodesic or degenerate cylinder c, $\psi(c)$ is freely homotopic to the core curves of a cylinder (possibly degenerate) on X, then ψ is isotopic to an affine automorphism of (X,ω).

The proof of this proposition essentially follows from the arguments of [13, Lemma 22]. Before getting into the proof, let us recall some basic notions of Thurston's compactification of the Teichmüller space. Let $\mathcal{MF}(S)$ denote the space of *measured foliations* on S. The space of *projective measured foliations* denoted by $\mathcal{PMF}(S)$ is naturally the quotient of $\mathcal{MF}(S)$ by $\mathbb{R}^*_+$. Thurston showed that $\mathcal{PMF}(S)$ can be identified with the boundary of $\mathcal{T}(S)$. A foliation is *minimal* if none of its leaves is a closed curve. A (measured) foliation is *uniquely ergodic* if it is minimal and there exists a unique transverse measure up to scalar multiplication.

The set of (free homotopy classes of) simple closed curves in S (that is, the vertex set of $\mathcal{C}(S)$) is naturally embedded in $\mathcal{MF}(S)$ with the transverse measure being the counting measure of intersections. The geometric intersection number $\iota(\cdot\,,\,\cdot)$ defined on the set of pairs of simple closed curves extends to a continuous symmetric function $\iota\colon \mathcal{MF}(S)\times\mathcal{MF}(S)\to[0,+\infty)$ which satisfies $\iota(a\lambda,b\mu)=ab\iota(\lambda,\mu)$, for all $a,b\in[0,+\infty)$ and $\lambda,\mu\in\mathcal{MF}(S)$. It has been shown by Thurston that the set

$$\{(0,+\infty)\cdot\alpha \mid \alpha \text{ is a simple closed curve}\}$$

is dense in $\mathcal{MF}(S)$.

Two measured foliations are *topologically equivalent* if the corresponding topological foliations are the same up to isotopy and Whitehead moves.

Proposition 6.3 [44] *If λ is a minimal measured foliation, and $\iota(\lambda,\mu)=0$, then λ and μ are topologically equivalent.*

Measured foliations are a special case of more general objects called *geodesic currents* which were introduced by Bonahon; see [3; 4]. We refer to [13] for an introduction to this concept with more details. While the space of measure foliations is the completion of the set of *simple* closed curves, the space of geodesic currents, denoted by $\mathscr{C}(S)$, can be viewed as the completion of closed curves on S. In particular, we have a continuous extension of the intersection number function ι to $\mathscr{C}(S)\times\mathscr{C}(S)$. A characterization of measured foliations in the space of geodesic currents was given by Bonahon:

Proposition 6.4 [3, Proposition 4.8] *$\mathcal{MF}(S)$ is exactly the set of geodesic currents with zero self-intersection, that is,*

$$\mathcal{MF}(S)=\{\lambda\in\mathscr{C}(S)\mid\iota(\lambda,\lambda)=0\}.$$

We will also need the following important feature of geodesic currents, due to Bonahon:

Proposition 6.5 [4, Proposition 4] *Let α be a geodesic current with the following property: every geodesic in $\widetilde{S}$ transversely meets another geodesic in the support of α. Then the set $\beta\in\mathscr{C}(S)$ such that $\iota(\alpha,\beta)\le 1$ is compact in $\mathscr{C}(S)$.*

Note that if λ is a minimal foliation, then the corresponding geodesic current satisfies the hypothesis of Proposition 6.5.

Every holomorphic 1–form (X,ω) (or more generally every holomorphic quadratic differential) defines naturally two measured foliations on X. The leaves of these foliations are respectively vertical and horizontal geodesic lines with the transverse measures given by $|\mathrm{Re}\,\omega|$ and $|\mathrm{Im}\,\omega|$. It is also a well-known fact that, if λ and μ are two uniquely ergodic measured foliations jointly filling up S, that is, for any $\nu \in \mathcal{MF}(S)$, we have $\iota(\nu,\lambda)+\iota(\nu,\mu) > 0$, then there is a unique Teichmüller geodesic g that joins $[\lambda]$ and $[\mu]$, where $[\lambda]$ and $[\mu]$ are the projections of λ and μ in $\mathcal{PMF}(S)$. As a consequence, assume that (X_1,ω_1) and (X_2,ω_2) are two holomorphic 1–forms that both satisfy the following condition: the vertical foliation of ω_i is topologically equivalent to λ, and the horizontal foliation is topologically equivalent to μ. Then there exists a diagonal matrix

$$A = \begin{pmatrix} \mathrm{e}^t & 0 \\ 0 & \mathrm{e}^s \end{pmatrix} \in \mathrm{GL}^+(2,\mathbb{R})$$

such that $(X_2,\omega_2) = A\cdot(X_1,\omega_1)$.

Proof of Proposition 6.1 By definition, $\phi\cdot[X,\omega,f]=[X,\omega,f\circ\phi^{-1}]$. Equivalently, we can write $\phi\cdot[X,\omega,f]=[X',\omega',f']$, where $f'\colon S\to X'$ satisfies the following condition: there exists an isomorphism $\hat{\phi}\colon X'\to X$ such that $\hat{\phi}^*\omega=\omega'$, and $f\circ\phi^{-1}$ is isotopic to $\hat{\phi}\circ f'$. Using this identification, we have

$$\hat{\mathcal{C}}_{\mathrm{cyl}}(X',\omega',f') = \phi(\hat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)).$$

Thus, by assumption, we have $\hat{\mathcal{C}}_{\mathrm{cyl}}(X',\omega',f') \subset \hat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$.

Via the maps $f\colon S\to X$, $f'\colon S\to X'$, for any direction $\theta\in\mathbb{RP}^1$, we denote by ν^θ and ν'^θ the measured foliations on S corresponding to the vertical foliations defined by $\mathrm{e}^{i\theta}\omega$ and $\mathrm{e}^{i\theta}\omega'$, respectively. The leaves of ν^θ and ν'^θ are geodesic lines in the direction of $\pm(\pi/2-\theta)$. Observe that if $\{\theta_k\}$ is a sequence of angles converging to θ, then ν^{θ_k} converges to ν^θ, and ν'^{θ_k} converges to ν'^θ in $\mathcal{MF}(S)$.

It follows from a classical result of Kerckhoff, Masur and Smillie [26] that for almost all directions $\theta\in\mathbb{RP}^1$, ν^θ and ν'^θ are uniquely ergodic. Set

$$\mathcal{UE}(\omega) := \{[\nu^\theta]\in\mathcal{PMF}(S) \mid \nu^\theta \text{ is uniquely ergodic, } \theta\in\mathbb{RP}^1\} \subset \mathcal{PMF}(S).$$

We define $\mathcal{UE}(\omega')$ in the same manner.

We will show that $\mathcal{UE}(\omega')\subset\mathcal{UE}(\omega)$. Let θ be a direction such that ν'^θ is uniquely ergodic. Without loss of generality, we can assume that $\mathrm{Area}(X)=1$. For any $t\in\mathbb{R}$, set

$$(X_t'^{\theta}, \omega_t'^{\theta}) := \begin{pmatrix} \mathrm{e}^t & 0 \\ 0 & \mathrm{e}^{-t} \end{pmatrix} \cdot (X', \mathrm{e}^{i\theta}\omega').$$

It follows from Theorem 5.2 that there exists a constant $R > 0$ such that for any $t \in \mathbb{R}$, $X_t'^{\theta}$ has a cylinder C_t' with circumference bounded by R. Let c_t' be a core curve of C_t', and consider the sequence $\{c_k'\}_{k\in\mathbb{N}}$. By definition, the length of c_k' with respect to $\omega_k'^{\theta}$, denoted by $\ell_{\omega_k'^{\theta}}(c_k')$, is bounded by R. Thus we have

$$\iota(\mathrm{e}^k \nu'^{\theta}, c_k') = \mathrm{e}^k \iota(\nu'^{\theta}, c_k') \le \ell_{\omega_k'^{\theta}}(c_k') \le R.$$

It follows that

$$\lim_{k\to+\infty} \iota(\nu'^{\theta}, c_k') = 0.$$

By Proposition 6.5, up to extracting a subsequence, we can assume that $\{c_k'\}$ converges to a geodesic current $\mu' \in \mathscr{C}(S)$. Since c_k' has zero self-intersection, it follows that $\iota(\mu', \mu') = 0$, hence $\mu' \in \mathcal{MF}(S)$ by Proposition 6.4. By continuity of ι, we have $\iota(\nu'^{\theta}, \mu') = 0$. Since ν'^{θ} is uniquely ergodic (so, in particular, it is minimal), it follows from Proposition 6.3 that μ' and ν'^{θ} are topologically equivalent. Hence μ' is also uniquely ergodic.

By definition, $\{c_k'\}_{k\in\mathbb{N}}$ are vertices of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X', \omega', f')$. By assumption, we have $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X', \omega', f') \subset \widehat{\mathcal{C}}_{\mathrm{cyl}}(X, \omega, f)$. Therefore, $\{c_k'\}_{k\in\mathbb{N}}$ are also vertices of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X, \omega, f)$, which means that c_k' is freely homotopic to either a simple closed geodesic or a degenerate cylinder in X. In particular, we see that each c_k' has a well-defined direction $\theta_k \in \mathbb{RP}^1$ with respect to ω. Again, by extracting a subsequence, we can assume that $\{\theta_k\}$ converges to $\hat{\theta}$. Thus, $\{\nu^{\theta_k}\}$ converges to $\nu^{\hat{\theta}}$. Since we have $\iota(\nu^{\theta_k}, c_k') = 0$, by continuity, it follows that $\iota(\nu^{\hat{\theta}}, \mu') = 0$. Since μ' is uniquely ergodic, so is $\nu^{\hat{\theta}}$, and we have $[\nu'^{\theta}] = [\mu'] = [\nu^{\hat{\theta}}] \in \mathcal{PMF}(S)$. We can then conclude that $\mathcal{UE}(\omega') \subset \mathcal{UE}(\omega)$.

Now pick a pair of projective uniquely ergodic measured foliations $([\lambda], [\mu]) \in \mathcal{UE}(\omega') \subset \mathcal{UE}(\omega)$ that jointly fill up S. There exist two matrices M and M' such that the vertical and horizontal foliations of $M \cdot [X, \omega, f]$ and $M' \cdot [X', \omega', f']$ are topologically equivalent to λ and μ, respectively. Since there is a unique Teichmüller geodesic joining $[\lambda]$ and $[\mu]$, there exists a diagonal matrix $A \in \mathrm{GL}^+(2, \mathbb{R})$ such that $M' \cdot [X', \omega', f'] = AM \cdot [X, \omega, f]$, implying that ϕ is represented by an affine automorphism of (X, ω). $\square$

Remark 6.6 This proof actually works for translation surfaces in any genus with $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ replaced by the subgraph consisting of nondegenerate cylinders.

7 Hyperbolicity

A translation surface (X,ω) is said to be *completely periodic* (in the sense of Calta) if the direction of any nondegenerate cylinder in X is periodic, which means that whenever we find a simple closed geodesic on X, the surface decomposes as union of (finitely many) cylinders in the same direction; see [10; 11]. It follows from [10] and [38] that, in $\mathcal{H}(2)$, a surface is completely periodic if and only if it is a Veech surface. In $\mathcal{H}(1,1)$, a surface is completely periodic if and only if it is *an eigenform for a real multiplication of a quadratic order*. In particular, there are completely periodic surfaces in $\mathcal{H}(1,1)$ that are not Veech surfaces.

Let us denote by $\mathcal{E}_D$, where D is a natural number such that $D \equiv 0$ or $1 \mod 4$, the locus of eigenforms for the real multiplication by the quadratic order $\mathcal{O}_D$ in $\Omega\mathcal{M}_2$. Each $\mathcal{E}_D$ is a 3–dimensional irreducible (algebraic) subvariety of $\Omega\mathcal{M}_2$ which is invariant by the $\mathrm{SL}(2,\mathbb{R})$–action. The set of eigenforms in $\Omega\mathcal{M}_2$ is then (see [38])

$$\mathcal{E} = \bigcup_{D\equiv 0,1 \bmod 4} \mathcal{E}_D.$$

Even though complete periodicity is initially defined for directions of nondegenerate cylinders, it is not difficult to show that in the case of genus two, this property actually implies the periodicity for directions of degenerate cylinders; see Lemma B.1. Alternatively, one can also use the argument in [51] to get the same result in more general contexts; see [52]. In what follows, by a *completely periodic* surface we will mean a surface for which the direction of any cylinder (degenerate or not) is periodic. By Lemma B.1, this apparently new definition agrees with the usual one by Calta. Our goal in this section is to show the following theorem:

Theorem 7.1 *If $(X,\omega) \in \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$ is completely periodic then $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ is Gromov hyperbolic.*

To prove this, we will use Masur and Schleimer's hyperbolicity criterion (see also [7, Proposition 3.1] and [15]), and follow Bowditch's approach in [6].

Theorem 7.2 (Masur and Schleimer [34, Theorem 3.13]) *Suppose that $\mathcal{X}$ is a graph with all edge lengths equal to one. Then $\mathcal{X}$ is Gromov hyperbolic if there is a constant $M \geq 0$ such that for all unordered pairs of vertices x,y in $\mathcal{X}^0$, there is a connected subgraph $g_{x,y}$ containing x and y with the following properties:*

- (local) *If $d_{\mathcal{X}(x,y)} \leq 1$ then $g_{x,y}$ has diameter at most M.*
- (slim triangle) *For any $x,y,z \in \mathcal{X}^0$, the subgraph $g_{x,y}$ is contained in the M–neighborhood of $g_{x,z} \cup g_{z,y}$.*

Let us fix $[X,\omega,f]\in\Omega\mathcal{T}_2$, where $(X,\omega)\in\mathcal{E}$ and $\mathrm{Area}(X,\omega)=1$. We will write $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ instead of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$. We know from Corollary 4.2 that $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is connected, and by definition the edges of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ have length equal to one. Let K be the constant in Theorem 5.2, and C be a cylinder of width bounded below by K in X. Note that the circumference of C is bounded above by $1/K$. Recall that from Theorem 4.1, we know that there are two constants K_1, K_2 such that for any pair of cylinders C, D in X, we have

$$\boldsymbol{d}(C,D)\le K_1\iota(C,D)+K_2,$$

where $\boldsymbol{d}$ is the distance in $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$, and $\iota(C,D)$ is the number of intersections of a core curve of C and a core curve of D.

7.1 Construction of subgraphs connecting pairs of vertices

We will now construct for each unordered pair of cylinders C, D a subgraph $\widehat{\mathcal{L}}_{C,D}$ of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ that satisfies the conditions of Theorem 7.2 with a constant M which will be derived along the way.

Let us first consider the case that C and D are parallel. If C or D is nondegenerate then $\iota(C,D)=0$ hence $\boldsymbol{d}(C,D)=1$, which means that C and D are connected by an edge in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$. We define $\widehat{\mathcal{L}}_{C,D}$ to be this edge. If both C and D are degenerate then it may happen that $\iota(C,D)>0$. Since (X,ω) is completely periodic, there is a nondegenerate cylinder E parallel to C and D. Since $\iota(C,E)=\iota(D,E)=0$, there are in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ two edges connecting E to C and to D. In this case, we define $\widehat{\mathcal{L}}_{C,D}$ to be the union of these two edges.

Assume from now on that C and D are not parallel. By applying an appropriate element of $\mathrm{SL}(2,\mathbb{R})$, we can assume that C is horizontal, D is vertical, and C and D have the same circumference. For any $t\in\mathbb{R}$, set

$$a_t=\begin{pmatrix}\mathrm{e}^t & 0\\ 0 & \mathrm{e}^{-t}\end{pmatrix}\quad\text{and}\quad(X_t,\omega_t)=a_t\cdot(X,\omega).$$

For any saddle connection s in (X,ω), we will denote by $\ell_t(s)$ its Euclidean length in (X_t,ω_t). If E is a cylinder in (X,ω), then $c_t(E)$ and $w_t(E)$ are respectively its circumference and width in (X_t,ω_t).

For any $R\in\mathbb{R}_{>0}$, let $\mathcal{L}^*_{C,D}(t,R)$ denote the set of cylinders (possibly degenerate) of circumference bounded above by R in (X_t,ω_t). Note that this set is finite. Let us choose a constant L_1 such that

$$L_1>\max\{1/K,9\},\tag{3}$$

and define

$$\mathcal{L}^*_{C,D}(L_1) = \bigcup_{t\in\mathbb{R}} \mathcal{L}^*_{C,D}(t, L_1).$$

We regard $\mathcal{L}^*_{C,D}(t, R)$ and $\mathcal{L}^*_{C,D}(L_1)$ as subsets of $\widehat{\mathcal{C}}^{(0)}_{\mathrm{cyl}}$. Observe that $\mathcal{L}^*_{C,D}(t, L_1)$ contains C when t tends to $-\infty$, and contains D when t tends to $+\infty$; therefore $\mathcal{L}^*_{C,D}$ contains C and D.

For each $t \in \mathbb{R}$, consider now the set $\mathcal{L}^*_{C,D}(t, 2L_1)$. From Theorem 5.2, $\mathcal{L}^*_{C,D}(t, 2L_1)$ contains a vertex corresponding to a cylinder $C_{0,t}$ of width bounded below by K. Set

$$M_1 := \max\{2(2K_1L_1/K + K_2), 2\}. \tag{4}$$

Then we have the following lemma:

Lemma 7.3 *As subset of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$, $\mathcal{L}^*_{C,D}(t, 2L_1)$ has diameter bounded by M_1.*

Proof Let E be a cylinder in $\mathcal{L}^*_{C,D}(t, 2L_1)$. If $\iota(E, C_{0,t}) = 0$, then we have $\boldsymbol{d}(C_{0,t}, E) = 1$. Otherwise we have $K\iota(E, C_{0,t}) \le \ell_t(E) \le 2L_1$. Hence, from (1) we get

$$\boldsymbol{d}(C_{0,t}, E) \le 2K_1L_1/K + K_2,$$

and the lemma follows. □

Moreover, we have the following lemma as well:

Lemma 7.4 *Assume that the surface (X, ω) admits cylinder decompositions in both vertical and horizontal directions. Then there exists a constant $T > 0$ such that the following hold:*

- *If $t > T$, then $\mathcal{L}^*_{C,D}(t, 2L_1)$ only contains the vertical cylinders in (X, ω) and $\mathcal{L}^*_{C,D}(t, 2L_1)$ has diameter at most 2.*
- *If $t < -T$, then $\mathcal{L}^*_{C,D}(t, 2L_1)$ only contains the horizontal cylinders in (X, ω) and $\mathcal{L}^*_{C,D}(t, 2L_1)$ has diameter at most 2.*

Proof We only give the proof of the first assertion as the second one follows from the same argument. By assumption, X decomposes into the union of some nondegenerate vertical cylinders $D_1, \dots, D_k$. Let $w_t(D_i)$ denote the width of D_i in (X_t, ω_t). Let $w_t = \min\{w_t(D_i) \mid i = 1, \dots, k\}$. A nonvertical cylinder must cross one of D_i, thus its circumference is bounded below by w_t in (X_t, ω_t). Since we have $w_t = \mathrm{e}^t w_0$; if t is large enough, any nonvertical cylinder has circumference at least $2L_1$ in (X_t, ω_t). Hence $\mathcal{L}^*_{C,D}(t, 2L_1)$ only contains the vertical cylinders. Since any vertical cylinder is of distance one from D_1 in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$, $\mathcal{L}^*_{C,D}(t, 2L_1)$ has diameter at most two. □

Lemma 7.5 *If $t-\log(2) \le t' \le t+\log(2)$ then $\mathcal{L}^*_{C,D}(t', R) \subset \mathcal{L}^*_{C,D}(t, 2R)$ for any $R \in \mathbb{R}_{>0}$. In particular, $C_{0,t'} \in \mathcal{L}^*_{C,D}(t, 2L_1)$.*

Proof Let s be a saddle connection or a regular geodesic in $(X_{t'}, \omega_{t'})$. Let $x+iy$ be the period of s in $(X_{t'}, \omega_{t'})$. Note that $(X_t, \omega_t) = a_{t-t'} \cdot (X_{t'}, \omega_{t'})$. Thus the period of s in (X_t, ω_t) is $(\mathrm{e}^{t-t'}x, \mathrm{e}^{t'-t}y)$. Therefore,

$$\ell_t(s) = \sqrt{\mathrm{e}^{2(t-t')}x^2 + \mathrm{e}^{2(t'-t)}y^2} \le 2\sqrt{x^2+y^2} = 2\ell_{t'}(s). \qquad \square$$

Set

$$\bar{\mathcal{L}}_{C,D}(2L_1) := \bigcup_{k\in\mathbb{Z}} \mathcal{L}^*_{C,D}(k\log(2), 2L_1) \subset \hat{\mathcal{C}}^{(0)}_{\mathrm{cyl}}.$$

It follows from Lemma 7.4 that if $n \in \mathbb{N}$ is large enough, then for any $m > n$, $\mathcal{L}^*_{C,D}(m, L_1) = \mathcal{L}^*_{C,D}(n, 2L_1)$, and $\mathcal{L}^*_{C,D}(-m, 2L_1) = \mathcal{L}^*_{C,D}(-n, 2L_1)$. Therefore, the set $\bar{\mathcal{L}}_{C,D}(2L_1)$ is actually finite. For each unordered pair (x, y) of vertices in $\bar{\mathcal{L}}_{C,D}(2L_1)$, let $\Gamma(x, y)$ be a path of minimal length in $\hat{\mathcal{C}}_{\mathrm{cyl}}$ joining x to y. Set

$$\hat{\mathcal{L}}_{C,D}(2L_1) = \bigcup_{x,y\in\bar{\mathcal{L}}_{C,D}(2L_1)} \Gamma(x, y).$$

As a direct consequence of Lemma 7.5, we get the following:

Corollary 7.6 (a) *If $x \in \mathcal{L}^*_{C,D}(t, 2L_1)$ and $y \in \mathcal{L}^*_{C,D}(t', 2L_1)$, then $\boldsymbol{d}(x, y) \le M_1(2 + |t-t'|/\log(2))$.*

(b) *The set $\mathcal{L}^*_{C,D}(L_1)$ is contained in $\bar{\mathcal{L}}_{C,D}(2L_1)$ and $\bar{\mathcal{L}}_{C,D}(2L_1)$ is contained in the M_1–neighborhood of $\mathcal{L}^*_{C,D}(L_1)$.*

(c) *For any pair of vertices $(x, y) \in \mathcal{L}^*_{C,D}(L_1) \times \mathcal{L}^*_{C,D}(L_1)$, there is a path $\Gamma(x, y)$ in $\hat{\mathcal{L}}_{C,D}(2L_1)$ from x to y of length equal to $\boldsymbol{d}(x, y)$.*

7.2 The local property for $\hat{\mathcal{L}}_{C,D}$

We will now show that the subgraphs $\hat{\mathcal{L}}_{C,D}(2L_1)$ constructed above satisfy the local property of Theorem 7.2.

Proposition 7.7 *There exists a constant M_2 such that if $(X, \omega) \in \mathcal{E}$ then for any pair of cylinders C, D in (X, ω) such that $\iota(C, D) = 0$, we have* $\operatorname{diam} \hat{\mathcal{L}}_{C,D}(2L_1) \le M_2$.

To prove this proposition, we make use of an elementary result on slit tori, Lemma B.3, and the fact that if C and D are not parallel, then there always exists a splitting of X into two subsurfaces, each of which contains one of C and D. Those auxiliary results are proved in Appendix B. The main technical difficulties arise when we have to deal with degenerate cylinders.

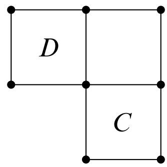


Figure 8: Disjoint simple cylinders on surfaces in $\mathcal{H}(2)$

Proof We split this proof into two cases: $(X,\omega)\in\mathcal{H}(2)$ and $(X,\omega)\in\mathcal{H}(1,1)$.

Case $(X,\omega)\in\mathcal{H}(2)$ If C and D are parallel then $\hat{\mathcal{L}}_{C,D}(2L_1)$ has diameter bounded by 2 and we have nothing to prove. Suppose from now on that C is horizontal, D is vertical, C and D have the same circumference equal to ℓ, and $\hat{\mathcal{L}}_{C,D}(2L_1)$ is the graph constructed above. Note that in this case (X,ω) is a Veech surface, thus both horizontal and vertical directions are periodic.

Case 1 One of C or D is nondegenerate. Assume that C is nondegenerate. Let c be a core curve of C and d a core curve of D. Note that c is a regular simple closed geodesic. By Lemma 3.6, the condition $\iota(C,D)=0$ implies that $c\cap d=\varnothing$. Clearly, C cannot fill X. If C is not simple then the complement of $\overline{C}$ is a horizontal simple cylinder C' whose boundary is contained in the boundary of C. Since D is disjoint from C, it must be contained in C'. But this is impossible since C' is horizontal and D is vertical. Therefore, C must be a simple cylinder.

The complement of C is then a slit torus with the slit corresponding to the boundary of C. We remark that a core curve of D must be disjoint from the interior of the slit, otherwise it would cross C entirely. Thus, we have in the slit torus an embedded square bounded by the boundary of D and the slit (which is actually the boundary of C); see Figure 8. By assumption, the length of the sides of this square is ℓ. Since this square has area less than one, we must have $\ell<1$. Therefore $C\in\mathcal{L}^*_{C,D}(t,L_1)$ for all $t\le 0$, and $D\in\mathcal{L}^*_{C,D}(t,L_1)$ for any $t\ge 0$. Hence any $E\in\bar{\mathcal{L}}_{C,D}(2L_1)$ is of distance at most M_1 from C or from D. Thus $\operatorname{diam}\hat{\mathcal{L}}_{C,D}(2L_1)\le 2M_1+1$.

Case 2 Both of C and D are degenerate. From Lemma 3.4, for any $\epsilon>0$ small enough, we can deform (X,ω) into another surface (X',ω') such that

- C corresponds to a simple horizontal cylinder C' in X' of width ϵ,
- D corresponds to a vertical cylinder in X'.

Since $\iota(C',D')=\iota(C,D)=0$, it follows from Lemma 3.6 that D' must be disjoint from C'. It follows in particular that D and D' have the same circumference ℓ. By construction C' has the same circumference as C, and

$$\operatorname{Area}(X',\omega')=\operatorname{Area}(X,\omega)+\epsilon\ell=1+\epsilon\ell.$$

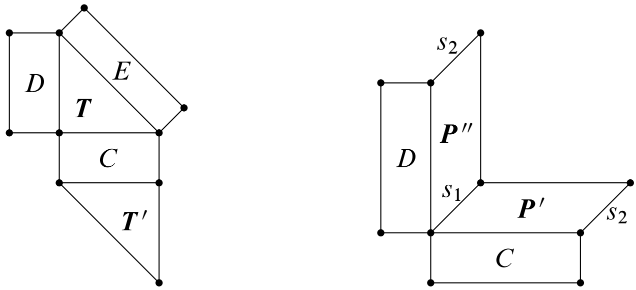


Figure 9: Disjoint cylinders on surfaces in $\mathcal{H}(1,1)$: one of C and D is simple

Applying the same arguments as above to (X',ω'), we see that X' contains an embedded square of size ℓ disjoint from C'. Therefore we have $\ell^2 < 1+\epsilon\ell$. Since ϵ can be chosen arbitrarily, we derive that $\ell \le 1$. We can then conclude by the same arguments as the previous case.

Case $(X,\omega)\in\mathcal{H}(1,1)$ Again, we only have to consider the case that C and D are not parallel. Thus we can assume that C is horizontal and D is vertical. We first choose a positive real number $L>\sqrt{2}$ such that

$$L_1 \ge 3f(\sqrt{2}L), \tag{5}$$

where $f(x)=\sqrt{x^2+1/x^2}$; see Lemma B.3.

Case 1 One of C and D is a simple cylinder. By Lemma B.2, we need to consider two cases (see Figure 9):

(i) There is a simple cylinder E disjoint from $C\cup D$ and the complement of $C\cup D\cup E$ is the union of two triangles $\boldsymbol{T}$, $\boldsymbol{T}'$; see Figure 9 (left). Since we have

$$\operatorname{Area}(\boldsymbol{T})+\operatorname{Area}(\boldsymbol{T}')=\ell^2<\operatorname{Area}(X,\omega)=1,$$

it follows that $\ell<1$. Hence we can use the same argument as in the case $(X,\omega)\in\mathcal{H}(2)$ to conclude that $\operatorname{diam}\hat{\mathcal{L}}_{C,D}(2L_1)\le 2M_1+1$.

(ii) There is a pair of homologous saddle connections s_1,s_2 that decompose X into a connected sum of two slit tori, (X',ω',s') containing C and (X'',ω'',s'') containing D; see Figure 9 (right).

By construction, the complement of C in X' is an embedded parallelogram $\boldsymbol{P}'$ bounded by s_1,s_2 and the boundary of C. Similarly, the complement of D in X'' is also an embedded parallelogram $\boldsymbol{P}''$ bounded by s_1,s_2 and the boundary of D. If $\ell\le 1$ then we can conclude using the argument above. Suppose that we have $\ell\ge 1$.

Let $\omega(s_i)=x+iy$. Since we have $\mathrm{Area}(\boldsymbol{P}')=|y|\ell$, and $\mathrm{Area}(\boldsymbol{P}'')=|x|\ell$, it follows that
$$\max\{|x|,|y|\}\leq 1/\ell\leq 1 \quad\text{and}\quad |s_i|=\sqrt{x^2+y^2}\leq\sqrt{2}/\ell\leq\sqrt{2}.$$
Set $A_1=\mathrm{Area}(X',\omega')$, $A_2=\mathrm{Area}(X'',\omega'')$; we have $A_1+A_2=1$. Without loss of generality, let us suppose that $A_1\geq\frac{1}{2}$. For any $t\in\mathbb{R}$, the period of s_i in (X_t,ω_t) is $(\mathrm{e}^t x,\mathrm{e}^{-t}y)$. Let (X'_t,ω'_t,s'_t) be the slit torus corresponding to (X',ω',s') in (X_t,ω_t). Recall that we have chosen $L>\sqrt{2}$ and L_1 satisfies (5). Let us choose a positive real number $L'\geq 1$ such that
$$L\geq\sqrt{L'^2+1}.$$

- For $0\leq t\leq\log(\ell L')$, we have $e^t|x|\leq L'$ and $\mathrm{e}^{-t}|y|\leq|y|\leq 1$, thus
$$\ell_t(s_1)\leq\sqrt{L'^2+1}\leq L.$$
Rescaling (X'_t,ω'_t,s'_t) by $1/\sqrt{A_1}$, we get a torus of area one with a slit of length bounded by $\sqrt{2}L$. Using Lemma B.3, we see that there exists in $(1/\sqrt{A_1})\cdot X'_t$ a cylinder E'_t disjoint from the slit of circumference bounded by L_1. Note that in X'_t, the circumference of E'_t is at most $\sqrt{A_1}L_1\leq L_1$. We have $\boldsymbol{d}(D,E'_t)=1$ and $E'_t\in\mathcal{L}^*_{C,D}(t,2L_1)$. Thus for any $E\in\mathcal{L}^*_{C,D}(t,2L_1)$ we have $\boldsymbol{d}(D,E)\leq M_1+1$.
- For $-\log(\ell L')\leq t\leq 0$, we have $\mathrm{e}^t|x|\leq|x|\leq 1$ and $\mathrm{e}^{-t}|y|\leq L'$, thus
$$\ell_t(s_i)\leq\sqrt{L'^2+1}\leq L.$$
The same argument as the previous case then shows that $\boldsymbol{d}(D,E)\leq M_1+1$, for any $E\in\mathcal{L}^*_{C,D}(t,2L_1)$.
- For $t\geq\log(\ell L')$, we have $\ell_t(D)=\mathrm{e}^{-t}\ell\leq 1/L'\leq 1\leq 2L_1$. Thus D is in $\mathcal{L}^*_{C,D}(t,2L_1)$, which implies that $\boldsymbol{d}(D,E)\leq M_1$ for any $E\in\mathcal{L}^*_{C,D}(t,2L_1)$.
- For $t\leq-\log(\ell L')$, we have $\ell_t(C)\leq 1/L'\leq 2L_1$, so for any $E\in\mathcal{L}^*_{C,D}(t,2L_1)$, $\boldsymbol{d}(C,E)\leq M_1$, which implies that $\boldsymbol{d}(D,E)\leq M_1+1$.

We can then conclude that for any $t\in\mathbb{R}$, and any $E\in\mathcal{L}^*_{C,D}(t,2L_1)$, we have $\boldsymbol{d}(D,E)\leq M_1+1$. Hence $\operatorname{diam}\hat{\mathcal{L}}_{C,D}\leq 2(M_1+1)$.

Case 2 One of C,D is nondegenerate and not simple. Without loss of generality, we can assume that C is neither simple nor degenerate. Lemma 3.6 implies that D is disjoint from C. Since C is not simple, the complement of $\overline{C}$ is either (a) empty, (b) a horizontal simple cylinder, (c) the union of two simple horizontal cylinders, or (d) another horizontal cylinder whose closure is a slit torus. Since there exists a vertical cylinder disjoint from C (namely D), only (d) can occur. In this case, there is a pair of horizontal homologous saddle connection $\{s_1,s_2\}$ contained in the boundary of C that

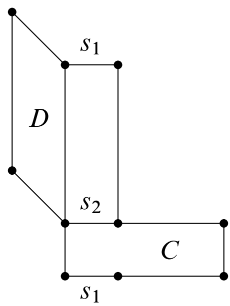


Figure 10: Disjoint cylinders on surfaces in $\mathcal{H}(1,1)$; C is neither simple nor degenerate.

decompose (X,ω) into the connected sum of two slit tori. Let (X',ω',s') be the slit torus which is the closure of C, and (X'',ω'',s'') be the other one that contains D; see Figure 10.

Let $x=|s_1|=|s_2|$. Observe that X'' contains a rectangle bounded by s_1, s_2 and the saddle connections bordering D. Therefore we have $x\ell\leq 1$, equivalently $0\leq x\leq 1/\ell$. By the same arguments as the previous case, we also get $\operatorname{diam}\widehat{\mathcal{L}}_{C,D}\leq 2(M+1)$.

Case 3 One of C and D is degenerate. Let us assume that C is degenerate. Using Lemma 3.4, we can find a family $(X_t,\omega_t), t\in[0,\epsilon)$, of surfaces in $\mathcal{H}(1,1)$ that are deformations of (X,ω), such that C corresponds to a simple horizontal cylinder C_t on X_t, for $t>0$, which has the same circumference. Note that the width of C_t is t. Therefore, $\operatorname{Area}(X_t,\omega_t)=\operatorname{Area}(X,\omega)+t\ell$.

By construction, D corresponds to a cylinder D_t on X_t which is disjoint from C_t (since we have $\iota(C_t,D_t)=\iota(C,D)=0$). By Lemma B.2 we know that either (i) (X_t,ω_t) contains two embedded triangles $\boldsymbol{T},\boldsymbol{T}'$ disjoint from C_t and D_t, or (ii) there is a splitting of (X_t,ω_t) into two slit tori (X'_t,ω'_t,s'_t) and (X''_t,ω''_t,s''_t) such that $C_t\subset X'_t$ and $D_t\subset X''_t$.

If (i) occurs, then we have $\operatorname{Area}(\boldsymbol{T})=\operatorname{Area}(\boldsymbol{T}')=\frac{1}{2}\ell^2\leq\frac{1}{2}$, which implies that $\ell\leq 1$. If (ii) occurs, then since the slits (s' and s'') are disjoint from C_t, they persist as we collapse C_t to get back (X,ω). Thus, we have the same splitting on (X,ω). In conclusion, we can use the same arguments as in Case 1 to handle this case. The proof of Proposition 7.7 is now complete. □

7.3 The slim triangle property for $\widehat{\mathcal{L}}_{C,D}$

We now prove that the subgraphs $\widehat{\mathcal{L}}_{C,D}(2L_1)$ satisfy the slim triangle property of Theorem 7.2. The idea of the proof can found in [6, Lemma 4.4]. To lighten notation, in what follows we will write $\widehat{\mathcal{L}}_{C,D}$ instead of $\widehat{\mathcal{L}}_{C,D}(2L_1)$.

Proposition 7.8 *There exists a constant M_3 such that for any triple of cylinders $\{C, D, E\}$ in (X,ω), we have that $\widehat{\mathcal{L}}_{C,D}$ is contained in the M_3–neighborhood of $\widehat{\mathcal{L}}_{C,E} \cup \widehat{\mathcal{L}}_{E,D}$ in $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$.*

Proof If C and D are parallel then $\widehat{\mathcal{L}}_{C,D}$ is contained in the 2–neighborhood of $\widehat{\mathcal{L}}_{C,E} \cup \widehat{\mathcal{L}}_{D,E}$. From now on we assume that C and D are not parallel.

By Corollary 7.6, we only need to show that $\mathcal{L}^*_{C,D}(L_1)$ is contained in the M_3–neighborhood of $\mathcal{L}^*_{C,E}(L_1) \cup \mathcal{L}^*_{E,D}(L_1)$. To define $\overline{\mathcal{L}}_{C,D}(2L_1)$ and $\widehat{\mathcal{L}}_{C,D}(2L_1)$ one needs to specify an origin for the time t by the condition that the circumferences of C and D are equal. On the other hand to define $\mathcal{L}^*_{C,D}(L_1)$, this normalization is not required. If E is parallel to C then $\mathcal{L}^*_{C,D}(L_1) = \mathcal{L}^*_{E,D}(L_1)$, and if E is parallel to D then $\mathcal{L}^*_{C,D}(L_1) = \mathcal{L}^*_{C,E}(L_1)$. In both of these cases we have nothing to prove.

Let us now assume that E is neither parallel to C nor to D. We can then renormalize (using $\mathrm{SL}(2,\mathbb{R})$) such that C is horizontal, D is vertical, and E has slope equal to 1. Recall that for any $t \in \mathbb{R}$, $(X_t,\omega_t) = a_t \cdot (X,\omega)$, $C_{0,t}$ is a cylinder of width bounded below by K in (X_t,ω_t), and the constant L_1 is chosen so that $L_1 > 1/K$; see (3).

Claim *If $t \le 0$ then $C_{0,t}$ is contained in the M_1–neighborhood of $\mathcal{L}^*_{C,E}(L_1)$.*

Proof of the claim Since (X,ω) is completely periodic, it decomposes into cylinders in both directions of C and E. Let us denote by $C = C_1,\dots,C_m$ the horizontal cylinders, and by $E = E_1,\dots,E_n$ the cylinders in the direction of E. As usual we denote by $\ell_t(C_i)$ (resp. $\ell_t(E_j)$) the circumference of C_i (resp. of E_j) in (X_t,ω_t). Let $u_i(t)$ be the width of C_i, and $v_j(t)$ be the width of E_j in (X_t,ω_t). We have

$$\begin{aligned} \ell_t(C_i) &= \mathrm{e}^t \ell(C_i), & u_i(t) &= \mathrm{e}^{-t} u_i, \\ \ell_t(E_j) &= \sqrt{\cosh(2t)}\,\ell(E_j), & v_j(t) &= \frac{v_j}{\sqrt{\cosh(2t)}}. \end{aligned}$$

Since (X,ω) has area 1, we also have

$$1 = \sum u_i \ell(C_i) = \sum v_j \ell(E_j). \tag{6}$$

Let x_j (resp. y_i) be the intersection number of a core curve of $C_{0,t}$ and a core curve of E_j (resp. of C_i). Since the circumference of $C_{0,t}$ is bounded by $1/K < L_1$, we have

$$\sum y_i u_i(t) = \mathrm{e}^{-t} \sum y_i u_i \le \ell(C_{0,t}) \le L_1 \quad \Rightarrow \quad \sum y_i u_i \le \mathrm{e}^t L_1. \tag{7}$$

Since the width of $C_{0,t}$ is bounded below by K, $x_j K \le \ell_t(E_j) = \sqrt{\cosh(2t)}\,\ell(E_j)$. Since $t \le 0$, it follows that

$$x_j \le \frac{\sqrt{\cosh(2t)}}{K} \ell(E_j) \le \frac{\mathrm{e}^{-t}}{K} \ell(E_j). \tag{8}$$

Let $(X',\omega') := U\cdot(X,\omega)$, where $U = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$. Let $\ell'(C_i)$ and u'_i (resp. $\ell'(E_j)$ and v'_j) be the circumference and the width of C_i (resp. of E_j) in (X',ω'). Note that C_i is horizontal, and E_j is vertical in (X',ω'). Thus, $\ell'(C_i) = \ell(C_i)$, $u'_i = u_i$, and $\ell'(E_j) = \ell(E_j)/\sqrt{2}$, $v'_j = \sqrt{2}v_j$.

For any $s\in\mathbb{R}$, let $(X'_s,\omega'_s) := a_s\cdot(X',\omega')$. Let $\ell'_s(C_i)$ and $u'_i(s)$ (resp. $\ell'_s(E_j)$ and $v'_j(s)$) be the circumference and the width of C_i (resp. of E_j) in (X'_s,ω'_s).

Let $x+iy$ be the period of the core curves of $C_{0,t}$ in (X'_s,ω'_s). From (8) we get

$$|x| = \sum x_j v'_j(s) = \mathrm{e}^s \sum x_j v'_j \le e^s \frac{\sqrt{2}\mathrm{e}^{-t}}{K} \sum \ell(E_j) v_j = \frac{\sqrt{2}\mathrm{e}^{s-t}}{K}. \tag{9}$$

From (7), we get

$$|y| = \sum y_i u'_i(s) = \mathrm{e}^{-s}\sum y_i u_i \le \mathrm{e}^{t-s} L_1. \tag{10}$$

Thus for $s=t$, the circumference of $C_{0,t}$ in (X'_s,ω'_s) is at most $\sqrt{3}L_1 < 2L_1$. Let ${C'}_{0,s}$ be a cylinder of width bounded below by K in (X'_s,ω'_s). By Lemma 7.3, we have $\boldsymbol{d}({C'}_{0,s}, C_{0,t}) \le M_1$ which means that $C_{0,t}$ is contained in the M_1–neighborhood of $\mathcal{L}^*_{C,E}(L_1)$. □

It follows immediately from the claim that $\mathcal{L}^*_{C,D}(t,L_1)$ is contained in the $2M_1$–neighborhood of $\mathcal{L}^*_{C,E}(L_1)$ if $t\le 0$. By similar arguments, one can also show that $\mathcal{L}^*_{C,D}(t,L_1)$ is contained in the $2M_1$–neighborhood of $\mathcal{L}^*_{E,D}(L_1)$ if $t\ge 0$. Therefore, we can conclude that $\mathcal{L}^*_{C,D}(L_1) = \cup_{t\in\mathbb{R}}\mathcal{L}^*_{C,D}(t,L_1)$ is contained in the $2M_1$–neighborhood of $\mathcal{L}^*_{C,E}(L_1)\cup\mathcal{L}^*_{E,D}(L_1)$, which implies that $\widehat{\mathcal{L}}_{C,D}$ is contained in the $3M_1$–neighborhood of $\mathcal{L}^*_{C,E}(L_1)\cup\mathcal{L}^*_{E,D}(L_1)$. □

7.4 Proof of Theorem 7.1

From Proposition 7.7, and Proposition 7.8, we see that $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ with the family of subgraphs $\widehat{\mathcal{L}}_{C,D}$ satisfies the two conditions of Theorem 7.2 with $M = \max\{M_2, M_3\}$. Therefore, $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ is Gromov hyperbolic. □

8 The quotient by affine automorphisms

In this section we investigate the quotient of $\widehat{\mathcal{C}}_{\mathrm{cyl}}(X,\omega,f)$ by the group $\mathrm{Aff}^+(X,\omega)$. Our main focus is the case where (X,ω) is a Veech surface, that is, when $\mathrm{SL}(X,\omega)$ is a lattice in $\mathrm{SL}(2,\mathbb{R})$. Throughout this section (X,ω) is a fixed translation surface in $\mathcal{H}(2)\sqcup\mathcal{H}(1,1)$, and $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ is the cylinder graph of (X,ω) with some marking map. We denote by $\mathcal{G}$ the quotient graph $\widehat{\mathcal{C}}_{\mathrm{cyl}}/\mathrm{Aff}^+(X,\omega)$, and by $\mathcal{V}$ and $\mathcal{E}$ the sets of vertices and edges of $\mathcal{G}$, respectively. Notice that an edge may join a vertex to itself (we then have a loop), and there may be more than one edge with the same endpoints. We use the notations $|\mathcal{V}|$ and $|\mathcal{E}|$ to designate the cardinalities of $\mathcal{E}$ and $\mathcal{V}$. We will show the following theorem:

Theorem 8.1 *Let (X,ω) be a surface in $\mathcal{H}(2)\sqcup\mathcal{H}(1,1)$. Then (X,ω) is a Veech surface if and only if $|\mathcal{V}|$ is finite.*

Theorem 8.1 does not mean, when (X,ω) is a Veech surface, that the quotient graph $\mathcal{G}$ is a finite graph, as we have the following:

Proposition 8.2 *If (X,ω) is Veech surface in $\mathcal{H}(2)$ then $\mathcal{G}$ is a finite graph, that is, $|\mathcal{V}|$ and $|\mathcal{E}|$ are both finite. There exist Veech surfaces in $\mathcal{H}(1,1)$ such that $|\mathcal{V}|<\infty$ but $|\mathcal{E}|=\infty$.*

8.1 Proof of Theorem 8.1

Recall that the $\mathrm{SL}(2,\mathbb{R})$–orbit of a Veech surface (X,ω) projects to an algebraic curve in $\mathcal{M}_2$ isomorphic to $\mathcal{X}:=\mathbb{H}/\mathrm{SL}(X,\omega)$; this curve is called a *Teichmüller curve*. The direction of any saddle connection on X is periodic, that is, X is decomposed into finitely many cylinders in this direction. Moreover, there is a parabolic element in $\mathrm{SL}(X,\omega)$ that fixes this direction. Thus each cylinder in X corresponds to a cusp in $\mathcal{X}$.

Let θ be a periodic direction for X. Let $C_1,\dots,C_k$ be the cylinders of X in the direction θ, and T_i be the Dehn twist about the core curves of C_i. Let γ be the generator of the parabolic subgroup of $\mathrm{SL}(X,\omega)$ that fixes θ. Then there exist some integers $m_1,\dots,m_k$ such that γ is the derivative of an element of $\mathrm{Aff}^+(X,\omega)$ isotopic to $T_1^{m_1}\circ\dots\circ T_k^{m_k}$.

8.1.1 Proof that (X,ω) is Veech implies that $\mathcal{V}$ is finite If $(X,\omega)\in\mathcal{H}(2)$, then X has one or two cylinders in the direction θ. In the first case, we have three more degenerate ones, and in the second case there is no degenerate cylinder. Thus the total number of cylinders (degenerate or not) in a periodic direction is at most 4. If $(X,\omega)\in\mathcal{H}(1,1)$, then by similar arguments, we see that X has at most 5 cylinders in the direction θ. We have seen that θ corresponds to a cusp of $\mathcal{X}$. Since $\mathcal{X}$ has finitely many cusps, it follows that X has finitely many cylinders up to action of $\mathrm{Aff}^+(X,\omega)$. Therefore, $\mathcal{V}$ is finite. □

8.1.2 Proof that $\mathcal{V}$ is finite implies that (X,ω) is Veech In what follows, by an *embedded triangle* in X, we mean the image of a triangle $\boldsymbol{T}$ in the plane by a map $\varphi\colon \boldsymbol{T}\to X$ which is locally isometric, injective in the interior of $\boldsymbol{T}$, and which sends the vertices of $\boldsymbol{T}$ to the singularities of X. Note that φ maps a side of $\boldsymbol{T}$ to a concatenation of some saddle connections. By a slight abuse of notation, we will also denote by $\boldsymbol{T}$ the image of φ in X. To show that (X,ω) is a Veech surface, we will use the following characterization of Veech surfaces by Smillie and Weiss [47].

Theorem 8.3 (Smillie and Weiss) *(X,ω) is a Veech surface if and only if there exists an $\epsilon > 0$ such that the area of any embedded triangle $\boldsymbol{T}$ in X is at least ϵ.*

We now assume that $|\mathcal{V}|$ is finite. If v is a vertex of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$, we denote by $\bar{v}$ its equivalence class in $\mathcal{V}$. Clearly, the group $\mathrm{Aff}^+(X,\omega)$ preserves the areas of the cylinders in X. Therefore, each element of $\mathcal{V}$ has a well-defined area (a degenerate cylinder has zero area). Since $\mathcal{V}$ is finite, we can write $\mathcal{V} = \{\bar{v}_1,\dots,\bar{v}_n\}$, where $n = |\mathcal{V}|$. Using $\mathrm{GL}^+(2,\mathbb{R})$, we can normalize so that $\mathrm{Area}(X,\omega) = 1$. Let $a_i = \mathrm{Area}(\bar{v}_i)$, and define

$$\begin{aligned}
\mathcal{A}_1 &= \{a_1,\dots,a_n\},\\
\mathcal{A}_2 &= \{|a_i - a_j| : a_i \neq a_j\},\\
\mathcal{A}_3 &= \{1-(a_i+a_j) : a_i + a_j < 1\},\\
\mathcal{A}_4 &= \{1-(a_i+a_j+a_k) : a_i + a_j + a_k < 1\}.
\end{aligned}$$

Set $\epsilon = \min\{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4\}$. We will need the following lemma on slit tori.

Lemma 8.4 *Let $(\widehat{X},\widehat{\omega},\hat{s})$ be a slit torus. By a* cylinder *in $\widehat{X}$, we will mean a connected component of X that is cut out by a pair of parallel simple closed geodesics passing through the endpoints of $\hat{s}$.*

Assume that $\hat{s}$ is not parallel to any simple closed geodesic of $\widehat{X}$. Then there exists a sequence of cylinders $\{\widehat{C}_k\}_{k\in\mathbb{N}}$ such that $\widehat{C}_k$ is disjoint from the slit $\hat{s}$ for all $k \in \mathbb{N}$, and $\mathrm{Area}(\widehat{C}_k) \to \mathrm{Area}(\widehat{X})$ as $k \to +\infty$.

Proof Using $\mathrm{GL}^+(2,\mathbb{R})$, we can normalize so that $(\widehat{X},\widehat{\omega}) = (\mathbb{C}/(\mathbb{Z}\oplus i\mathbb{Z}), dz)$. The slit $\hat{s}$ is then represented by a segment $[0,(1+i\alpha)t]$, with $t\in(0,\infty)$ and $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. In this setting, each simple closed geodesic c of $\widehat{X}$ corresponds to a vector $p+iq$ with $p,q\in\mathbb{Z}$ and $\gcd(p,q)=1$. Let c_1 and c_2 be the simple geodesics parallel to c which pass through the endpoints of $\hat{s}$. Note that c_1, c_2 cut $\widehat{X}$ into two cylinders. By [41, Lemma 4.1], we know that one of the two cylinders is disjoint from $\hat{s}$ if and only if

$$t\left|\det\begin{pmatrix} p & 1\\ q & \alpha\end{pmatrix}\right| = t|p\alpha - q| < 1.$$

Note that the quantity $t|p\alpha - q|$ is precisely the area of the cylinder that contains $\hat{s}$. Since α is an irrational number, one can find a sequence $\{(p_k,q_k)\}_{k\in\mathbb{N}}$ such that

$$\gcd(p_k,q_k) = 1, \quad t|\alpha p_k - q_k| < 1 \quad \text{and} \quad \lim_{k\to\infty} |\alpha p_k - q_k| = 0.$$

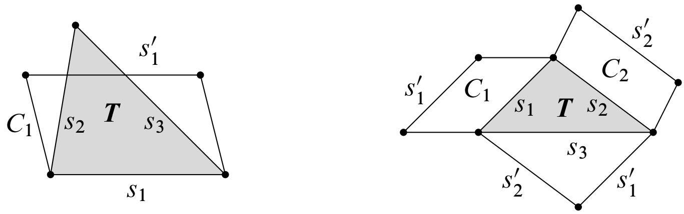


Figure 11: Embedded triangles in a surface in $\mathcal{H}(2)$ in Case 1 (left) and Case 2 (right)

For each (p_k, q_k) in this sequence, we have a cylinder $\widehat{C}_k$ in direction of $p_k + iq_k$ disjoint from $\hat{s}$ such that

$$\mathrm{Area}(\widehat{C}_k) = 1 - t|\alpha p_k - q_k|.$$

In particular, we have $\lim_{k\to\infty} \mathrm{Area}(\widehat{C}_k) = 1$, which proves the lemma. □

As a consequence of this lemma, we get the following.

Corollary 8.5 *Let (s_1, s_2) be a pair of homologous saddle connections in X that are exchanged by the hyperelliptic involution τ. If one of the connected components cut out by (s_1, s_2) is a slit torus, then the direction of s_1, s_2 is periodic.*

Proof If $(X,\omega) \in \mathcal{H}(2)$ then X is decomposed by (s_1, s_2) into a simple cylinder and a slit torus, if $(X,\omega) \in \mathcal{H}(1,1)$ then X is decomposed into two slit tori. Thus, it suffices to show that s_i is parallel to a closed geodesic in each slit torus. If this is not the case, then by Lemma 8.4, we can find in this slit torus a sequence of cylinders disjoint from the slit whose area converges to the area of the torus. Note that such cylinders are also cylinders of X. Thus their areas belong to $\mathscr{A}_1$. Since $\mathscr{A}_1$ is finite, it cannot contain a nonconstant converging sequence. Therefore, we can conclude that the direction of (s_1, s_2) is periodic. □

Let $\boldsymbol{T}$ be an embedded triangle in X. We will show that $\mathrm{Area}(\boldsymbol{T}) > \frac{1}{2}\epsilon$. We first remark that it suffices to consider the case where each side of $\boldsymbol{T}$ is a saddle connection, since otherwise there is another embedded triangle contained in $\boldsymbol{T}$ with this property. Let τ denote the hyperelliptic involution of X, and $\boldsymbol{T}' = \tau(\boldsymbol{T})$. Let s_1, s_2, s_3 be the sides of $\boldsymbol{T}$ and s'_i be the image of s_i by τ. The proof that $\mathrm{Area}(\boldsymbol{T}) > \frac{1}{2}\epsilon$ naturally splits into two cases depending on the stratum of (X,ω).

Case $(X,\omega) \in \mathcal{H}(2)$ We need to consider the following two situations:

Case 1 None of the sides of $\boldsymbol{T}$ is invariant by τ. From Lemma 2.4, s_i and s'_i bound a simple cylinder denoted by C_i. Let h_i be length of the perpendicular segment from the opposite vertex of s_i in $\boldsymbol{T}$ to s_i. If $\mathrm{int}(\boldsymbol{T}) \cap \mathrm{int}(C_1) \neq \varnothing$, then both s_2 and s_3 cross C_1

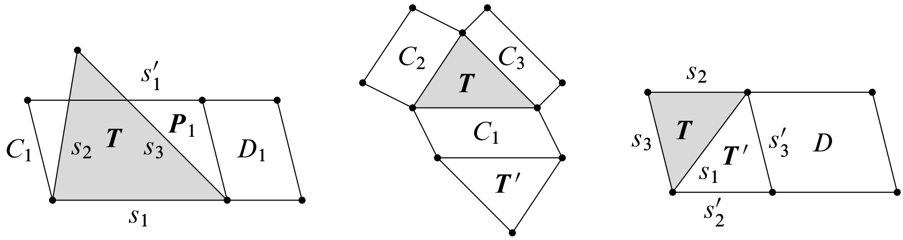


Figure 12: Embedded triangles in a surface in $\mathcal{H}(1,1)$; from left to right we have Cases 1–3.

entirely, which implies that the width of C_1 is at most h_1; see Figure 11 (left). It follows that $\mathrm{Area}(\boldsymbol{T}) \geq \frac{1}{2}\,\mathrm{Area}(C_1) > \min \frac{1}{2}\mathscr{A}_1$. The same arguments apply in the cases that $\mathrm{int}(\boldsymbol{T})$ intersects $\mathrm{int}(C_2)$ or $\mathrm{int}(C_3)$. If $\mathrm{int}(\boldsymbol{T})$ is disjoint from $\mathrm{int}(C_i)$, $i = 1,2,3$, then we have three disjoint cylinders in X (if $\mathrm{int}(C_i) \cap \mathrm{int}(C_j) \neq \varnothing$ then s_i must cross C_j entirely hence $\mathrm{int}(\boldsymbol{T}) \cap \mathrm{int}(C_j) \neq \varnothing$). Since $(X,\omega) \in \mathcal{H}(2)$, this situation cannot occur; see Theorem 2.6. Hence, we can conclude that $\mathrm{Area}(\boldsymbol{T}) \geq \frac{1}{2}\epsilon$ here.

Case 2 One of the sides of $\boldsymbol{T}$ is invariant by τ. In this case, the union of $\boldsymbol{T}$ and its image by τ is an embedded parallelogram; see Lemma 2.1. This means that there is a parallelogram $\boldsymbol{P}$ in the plane such that $\boldsymbol{T}$ is one of the two triangles cut out by a diagonal of $\boldsymbol{P}$, and there is a map $\varphi\colon \boldsymbol{P} \to X$ locally isometric, injective in $\mathrm{int}(\boldsymbol{T})$, mapping the vertices of $\boldsymbol{P}$ to the singularity of X. We remark that all the sides of $\boldsymbol{T}$ cannot be invariant by τ because this would imply that $X = \varphi(\boldsymbol{P})$ is a torus. If there are two sides of $\boldsymbol{T}$ that are invariant by τ, then $\varphi(\boldsymbol{P})$ is a simple cylinder in X, hence $\mathrm{Area}(\boldsymbol{T}) \geq \min \frac{1}{2}\mathscr{A}_1$. If there is only one side invariant by τ, then the complement of $\varphi(\boldsymbol{P})$ is the union of two disjoint simple cylinders C_1, C_2 (see Figure 11, right), which implies $\mathrm{Area}(\boldsymbol{P}) = 1 - (\mathrm{Area}(C_1) + \mathrm{Area}(C_2))$. Therefore, we have $\mathrm{Area}(\boldsymbol{T}) > \min \frac{1}{2}\mathscr{A}_3 \geq \frac{1}{2}\epsilon$. This completes the proof of Theorem 8.1 for the case $(X,\omega) \in \mathcal{H}(2)$.

Case $(X,\omega) \in \mathcal{H}(1,1)$ We consider the following situations:

Case 1 There exists i such that s_i' intersects $\mathrm{int}(\boldsymbol{T})$. Note that we must have $s_i' \neq s_i$. Let us assume that $i = 1$. Recall that s_1 and s_1' either bound a simple cylinder or decompose X into two tori. In the first case, the same argument as in the case $(X,\omega) \in \mathcal{H}(2)$ shows that $\mathrm{Area}(\boldsymbol{T}) \geq \min \frac{1}{2}\mathscr{A}_1$. For the second case, observe that the intersection of $\boldsymbol{T}$ with one of the slit tori consists of a domain bounded by s_1 and some subsegments of s_2, s_3 and s_1'; see Figure 12. Let $(X_1,\omega_1,\tilde{s}_1)$ denote this slit torus.

We can assume that s_1 is horizontal. By Corollary 8.5 we know that the horizontal direction is periodic for X_1, thus X_1 is the closure of a horizontal cylinder C_1. We remark that X_1 contains a transverse simple cylinder D_1 disjoint from $s_1 \cup s_1'$, whose core curves cross C_1 once. The complement of D_1 in X_1 is an embedded parallelogram $\boldsymbol{P}_1$ bounded by s_1, s_1' and the boundary of D_1. Clearly, we have $\mathrm{Area}(\boldsymbol{T}) \geq \frac{1}{2}\,\mathrm{Area}(\boldsymbol{P}_1)$. By definition, we have

$$\mathrm{Area}(\boldsymbol{P}_1) = \mathrm{Area}(C_1) - \mathrm{Area}(D_1) \geq \min \mathscr{A}_2.$$

Thus we have $\mathrm{Area}(\boldsymbol{T}) \geq \frac{1}{2}\epsilon$.

Case 2 None of s_i' intersects $\mathrm{int}(\boldsymbol{T})$, and $s_i' \neq s_i$, $i = 1, 2, 3$. It is not difficult to show that this case only happens when s_i and s_i' bound a simple cylinder C_i disjoint from $\mathrm{int}(\boldsymbol{T}) \cup \mathrm{int}(\boldsymbol{T}')$. Therefore, X is decomposed into the union of three cylinders C_1, C_2, C_3, and $\boldsymbol{T} \cup \boldsymbol{T}'$; see Figure 12. Thus in this case, we have

$$\mathrm{Area}(\boldsymbol{T}) = \tfrac{1}{2}\big(1 - (\mathrm{Area}(C_1) + \mathrm{Area}(C_2) + \mathrm{Area}(C_3))\big) \geq \min \tfrac{1}{2}\mathscr{A}_4 \geq \tfrac{1}{2}\epsilon.$$

Case 3 None of s_i' intersects $\mathrm{int}(\boldsymbol{T})$ and one of s_1, s_2, s_3 is invariant by τ. Assume that $s_1' = s_1$. It follows that $\boldsymbol{T} \cup \boldsymbol{T}'$ is an embedded parallelogram $\boldsymbol{P}$. If both (s_2, s_2') and (s_3, s_3') are the boundaries of some simple cylinders C_2 and C_3, respectively, then C_2 and C_3 are disjoint, and $C_2 \cup C_3$ is disjoint from $\boldsymbol{P}$. By construction we must have $X = \overline{\boldsymbol{P}} \cup \overline{C}_2 \cup \overline{C}_3$, which is impossible since $(X, \omega) \in \mathcal{H}(1,1)$. Therefore, we can assume that (s_2, s_2') decompose X into two slit tori. Let X_1 be the slit torus that contains $\boldsymbol{P}$. By Corollary 8.5, we know that the direction of (s_2, s_2') is periodic, which means that X_1 is the closure of a cylinder C. Observe that the complement of $\boldsymbol{P}$ in X_1 must be a cylinder D bounded by (s_3, s_3'); see Figure 12. Therefore,

$$\mathrm{Area}(\boldsymbol{T}) = \tfrac{1}{2}\,\mathrm{Area}(\boldsymbol{P}) = \tfrac{1}{2}(\mathrm{Area}(C) - \mathrm{Area}(D)) \geq \tfrac{1}{2}\min \mathscr{A}_2 \geq \tfrac{1}{2}\epsilon.$$

Case 4 None of s_i' intersects $\mathrm{int}(\boldsymbol{T})$ and two of s_1, s_2, s_3 are invariant by τ. In this case $\boldsymbol{T} \cup \boldsymbol{T}'$ is a simple cylinder. Therefore, $\mathrm{Area}(\boldsymbol{T}) \geq \min \frac{1}{2}\mathscr{A}_1 \geq \frac{1}{2}\epsilon$.

In all cases $\mathrm{Area}(\boldsymbol{T}) \geq \frac{1}{2}\epsilon$, thus Theorem 8.3 implies that (X, ω) is a Veech surface. □

8.2 Proof of Proposition 8.2

Case $(X, \omega) \in \mathcal{H}(2)$ We have shown that $\mathscr{V}$ is finite; it remains to show that $\mathscr{E}$ is also finite. Let v be a vertex of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$, and C be the corresponding cylinder in X. We denote by $\bar{v}$ the equivalence class of v in $\mathscr{G}$. Using $\mathrm{SL}(2, \mathbb{R})$, we can suppose that C is horizontal.

If C is a nondegenerate cylinder, then we have three cases: (a) C is the unique horizontal cylinder, (b) X has two horizontal cylinders and C is not simple, and (c) C is a simple cylinder. In case (a), there are three edges in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ that have v as an endpoint, those edges connect v to three degenerate cylinders contained in the boundary of C. In case (b), there is only one edge in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ having v as an endpoint, this edge connects C

to the other horizontal simple cylinder. Thus in cases (a) and (b), there are only finitely many edges having $\bar{v}$ as an endpoint.

Assume now that we are in case (c). Let D be the other horizontal cylinder of X. Observe that the closure of D is a slit torus (X', ω', s') where s' corresponds to the boundary of C. Let d be a core curve of D, and e be a simple closed geodesic in X' disjoint from the slit s' and crossing d once. We consider $\{d, e\}$ as a basis of $H_1(X', \mathbb{Z})$. If C' is a cylinder in X disjoint from C, then C' must be entirely contained in $\overline{D}$. Thus the core curves of C' are determined by a unique element of $H_1(X', \mathbb{Z})$, and we can write $C' = md + ne$ with $m, n \in \mathbb{Z}$.

By assumption, a core curve c' of C' cannot cross the slit s'. The necessary and sufficient condition for this is that $|\omega'(c') \wedge \omega'(s')| \leq \mathrm{Area}(X') = \mathrm{Area}(D)$; see [41, Lemma 4.1]. But $|\omega'(c') \wedge \omega'(s')| = |n||\omega'(e) \wedge \omega'(s')|$. Thus we can conclude that $|n|$ is bounded by some constant n_0.

We have seen that $\mathrm{Aff}^+(X, \omega)$ contains an element $\phi = T_1^{m_1} \circ T_2^{m_2}$, where T_1 and T_2 are the Dehn twists about the core curves of C and D, respectively. Observe that ϕ fixes the vertices of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ corresponding to C and D. The action of ϕ on the curves contained in $\overline{D}$ is given by

$$\phi(md + ne) = (m \pm m_2 n)d + ne.$$

Thus up to action of $\{\phi^k\}_{k \in \mathbb{Z}}$, any cylinder C' contained in $\overline{D}$ belongs to the equivalence class of a cylinder C'' also contained in $\overline{D}$ whose core curves are represented by $md + nc$ with $|n| \leq |n_0|$ and $|m| \leq |m_2 n| \leq |m_2||n_0|$. We can then conclude that there are finitely many edges in $\mathscr{E}$ which contain $\bar{v}$ as an endpoint.

It remains to consider the case that C is degenerate. In this case X has a unique nondegenerate cylinder in the horizontal direction, which contains C in its boundary. Note that the complement of C in X can be isometrically identified with a flat torus with an embedded geodesic segment removed. Therefore, the arguments above also hold in this case. Since we have proved that the set of vertices of $\mathscr{G}$ is finite, it follows that the set of edges of $\mathscr{G}$ is also finite. □

Case $(X, \omega) \in \mathcal{H}(1,1)$ Let (X, ω) be the surface constructed from 6 squares as shown in Figure 13. This surface has 3 horizontal cylinders denoted by C_1, C_2, C_3, where C_i is the cylinder with i squares. It has two vertical cylinders denoted by D_1 and D_2, where the core curves of D_1 cross C_1 and C_3. Let v be the vertex of $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ corresponding to C_1, and w be the vertex corresponding to C_2. The fact that $\mathscr{G}$ has finitely many vertices follows from Theorem 8.1. We will show that $\mathscr{G}$ has infinitely many edges.

Given a cylinder C on X, we denote by T_C the Dehn twist about the core curves of C. Observe that $f = T_{C_1}^6 \circ T_{C_2}^3 \circ T_{C_3}^2$ and $g = T_{D_1} \circ T_{D_2}^2$ are two elements of

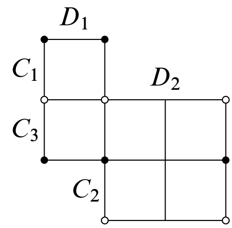


Figure 13: Example of a square-tiled surface in $\mathcal{H}(1,1)$

$\mathrm{Aff}^+(X,\omega)$ whose derivatives are

$$\begin{pmatrix} 1 & 6 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix},$$

respectively. If h is an element of $\mathrm{Aff}^+(X,\omega)$ that preserves the horizontal direction, then h must map a horizontal cylinder to a horizontal cylinder. Since C_1, C_2, C_3 have different circumferences, h must preserve each of them, which implies that $h = f^k$, $k \in \mathbb{Z}$. We derive in particular that there is no affine homeomorphism that maps C_2 to C_1.

For any $n \in \mathbb{N}$, let E_n be the image of C_2 by g^n. We remark that $E_n = T_{D_2}^{2n}(C_2)$, hence E_n is contained in the closure $\overline{D}_2$ of D_2. In particular, E_n is disjoint from C_1. Thus, there is an edge $\boldsymbol{e}_n$ in $\widehat{\mathcal{C}}_{\mathrm{cyl}}$ connecting v to the vertex w_n corresponding to E_n. By definition, all the vertices w_n belong to the equivalence class $\overline{w}$ of w in $\mathcal{G}$. We will show that the edges $\{\boldsymbol{e}_n\}_{n\in\mathbb{N}}$ are all distinct up to action of $\mathrm{Aff}^+(X,\omega)$, which means that there are infinitely many edges in $\mathcal{E}$ between $\bar{v}$ and $\overline{w}$.

Assume that there is an affine automorphism $h \in \mathrm{Aff}^+(X,\omega)$ such that $h(\boldsymbol{e}_{n_1}) = \boldsymbol{e}_{n_2}$, for some $n_1, n_2 \in \mathbb{N}$. If $h(w_{n_i}) = v$, then there is an element of $\mathrm{Aff}^+(X,\omega)$ that sends w to v, or equivalently C_2 to C_1. But we have already seen that such an element does not exist, thus this case cannot occur. Therefore, we must have $h(v) = v$ and $h(w_{n_1}) = w_{n_2}$. Since any element of $\mathrm{Aff}^+(X,\omega)$ preserving C_1 belongs to the subgroup generated by f, we derive that h also preserves C_2 and C_3. Observe that a core curve of E_{n_i} crosses C_2 $2n_i$ times. Therefore, if $n_1 \neq n_2$, then h cannot exist. We can then conclude that the projections of all the edges $\boldsymbol{e}_n$ are distinct in $\mathcal{G}$, which proves the proposition. □

9 Quotient graphs and McMullen's prototypes

By the works of McMullen [38; 37], we know that closed $\mathrm{GL}^+(2,\mathbb{R})$–orbits in $\mathcal{H}(2)$ are indexed by the discriminant D, that is, a natural number $D \in \mathbb{N}$ such that $D \equiv 0, 1 \mod 4$, together with the parity of the spin structure when $D \equiv 1 \mod 8$ and $D \neq 9$.

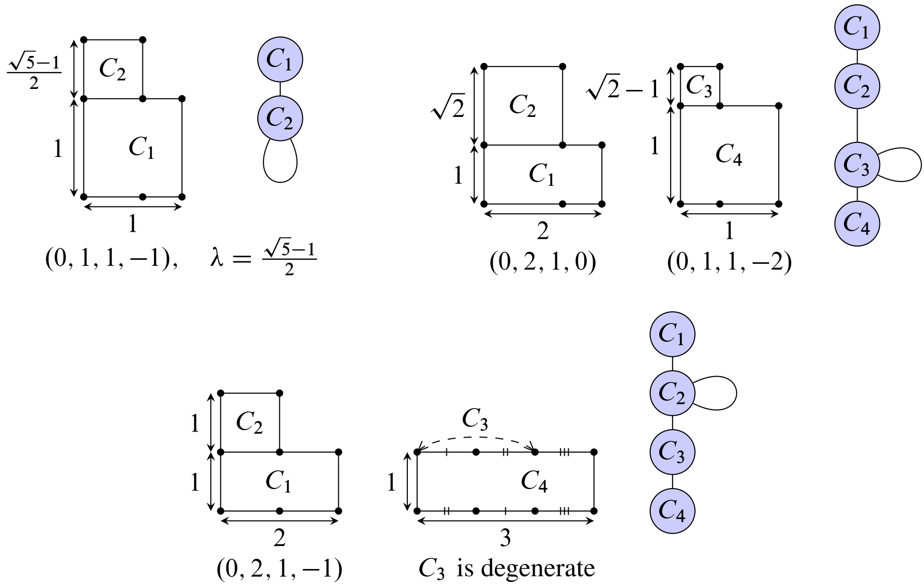


Figure 14: Examples of $\mathcal{G}$ for $D=5$ (top left), $D=8$ (top right) and $D=9$ (bottom). For each two-cylinder decomposition, we provide the corresponding prototype (a,b,c,e). A loop at some vertex represents a butterfly move that does not change the prototype.

Let (X,ω) be an eigenform in $\mathcal{E}_D \cap \mathcal{H}(2)$ for some fixed D. Following [37], every two-cylinder decomposition of X is encoded by a quadruple of integers $(a,b,c,e)\in\mathbb{Z}^4$ called a *prototype* satisfying the following conditions:

$$(\mathcal{P}_D) \qquad \begin{array}{lll} b>0, & c>0, & \gcd(b,c)>a\geq 0, \\ D=e^2+4bc, & b>c+e, & \gcd(a,b,c,e)=1. \end{array}$$

Set $\lambda=\frac{1}{2}(e+\sqrt{D})$. Up to action of $\mathrm{GL}^+(2,\mathbb{R})$, the decomposition of X consists of two horizontal cylinders. The first one is simple and represented by a square of size λ. The other one is nonsimple and represented by a parallelogram constructed from the vectors $(b,0)$ and (a,c). Note that we always have $b>\lambda$.

The quotient graph $\mathcal{G}$ turns out to be closely related to the set of McMullen's prototypes. Namely, each prototype corresponds to a cluster of two vertices of $\mathcal{G}$ which represent the cylinders in the corresponding cylinder decomposition. Let C_1,C_2 be the cylinders in this decomposition, where C_1 is the simple one. Then the vertex corresponding to C_2 is only adjacent to the one corresponding to C_1 in $\mathcal{G}$. This is because any other cylinder of X must cross C_2.

On the other hand, if there is an edge in $\mathcal{G}$ between two vertices representing two simple cylinders which are not parallel, then the two cylinder decompositions are related by a "butterfly move"; see [37, Section 7] for the precise definitions. In other words, $\mathcal{G}$ can be viewed as a geometric object representing $\mathcal{P}_D$: each prototype is represented by two vertices connected by an edge, and all the other edges of $\mathcal{G}$ represent butterfly moves.

There is nevertheless a slight difference between the two notions. The set $\mathcal{P}_D$ only parametrizes two-cylinder decompositions of X, while in $\mathcal{G}$ we also have one-cylinder decompositions. If $\sqrt{D} \notin \mathbb{N}$, then any cylinder in X is contained in a two-cylinder decomposition. Thus, the set of prototypes exhausts all the equivalence classes of cylinders in X (hence it provides the complete list of cusps of the corresponding Teichmüller curve). But when D is a square (eg $D = 9$), we need to take into account one-cylinder decompositions as well as degenerate cylinders. In Figure 14, we draw the quotient cylinder graphs of surfaces corresponding to some small values of D.

Appendix A: Triangulations

In this section we construct triangulations of (X,ω) that are invariant by the hyperelliptic involution. The aim of these triangulations is to provide a preferred way to represent (X,ω) as a polygon in $\mathbb{R}^2$ when we have a horizontal simple cylinder on X. The results of this section are certainly not new and are known to most people in the field; see eg [49]. We present them here only for the sake of completeness.

In what follows, for any saddle connection s, we will denote by $\boldsymbol{h}(s)$ the length of the horizontal component of s, that is, $\boldsymbol{h}(s) = |\mathrm{Re}(\omega(s))|$. If Δ is a triangle bounded by the saddle connections s_1, s_2, s_3, we define $\boldsymbol{h}(\Delta) = \max\{\boldsymbol{h}(s_i) \mid i = 1,2,3\}$. Our main result in this section is the following:

Proposition A.1 *Let (X,ω) be a translation surface in $\mathcal{H}(2) \sqcup \mathcal{H}(1,1)$ having a simple horizontal cylinder C. Assume that every regular leaf of the vertical foliation of (X,ω) crosses C.*

(i) *If $(X,\omega) \in \mathcal{H}(2)$, then (X,ω) can be obtained by identifying the pairs of opposite sides of an octagon $\boldsymbol{P} = (P_0 \cdots P_3 Q_0 \cdots Q_3) \subset \mathbb{R}^2$* (see Figure 15), *where the vertices are labeled clockwise, such that the following hold:*

- $\overrightarrow{P_iP_{i+1}} = -\overrightarrow{Q_iQ_{i+1}}$, $i = 0,1,2$, *and* $\overrightarrow{P_3Q_0} = -\overrightarrow{Q_3P_0}$.
- *The diagonals $\overline{P_0P_3}$ and $\overline{Q_0Q_3}$ are horizontal, the parallelogram $(P_0P_3Q_0Q_3)$ is contained in $\boldsymbol{P}$ and projects to $C \subset X$.*
- *For $i = 1,2$, the vertical line through P_i (resp. Q_i) intersects $\overline{P_0P_3}$ (resp. $\overline{Q_0Q_3}$), and the vertical segment from P_i (resp. from Q_i) to the intersection is contained in $\boldsymbol{P}$.*

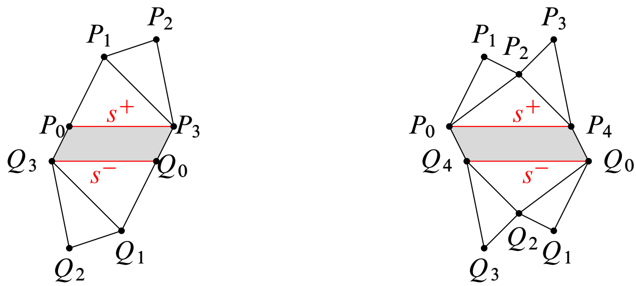


Figure 15: Representations of surfaces (X,ω) in $\mathcal{H}(2)$ (left) and $\mathcal{H}(1,1)$ (right) with symmetric polygons. The simple horizontal cylinder is represented by the highlighted parallelogram.

(ii) *If $(X,\omega)\in\mathcal{H}(1,1)$, then (X,ω) can be obtained by identifying the pairs of opposite sides of a decagon $\boldsymbol{P}=(P_0\cdots P_4Q_0\cdots Q_4)$* (see Figure 15), *where the vertices are labeled clockwise, such that the following hold:*

- $\overrightarrow{P_iP_{i+1}}=-\overrightarrow{Q_iQ_{i+1}}$, $i=0,\dots,3$, *and* $\overrightarrow{P_4Q_0}=-\overrightarrow{Q_4P_0}$.
- *The diagonals $\overline{P_0P_4}$ and $\overline{Q_0Q_4}$ are horizontal, the parallelogram $(P_0P_4Q_0Q_4)$ is contained in $\boldsymbol{P}$ and projects to $C\subset X$.*
- *For $i=1,2,3$, the vertical line through P_i (resp. Q_i) intersects $\overline{P_0P_4}$ (resp. $\overline{Q_0Q_4}$), and the vertical segment from P_i (resp. from Q_i) to the intersection is contained in $\boldsymbol{P}$.*

Proof Cut off C from X, and identify the geodesic segments in the boundary of the resulting surface, we then obtain either a slit torus (if $(X,\omega)\in\mathcal{H}(2)$) or a surface in $\mathcal{H}(2)$ with a marked saddle connection (if $(X,\omega)\in\mathcal{H}(1,1)$). Let (X',ω') denote the new surface, and s' the marked saddle connection. If (X',ω') is a slit torus, then there is a unique involution of X' that acts by $-\mathrm{Id}$ on $H_1(X',\mathbb{Z})$ and exchanges the endpoints of s'. By a slight abuse of notation, we will call this involution the hyperelliptic involution of X'. Thus, in both cases, s' is invariant by the hyperelliptic involution.

By assumption all the regular vertical leaves of X' intersect s'. Let $\{\Delta_i^{\pm}\mid i=1,\dots,k\}$ be the triangulation of X' provided by Lemmas A.2 and A.3; if $(X',\omega')\in\mathcal{H}(0,0)$, $k=2$, if $(X',\omega')\in\mathcal{H}(2)$, $k=3$. We can represent C by a parallelogram in $\mathbb{R}^2$. The polygon $\boldsymbol{P}$ is obtained from this parallelogram by gluing successively the triangles $\Delta_1^+,\dots,\Delta_k^+$, then $\Delta_1^-,\dots,\Delta_k^-$. □

Lemma A.2 *Let (X,ω,s) be a slit torus. Let τ be the elliptic involution of X that exchanges the endpoints P_1, P_2 of s. Assume that all the leaves of vertical foliation*

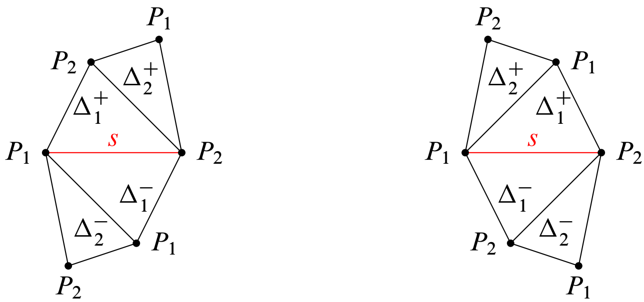


Figure 16: Triangulation of a slit torus

meet s. Then there exists a unique triangulation of X into four triangles $\Delta_1^\pm, \Delta_2^\pm$ with vertices in $\{P_1, P_2\}$, such that the following are satisfied:

- *Δ_i^+ and Δ_i^- are exchanged by τ.*
- *s is contained in both Δ_1^+ and Δ_1^-.*
- *For $i = 1, 2$, the union $\Delta_i^+ \cup \Delta_i^-$ is a cylinder in X.*
- *Δ_1^+ is adjacent to Δ_1^- and Δ_2^+, Δ_1^- is adjacent to Δ_1^+ and Δ_2^-.*
- *$\boldsymbol{h}(\Delta_1^\pm) = \boldsymbol{h}(s)$, and $\boldsymbol{h}(\Delta_2^\pm) = \boldsymbol{h}(c^+)$, where c^+ is the unique common side of Δ_2^+ and Δ_1^+.*

There are two possible configurations for this triangulation, shown in Figure 16.

Proof By Lemma 2.3, we know that there exists a pair of simple closed geodesics c^+, c^- passing through the endpoints of s that cut X into two cylinders satisfying $\boldsymbol{h}(c^\pm) \le \boldsymbol{h}(s)$. One of the cylinders cut out by $c^\pm$ contains s, we denote it by C_1, the other one is denoted by C_2. Note that we must have $\boldsymbol{h}(c^\pm) > 0$, otherwise there are vertical leaves that do not meet s. It is easy to see that we get the desired triangulation by adding some geodesic segments in C_1 and C_2 joining the endpoints of s. □

Lemma A.3 *Let (X, ω) be a surface in $\mathcal{H}(2)$ and s be a saddle connection on X, invariant by the hyperelliptic involution τ. We assume that s is horizontal and all the leaves of the vertical foliation meet s. Then we can triangulate X into six triangles $\Delta_i^\pm$, $i = 1, 2, 3$, whose sides are saddle connections, satisfying the following:*

- *$\tau(\Delta_i^+) = \Delta_i^-$, $i = 1, 2, 3$.*
- *Δ_1^+ and Δ_1^- contain s, and $\boldsymbol{h}(\Delta_1^\pm) = \boldsymbol{h}(s)$.*
- *Δ_2^+ has a unique common side with Δ_1^+ which will be denoted by a^+, and $\boldsymbol{h}(\Delta_2^+) = \boldsymbol{h}(a^+)$.*
- *Δ_3^+ either has a unique common side b^+ with Δ_1^+ and $\boldsymbol{h}(\Delta_3^+) = \boldsymbol{h}(b^+)$ or Δ_3^+ has a unique common side c^+ with Δ_2^+ and $\boldsymbol{h}(\Delta_3^+) = \boldsymbol{h}(c^+)$.*

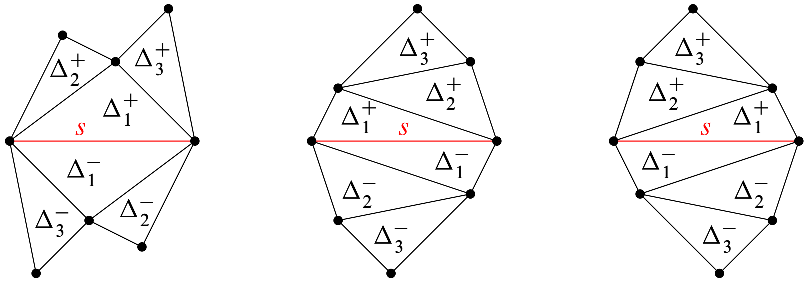


Figure 17: Triangulation of surfaces in $\mathcal{H}(2)$

This triangulation is unique. The configurations of the triangles $\Delta_i^{\pm}$, $i = 1, 2, 3$, *are shown in Figure 17.*

Proof From Lemma 2.1, we see that there exist a parallelogram $\boldsymbol{P} \subset \mathbb{R}^2$ and a locally isometric map $\varphi\colon \boldsymbol{P} \to X$ that maps a diagonal of $\boldsymbol{P}$ to s. By construction, $\varphi(\boldsymbol{P})$ is decomposed into two embedded triangles $\Delta_1^{\pm}$, where Δ_1^+ is the one above s, both of which satisfy $\boldsymbol{h}(\Delta_1^{\pm}) = \boldsymbol{h}(s) = |s|$. Note also that $\tau(\Delta_1^+) = \Delta_1^-$.

Let us denote the nonhorizontal sides of Δ_1^+ by a^+ and b^+, and their images by τ by a^- and b^-, respectively. If both of a^+ and b^+ are invariant by τ then $X = \varphi(\boldsymbol{P})$, which implies that X is a torus, and we have a contradiction. Therefore, we only have two cases:

(a) None of a^+, b^+ is invariant by τ. In this cases, by Lemma 2.4 the complement of $\varphi(\boldsymbol{P})$ is the disjoint union of two cylinders bounded by $a^{\pm}$ and $b^{\pm}$, respectively. Note that none of a^+ and b^+ is vertical, otherwise there would be vertical leaves that do not meet s. We can then triangulate the cylinders bounded by $a^{\pm}$ and $b^{\pm}$ in the same way as in Lemma A.2.

(b) One of a^+, b^+ is invariant by τ. We can assume that b^+ is invariant by τ. In this case, $\varphi(\boldsymbol{P})$ is a simple cylinder bounded by $a^{\pm}$. The complement of $\varphi(\boldsymbol{P})$ is then a slit torus (X_1, ω_1, s_1), where s_1 is the identification of $a^{\pm}$. From the assumption that all the vertical leaves meet s, we derive that $a^{\pm}$ are not vertical. Thus we can follow the same argument as in Lemma A.2 to get the desired triangulation. □

Appendix B: Cylinders and decompositions

In this section, we give the proofs of some lemmas which are used in Section 7.

Lemma B.1 *Let* $(X, \omega) \in \mathcal{H}(2) \sqcup \mathcal{H}(1,1)$ *be a completely periodic surface in the sense of Calta. If* C *is a degenerate cylinder in* X, *then the direction of* C *is periodic, that is,* X *is decomposed into cylinders in the direction of* C.

Proof If (X,ω) is in $\mathcal{H}(2)$ then (X,ω) is a Veech surface, thus the direction of any saddle connection is periodic and we are done. Assume now that (X,ω) is in $\mathcal{H}(1,1)$. In $\mathcal{H}(1,1)$, we have a local action of $\mathbb{C}$ which only changes the relative periods and leaves the absolute periods invariant. Orbits of this local action are leaves of the kernel foliation. It is well known that the any eigenform locus is invariant by this local action.

Let us label the zeros of ω by x_1, x_2 and define the orientation of any path connecting x_1 and x_2 to be from x_1 to x_2. Using this local action of $\mathbb{C}$, we can collapse the two zeros of ω as follows. Let s be a saddle connection invariant by the hyperelliptic involution satisfying the following condition, which we will call condition $(\mathcal{S})$: if there exists another saddle connection s' joining x_1 and x_2 such that $\omega(s') = \lambda\omega(s)$ with $\lambda \in (0;+\infty)$, then we have $\lambda > 1$.

We can then reduce the length of s to zero by moving in the kernel foliation leaf of (X,ω), the resulting surface is an eigenform in $\mathcal{H}(2)$ having the same absolute periods as (X,ω). The condition on s implies that x_1 and x_2 do not collide before s is reduced to a point, for a proof of this fact, we refer to [28; 29]. We remark that the new surface in $\mathcal{H}(2)$ is a Veech surface.

Without loss of generality, we can assume that C is horizontal. By definition, C is the union of two saddle connections s_1, s_2 both invariant by the hyperelliptic involution, and up to a renumbering we have $\omega(s_1) \in \mathbb{R}_{>0}, \omega(s_2) \in \mathbb{R}_{<0}$.

Assume that neither of s_1, s_2 satisfies $(\mathcal{S})$, then there exist two other saddle connections s'_1, s'_2 such that $\omega(s'_i) = \lambda_i\omega(s_i)$, with $\lambda_i \in (0;1)$. This implies that there are four horizontal saddle connections on X. Since $(X,\omega) \in \mathcal{H}(1,1)$, there are at most 4 saddle connections in a fixed direction, and this maximal number is realized if and only if the direction is periodic. Thus, in this case we can conclude that X is horizontally periodic.

Let us now assume that one of s_1, s_2, say s_1, satisfies the condition $(\mathcal{S})$. We can then collapse x_1, x_2 along s_1 to get a Veech surface $(X_0,\omega_0) \in \mathcal{H}(2)$. Since $\omega(s_2) - \omega(s_1)$ is an absolute period, it stays unchanged along the collapsing procedure. Therefore, s_2 persists in X_0, and we have $\omega_0(s_2) = \omega(s_2) - \omega(s_1) \in \mathbb{R}$. In particular, (X_0,ω_0) has a horizontal saddle connection, and because (X_0,ω_0) is a Veech surface, it must be horizontally periodic. It follows that (X,ω) is also horizontally periodic. This completes the proof of the lemma. □

Lemma B.2 *Let $(X,\omega) \in \mathcal{H}(1,1)$. Let C be a horizontal (possibly degenerate) cylinder in X, and D be a vertical simple cylinder disjoint from C. Then either*

(a) *there is another simple cylinder E disjoint from $C \cup D$ such that the complement of $C \cup D \cup E$ is the union of two embedded triangles, or*

(b) *there exists a pair of homologous saddle connections s_1, s_2 that decompose X into two slit tori (X', ω', s') and (X'', ω'', s'') such that C is contained in X' and D is contained in X''.*

Proof We first consider the case that C is not degenerate. In this case, the complement of $\overline{C}$ in X is either (1) empty, (2) a horizontal simple cylinder, (3) the disjoint union of two horizontal simple cylinders, (4) a torus with a horizontal slit, or (5) a surface $(\hat{X}, \hat{\omega}) \in \mathcal{H}(2)$ with a marked horizontal saddle connection s. Since we have a vertical simple cylinder disjoint from C, only (4) and (5) can occur. In case (4), we automatically have two slit tori, one of which is the closure of C, and the other one must contain D. Therefore we get case (b) of the statement of the lemma.

Let us now assume that we are in case (5). In this case C must be a simple horizontal cylinder, and the saddle connection s in $\hat{X}$ corresponds to the boundary of C. Note that s is invariant by the hyperelliptic involution $\hat{\tau}$ of $\hat{X}$. Let $\varphi\colon \boldsymbol{P} \to \hat{X}$ be the embedded parallelogram associated to s. Let $a^{\pm}$ and $b^{\pm}$ be the images by φ of the sides of $\boldsymbol{P}$, where $\hat{\tau}(a^+) = a^-$ and $\hat{\tau}(b^+) = b^-$. Note that D must be disjoint from $\varphi(\boldsymbol{P})$ since any vertical geodesic intersecting $\varphi(\mathrm{int}(\boldsymbol{P}))$ must intersect $\mathrm{int}(s)$, and hence C, but we have assumed that D is disjoint from C.

If $a^+ = a^-$ and $b^+ = b^-$ then $\hat{X}$ must be a torus, and we have a contradiction. Therefore, we only have two cases:

- $\boldsymbol{a^+ \neq a^-}$ **and** $\boldsymbol{b^+ \neq b^-}$ In this case, the complement of $\varphi(\boldsymbol{P})$ is the disjoint union of two simple cylinders. Since D is contained in this union, D must be one of the two. Let us denote the other one by E. To obtain (X, ω) from $(\hat{X}, \hat{\omega})$, we need to slit open s and glue back C. Consequently, we see that (X, ω) has three disjoint simple cylinders C, D, E. The complement of $C \cup D \cup E$ is the union of two embedded triangles, which are the images of the triangles in $\boldsymbol{P}$ cut out by s. Thus, we get case (a) of the statement of the lemma.

- $\boldsymbol{a^+ = a^-}$ **and** $\boldsymbol{b^+ \neq b^-}$ In this case, $\varphi(\boldsymbol{P})$ is a simple cylinder bounded by $b^{\pm}$. The complement of $\varphi(\boldsymbol{P})$ is then a slit torus (X'', ω'', s'') with the slit s'' corresponding to $b^{\pm}$. We can view (X'', ω'', s'') as a subsurface of X. Observe that D must be contained in (X'', ω'') and disjoint from the slit s'', since otherwise a core curve of D must cross C. The complement of (X'', ω'', s'') is another slit torus (X', ω', s') which is obtained by slitting $\varphi(\boldsymbol{P})$ along s and gluing back C. Therefore, we get case (b) of the statement of the lemma.

Assume now C is degenerate. By Lemma 3.4, there exist deformations (X_t, ω_t), $t \in [0, \epsilon)$, of (X, ω) such that C corresponds to a simple horizontal cylinder C_t in X_t, which has the same circumference as C and height equal to t. By construction, D

corresponds to a simple vertical cylinder D_t in X_t which is disjoint from C_t. Observe that C_t and D_t satisfy case (5) above. Therefore, by the preceding arguments, the conclusion of the lemma is true for C_t and D_t. In either case, the corresponding decomposition of X_t persists as $t \to 0$, which implies that we have the same decomposition on (X, ω). □

In what follows, if $u = (u_1, u_2)$ and $v = (v_1, v_2)$ are two vectors in $\mathbb{R}^2$, we denote

$$u \wedge v := \det \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix},$$

and $|u|, |v|$ are the Euclidean norms of u and v, respectively.

Lemma B.3 *Given a constant $L > 0$, let*

$$L_1 := 3 \max\{f(L), f(2\delta)\}, \tag{11}$$

where $f(x) = \sqrt{x^2 + 1/x^2}$, and $\delta := \left(\frac{3}{4}\right)^{\frac{1}{4}}$. Then for any slit torus (X, ω, s) with $\mathrm{Area}(X, \omega) = 1$*, and $|s| < L$, there exists in X a cylinder disjoint from s with area at least $\frac{1}{2}$ and circumference bounded above by L_1.*

Proof Let Λ be the lattice in $\mathbb{C}$ such that (X, ω) can be identified with $(\mathbb{C}/\Lambda, dz)$. Since Λ has covolume 1, there exists a vector $v \in \Lambda$ such that $|v| \le \delta$. Define $u = \omega(s) \in \mathbb{C} \simeq \mathbb{R}^2$.

Let us first consider the case that $|u| \le \frac{1}{2\delta}$. We then have

$$|u \wedge v| \le |u|\,|v| \le \tfrac{1}{2}.$$

The vector v corresponds to a simple closed geodesic c on X. The inequality above implies that there exist a pair of simple closed geodesics parallel to c cutting X into two cylinders, one of which contains s denoted by C, the other one denoted by C' consists of closed geodesics parallel to c that do not intersect s; see [41, Lemma 4.1] or [37, Theorem 7.2]. Note that the circumferences of both C and C' are $|v| \le \delta$. Since $\mathrm{Area}(C) = |u \wedge v| \le \frac{1}{2}$, we have $\mathrm{Area}(C') \ge \frac{1}{2}$. Thus C' has the required properties.

We can now turn to the case that $\frac{1}{2\delta} \le |s| \le L$. By definition, we have $f(|s|) \le \frac{1}{3} L_1$. By multiplying ω by a complex number of modulus 1, which does not change the area of X and the length of s, we can assume that s is horizontal. From Lemma 2.1, we know that there exists a local isometry φ from a parallelogram $\boldsymbol{P} \subset \mathbb{R}^2$ into X such that a horizontal diagonal of $\boldsymbol{P}$ is mapped to s. Since X is a torus, $C := \varphi(\boldsymbol{P})$ is actually a cylinder in X. Let η be the distance from the highest point of $\boldsymbol{P}$ to its horizontal diagonal. By construction, we have $\mathrm{Area}(C) = \mathrm{Area}(\boldsymbol{P}) = \eta |s| \le \mathrm{Area}(X, \omega) = 1$.

Thus $\eta \leq 1/|s|$. Note that the boundary components of C are the images by φ of two opposite sides of $\boldsymbol{P}$. Hence the circumference of C is bounded by

$$\sqrt{|s|^2+\eta^2} \leq f(|s|) \leq \tfrac{1}{3}L_1.$$

Observe that the complement of C is another cylinder C' in X sharing the same boundary with C. If $\mathrm{Area}(C') \geq \frac{1}{2}$ then we are done. Let us consider the case that $\mathrm{Area}(C') < \frac{1}{2}$, which means that $\mathrm{Area}(C) > \frac{1}{2} > \mathrm{Area}(C')$. By cutting and pasting, we can also realize C as a parallelogram $\boldsymbol{Q} = (P_1P_2P_3P_4)$ with two horizontal sides $\overline{P_1P_2}$ and $\overline{P_4P_3}$ identified with s. Note that the distance between $\overline{P_1P_2}$ and $\overline{P_4P_3}$ is η. We can then realize C' as a parallelogram $\boldsymbol{Q}' = (P_2P_3P_5P_6)$ adjacent to $\boldsymbol{Q}$, where P_5 is contained in the horizontal stripe bounded by the lines supporting $\overline{P_1P_2}$ and $\overline{P_4P_3}$; see Figure 18. Let P_6' and P_5' be the intersections of the line supporting $\overline{P_5P_6}$ and the lines supporting $\overline{P_1P_2}$ and $\overline{P_4P_3}$, respectively.

Clearly we have $\mathrm{Area}(C') = \mathrm{Area}(\boldsymbol{Q}') = \mathrm{Area}((P_2P_3P'_5P'_6))$. Since $\mathrm{Area}(C') < \mathrm{Area}(C)$, we have $|\overline{P_2P_6'}| < |\overline{P_1P_2}|$, and $|\overline{P_1P_6'}| < 2|\overline{P_1P_2}| \leq 2L$. If $P_6' \equiv P_6$, then X has a horizontal cylinder C_0 with circumference equal $|\overline{P_1P_6'}|$ and area equal 1. Clearly the core curves of C_0 do not intersect s, therefore C_0 has the required properties. If $P_6 \neq P_6'$, then by construction, $\overline{P_1P_5}$ and $\overline{P_4P_5}$ project to two simple closed geodesics in X, denoted by c_1 and c_2, respectively. These closed geodesics meet s only at one of its endpoints. Let d_1 and d_2 be respectively the simple closed geodesics parallel to c_1 and c_2 passing through the other endpoint of s. Observe that c_1 and d_1 (resp. c_2 and d_2) cut X into two cylinders, one of which contains s and will be denoted by C_1 (resp. C_2), and the other is denoted by C_1' (resp. C_2'). Now, we remark that

$$\mathrm{Area}(C_1) = |\overrightarrow{P_1P_5} \wedge \overrightarrow{P_1P_2}| \quad \text{and} \quad \mathrm{Area}(C_2) = |\overrightarrow{P_4P_5} \wedge \overrightarrow{P_4P_3}|.$$

Since

$$|\overrightarrow{P_1P_5} \wedge \overrightarrow{P_1P_2}| + |\overrightarrow{P_4P_5} \wedge \overrightarrow{P_4P_3}| = |\overrightarrow{P_1P_2} \wedge \overrightarrow{P_1P_4}| = \mathrm{Area}(C) \leq 1,$$

we have either $\mathrm{Area}(C_1) \leq \frac{1}{2}$ or $\mathrm{Area}(C_2) \leq \frac{1}{2}$. Assume that $\mathrm{Area}(C_1) \leq \frac{1}{2}$, so that $\mathrm{Area}(C_1') \geq \frac{1}{2}$. We have

$$|c_1| = |\overline{P_1P_5}| \leq |\overline{P_1P_6'}| + |\overline{P_6'P_5}| \leq \tfrac{2}{3}L_1 + \tfrac{1}{3}L_1 = L_1.$$

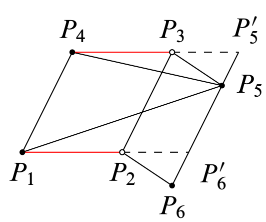


Figure 18: Cylinder with bounded circumference and area at least $\frac{1}{2}$ in a slit torus

Thus we can conclude that C_1' satisfies the statement of the lemma. In the case that $\mathrm{Area}(C_2) \leq \frac{1}{2}$, the same argument shows that the complement C_2' of C_2 has the required properties. The proof of the lemma is now complete. □

References

[1] **T Aougab**, *Uniform hyperbolicity of the graphs of curves*, Geom. Topol. 17 (2013) 2855–2875 MR

[2] **M Bestvina**, **K Fujiwara**, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. 6 (2002) 69–89 MR

[3] **F Bonahon**, *Bouts des variétés hyperboliques de dimension 3*, Ann. of Math. 124 (1986) 71–158 MR

[4] **F Bonahon**, *The geometry of Teichmüller space via geodesic currents*, Invent. Math. 92 (1988) 139–162 MR

[5] **B H Bowditch**, *Hyperbolic 3–manifolds and the geometry of the curve complex*, from "European Congress of Mathematics" (A Laptev, editor), Eur. Math. Soc., Zürich (2005) 103–115 MR

[6] **B H Bowditch**, *Intersection numbers and the hyperbolicity of the curve complex*, J. Reine Angew. Math. 598 (2006) 105–129 MR

[7] **B H Bowditch**, *Uniform hyperbolicity of the curve graphs*, Pacific J. Math. 269 (2014) 269–280 MR

[8] **J F Brock**, **R D Canary**, **Y N Minsky**, *The classification of Kleinian surface groups, II: The ending lamination conjecture*, Ann. of Math. 176 (2012) 1–149 MR

[9] **P Buser**, *Geometry and spectra of compact Riemann surfaces*, Progress in Mathematics 106, Birkhäuser, Boston (1992) MR

[10] **K Calta**, *Veech surfaces and complete periodicity in genus two*, J. Amer. Math. Soc. 17 (2004) 871–908 MR

[11] **K Calta**, **J Smillie**, *Algebraically periodic translation surfaces*, J. Mod. Dyn. 2 (2008) 209–248 MR

[12] **M Clay**, **K Rafi**, **S Schleimer**, *Uniform hyperbolicity of the curve graph via surgery sequences*, Algebr. Geom. Topol. 14 (2014) 3325–3344 MR

[13] **M Duchin**, **C J Leininger**, **K Rafi**, *Length spectra and degeneration of flat metrics*, Invent. Math. 182 (2010) 231–277 MR

[14] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR

[15] **R H Gilman**, *On the definition of word hyperbolic groups*, Math. Z. 242 (2002) 529–541 MR

[16] **U Hamenstädt**, *Train tracks and the Gromov boundary of the complex of curves*, from "Spaces of Kleinian groups" (Y N Minsky, M Sakuma, C Series, editors), London Math. Soc. Lecture Note Ser. 329, Cambridge Univ. Press (2006) 187–207 MR

[17] **U Hamenstädt**, *Geometry of the complex of curves and of Teichmüller space*, from "Handbook of Teichmüller theory, I" (A Papadopoulos, editor), IRMA Lect. Math. Theor. Phys. 11, Eur. Math. Soc., Zürich (2007) 447–467 MR

[18] **U Hamenstädt**, *Stability of quasi-geodesics in Teichmüller space*, Geom. Dedicata 146 (2010) 101–116 MR

[19] **J L Harer**, *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent. Math. 84 (1986) 157–176 MR

[20] **W J Harvey**, *Boundary structure of the modular group*, from "Riemann surfaces and related topics: proceedings of the 1978 Stony Brook conference" (I Kra, B Maskit, editors), Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 245–251 MR

[21] **S Hensel**, **P Przytycki**, **R C H Webb**, *1–slim triangles and uniform hyperbolicity for arc graphs and curve graphs*, J. Eur. Math. Soc. 17 (2015) 755–762 MR

[22] **P Hubert**, **E Lanneau**, *Veech groups without parabolic elements*, Duke Math. J. 133 (2006) 335–346 MR

[23] **P Hubert**, **T A Schmidt**, *Infinitely generated Veech groups*, Duke Math. J. 123 (2004) 49–69 MR

[24] **N V Ivanov**, *Automorphism of complexes of curves and of Teichmüller spaces*, Internat. Math. Res. Not. 1997 (1997) 651–666 MR

[25] **R Kenyon**, **J Smillie**, *Billiards on rational-angled triangles*, Comment. Math. Helv. 75 (2000) 65–108 MR

[26] **S Kerckhoff**, **H Masur**, **J Smillie**, *Ergodicity of billiard flows and quadratic differentials*, Ann. of Math. 124 (1986) 293–311 MR

[27] **E Klarreich**, *The boundary at infinity of the curve complex and the relative Teichmüller space*, preprint (1999) Available at `http://www.math.unicaen.fr/~levitt/klarreich.pdf`

[28] **M Kontsevich**, **A Zorich**, *Connected components of the moduli spaces of Abelian differentials with prescribed singularities*, Invent. Math. 153 (2003) 631–678 MR

[29] **E Lanneau**, **D-M Nguyen**, *Connected components of Prym eigenform loci in genus three*, Math. Ann. (online publication April 2017) 41 pages

[30] **R Lehnert**, *On the critical exponent of infinitely generated Veech groups*, Math. Ann. (online publication August 2016) 42 pages

[31] **F Luo**, *Automorphisms of the complex of curves*, Topology 39 (2000) 283–298 MR

[32] **H Masur**, *Closed trajectories for quadratic differentials with an application to billiards*, Duke Math. J. 53 (1986) 307–314 MR

[33] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. 138 (1999) 103–149 MR

[34] **H Masur**, **S Schleimer**, *The geometry of the disk complex*, J. Amer. Math. Soc. 26 (2013) 1–62 MR

[35] **H Masur**, **S Tabachnikov**, *Rational billiards and flat structures*, from "Handbook of dynamical systems, 1A" (B Hasselblatt, A Katok, editors), North-Holland, Amsterdam (2002) 1015–1089 MR

[36] **C T McMullen**, *Teichmüller geodesics of infinite complexity*, Acta Math. 191 (2003) 191–223 MR

[37] **C T McMullen**, *Teichmüller curves in genus two: discriminant and spin*, Math. Ann. 333 (2005) 87–130 MR

[38] **C T McMullen**, *Dynamics of $\mathrm{SL}_2(\mathbb{R})$ over moduli space in genus two*, Ann. of Math. 165 (2007) 397–456 MR

[39] **Y N Minsky**, *Teichmüller geodesics and ends of hyperbolic 3–manifolds*, Topology 32 (1993) 625–647 MR

[40] **M Möller**, *Affine groups of flat surfaces*, from "Handbook of Teichmüller theory, II" (A Papadopoulos, editor), IRMA Lect. Math. Theor. Phys. 13, Eur. Math. Soc., Zürich (2009) 369–387 MR

[41] **D-M Nguyen**, *Parallelogram decompositions and generic surfaces in $\mathcal{H}^{\mathrm{hyp}}(4)$*, Geom. Topol. 15 (2011) 1707–1747 MR

[42] **D-M Nguyen**, *On the topology of $\mathcal{H}(2)$*, Groups Geom. Dyn. 8 (2014) 513–551 MR

[43] **K Rafi**, *Hyperbolicity in Teichmüller space*, Geom. Topol. 18 (2014) 3025–3053 MR

[44] **M Rees**, *An alternative approach to the ergodic theory of measured foliations on surfaces*, Ergodic Theory Dynamical Systems 1 (1981) 461–488 MR

[45] **J Smillie**, *The dynamics of billiard flows in rational polygons*, from "Dynamical systems, ergodic theory and applications" (Y G Sinai, editor), 2nd edition, Encyclopaedia of Mathematical Sciences 100, Springer (2000) 360–382 MR

[46] **J Smillie**, **B Weiss**, *Minimal sets for flows on moduli space*, Israel J. Math. 142 (2004) 249–260 MR

[47] **J Smillie**, **B Weiss**, *Characterizations of lattice surfaces*, Invent. Math. 180 (2010) 535–557 MR

[48] **M Troyanov**, *Prescribing curvature on compact surfaces with conical singularities*, Trans. Amer. Math. Soc. 324 (1991) 793–821 MR

[49] **W A Veech**, *Geometric realizations of hyperelliptic curves*, from "Algorithms, fractals, and dynamics" (Y Takahashi, editor), Plenum (1995) 217–226 MR

[50] **Y Vorobets**, *Periodic geodesics on generic translation surfaces*, from "Algebraic and topological dynamics" (S Kolyada, Y Manin, T Ward, editors), Contemp. Math. 385, Amer. Math. Soc. (2005) 205–258 MR

[51] **A Wright**, *Cylinder deformations in orbit closures of translation surfaces*, Geom. Topol. 19 (2015) 413–438 MR

[52] **A Wright**, *Translation surfaces and their orbit closures: an introduction for a broad audience*, EMS Surv. Math. Sci. 2 (2015) 63–108 MR

[53] **A Zorich**, *Flat surfaces*, from "Frontiers in number theory, physics, and geometry, I" (P Cartier, B Julia, P Moussa, P Vanhove, editors), Springer (2006) 437–583 MR

Institut de Mathématiques de Bordeaux, Université de Bordeaux, CNRS UMR 5251
351, Cours de la Libération, F-33405 Talence Cedex, France

duc-manh.nguyen@math.u-bordeaux.fr

Received: 31 March 2016 Revised: 30 September 2016

Algebraic & Geometric Topology 17 (2017) 2239–2282

The diagonal slice of Schottky space

CAROLINE SERIES
SER PEOW TAN
YASUSHI YAMASHITA

An irreducible representation of the free group on two generators X, Y into $\mathrm{SL}(2,\mathbb{C})$ is determined up to conjugation by the traces of X, Y and XY. If the representation is faithful and discrete, the resulting manifold is in general a genus-2 handlebody. We study the diagonal slice of the representation variety in which $\operatorname{Tr} X = \operatorname{Tr} Y = \operatorname{Tr} XY$. Using the symmetry, we are able to compute the Keen–Series pleating rays and thus fully determine the locus of faithful discrete representations. We also computationally determine the "Bowditch set" consisting of those parameter values for which no primitive elements in $\langle X, Y\rangle$ have traces in $[-2,2]$, and at most finitely many primitive elements have traces with absolute value at most 2. The graphics make clear that this set is both strictly larger than, and significantly different from, the discreteness locus.

30F40; 57M50

1 Introduction

It is well known that an irreducible representation ρ of the free group F_2 on two generators X, Y into $\mathrm{SL}(2,\mathbb{C})$ is determined up to conjugation by the traces of $\rho(X)$, $\rho(Y)$ and $\rho(XY)$. More generally, if we take the GIT quotient of all (not necessarily irreducible) representations, then the resulting $\mathrm{SL}(2,\mathbb{C})$ character variety of F_2 can be identified with $\mathbb{C}^3$ via these traces using an old result of Vogt; see for example Goldman [10]. If the representation is faithful, discrete, purely loxodromic and geometrically finite, the resulting manifold is a genus-2 handlebody; see Section 3. The collection of all such representations is known as *Schottky space*, denoted by $\mathcal{SCH}$. It is a consequence of Bers' density theorem that $\mathcal{SCH}$ is the interior of the faithful discreteness locus; see for example Canary [4]. It is natural to ask: for which values of $x = \operatorname{Tr}\rho(X)$, $y = \operatorname{Tr}\rho(Y)$, $z = \operatorname{Tr}\rho(XY)$ is the corresponding representation in $\mathcal{SCH}$?

Let $\mathcal{P}$ denote the set of primitive elements of F_2 modulo conjugation and inverse. For $(x,y,z) \in \mathbb{C}^3$, let $\rho_{(x,y,z)}$ denote a choice of representation $F_2 \to \mathrm{SL}(2,\mathbb{C})$ in the conjugacy class determined by the trace triple. The *Bowditch set* (or BQ–set) $\mathcal{B}$ is defined in Tan, Wong and Zhang [30] as the set of $(x,y,z)\in\mathbb{C}^3$ corresponding to irreducible representations for which

$$\operatorname{Tr}\rho_{(x,y,z)}(g)\notin[-2,2] \text{ for all } g\in\mathcal{P} \quad\text{and}\quad \{g\in\mathcal{P}:|\operatorname{Tr}\rho_{(x,y,z)}(g)|\le 2\} \text{ is finite.}$$

(The exceptional case in which $\operatorname{Tr}\rho([X,Y])=2$ corresponds to reducible representations and is excluded from the discussion; see Remark 2.1.) The Bowditch set is open, and the set of outer automorphisms $\operatorname{Out}(F_2)$ of F_2 acts properly discontinuously on it. Thus it is essentially the domain of discontinuity for the mapping class group acting on traces of primitive words. Clearly, $\mathcal{SCH}\subset\mathcal{B}$.

Bowditch's original work [3] was on the case in which the image of the commutator $[X,Y]=XYX^{-1}Y^{-1}$ is parabolic and $\operatorname{Tr}\rho([X,Y])=-2$. He conjectured that the subsets of $\mathcal{SCH}$ and $\mathcal{B}$ corresponding to this restriction coincide. Although this has not been proven, computer pictures indicate his conjecture may well be true.

In this paper, we restrict to the special case in which $x=y=z$, which we call the *diagonal slice* of the character variety, denoted by Δ and parametrised by the single complex variable x. We show that in this slice, the analogue of Bowditch's conjecture is far from being true. This is illustrated in Figure 1, which compares the intersections of Δ with $\mathcal{SCH}$ and $\mathcal{B}$. The discreteness locus is the outer region foliated by rays; these are the Keen–Series pleating rays which relate to the geometry of the convex hull boundary as explained in Section 4.2 and whose closure is known to be $\overline{\Delta\cap\mathcal{SCH}}$; see Theorem 4.23. The Bowditch set, by contrast, is the complement of the black part. It is clear that $\mathcal{B}\cap\Delta$ contains a large open region not in $\Delta\cap\mathcal{SCH}$, and also has different symmetries. In particular, it is not hard to show that the interval $(2,3)$ is contained in $\mathcal{B}\setminus\mathcal{SCH}$; see the discussion in Section 2.2.2.

We would like to emphasise that there are two problems at issue here; namely, to find the locus of discrete faithful representations, and to find the domain of discontinuity for the automorphism group $\operatorname{Out}(F_2)$ acting on traces of primitive words. Both of these problems are quite difficult and subtle with not many previous results. Moreover, while there appeared to be some evidence from earlier studies that the two sets might, modulo some minor caveats, coincide, our results indicate that on the contrary they are unlikely to be related, or at least that their relationship is not obvious.

The main content of this paper is an explanation and justification of how these plots were made, in particular, to explain how we enumerated and computed the pleating rays for the symmetric genus-2 handlebody corresponding to the trace triple (x,x,x).

To compute the Bowditch set $\mathcal{B}$ we use an algorithm based on the ideas in [3] and developed further in [30]. This is explained in Section 2.2.1.

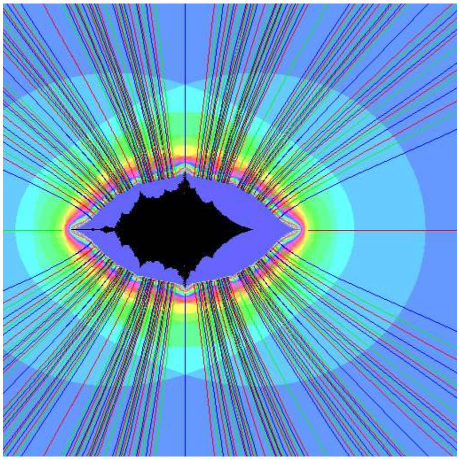


Figure 1: Superposition of the discreteness locus for $\pi_1(\mathcal{H})$ and the Bowditch set in the x–plane. The Bowditch set for the (x,x,x)–triple is the complement of the central black region, while the discreteness locus is the closure of the region foliated by rays. The rays are actually computed as the pleating rays for the quotient orbifold $\mathcal{S}$. The vertical ray is at $x=\frac{1}{2}$, and the discreteness locus intersects $\mathbb{R}$ in $(-\infty,-2]$ and $[3,\infty)$; see Section 4.5.

The discreteness problem is tackled as follows. If $(x,x,x)\in\mathcal{SCH}$, then the quotient 3–manifold $\mathbb{H}^3/G$ is a handlebody $\mathcal{H}$ with order-3 symmetry. We use the symmetry to reduce the problem of finding $\Delta\cap\mathcal{SCH}$ to a problem very similar to that of determining the so-called *Riley slice of Schottky space*. This is actually a space of groups on the boundary of $\mathcal{SCH}$, consisting of those free, discrete and geometrically finite groups for which the two generators $\rho(X)$, $\rho(Y)$ are parabolic, thus contained in the slice $(2,2,z)\subset\mathbb{C}^3$. The corresponding manifold is a handlebody whose conformal boundary is a sphere with four parabolic points. The problem of finding those z–values for which such a group is free, discrete and geometrically finite was solved using the method of pleating rays in Keen and Series [15]. In the present case, the quotient of $\mathcal{H}$ by the symmetry is an orbifold $\mathcal{S}$ with two order-3 cone axes, whose conformal boundary is a sphere with four order-3 cone points. Thus similar methods enable us to find $\Delta\cap\mathcal{SCH}$ here.

Although Figure 1 shows that in Δ, the analogue of Bowditch's conjecture fails since $\mathcal{B}$ and the interior of the discreteness locus are plainly distinct, in many other slices (see for example Figure 8), the (modified) Bowditch set and the interior of the discreteness locus appear to coincide. This is connected to the dynamics of the action of a suitable mapping class group on representations and raises many interesting questions which we hope to address elsewhere.

The plan of the paper is as follows. We begin in Section 2 with a discussion of the Markoff tree and the algorithm used to compute the Bowditch set. In Section 3, we introduce a basic geometrical construction which conveniently encapsulates the 3–fold symmetry. The quotient of the original handlebody $\mathcal{H}$ by the symmetry is a ball with two order-3 cone axes. This orbifold $\mathcal{S}$ has a further 4–fold symmetry group whose quotient is again a topological ball. Our construction allows us to write down specific $\mathrm{SL}(2,\mathbb{C})$ representations (in some cases, $\mathrm{PSL}(2,\mathbb{C})$ representations; see the discussion in Section 3.1 and in particular Remark 3.2) of all the groups involved with ease. In Section 4, we turn to the discreteness question. After reducing the problem to one on $\mathcal{S}$, we briefly review material from the Keen–Series theory of pleating rays and recall what is needed from [15], allowing us to apply a similar proof in the present context. Section 5, not strictly logically necessary for our development, explains how we did our trace computations in practice, by relating the problem to one on a commensurable torus with a single cone point of angle $\frac{4\pi}{3}$.

Acknowledgements We would like to thank the referee for a very careful reading of the paper which resulted in significant improvements.

Tan is partially supported by the National University of Singapore academic research grant R-146-000-186-112. Yamashita is partially supported by JSPS KAKENHI grant number 23540088.

2 The Markoff tree and the Bowditch set

Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{C})$ so that $ad-bc=1$. As usual we define its trace $\mathrm{Tr}\, A = a+d$.

Let $F_2 = \langle X, Y \mid - \rangle$ be the free group on two generators. It is well known that a representation $\rho\colon F_2 \to \mathrm{SL}(2,\mathbb{C})$ is determined up to conjugation (modulo taking the GIT quotient under the conjugation action; see [10]) by the three traces $x = \mathrm{Tr}\,\rho(X)$, $y = \mathrm{Tr}\,\rho(Y)$, $z = \mathrm{Tr}\,\rho(XY)$. In fact, given $x, y, z \in \mathbb{C}$, we can define a representation

$$\rho_{x,y,z}\colon F_2 \to \mathrm{SL}(2,\mathbb{C}), \quad \rho(X) = \begin{pmatrix} x & 1 \\ -1 & 0 \end{pmatrix}, \ \rho(Y) = \begin{pmatrix} 0 & \eta \\ -\eta^{-1} & y \end{pmatrix},$$

where $z = -(\eta+\eta^{-1})$; see [9]. Clearly, with this definition, $\mathrm{Tr}\,\rho(X) = x$, $\mathrm{Tr}\,\rho(Y) = y$ and $\mathrm{Tr}\,\rho(XY) = z$.

2.1 The Markoff tree

For matrices $\widehat{U}, \widehat{V} \in \mathrm{SL}(2,\mathbb{C})$, set $u = \mathrm{Tr}\,\widehat{U}$, $v = \mathrm{Tr}\,\widehat{V}$, $w = \mathrm{Tr}\,\widehat{U}\widehat{V}$ (where we use the notation $\widehat{U}, \widehat{V}$ to distinguish from generators U, V of F_2). Recall the trace

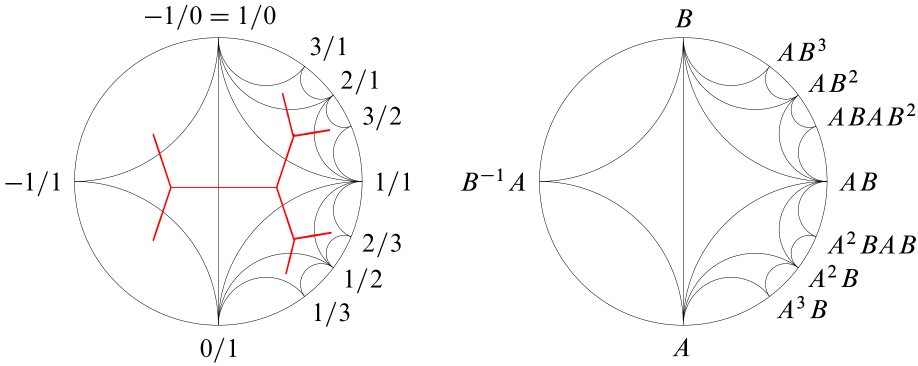


Figure 2: The Farey diagram, showing the arrangement of rational numbers on the left with the corresponding primitive words on the right

relations

$$\operatorname{Tr}\widehat{U}\widehat{V}^{-1} = uv - w, \tag{2-1}$$

$$u^2 + v^2 + w^2 = uvw + \operatorname{Tr}[\widehat{U}, \widehat{V}] + 2. \tag{2-2}$$

Setting $\mu = \operatorname{Tr}[\widehat{U}, \widehat{V}] + 2$, this last equation takes the form

$$u^2 + v^2 + w^2 - uvw = \mu.$$

Let $F_2 = \langle X, Y \mid - \rangle$ as above. An element $U \in F_2$ is *primitive* if it is a member of a generating pair; we denote the set of all primitive elements by $\mathcal{P}$. The conjugacy classes of primitive elements are enumerated by $\widehat{\mathbb{Q}} = \mathbb{Q} \cup \infty$ and are conveniently organised relative to the Farey diagram $\mathcal{F}$ as shown in Figure 2. This consists of the images of the ideal triangle with vertices at $1/0, 0/1$ and $1/1$ under the action of $\mathrm{SL}(2,\mathbb{Z})$ on the upper half plane, suitably conjugated to the position shown in the disk. The label p/q in the disk is just the conjugated image of the actual point $p/q \in \mathbb{R}$.

Since the rational points are precisely the images of ∞ under $\mathrm{SL}(2,\mathbb{Z})$, they correspond bijectively to the vertices of $\mathcal{F}$. A pair $p/q, r/s \in \widehat{\mathbb{Q}}$ are the endpoints of an edge if and only if $pr - qs = \pm 1$; such pairs are called *neighbours*. A triple of points in $\widehat{\mathbb{Q}}$ are the vertices of a triangle precisely when they are the images of the vertices of the initial triangle $(1/0, 0/1, 1/1)$; such triples are always of the form $(p/q, r/s, (p+r)/(q+s))$ where p/q, r/s are neighbours. In other words, if p/q, r/s are the endpoints of an edge, then the vertex of the triangle on the side away from the centre of the disk is found by "Farey addition" to be $(p+r)/(q+s)$. Starting from $1/0 = -1/0 = \infty$ and $0/1$, all points in $\widehat{\mathbb{Q}}$ are obtained recursively in this way. (Note we need to start with $-1/0 = \infty$ to get the negative fractions on the left side of the left-hand diagram in Figure 2.)

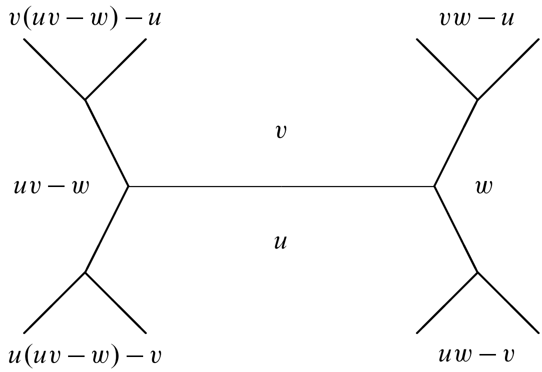


Figure 3: The Markoff tree used to compute traces with an initial triple (u, v, w)

The right-hand picture in Figure 2 shows a corresponding arrangement of primitive elements in F_2, one in each conjugacy class, starting with initial triple (A, B, AB). Each vertex is labelled by a certain cyclically reduced representative of the corresponding word. Pairs of primitive elements form a generating pair if and only if they are at the endpoints of an edge. Triples at the vertices of a triangle correspond to a generator triple of the form (U, V, UV). Corresponding to the process of Farey addition, successive words can be found by juxtaposition as indicated on the diagram. Note that for this to work, it is important to preserve the order: if U, V are the endpoints of an edge with U before V in the anticlockwise order round the circle, the correct concatenation is UV. Note also that the words on the left side of the diagram involve $B^{-1}A$, corresponding to starting with $\infty = -1/0$. We denote the particular representative of the conjugacy class corresponding to $p/q \in \hat{\mathbb{Q}}$ found by concatenation by $W_{p/q}$. Its word length in the generators A, B is a function $F(p/q)$ of p/q. A function on $\hat{\mathbb{Q}}$ is said to have *Fibonacci growth* if it is comparable with uniform upper and lower bounds to F.

In this paper, we are largely interested in computing traces of primitive elements. Following Bowditch [3], these can also be easily computed by using the trivalent tree $\mathbb{T}$ dual to $\mathcal{F}$; see the left frame of Figure 2 and Figure 3. Let Ω denote the set of complementary regions of $\mathbb{T}$; abstractly, a complementary region is the closure of a connected component of the complement of $\mathbb{T}$. As is apparent from Figure 2, there is a bijection between Ω and the set of vertices of $\mathcal{F}$. Thus the set Ω can be identified with conjugacy classes of primitive elements and hence with $\hat{\mathbb{Q}}$.

Given a representation $\rho\colon F_2 \to \mathrm{SL}(2, \mathbb{C})$, each $U \in \Omega$ is labelled by $u = \operatorname{Tr} \rho(U)$, the trace of the corresponding generator, as shown in Figure 3. Labels on opposite sides of an edge of $\mathbb{T}$ correspond to traces of a generator pair: the three labels round a vertex correspond to a generator triple (U, V, UV).

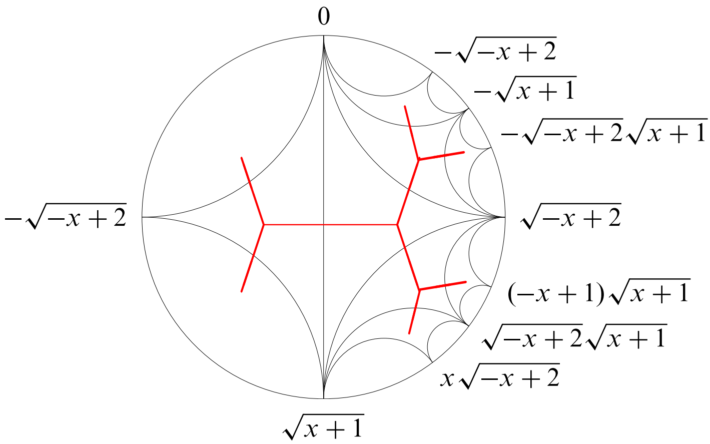


Figure 4: The Farey tessellation used to compute traces. See Section 5.0.2 for a discussion of the choice of sign of the square roots.

Suppose that (U,V,W) are the labels of regions round a vertex with $u=\operatorname{Tr}\rho(U)$, $v=\operatorname{Tr}\rho(V)$, $w=\operatorname{Tr}\rho(W)$. By (2-2), we have $u^2+v^2+w^2-uvw=\mu$. By (2-1), the two vertices opposite the ends of the edge between regions (U,V) correspond to regions UV, UV^{-1} with labels $w, uv-w$, respectively. Moving in this way, along the three edges which meet at the vertex with labels (u,v,w) to the three adjacent vertices, gives rise to the three basic moves $(u,v,w)\to(u,v,uv-w)$, $(u,v,w)\to(u,uw-v,w)$, $(u,v,w)\to(vw-u,v,w)$ which generate traces of all possible elements in Ω (and hence $\mathcal{P}$). Note that any of these three moves leaves $\operatorname{Tr}\rho([U,V])$ and hence μ invariant; in other words, μ is an invariant of the tree. Bowditch's original paper was mostly confined to the case $\mu=0$.

In this way, the Markoff tree provides a fast way to compute traces of elements in $\mathcal{P}$ starting from an initial triple (u,v,w). This is illustrated in Figure 4 with the initial triple $(\sqrt{x+1}, 0, \sqrt{-x+2})$ which is used in Section 5.0.2. We denote the tree of traces associated to an initial triple (u,v,w) by $\mathbb{T}_{(u,v,w)}$. Later we will use a variant of this construction to compute traces of curves on a four-pointed sphere; see Section 4.4.

2.2 The Bowditch set

It is convenient to rephrase the above discussion using the terminology introduced in [3]. As above, let Ω denote the set of complementary regions of the tree $\mathbb{T}$. Define a *Markoff map* to be a map $\phi\colon \Omega\to\mathbb{C}$ such that ϕ satisfies the trace relations (2-1) and (2-2). The set of all Markoff maps is denoted by Φ. Since traces depend only on conjugacy classes, a representation $\rho\colon F_2\to \mathrm{SL}(2,\mathbb{C})$ defines a Markoff map by setting $\phi(U) = \operatorname{Tr}\rho(U)$ for $U \in \Omega$. Fixing once and for all an identification of Ω with $\widehat{\mathbb{Q}}$ (and recalling that Ω is identified with conjugacy classes of elements in $\mathcal{P}$), we have $\phi(p/q) = \operatorname{Tr}\rho(W_{p/q})$ for $p/q \in \widehat{\mathbb{Q}}$, where $W_{p/q}$ is the special word in the conjugacy class corresponding to $p/q \in \Omega$.

Thus as explained above, using the trace relations (2-1) and (2-2), an initial triple $(x, y, z) \in \mathbb{C}^3$ uniquely determines a Markoff map $\phi = \phi_{x,y,z}$ together with a corresponding labelling of $\mathbb{T}$. Conversely a Markoff map $\phi \in \Phi$ determines $(x, y, z) \in \mathbb{C}^3$ by setting $x = \phi(0/1)$, $y = \phi(1/0)$, $z = \phi(1/1)$. In this way, we can identify Φ with $\mathbb{C}^3$. For $\phi \in \Phi$, denote the corresponding tree endowed with the labelling given by ϕ by $\mathbb{T}_\phi = \mathbb{T}_{(\phi(0/1),\phi(1/0),\phi(1/1))}$.

The Bowditch set $\mathcal{B}$ is the set of all $\phi \in \Phi$ with $\mu \neq 4$ which satisfy the conditions

$$\phi(U) \notin [-2, 2] \text{ for all } U \in \Omega, \tag{2-3}$$

$$\{U \in \Omega : |\phi(U)| \le 2\} \text{ is finite.} \tag{2-4}$$

The Bowditch set $\mathcal{B}$ is open in $\mathbb{C}^3$ and $\operatorname{Out}(F_2)$ acts properly discontinuously on $\mathcal{B}$. Furthermore, if $\phi \in \mathcal{B}$, then $\log^+ |\phi(U)| = \max\{0, \log |\phi(U)|\}$ has Fibonacci growth on Ω; see [30].

Remark 2.1 The maps ϕ for which $\mu = 4$ correspond to the reducible representations: our definition above automatically excludes them from $\mathcal{B}$. For such ϕ, there are infinitely many $U \in \Omega$ such that $|\phi(U)| < m$ for $m > 2$, they can alternatively be excluded from $\mathcal{B}$ by relaxing condition (2-4) to the condition that $\{U \in \Omega : |\phi(U)| \le 2 + \epsilon\}$ be finite for any $\epsilon > 0$. As is easily checked from the trace relation (2-2), such representations occur in Δ precisely at the points $x = -1$, $x = 2$.

2.2.1 Background to the algorithm Our algorithm for computing which points lie in $\mathcal{B}$ is based on results from [3; 30] which we summarise here. We consider only ϕ for which $\mu \neq 4$. Following Bowditch [3], we orient the edges of $\mathbb{T}_\phi$ in the following way. Suppose that labels of the regions adjacent to some edge e are u, v, and the labels of the two remaining regions at the two end vertices are w, t; see Figure 3. From the trace relations, $t = uv - w$. Orient e by putting an arrow from t to w whenever $|t| > |w|$ and vice versa. If both moduli are equal, make either choice; if the inequality is strict, say that the edge is *oriented decisively*.

A *sink region* of $\mathbb{T}_\phi$ is a connected nonempty subtree T such that the arrow on any edge not in T points towards T decisively. A sink region may consist of a single *sink vertex* v (the three edges adjacent to v point towards v) and no edges. Clearly a sink region is not unique: one can always add further vertices and edges around the boundary of T.

For any $m \geq 0$ and $\phi \in \Phi$, define $\Omega_\phi(m) = \{U \in \Omega : |\phi(U)| \leq m\}$. The following lemmas from [30] show that $\Omega_\phi(2)$ is connected, and that from any initial vertex not adjacent to regions in $\Omega_\phi(2)$, the arrows determine a descending path through $\mathbb{T}$ which either runs into a sink, or meets vertices adjacent to regions in $\Omega_\phi(2)$. Furthermore, if $\phi(U)$ takes values away from the exceptional set $E = [-2,2] \cup \{\pm\sqrt{\mu}\} \subset \mathbb{C}$, then there exists a finite segment of ∂U such that the edges adjacent to U not in this segment are directed towards this segment.

Lemma 2.2 [30, Lemma 3.7] *Suppose $U, V, W \in \Omega$ meet at a vertex v with the arrows on both the edges adjacent to U pointing away from v. Then either $|\phi(U)| \leq 2$ or $\phi(V) = \phi(W) = 0$.*

Corollary 2.3 [30, Theorem 3.1(2)] *Let $\phi \in \Phi$. Then $\Omega_\phi(2)$ (more generally, $\Omega_\phi(m)$ for $m \geq 2$) is connected.*

Lemma 2.4 [30, Lemma 3.11 and following comment] *Suppose β is an infinite ray consisting of a sequence of edges of $\mathbb{T}_\phi$ all of whose arrows point away from the initial vertex. Then β meets at least one region $U \in \Omega$ with $|\phi(U)| < 2$. Furthermore, if the ray does not follow the boundary of a single region, it meets infinitely many regions with this property.*

Lemma 2.5 [30, Lemma 3.20] *Suppose that $\phi(U) \notin E$, and consider the regions $V_i, i \in \mathbb{Z}$ adjacent to U in order round ∂U. Then away from a finite subset, the values $|\phi(V_i)|$ are increasing and approach infinity as $i \to \infty$ in both directions. Hence there exists a finite segment of ∂U such that the edges adjacent to U not in this segment are directed towards this segment.*

We remark that if $\phi(U) = \pm\sqrt{\mu}$ and $\sqrt{\mu} \notin [-2,2]$, then the values of $|\phi(V_i)|$ in Lemma 2.5 approach zero in one direction round ∂U [30, Lemma 3.10], and hence $\phi \notin \mathcal{B}$ since condition (2-4) will not be satisfied. Hence, for $\phi \in \mathcal{B}$, we have $\phi(U) \notin E$ for all $U \in \Omega$.

The set $\Omega_\phi(2)$ can be used to construct a sink region T (which depends of course on ϕ) which is finite if and only if $\phi \in \mathcal{B}$. Essentially, if $\phi \in \mathcal{B}$, then T consists of finite segments of the boundaries of the (finite number of) elements of $\Omega_\phi(2)$. These are the segments alluded to in Lemma 2.5; they have to be large enough so the conclusion of the lemma holds, and also to contain all edges adjacent to any pair U, V, both of which are in $\Omega_\phi(2)$, so that the union is connected. To do this, an explicit function $H_\mu \colon \mathbb{C} \to \mathbb{R}^+ \cup \{\infty\}$ is constructed (see Lemma 3.20, the following remark and Lemma 3.23 in [30]) as follows:

(1) If $x \in E$, define $H_\mu(x) = \infty$.

(2) For $x \notin E$, let $x = \lambda + \lambda^{-1}$ with $|\lambda| > 1$ (note that $|\lambda| \neq 1$ since $x \notin [-2, 2]$). Define

$$H_\mu(x) = \max\left\{2,\ \sqrt{\left|\frac{x^2-\mu}{x^2-4}\right| \frac{2|\lambda|^2}{|\lambda|-1}}\right\}. \tag{2-5}$$

Then H_μ is continuous on $\mathbb{C} \setminus E$. Now we can define a specific attracting subtree:

Definition 2.6 Given $\phi \in \Phi$, let $T = T_\phi$ be the subset of $\mathbb{T}_\phi$ defined as follows:

(1) An edge with adjacent regions U, V is in T if and only if either $|\phi(U)| \leq 2$ and $|\phi(V)| \leq H_\mu(\phi(U))$, or vice versa.

(2) Any sink vertex is in T, as are any vertices which are the endpoints of two edges in T.

Based on the above lemmas, we have the following theorem (see also the special properties of the function H_μ and Lemmas 3.21–3.24 in [30]).

Theorem 2.7 *Given $\phi \in \Phi$ (with $\mu \neq 4$), the set $T = T_\phi$ in Definition 2.6 is a nonempty, connected subtree of $\mathbb{T}_\phi$. Moreover, T is a sink region for $\mathbb{T}_\phi$; that is, all edges not in T are directed decisively towards T. Furthermore, T is finite if and only if $\phi \in \mathcal{B}$.*

2.2.2 The algorithm Based on the above discussion, our algorithm to decide whether or not $\phi \in \mathcal{B}$ is as follows.

Step 1 Starting at any vertex, follow the direction of decreasing arrows. On reaching a sink vertex, stop. This vertex is in T by Definition 2.6. If the input is $\mathcal{B}$, then this method always finds a sink vertex in finite time because there is a finite sink region. Otherwise, the process may not terminate in (prespecified) finite time, and the algorithm is indecisive.

Step 2 Assuming a stopping point is found in Step 1, starting from this point, search outwards by a depth first search using Definition 2.6 to identify whether or not an edge is in T. This works because of the connectedness of T. If this search terminates in (prespecified) finite time, then $\phi_{x,y,z} \in \mathcal{B}$. Otherwise, the algorithm is indecisive.

Note that if the starting point is a sink vertex and the three adjacent edges are not in T, then T consists of just the sink vertex by the connectedness of T, hence $\phi_{x,y,z} \in \mathcal{B}$. This occurs for example for the tree $\mathbb{T}_{(x,x,x)}$ with $x \in (2, 3)$.

Figure 5 shows the Bowditch set in the diagonal slice Δ as determined by this algorithm.

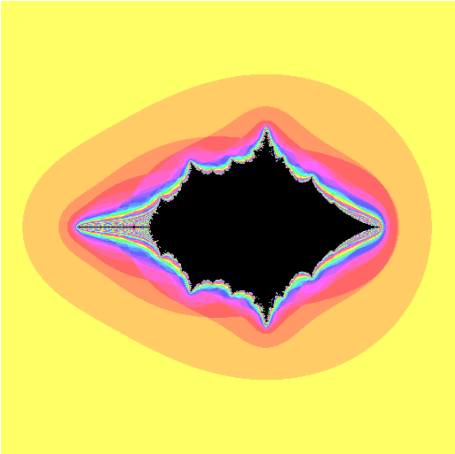


Figure 5: The Bowditch set $\mathcal{B}$ for the Markoff maps $\phi_{(x,x,x)}$, plotted in the x–plane. The grey (coloured) points are in $\mathcal{B}$ and the black ones are outside. The shades of grey (colours) indicate the size of the sink region T.

Remark 2.8 We do not have an algorithm whose output is $\phi_{x,y,z} \notin \mathcal{B}$. When $\mu = 0$, it was shown in [25] that if $|\phi(U)| \le 0.5$ for some $U \in \Omega$, then $\phi_{x,y,z} \notin \mathcal{B}$. Hence in Step 1 above, if $\mu = 0$, we can stop when we hit a region satisfying this condition and conclude that $\phi_{x,y,z} \notin \mathcal{B}$. Using the same methods, a similar upper bound can be found for μ close to 0. In particular, there is a neighbourhood of $(0,0,0)$ which is disjoint from $\mathcal{B}$, as clearly illustrated in Figure 5. However, as shown in [11], no such universal positive bound exists for all μ: precisely, for any $\epsilon > 0$ and $\mu > 4$, there exist $\phi \in \mathcal{B}_\mu$ and $U \in \Omega$ such that $|\phi(U)| < \epsilon$. Another issue is that the sink region may be extremely large so may not be detected in a program with a given finite number of steps, this occurs when we approach the boundary of $\mathcal{B}$. Thus the algorithm is not completely decisive although it appears to give nice results. In particular, there may be false negatives; however points which are determined to be in $\mathcal{B}$ are correctly marked.

3 Groups, manifolds, symmetries and quotients

In this section we detail a construction which allows us conveniently to exploit the three-fold symmetry of groups in the diagonal slice Δ. We denote hyperbolic 3–space by $\mathbb{H}^3$ and identify its group $\mathrm{Isom}^+ \mathbb{H}^3$ of orientation-preserving isometries with $\mathrm{PSL}(2,\mathbb{C})$. As is well known, if the image of a representation ρ: $F_2 \to \mathrm{SL}(2,\mathbb{C})$ is faithful, discrete and geometrically finite without parabolics, then $\mathbb{H}^3/\rho(F_2)$ is a genus-two handlebody $\mathcal{H}$; see [24, Corollary X.H.6] and also [12, Theorem 5.2]. (To apply Hempel's result, note that a hyperbolic 3–manifold is irreducible, hence prime, and that $\pi_2(M) = 0$.) Rather than working with $\mathcal{H}$, however, it is much easier to work with the quotient $\mathcal{S}$ of $\mathcal{H}$ by the order-3 symmetry κ corresponding to cyclic permutation of the parameters. We also introduce a commensurable orbifold $\mathcal{T}$ with a torus boundary $\partial\mathcal{T}$.

Both $\mathcal{S} = \mathcal{H}/\kappa$ and $\mathcal{T}$ surject to a 3–orbifold $\mathcal{U}$ with fundamental group a so-called (P, Q, R)*–group*. Its boundary $\partial\mathcal{U}$ is a sphere with three order-2 and one order-3 cone points. A similar construction has been used extensively by Akiyoshi et al (see for example [1]), and is the basis of Wada's program OPTi [31; 32], hence was convenient for our computations. In this section we explain these constructions in detail, using them to find explicit representations of all four groups.

3.1 The handlebody and related orbifolds

The symmetric handlebody $\mathcal{H}$ can be thought of as made by gluing two solid pairs of pants each with order-3 symmetry. More precisely, take a 3–ball and choose three closed disks on the boundary, placed so as to have order-3 rotational symmetry. Gluing two such balls along the closed disks produces a handlebody $\mathcal{H}$ with the required order-three symmetry κ. Rather than write down a suitably symmetric representation of $\pi_1(\mathcal{H})$ directly, we consider first the quotient orbifold $\mathcal{S} = \mathcal{H}/\kappa$. As will be justified in retrospect when we have identified the representations explicitly, this is a ball with two cone axes around each of which the angle is $\frac{2\pi}{3}$. Its boundary $\partial\mathcal{S}$ is a sphere $\Sigma_{0;3,3,3,3}$ with four order-3 cone points. We will call $\mathcal{S}$ the *large coned ball*.

The ball $\mathcal{S}$ has a further order-4 symmetry group. Consider the two cone axes which form the singular locus of $\mathcal{S}$, together with their common perpendicular. Lifting to $\mathbb{H}^3$, we obtain a configuration invariant under the π–rotation about C, the common perpendicular to the two lifted cone axes, and also under π–rotations about a unique pair of orthogonal lines in the plane orthogonal to C passing through its midpoint O; see Section 3.2.1. Denoting these latter rotations $\bar{P}, \bar{Q} \in \mathrm{Isom}^+\,\mathbb{H}^3$, the π–rotation about C is $\bar{P}\bar{Q}$ and the entire configuration is invariant under $\langle \bar{P}, \bar{Q}\rangle = \mathbb{Z}_2 \times \mathbb{Z}_2$. Thus we obtain a further quotient orbifold $\mathcal{U} = \mathcal{S}/(\mathbb{Z}_2 \times \mathbb{Z}_2)$, also topologically a ball, which we call the *small coned ball*. The singular locus of $\mathcal{U}$ is as follows. Let $\hat{O}$ and $\hat{E}$ be the images in $\mathcal{U}$ of the midpoint O of C and the point where C meets the axis of $\bar{K}$, respectively, where $\bar{K} \in \mathrm{Isom}^+\,\mathbb{H}^3$ is one of the two order-three rotations about the pair of lifted cone axes. Let $\hat{C}$ be the image in $\mathcal{U}$ of C, so that $\hat{C}$ is a line from $\hat{O}$ to $\hat{E}$. From $\hat{O}$ emanate three mutually orthogonal lines corresponding to the order-2 axes of $\bar{P}$, $\bar{Q}$ and $\bar{P}\bar{Q}$. One of these is the line $\hat{C}$ corresponding to $\bar{P}\bar{Q}$ which ends at $\hat{E}$. From $\hat{E}$ also emanates an order-3 singular line, the axis of $\bar{\bar{K}}$, perpendicular to $\hat{C}$. The boundary $\partial\mathcal{U}$ is a sphere $\Sigma_{0;2,2,2,3}$ with 3 cone points of order 2 and one of order 3. The order-3 cone point is the image of the endpoint of the order-3 singular line and the order-2 cone points correspond to the endpoints on $\partial\mathcal{U}$ of the axes of $\bar{P}$, $\bar{Q}$ and a third involution $\bar{R}$ defined below.

Finally, there is a double cover of the small coned ball $\mathcal{U}$ by an cone manifold $\mathcal{T}$ which is topologically a solid torus. Its boundary is a torus $\partial\mathcal{T}$ with a single cone point of angle $\frac{4\pi}{3}$. Just as the quotient of a once punctured torus $\Sigma_{1;\infty}$ by the hyperelliptic involution is the surface $\Sigma_{0;2,2,2,\infty}$, so the quotient of $\partial\mathcal{T}$ by the hyperelliptic involution ι is the surface $\partial\mathcal{U} = \Sigma_{0;2,2,2,3}$. The involution ι extends to an involution, also denoted by ι, of $\mathcal{T}$ such that $\mathcal{T}/\iota = \mathcal{U}$.

The group $\pi_1(\mathcal{U})$ is generated by $(\bar{P}, \bar{Q}, \bar{K})$, where $(\bar{P}, \bar{Q}, \bar{K})$ are regarded as elements of $\mathrm{Isom}^+ \mathbb{H}^3 = \mathrm{PSL}(2,\mathbb{C})$. We can replace $\bar{K}$ by a further involution $\bar{R}$ such that $\bar{R}\bar{Q}\bar{P} = \bar{K}$. To do this, let $\bar{R}$ be an order-2 rotation about an axis contained in the plane through E orthogonal to $\mathrm{Ax}\,\bar{K}$, such that the axis makes an angle $\frac{1}{3}\pi$ with C. (We will fix orientations more precisely below.) Then $\bar{R}(\bar{Q}\bar{P})$ is a $\frac{2\pi}{3}$–rotation about $\mathrm{Ax}\,\bar{K}$, in other words, provided orientations have been chosen correctly, we can identify $\pi_1(\mathcal{U})$ with a group

$$\langle \bar{P}, \bar{Q}, \bar{R} \mid \bar{P}^2 = \bar{Q}^2 = \bar{R}^2 = (\bar{R}\bar{Q}\bar{P})^3 = \mathrm{id},\ \bar{P}\bar{Q} = \bar{Q}\bar{P} \rangle \subset \mathrm{PSL}(2,\mathbb{C}).$$

As discussed in Remarks 3.1 and 3.2 below, the above group $\pi_1(\mathcal{U})$ cannot be lifted to a subgroup of $\mathrm{SL}(2,\mathbb{C})$ since it contains elements of order two. Nevertheless, we shall find lifts $P, Q, R \in \mathrm{SL}(2,\mathbb{C})$ of $\bar{P}, \bar{Q}, \bar{R} \in \mathrm{PSL}(2,\mathbb{C})$ for which

$$\Gamma_{\mathcal{U}} = \langle P, Q, R \mid P^2 = Q^2 = R^2 = (RQP)^3 = -\mathrm{id},\ PQ = -QP \rangle \subset \mathrm{SL}(2,\mathbb{C}),$$

so that $\Gamma_{\mathcal{U}}$ projects to $\pi_1(\mathcal{U})$.

To do this, we recall that in [1] and other papers by the same authors, groups generated by three involutions $P, Q, R \in \mathrm{SL}(2,\mathbb{C})$ with RQP parabolic, are used as a convenient way of parametrising representations of once punctured tori, where the torus in question is now a two-fold cover of the orbifold with fundamental group $\langle P, Q, R \rangle$ with quotient induced by the hyperelliptic involution. A small modification of their parametrisation allows us to write down a convenient general form for a representation of the group $\Gamma_{\mathcal{U}}$ with the presentation above into $\mathrm{SL}(2,\mathbb{C})$, from which we obtain explicit $\mathrm{SL}(2,\mathbb{C})$ representations of $\pi_1(\mathcal{H})$, $\pi_1(\mathcal{S})$, together with groups in $\mathrm{SL}(2,\mathbb{C})$ which project to $\mathrm{PSL}(2,\mathbb{C})$ representations of $\pi_1(\mathcal{U})$ and $\pi_1(\mathcal{T})$ as above. This we do in the next section.

3.2 The basic configuration and the small coned ball

We start with a general construction for representations $\Gamma_{\mathcal{U}} \to \mathrm{SL}(2,\mathbb{C})$, that is, of subgroups $\langle P, Q, R \mid P^2 = Q^2 = R^2 = (RQP)^3 = -\mathrm{id},\ PQ = -QP \rangle \subset \mathrm{SL}(2,\mathbb{C})$.

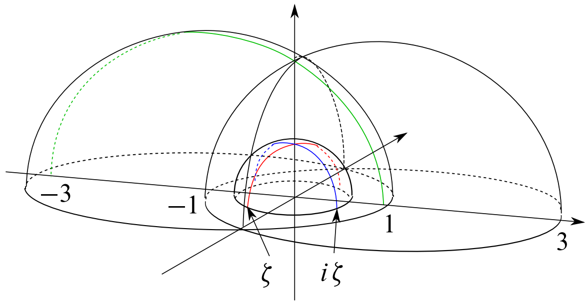


Figure 6: The basic configuration for the (P, Q, R)–group $\pi_1(\mathcal{U})$

For convenience we refer to such a group (or its image in $\mathrm{PSL}(2,\mathbb{C})$) as a (P, Q, R)–*group*. The elements P, Q, R, K we construct will project to the $\mathrm{PSL}(2,\mathbb{C})$ elements $\overline{P}, \overline{Q}, \overline{R}, \overline{K}$ discussed above.

We will make our calculations using *line matrices* following [8]. Note this will define representations into $\mathrm{SL}(2,\mathbb{C})$, thus fixing the signs of traces. Let $u, u' \in \widehat{\mathbb{C}}$, and denote the oriented line from u to u' by $[u, u']$. The associated line matrix $M([u,u']) \in \mathrm{SL}(2,\mathbb{C})$ is a matrix which induces an order two rotation about $[u,u']$ and such that $M([u,u'])^2 = -\mathrm{id}$, so that in particular,

$$M([0,\infty]) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

By [8, page 64, equation (1)], we have, if $u, u' \in \mathbb{C}$,

$$M([u,u']) = \frac{i}{u'-u}\begin{pmatrix} u+u' & -2uu' \\ 2 & -u-u' \end{pmatrix}.$$

The representation we require is derived from a basic configuration shown in Figure 6. It depends on a single parameter $\zeta \in \mathbb{C}$ which we will relate to the original parameter x in Section 3.2.3 below.

Let $\zeta \in \mathbb{C}$ and $P, Q, R \in \mathrm{SL}(2,\mathbb{C})$ be π–rotations about the oriented lines $[\zeta, -\zeta]$, $[i\zeta, -i\zeta]$ and $[1, -3]$, respectively. By construction $P^2 = Q^2 = R^2 = -\mathrm{id}$. Moreover, $\mathrm{Ax}\,P$ and $\mathrm{Ax}\,Q$ intersect at the point $|\zeta|j \in \mathbb{H}^3$ on the hemisphere of radius $|\zeta|$ and centre $0 \in \mathbb{C}$, where $z + tj$ represents the point at height $t > 0$ above $z \in \mathbb{C}$ in the upper half space model of $\mathbb{H}^3$. Thus $PQ = -QP$ and PQ is an order-2 rotation about the vertical axis $0 + tj, t > 0$.

Let V be the vertical plane above the real axis in $\mathbb{H}^3$. Note that the oriented axes of the order two rotations PQ and R both lie in V, intersecting in the point $\sqrt{3}j$ at angle $\frac{1}{3}\pi$. The line $[\sqrt{3}i, -\sqrt{3}i]$ passes through this point and is orthogonal to V. It follows that $RPQ = -RQP$ is anticlockwise rotation through $\frac{2\pi}{3}$ about the line $[\sqrt{3}i, -\sqrt{3}i]$. Using line matrices as above, we can now easily write down the corresponding representation in $\mathrm{SL}(2,\mathbb{C})$:

$$\begin{aligned} P = M([\zeta,-\zeta]) &= -\frac{i}{2\zeta}\begin{pmatrix}0 & 2\zeta^2\\ 2 & 0\end{pmatrix} = \begin{pmatrix}0 & -i\zeta\\ -i/\zeta & 0\end{pmatrix},\\ Q = M([i\zeta,-i\zeta]) &= -\frac{1}{2\zeta}\begin{pmatrix}0 & -2\zeta^2\\ 2 & 0\end{pmatrix} = \begin{pmatrix}0 & \zeta\\ -1/\zeta & 0\end{pmatrix},\\ R = M([1,-3]) &= -\frac{i}{4}\begin{pmatrix}-2 & 6\\ 2 & 2\end{pmatrix} = \begin{pmatrix}i/2 & -3i/2\\ -i/2 & -i/2\end{pmatrix}.\end{aligned}$$

Let $K = RPQ$. Then

$$K = \begin{pmatrix}-1/2 & -3/2\\ 1/2 & -1/2\end{pmatrix}, \quad K^3 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},$$

so that as expected, K is a anticlockwise rotation about $[\sqrt{3}i, -\sqrt{3}i]$ by $\frac{2\pi}{3}$.

Note that $P^2 = Q^2 = -\mathrm{id}$ and $PQ = -QP$ as matrices in $\mathrm{SL}(2,\mathbb{C})$. As isometries of $\mathbb{H}^3$, the signs are irrelevant. We could have chosen $K = RQP$, in which case $K^3 = -\mathrm{id}$, but see Remark 3.1 below. We denote the group generated by P, Q, R by $G_{\mathcal{U}}(\zeta)$ and the corresponding representation $\Gamma_{\mathcal{U}} \to \mathrm{SL}(2,\mathbb{C})$ by $\rho_{\mathcal{U}}(\zeta)$.

3.2.1 The large coned ball $\mathcal{S}$ To relate $\pi_1(\mathcal{U})$ to $\pi_1(\mathcal{S})$, start with two oriented axes A_0, A_1 about each of which we have order-3 anticlockwise rotations K_0, K_1, measured with respect to the orientation of the axes. Let C denoted the common perpendicular between A_0 and A_1, oriented from A_0 to A_1. We denote this configuration, which is clearly well defined up to isometry, by $\mathcal{CF}$. As described in Section 3.1, $\mathcal{CF}$ has a further $\mathbb{Z}_2 \times \mathbb{Z}_2$ group of symmetries generated by the π–rotations $\bar{P}, \bar{Q} \in \mathrm{PSL}(2,\mathbb{C})$ with axes through the mid-point of C: precisely, let Π be the plane through the mid-point of C and orthogonal to C. Then (working equivalently with the lifts $P, Q \in \mathrm{SL}(2,\mathbb{C})$) the axes of P, Q are the two lines in Π which bisect the angles between the projections of $\mathrm{Ax}\,K_0, \mathrm{Ax}\,K_1$ onto Π, chosen so that the angle bisected by $\mathrm{Ax}\,P$ is that between the projection of the lines $\mathrm{Ax}\,K_0, \mathrm{Ax}\,K_1$ with the same (say outward) orientation.

This choice of P ensures that $PK_0P^{-1} = K_1$ while $QK_0Q^{-1} = K_1^{-1}$. Also PQ is the order-2 rotation about C, and $PQK_iQ^{-1}P^{-1} = K_i^{-1}$ for $i = 0, 1$. As in Section 3.1, $\mathcal{U} = \mathcal{S}/(\mathbb{Z}_2 \times \mathbb{Z}_2)$, and we can take $\pi_1(\mathcal{U})$ to be the (P, Q, R)–group defined in Section 3.2. In terms of (P, Q, R), the generators of $\pi_1(\mathcal{S})$ are $K_0 = RPQ$, $K_1 = PK_0P^{-1}$. Thus

$$K_0 = -\begin{pmatrix}1/2 & 3/2\\ -1/2 & 1/2\end{pmatrix}, \quad K_1 = -\begin{pmatrix}1/2 & -\zeta^2/2\\ 3/(2\zeta^2) & 1/2\end{pmatrix}.$$

In terms of generators for $\pi_1(\partial\mathcal{S})$, we also have $K_2 = QK_0Q^{-1}$, $K_3 = RK_0R^{-1}$, where

$$K_2 = -\begin{pmatrix} 1/2 & \zeta^2/2 \\ -3/(2\zeta^2) & 1/2 \end{pmatrix}, \quad K_3 = -\begin{pmatrix} 1/2 & -3/2 \\ 1/2 & 1/2 \end{pmatrix},$$

so that $K_0K_3K_1K_2 = \mathrm{id}$.

We denote the group with generators K_0, K_1 by $G_{\mathcal{S}}(\zeta)$ and the corresponding representation $\pi_1(\mathcal{S}) \to \mathrm{SL}(2,\mathbb{C})$ by $\rho_{\mathcal{S}}(\zeta)$. From now on, we frequently drop the subscript and refer to K_0 as K.

3.2.2 The handlebody $\mathcal{H}$ We are now able to determine the images of generators X, Y of $\pi_1(\mathcal{H})$ as matrices in $\mathrm{SL}(2,\mathbb{C})$ under a suitable representation $\rho_{\mathcal{H}}(\zeta)$. To simplify notation, we shall from now on frequently identify generators of $\pi_1(\mathcal{H})$ with their images in $\mathrm{SL}(2,\mathbb{C})$, thus writing X, Y in place of $\rho_{\mathcal{H}}(\zeta)(X), \rho_{\mathcal{H}}(\zeta)(Y)$ and so on.

Observe that the generator $X \in \pi_1(\mathcal{H})$ projects to the loop represented by K_0K_1 in $\mathcal{H}/\kappa$. (This latter is a loop in $\partial\mathcal{H}/\kappa$ which separates one of each pair of the cone points of K_0, K_1 from the other pair.) We arrange that the action of κ is induced by conjugation by $K_0^{-1} = K^{-1}$, so the generators of $\pi_1(\mathcal{H})$ can be written in terms of the generators of $\pi_1(\mathcal{S})$ as $X = K_0K_1$, $Y = K_0^{-1}XK_0 = K_1K_0$. Thus we have

$$K^{-1}XK = Y, \quad K^{-1}YK = (XY)^{-1}, \quad K^{-1}(XY)^{-1}K = X.$$

Using the formulae from the previous section, this gives

$$X = \begin{pmatrix} 9/(4\zeta^2) + 1/4 & -\zeta^2/4 + 3/4 \\ 3/(4\zeta^2) - 1/4 & \zeta^2/4 + 1/4 \end{pmatrix}, \quad Y = \begin{pmatrix} \zeta^2/4 + 1/4 & -\zeta^2/4 + 3/4 \\ 3/(4\zeta^2) - 1/4 & 9/(4\zeta^2) + 1/4 \end{pmatrix}.$$

In particular this reveals the relation between the parameter ζ and x:

$$x = \mathrm{Tr}\,X = \mathrm{Tr}\,Y = \mathrm{Tr}\,XY = \frac{\zeta^2}{4} + \frac{9}{4\zeta^2} + \frac{1}{2}. \tag{3-1}$$

We denote the group with generators X, Y by $G_{\mathcal{H}}(\zeta)$ and the corresponding representation $\pi_1(\mathcal{H}) \to \mathrm{SL}(2,\mathbb{C})$ by $\rho_{\mathcal{H}}(\zeta)$; we explain in Section 3.2.5 why up to conjugation $\rho_{\mathcal{H}}(\zeta)$ in fact depends only on x.

Remark 3.1 In the above discussion, we made choices of sign so that $K^3 = \mathrm{id}$, $X = K_0K_1$ (where $K = K_0$ as above). To compute the discreteness locus of a family of representations only requires looking in $\mathrm{PSL}(2,\mathbb{C})$, however for computations involving traces we need a lift to $\mathrm{SL}(2,\mathbb{C})$.

By [6], any $\mathrm{PSL}(2,\mathbb{C})$ representation of a Kleinian group can be lifted to $\mathrm{SL}(2,\mathbb{C})$ provided there are no elements of order 2; in particular this applies to $\mathrm{PSL}(2,\mathbb{C})$ representations of $\pi_1(\mathcal{S})$ and $\pi_1(\mathcal{H})$. Since the product of the three generating loops corresponding to X, Y, Z is the identity in $\pi_1(\mathcal{H})$, we should make a choice of lift in which $XYZ = \mathrm{id}$ in $\mathrm{SL}(2,\mathbb{C})$. We could choose the element K which represents the 3–fold symmetry κ to be such that either $K^3 = \mathrm{id}$ or $K^3 = -\mathrm{id}$; however, since we intend to work with representations of $\pi_1(\mathcal{S}) \to \mathrm{SL}(2,\mathbb{C})$, we should make the choice $K^3 = \mathrm{id}$ because K corresponds to a loop round an order-3 cone axis in the quotient orbifold $\mathcal{S}$.

In the representation we have written down, we achieve $K^3 = \mathrm{id}$ with the choice $K = RPQ = \begin{pmatrix} -1/2 & -3/2 \\ 1/2 & -1/2 \end{pmatrix}$. It is easy to check that taking $K^3 = \mathrm{id}$, if we let $X = K_0 K_1$, we get $XYZ = \mathrm{id}$ as required, but if we choose $X = -K_0 K_1$, we get $XYZ = -\mathrm{id}$, which is wrong.

3.2.3 The singular solid torus $\mathcal{T}$ Finally we discuss the associated singular solid torus $\mathcal{T}$, which is constructed in a standard way from the (P, Q, R)–group. We do not logically need to use $\mathcal{T}$ in our further development, however as explained in Section 5, in practice we used $\mathcal{T}$ for computations, moreover the interpretation of the problem in the more familiar setting of a torus with a cone point may be helpful.

The boundary $\partial\mathcal{U}$ is a sphere with 4 cone points x_P, x_Q, x_R and x_K corresponding to P, Q, R and $K = RPQ$. Thus we can take as generators of $\pi_1(\mathcal{T})$ the element $B = PQ$ whose projection to $\mathcal{T}$ is a loop separating x_P, x_Q from x_R, x_K, and the element $A = RQ$ which projects to a loop separating x_R, x_Q from x_P, x_K. Since P, Q have a common fixed point, B is an order-2 elliptic, while since the axes of R, Q are (generically) disjoint, A is a loxodromic whose axis extends the common perpendicular to $\mathrm{Ax}\, R$ and $\mathrm{Ax}\, Q$.

Using the formulae above for the (P, Q, R)–group, we compute

$$RQ = A = \begin{pmatrix} 3i/(2\zeta) & i\zeta/2 \\ i/(2\zeta) & -i\zeta/2 \end{pmatrix}, \quad PQ = B = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

so that

$$\operatorname{Tr} A = \frac{3i}{2\zeta} - \frac{i\zeta}{2}, \quad \operatorname{Tr} B = 0, \quad \operatorname{Tr} AB = -\frac{\zeta}{2} - \frac{3}{2\zeta}. \tag{3-2}$$

Note that $AB = RP$ and $A^2 = -K_0 K_1$, $B^2 = -\mathrm{id}$. We also deduce that

$$ABA^{-1}B^{-1} = [A, B] = \begin{pmatrix} 1/2 & -3/2 \\ 1/2 & 1/2 \end{pmatrix}, \quad \text{so that } \operatorname{Tr}[A, B] = 1.$$

Note that $\operatorname{Tr} A \in [-2,2]$ if and only if $|\zeta| = \sqrt{3}$ or $\zeta = it$ with $1 \le |t| \le 3$, justifying the above remark that generically A is loxodromic. Note also that $A^2 = -K_0 K_1$ is consistent with the direct computation using (3-2) that $\operatorname{Tr}(A^2) = -(\zeta^2/4 + 9/(4\zeta^2) + 1/2)$. Also note that $[A, B] = -K^2$, so that the commutator is rotation by $\frac{4\pi}{3}$ about $\operatorname{Ax} K$. Since $\operatorname{Tr} K^2 = (\operatorname{Tr} K)^2 - 2$, we find also that $\operatorname{Tr}[A, B] = 1$ independently of the choice of sign for K. This is consistent with $\operatorname{Tr}[A, B] = -2\cos\left(\frac{2\pi}{3}\right)$, the sign being negative by analogy with the well known fact that for any irreducible representation of a once punctured torus group for which the commutator is parabolic, we have $\operatorname{Tr}[A, B] = -2$.

We denote the group with generators A, B by $G_{\mathcal{T}}(\zeta)$ and the corresponding representation $\pi_1(\mathcal{T}) \to \mathrm{SL}(2,\mathbb{C})$ by $\rho_{\mathcal{T}}(\zeta)$.

Remark 3.2 Once again there are questions of sign which this time are a little more subtle. If $\alpha \in \mathrm{PSL}(2,\mathbb{C})$ corresponds to an element of order 2 in $\pi_1(M)$, then the corresponding representation cannot be lifted to $\mathrm{SL}(2,\mathbb{C})$, because for nontrivial $\alpha \in \mathrm{SL}(2,\mathbb{C})$, necessarily $\alpha^2 = -\mathrm{id}$; see [20] and [6]. Since in $\pi_1(\mathcal{T})$ the element B^2 is trivial, a $\mathrm{PSL}(2,\mathbb{C})$ representation of $\pi_1(\mathcal{T})$ cannot be lifted to $\mathrm{SL}(2,\mathbb{C})$. Nevertheless, we can as above write down a group in $\mathrm{SL}(2,\mathbb{C})$ which projects to a $\mathrm{PSL}(2,\mathbb{C})$ representation for $\pi_1(\mathcal{T})$. See Section 5 for further discussion on this point.

3.2.4 More on the configuration for the large coned ball $\mathcal{S}$ The relation (3-1) can be given a geometrical interpretation in terms of the perpendicular distance between the axes of K_0, K_1 which sheds light on the symmetries of the configuration $\mathcal{CF}$ in Section 3.2.1. To measure complex distance, we use the conventions spelled out in detail in [28, Section 2.1]. The signed complex distance $\boldsymbol{d}_\alpha(L_1, L_2)$ between two oriented lines L_1, L_2 along their oriented common perpendicular α is defined as follows. The signed real distance $d_\alpha(L_1, L_2)$ is the positive real hyperbolic distance between L_1, L_2 if α is oriented from L_1 to L_2 and its negative otherwise. Let $\boldsymbol{v}_i$ for $i = 1, 2$ be unit vectors to L_i at the points $L_i \cap \alpha$ and let $\boldsymbol{w}_1$ be the parallel translate of $\boldsymbol{v}_1$ along α to the point $\alpha \cap L_2$. Then $\boldsymbol{d}_\alpha(L_1, L_2) = \delta_\alpha(L_1, L_2) + i\theta$ where θ is the angle, mod $2\pi i$, from $\boldsymbol{w}_1$ to $\boldsymbol{v}_2$ measured anticlockwise in the plane spanned by $\boldsymbol{w}_1$ to $\boldsymbol{v}_2$ and oriented by α.

Let σ be the signed complex distance from the *oriented* axis $\operatorname{Ax} K_0$ to the oriented axis $\operatorname{Ax} K_1$, measured along the common perpendicular C oriented from $\operatorname{Ax} K_0$ to $\operatorname{Ax} K_1$. Then $\operatorname{Ax} K_0, \operatorname{Ax} K_1$ together with $\operatorname{Ax} K_0 K_1$ form the alternate sides of a right angled skew hexagon whose other three sides are the common perpendiculars between the three axes taken in pairs. The cosine formula gives σ in terms of the complex half translation lengths $\lambda_0, \lambda_1, \lambda_2$ of K_0, K_1 and $K_0 K_1$, respectively. To get the sides oriented consistently round the hexagon, we have to reverse the orientation of

Ax K_0 so that the complex distance σ should be replaced by $\sigma' = \sigma + i\pi$ and λ_0 by $\lambda_0' = -\lambda_0$ (see [28]), so the formula gives

$$\cosh\sigma' = \frac{\cosh\lambda_2 - \cosh\lambda_0' \cosh\lambda_1}{\sinh\lambda_0' \sinh\lambda_1}.$$

As in Section 3.2.2, we have $X = K_0K_1$ so $x = \operatorname{Tr} K_0K_1 = 2\cosh\lambda_2$ while for $i = 0, 1$ we have $\cosh\lambda_i = \cos\frac{2\pi}{3} = -\frac{1}{2}$ and $\sinh\lambda_i = i\sin\frac{2\pi}{3} = \frac{1}{2}i\sqrt{3}$. (Note that since K_0, K_1 are conjugate we should take $\lambda_0 = \lambda_1$ so the possible additive ambiguity of $i\pi$ in the definition of the λ_i does not change the resulting equation.) Substituting, we find

$$-\cosh\sigma = \frac{x/2 - 1/4}{(\sqrt{3}/2)^2} = \tfrac{1}{3}(2x-1). \tag{3-3}$$

We can also relate σ directly to our parameter ζ. By construction Ax K_0 is the oriented line $[-\sqrt{3}i, \sqrt{3}i]$, while $K_1 = PK_0P^{-1}$ so that Ax K_1 is the oriented line $[i\zeta^2/\sqrt{3}, -i\zeta^2/\sqrt{3}]$ and C is the oriented line from ∞ to 0. Thus the real part of the hyperbolic distance from Ax K_0 to Ax K_1 is $2\log\sqrt{3}/|\zeta|$, and the anticlockwise angle, measured in the plane oriented *downwards* along the vertical axis C, is $-(\pi + 2\operatorname{Arg}\zeta)$. Hence

$$\sigma = 2\log\frac{\sqrt{3}}{|\zeta|} - 2i\operatorname{Arg}\zeta - i\pi = 2\log\frac{\sqrt{3}}{\zeta}.$$

Comparing to (3-3), we find

$$\left[\left(\frac{\sqrt{3}}{\zeta}\right)^2 + \left(\frac{\zeta}{\sqrt{3}}\right)^2\right] = 2\cosh(\sigma + i\pi) = \tfrac{2}{3}(2x-1),$$

or

$$x - \frac{1}{2} = \frac{3}{4}\left[\left(\frac{\sqrt{3}}{\zeta}\right)^2 + \left(\frac{\zeta}{\sqrt{3}}\right)^2\right], \tag{3-4}$$

recovering and giving a more satisfactory geometrical meaning to (3-1).

3.2.5 Dependence on x versus ζ It is not perhaps immediately obvious why the groups $G_{\mathcal{S}}(\zeta), G_{\mathcal{H}}(\zeta)$ as defined above depend up to conjugation only on our original parameter x. This is clarified by the above discussion, because up to conjugation $G_{\mathcal{S}}(\zeta)$ depends only on the configuration $\mathcal{CF}$ and hence on σ which is related to x as in (3-3). An alternative way to see this is the discussion on computing traces in Section 4.4. Thus from now on, we shall alternatively write $G_{\mathcal{S}}(x), G_{\mathcal{H}}(x)$ in place of $G_{\mathcal{S}}(\zeta), G_{\mathcal{H}}(\zeta)$.

3.2.6 Symmetries The discussion in Section 3.2.4 gives insight into various symmetries of the parameters x and ζ. Equation (3-1) shows that the map $\zeta \mapsto x$ is a 4–fold covering with branch points at $\zeta = \pm\sqrt{3}$, $\zeta = \pm i\sqrt{3}$ and $z = 0, \infty$. Correspondingly, we have a Klein 4–group $\mathbb{Z}_2 \times \mathbb{Z}_2$ of symmetries which change ζ but not x:

(1) Replacing ζ by $-\zeta$ leaves the basic construction unchanged but the line matrices defining P, Q change sign.

(2) Replacing ζ by $-3/\zeta$ is an order-2 rotation about the axis $[-\sqrt{3}i, \sqrt{3}i]$. This fixes K_0 and moves K_1 into a position on the opposite side of K_0 along the vertical line C. This changes nothing other than the position we choose for the basic configuration in Section 3.2. Note however that the line matrices defining P, Q change sign.

There is also a symmetry which changes x as well as ζ. Say we fix the orientation of one of the two axes $\operatorname{Ax} K_0, \operatorname{Ax} K_1$ while reversing the other. On the level of the configuration $\mathcal{CF}$ from Section 3.2.1, this interchanges P and Q. Since $PK_0P^{-1} = K_1$ while $QK_0Q^{-1} = K_1^{-1}$, this is equivalent to fixing the orientation of one of the two axes $\operatorname{Ax} K_0, \operatorname{Ax} K_1$ while reversing the other. This symmetry interchanges the marked group P, Q, R with the marked group Q, P, R, so that one group is discrete if and only if so is the other. In terms of our parameters, the complex distance σ between the axes changes to $\sigma + i\pi$, so that $\cosh\sigma \mapsto -\cosh\sigma$ giving the symmetry $\bigl(x-\frac{1}{2}\bigr) \mapsto -\bigl(x-\frac{1}{2}\bigr)$ of (3-3). Note that the diagonal slice of the Bowditch set $\Delta \cap \mathcal{B}$ does not possess this symmetry. Interchanging P and Q is induced by the map $\zeta \mapsto i\zeta$; more precisely this map sends P to Q and Q to $-P$. This clearly induces the same symmetry in Equation (3-1). Note that by the definition, in this symmetry R remains unchanged.

On the level of the torus group $\pi_1(\mathcal{T})$, we have by definition $RQ = A$, $PQ = B$ so that $AB = RP$. Thus sending P to Q and Q to $-P$ while fixing R sends B to $-B$ and A to $-AB$. (Recall that on the level of matrices, $PQ = -QP$.) The symmetry should therefore replace the trace triple $(\operatorname{Tr} A, \operatorname{Tr} B, \operatorname{Tr} AB)$ by the triple $(-\operatorname{Tr} AB, -\operatorname{Tr} B, \operatorname{Tr} A)$. It is easily checked from (3-2) that this is exactly the change effected by $\zeta \mapsto i\zeta$.

Finally, we have the symmetry of complex conjugation induced by $x \to \overline{x}$ or equivalently $\zeta \mapsto \overline{\zeta}$. This sends $\sigma \mapsto \overline{\sigma}$ thus replacing $G_{\mathcal{H}}(x)$ by a conjugate group in which the distance between $\operatorname{Ax} K_0$ and $\operatorname{Ax} K_1$ is unchanged but the angle measured along their common perpendicular changes sign. Clearly these are different groups but one is discrete if and only if the same is true of the other.

The diagonal slice of the Bowditch set obviously also enjoys the symmetry by conjugation, however, that is its only symmetry. In particular $(x, x, x) \to (-x, -x, -x)$ is not a symmetry and the corresponding $\mathrm{SL}(2, \mathbb{C})$ representations project to different representations of F_2 into $\mathrm{PSL}(2,\mathbb{C})$. This is because any two distinct lifts of a representation from $\mathrm{PSL}(2,\mathbb{C})$ to $\mathrm{SL}(2,\mathbb{C})$ differ by multiplying exactly two of the parameters x, y, z by -1. The allowed replacement $X \to -X$ and $Y \to -Y$ gives the group $(-x,-x,x)$ with parameters which are not in the diagonal slice Δ.

The symmetries can be seen in our plots by comparing Figure 5, the Bowditch set for the triple $\phi_{(x,x,x)}$ in the x–plane, with the right-hand frame of Figure 13, which shows the same set in the ζ–plane. Note the symmetry of complex conjugation in both pictures. In addition, Figure 13 is invariant under the maps $\zeta \mapsto -\zeta$ and $\zeta \mapsto -3/\zeta$, neither of which are seen in Figure 5. Thus the upper half plane in Figure 13 is a 4–fold covering of the upper half plane in Figure 5: as is easily checked from (3-1), the imaginary axis in Figure 13 maps to the negative real axis in Figure 5 while the real axis in Figure 13 maps to the positive real axis in Figure 5. In particular, note the following branch points and special values: if $x = 3$, then $\zeta = \pm 1, \pm 3$; if $x = 2$, then $\zeta = \pm\sqrt{3}$; if $x = -1$, then $\zeta = \pm\sqrt{3}i$; if $x = -2$, then $\zeta = \pm i, \pm 3i$.

Finally, the symmetry $\left(x - \frac{1}{2}\right) \mapsto -\left(x - \frac{1}{2}\right)$ is not visible in either picture because it does not preserve the property of lying in the Bowditch set. As we shall see later, this symmetry is visible in pictures of the discreteness locus; see the upper frame of Figure 8.

4 Discreteness

We now turn to the question of finding those values of the parameter x for which representation $\rho_x\colon F_2 \to \mathrm{SL}(2,\mathbb{C})$ is faithful with discrete geometrically finite image, where as usual $F_2 = \langle X, Y \mid - \rangle$. Let $\mathcal{D}_{\mathcal{S}}, \mathcal{D}_{\mathcal{H}} \subset \mathbb{C}$ denote the subsets of the complex x–plane on which the representations $\rho_{\mathcal{S}}(x), \rho_{\mathcal{H}}(x)$ are respectively faithful and $G_{\mathcal{S}}(x), G_{\mathcal{H}}(x)$ are discrete and geometrically finite. (See Section 3.2.5 for the replacement of $G_{\mathcal{S}}(\zeta), G_{\mathcal{H}}(\zeta)$ by $G_{\mathcal{S}}(x), G_{\mathcal{H}}(x)$.) We first show that $\mathcal{D}_{\mathcal{S}} = \mathcal{D}_{\mathcal{H}}$.

We begin with the easy observation that since all the groups in Section 3 are commensurable, they are either all discrete or all nondiscrete together:

Lemma 4.1 *Suppose that G, H are subgroups of $\mathrm{PSL}(2,\mathbb{C})$ with $G \supset H$ and that $[G : H]$ is finite. Then G is discrete (geometrically finite) if and only if the same is true of H.*

Proof If G is discrete, clearly so is H. Suppose that H is discrete but G is not. Then infinitely many distinct orbit points in $G \cdot O$ accumulate in some compact set $D \subset \mathbb{H}^3$. Label the cosets of $[G : H]$ as $g_1 H, \dots, g_k H$. Then for some i there are infinitely many points $g_i h_r \cdot O \in D$, which gives infinitely many distinct points $h_r \in g_i^{-1} D$. This contradicts discreteness of H. The proof for geometric finiteness is equally straightforward. $\square$

Lemma 4.2 *The representation $\rho_{\mathcal{S}}(x)\colon \pi_1(\mathcal{S}) \to G_{\mathcal{S}}(x)$ is faithful if and only if the same is true of $\rho_{\mathcal{H}}(x)\colon \pi_1(\mathcal{H}) \to G_{\mathcal{H}}(x)$.*

Proof Note that $\pi_1(\mathcal{S})$ is isomorphic to $\mathbb{Z}/3\mathbb{Z} * \mathbb{Z}/3\mathbb{Z} = \langle k_0, k_1 \mid k_0^3 = k_1^3 = \mathrm{id}\rangle$, while $\pi_1(\mathcal{H})$ is the subgroup of $\pi_1(\mathcal{S})$ generated by k_0k_1 and k_1k_0, and is isomorphic to a free group of rank 2. By construction, $\rho_{\mathcal{H}}(x)$ is the restriction of $\rho_{\mathcal{S}}(x)$ to $\pi_1(\mathcal{H})$. Thus, if $\rho_{\mathcal{S}}(x)$ is faithful, then so is $\rho_{\mathcal{H}}(x)$.

Now $\pi_1(\mathcal{H})$ has index three in $\pi_1(\mathcal{S})$ and $\pi_1(\mathcal{S}) = \pi_1(\mathcal{H}) \cup k_0\pi_1(\mathcal{H}) \cup k_0^{-1}\pi_1(\mathcal{H})$. Suppose that $\rho_{\mathcal{H}}(x)$ is faithful but $\rho_{\mathcal{S}}(x)$ is not. Then there exists $g \in \pi_1(\mathcal{S})$ such that $\rho_{\mathcal{S}}(x)(g) = \mathrm{id}$. Now $g = k_0^e h$, where $e = \pm 1$ and $h \in \pi_1(\mathcal{H})$. Thus $\mathrm{id} = \rho_{\mathcal{S}}(x)(g) = \rho_{\mathcal{S}}(x)(k_0^e)\rho_{\mathcal{H}}(x)(h)$ so that $\rho_{\mathcal{H}}(x)(h^3) = \rho_{\mathcal{S}}(x)(k_0^{-3e}) = \mathrm{id}$, contradicting the assumption that $\rho_{\mathcal{H}}(x)$ is faithful. □

Corollary 4.3 *The representations $\rho_{\mathcal{S}}(x), \rho_{\mathcal{H}}(x)$ are faithful, discrete and geometrically finite together; that is, $\mathcal{D}_{\mathcal{S}} = \mathcal{D}_{\mathcal{H}}$.*

Thus we may write $\mathcal{D} = \mathcal{D}_{\mathcal{S}} = \mathcal{D}_{\mathcal{H}}$. Our next aim is to find $\mathcal{D} \subset \mathbb{C}$.

4.1 Fundamental domains

We can make a rough estimate for $\mathcal{D}$ by exhibiting a fundamental domain for $G_{\mathcal{S}}(x)$ for sufficiently large x.

Proposition 4.4 *Writing $x = u + iv$, the region $\mathcal{D}$ contains the region outside the ellipse $\frac{1}{25}(2u-1)^2 + \frac{1}{4}v^2 = 1$ in the x–plane.*

Proof In view of Corollary 4.3, we can work with the large cone manifold $\mathcal{S}$ with generators K_0, K_1 of Section 3.2.1. The axis of K_0 is the line $[-i\sqrt{3}, i\sqrt{3}]$ passing through $j\sqrt{3}$. Let H, H' be the hemispheres which meet $\mathbb{R}$ orthogonally at points $-3, 1$ and $-1, 3$, respectively, and let E, E' be the closed half spaces they cut out which contain 0. Then H, H' intersect in $\operatorname{Ax} K_0$, moreover $E \cap E'$ is a fundamental domain for the group $\langle K_0\rangle$ acting on $\mathbb{H}^3$.

Recalling that P is the π–rotation about the line $[-\zeta, \zeta]$ which bisects the common perpendicular between $\operatorname{Ax} K_0$ and $\operatorname{Ax} K_1$, we see that the images of H, H' under P meet along $\operatorname{Ax} K_1$. Since $P(z) = \zeta^2/z$ for $z \in \mathbb{C}$, we have that $P(H), P(H')$ meet $P(\mathbb{R})$ orthogonally in points $-\frac{1}{3}\zeta^2, \zeta^2$ and $\frac{1}{3}\zeta^2, -\zeta^2$, respectively. If F, F' are the half spaces cut out by $P(H), P(H')$ which contain $\infty = P(0)$ (so that $F = P(E)$), then in a similar way, $F \cap F'$ is a fundamental domain for $\langle K_1\rangle$.

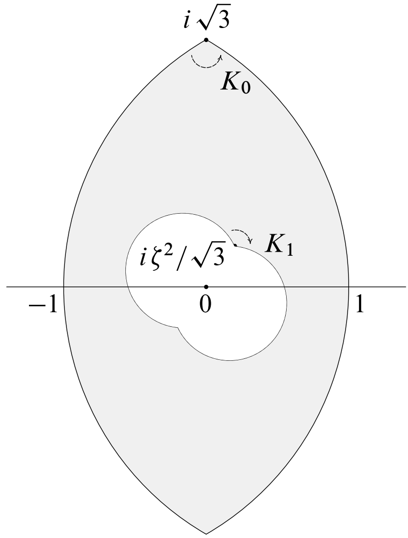


Figure 7: The shaded region illustrates the fundamental domain for $\pi_1(\mathcal{S})$ acting in its regular set in $\widehat{\mathbb{C}}$ when $|\zeta| < 1$, so that x is outside the ellipse of Proposition 4.4.

Thus if $|\zeta| < 1$, then the hemisphere of radius $|\zeta|$ centred at 0 separates the regions $(E \cap E')^C$ and $(F \cap F')^C$. We conclude by Poincaré's theorem (or a suitable simple version of the Klein–Maskit combination theorem) that in this situation the region $(E\cap E')\cap(F\cap F')$ is a fundamental domain for $\langle K_0, K_1\rangle$, which moreover is discrete with presentation $\langle K_0, K_1 \mid K_0^3 = K_1^3 = \mathrm{id}\rangle$. Thus the representation $\rho_{\mathcal{S}}(x)$ with $x = x(\zeta)$ as in (3-1), is faithful, and hence $x \in \mathcal{D}$.

Suppose that $\zeta = e^{i\phi}$. Then from (3-3), $\frac{1}{3}(2x-1) = \cosh\sigma = \frac{1}{2}\big(\frac{1}{3}e^{2i\phi} + 3e^{-2i\phi}\big)$ so that $x = u + iv$ lies on the ellipse $\frac{1}{25}(2u-1)^2 + \frac{1}{4}v^2 = 1$ as claimed. □

The configuration when $x \in \mathbb{R}$ is of particular interest since in this case $G_{\mathcal{H}}(x)$ is Fuchsian. The ellipse meets the real axis in points $-2, 3$ so that $G_{\mathcal{H}}(x)$ is discrete and the representation is faithful on $(\infty, -2]$ (corresponding to $|\zeta| > 1$, $\zeta \in i\mathbb{R}$) and $[3, \infty)$ (corresponding to $|\zeta| > 1, \zeta \in \mathbb{R}$). In these two cases the fundamental domains look the same; see Figure 11. Note that the interval $(-2, 3)$ is definitely not in $\mathcal{D}$: if $-2 < x < 2$, then K_0K_1 is elliptic since $x = \mathrm{Tr}\, K_0K_1$, while if $-1 < x < 3$, then $K_0K_1^{-1}$ is elliptic since $\mathrm{Tr}\, K_0K_1^{-1} = 1 - x$; see also Section 4.5.

In the general case, a fundamental domain can be found by a modification of Wada's program OPTi [31; 32]. This program allows one to compute the limit set and fundamental domains for the PQR–group $G_{\mathcal{U}}$. A short Python program for doing this is available at `http://vivaldi.ics.nara-wu.ac.jp/~yamasita/DiagonalSlice/`.

4.2 The method of pleating rays

To determine $\mathcal{D}$, we use the Keen–Series method of pleating rays applied to the large coned sphere $\mathcal{S}$. This is closely analogous to the problem of computing the Riley slice of Schottky space, that is the parameter space of discrete geometrically finite groups freely generated by two parabolics, which was solved in [15; 19].

We begin by briefly summarising the elements of pleating ray theory we need. For more details see various of the first author's papers, for example [15; 5].

Suppose that $G \subset \mathrm{SL}(2,\mathbb{C})$ is a geometrically finite Kleinian group with corresponding orbifold $M = \mathbb{H}^3/G$ and let $\mathcal{C}/G$ be its convex core, where $\mathcal{C}$ is the convex hull in $\mathbb{H}^3$ of the limit set of G; see [7]. Then $\partial\mathcal{C}/G$ is a convex pleated surface (see for example [7]) also homeomorphic to ∂M. The bending of this pleated surface is recorded by means of a measured geodesic lamination, the *bending lamination* $\beta = \beta(G)$, whose support forms the bending lines of the surface and whose transverse measure records the total bending angle along short transversals. We say β is *rational* if it is supported on closed curves: note that closed curves in the support of β are necessarily simple and pairwise disjoint. If a bending line is represented by a curve $\gamma \in \pi_1(\mathcal{S})$, then by definition it is the projection of a geodesic axis to $\partial\mathcal{C}/G$, so in particular β contains no peripheral curves in its support. Note that any two homotopically distinct nonperipheral simple closed curves on $\partial\mathcal{S}$ intersect. Thus in this case, β is rational only if its support is a single simple essential nonperipheral closed curve on $\partial\mathcal{C}/G$.

As above, we parametrise representations $\rho_{\mathcal{S}}(x)\colon \pi_1(\mathcal{S}) \to \mathrm{SL}(2,\mathbb{C})$ by $x \in \mathbb{C}$ and denote the image group by $G_{\mathcal{S}}(x)$. From now on, we frequently write ρ_x for $\rho_{\mathcal{S}}(x)$.

Definition 4.5 Let γ be a homotopy class of simple essential nonperipheral closed curves on $\partial\mathcal{S}$. The *pleating ray* $\mathcal{P}_\gamma$ of γ is the set of points $x \in \mathcal{D}$ for which $\beta(G_{\mathcal{S}}(x)) = \gamma$.

Such rays are called *rational pleating rays*; a similar definition can be made for general projective classes of bending lamination; see [5].

The following key lemma is proved in [5, Proposition 4.1]; see also [14, Lemma 4.6]. The essence is that because the two flat pieces of $\partial\mathcal{C}/G$ on either side of a bending line are invariant under translation along the line, the translation can have no rotational part.

Lemma 4.6 *If the axis of $g \in G$ is a bending line of $\partial\mathcal{C}/G_{\mathcal{S}}(x)$, then* $\operatorname{Tr} g(x) \in \mathbb{R}$.

Notice that the lemma applies even when the bending angle θ_γ along γ vanishes, so the corresponding surface is flat, or when the angle is π, in which case either γ is parabolic or $G_{\mathcal{S}}(x)$ is Fuchsian.

If $g \in G$ represents a curve γ on $\partial\mathcal{S}$, define the *real trace locus* $\mathbb{R}_\gamma$ of γ to be the locus of points in $\mathbb{C}$ for which $\operatorname{Tr} g \in (-\infty,-2] \cup [2,\infty)$. By the above lemma, $\mathcal{P}_\gamma \subset \mathbb{R}_\gamma$.

Our aim is to compute the locus of faithful discrete geometrically finite representations $\mathcal{D}_\mathcal{S} = \mathcal{D}$. In summary, we do this as follows:

(1) Show that up to homotopy in $\mathcal{S}$, the essential nonperipheral curves on $\partial\mathcal{S}$ are indexed by $\mathbb{Q}/\sim$, where $p/q \sim \pm(p+2kq)/q$ for $k \in \mathbb{Z}$ (Proposition 4.9).

(2) Given $\gamma \in \pi_1(\partial\mathcal{S})$, give an algorithm for computing $\operatorname{Tr} \rho_x(\gamma)$ as a polynomial in x, in particular identifying its two highest order terms in terms of p, q (Section 4.4 and Proposition 4.11).

(3) Show that $\mathcal{P}_{0/1} = (-\infty,-3]$ and $\mathcal{P}_{1/1} = [2,\infty)$ (where $\mathcal{P}_{p/q}$ denotes the pleating ray of the curve $\gamma_{p/q} \in \pi_1(\partial\mathcal{S})$ identified with p/q) (Section 4.5).

(4) Show $\mathcal{P}_{p/q}$ is a union of connected nonsingular branches of $\mathbb{R}_\gamma$ (Theorem 4.14).

(5) For $p, q \neq 0, 1$, identify $\mathcal{P}_{p/q}$ by showing it has two connected components, namely the branches of $\mathbb{R}_\gamma$ which are asymptotic to the directions $e^{\pm i\pi(p/q+1)}$ as $|x| \to \infty$ (Proposition 4.20).

(6) Prove that rational rays $\mathcal{P}_{p/q}$ are dense in $\mathcal{D}_\mathcal{S}$ (Theorem 4.23).

One could carry all this out following almost word for word the arguments in [15]. Rather than do this, we indicate as appropriate how more general results can be put together to provide a somewhat less ad hoc proof of the results. The claim that $\mathcal{P}_{p/q}$ has two connected components appears to contradict the results in [15]; see however the following remark and Proposition 4.20 below. The pleating rays are shown on the top in Figure 8 with the Riley slice rays from [15] below for comparison.

Remark 4.7 There were two rather subtle errors in [15]. The first was that, in the enumeration of curves on $\partial\mathcal{S}$, we omitted to note that $\gamma_{p/q}$ is homotopic to $\gamma_{-p/q}$ in $\mathcal{S}$. The second was, that we found only one of the two components of $\mathcal{P}_{p/q}$. Since $\mathcal{P}_{p/q} = \mathcal{P}_{-p/q}$, these two errors in some sense cancelled each other out. They were discussed at length and resolved in [19] and we make corresponding corrections here.

Remark 4.8 The space of all faithful discrete representations is known to be the closure of the geometrically finite ones by the tameness and density theorems; see [23] for a detailed overview of this and other facts about deformation spaces. However these issues are not the main point of concern to us here.

4.3 Step 1: Enumeration of curves on $\partial\mathcal{S}$

We need to enumerate essential nonperipheral unoriented simple curves on $\partial\mathcal{S}$ up to free homotopy equivalence in $\mathcal{S}$. As is well known, such curves on $\partial\mathcal{S}$ are, up to free

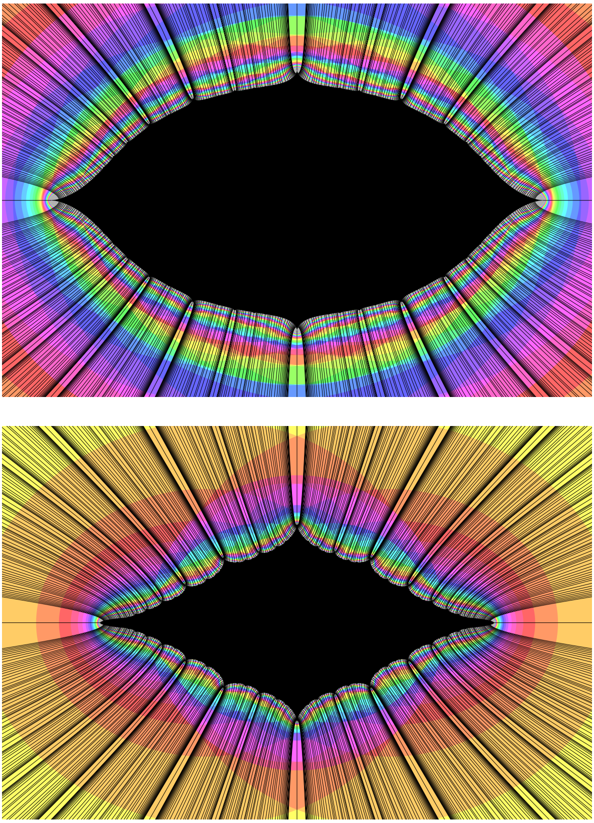


Figure 8: Top: pleating rays for $G_{\mathcal{S}}(x)$. Bottom: pleating rays for the Riley slice as described in [15]. The underlying colours indicate the Bowditch sets for the initial triples discussed in Section 5.0.3; conjecturally these coincide with the closure of the regions filled by the pleating rays. For a discussion of how the rays were actually computed; see Section 5.0.2

homotopy equivalence in $\partial\mathcal{S}$, in bijective correspondence with lines of rational slope in the plane, that is, with $\mathbb{Q} \cup \infty$; see for example [15; 19]. For (p,q) relatively prime and $q \geq 0$, denote the class corresponding to p/q by $\gamma_{p/q}$. We have:

Proposition 4.9 [19, Theorem 1.2] *The unoriented curves $\gamma_{p/q}, \gamma_{p'/q'}$ are freely homotopic in $\mathcal{S}$ if and only if $p'/q' = \pm p/q + 2k$ for $k \in \mathbb{Z}$.*

Missing the identification $\gamma_{p/q} \sim \gamma_{-p/q}$ was the first of the two errors in [15] referred to in Remark 4.7.

Before proving the proposition, we need to explain the identification of curves on $\partial\mathcal{S}$ with $\mathbb{Q} \cup \infty$. In [15; 19] this was done using the plane punctured at integer points as

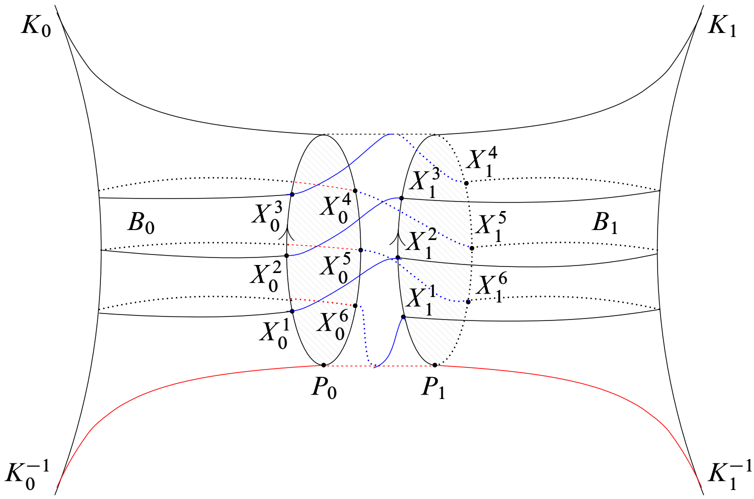


Figure 9: The arrangement of arcs on $\partial\mathcal{S}$. The curve shown illustrates the case $p=1, q=3$.

an intermediate covering between $\partial\mathcal{S}$ and its universal cover. The idea is sketched in Section 5.0.1. Here we give a slightly different description of the curve $\gamma_{p/q}$ which leads to a nice proof of the above result.

Cut $\mathcal{S}$ into two halves along the meridian disk m which is the projection of the plane which perpendicularly bisects the common perpendicular C to the two singular axes $\operatorname{Ax} K_i$ for $i=0,1$. Each half is a ball $\widehat{B}_i$ with a singular axis $\operatorname{Ax} K_i$. The boundary $\partial B_i = \partial \widehat{B}_i \cap \partial\mathcal{S}$ is a sphere with two cone points and a hole ∂m. Since the axes of K_i are oriented, we can distinguish one cone point on each ∂B_i as the positive end of $\operatorname{Ax} K_i$. Now $\partial\mathcal{S}$ has a hyperbolic structure inherited from the ordinary set (or from the pleated surface structure on $\partial\mathcal{C}/G_{\mathcal{S}}(x)$), in which ∂m is geodesic. With respect to such a structure, each ∂B_i has a reflectional symmetry ι in the (projection of the) plane containing $\operatorname{Ax} K_i$ and C, which maps the cone points to themselves, and which maps the "front" to the "back" as shown in Figure 9. There is a preferred base point P_i on ∂m, namely the foot of the perpendicular from the negative end of $\operatorname{Ax} K_i$ to ∂m.

Let γ be an essential nonperipheral simple curve on $\partial\mathcal{S}$, which we may assume has minimal intersection in its isotopy class with ∂m. Then $\gamma \cap \partial B_i$ consists of q arcs joining ∂m to itself for each $i=0,1$. On each B_i separately, after suitable isotopies, we may arrange the strands of γ symmetrically with respect to ι, that is, with front to back symmetry. However, these two isotopies may not be consistent, that is, they may not glue together to form an isotopy of ∂S. We reconstruct the gluing as follows.

Orient ∂m so that it points "upwards" on the front side of the figure. Lifting ∂m to its cyclic cover $\mathbb{R}$, enumerate in order the endpoints $X_i^k, i = 0, 1; k \in \mathbb{Z}$ of arcs of γ starting (say) with the arc meeting ∂m nearest P_i, and so that increasing order is in the direction of the upwards orientation of ∂m viewed from the front side in the figure. Since $X_i^k = X_i^{k+2q}$, the enumeration is really mod $2q\mathbb{Z}$.

To reconstruct γ we have to join the endpoints X_0^k on ∂B_0 to the endpoints $X_1^{k'}$ on ∂B_1. Since the arcs have to be matched in order round ∂m, if X_0^i is joined to X_1^j, then X_0^{i+k} is joined to X_1^{j+k} for all $k \in \mathbb{Z}$. Set $p = j - i$. Clearly this gluing can be implemented by an isotopy in an annular neighbourhood of ∂m, which can be extended to an isotopy of the whole of $\mathcal{S}$ compatible with the previous isotopies on ∂B_i.

It is not hard to see that the resulting curve $\gamma_{p/q}$ is connected if and only if (p, q) are relatively prime. Note that with this description, ∂m is the curve $q = 0$, that is, $\gamma_{1/0}$. The curve $\gamma_{0/1}$ is represented by $K_0 K_1$ and $\gamma_{1/1}$ by $K_0 K_1^{-1}$. We leave it to the reader to see that this description is the same as that obtained from the lattice picture in [15]; see also Section 5.0.1.

Notation From now on, to simplify notation, for $\gamma \in \pi_1(\mathcal{S})$ and $Z \in \mathrm{SL}(2, \mathbb{C})$, we write $\gamma \leftrightarrow Z$ to indicate that the matrix Z corresponds to the geodesic in the free homotopy class of γ under the representation $\rho_{\mathcal{S}}(x)$. Thus in particular, $\gamma_{0/1} \leftrightarrow K_0 K_1$ and $\gamma_{1/1} \leftrightarrow K_0 K_1^{-1}$.

Proof of Proposition 4.9 Write $\gamma_{p/q} \sim \gamma_{p'/q'}$ to indicate that $\gamma_{p/q}, \gamma_{p'/q'}$ are homotopic in $\mathcal{S}$. Since Dehn twisting round ∂m is trivial in $\mathcal{S}$ and sends $X_i^k \to X_i^{k+2q}$, we have $\gamma_{p/q} \sim \gamma_{p/q+2}$. To see why $\gamma_{p/q} \sim \gamma_{-p/q}$, first note that the result does not depend on the relative twisting between $\mathrm{Ax}\, K_0, \mathrm{Ax}\, K_1$. Thus we shall consider the case in which $\sigma \in \mathbb{R}$ (recall that σ is the complex distance between these two axes), so that $\mathrm{Ax}\, K_0, \mathrm{Ax}\, K_1$ are coplanar and point in the same "vertical" direction as in Figure 9.

Consider the orientation-reversing symmetry r of reflection in the "horizontal" plane of Figure 9, that is the plane containing C orthogonal to the two axes $\mathrm{Ax}\, K_0, \mathrm{Ax}\, K_1$. (This is where we use that $\sigma \in \mathbb{R}$.) Clearly, this symmetry sends $\gamma_{p/q}$ to $\gamma_{-p/q}$. Fixing an orientation on $\partial\mathcal{S}$, let $\alpha, \beta, \gamma, \delta$ be anticlockwise loops on $\partial\mathcal{S}$ around the four cone points represented by the projections of the positive endpoints of the (oriented) axes of $K_0, K_0^{-1}, K_1, K_1^{-1}$, respectively, so that $\alpha\beta\gamma\delta = \mathrm{id}$ and $\alpha, \beta, \gamma, \delta$ generate $\pi_1(\partial\mathcal{S})$. Since r reverses orientation on $\partial\mathcal{S}$ it sends an anticlockwise loop round the positive endpoint of K_0 to a clockwise loop round the negative endpoint of K_0, which is the positive endpoint of the oriented axis of K_0^{-1}. Thus $r(\alpha) = \beta^{-1}$, and likewise $r(\gamma) = \delta^{-1}$. Since $\alpha = \beta^{-1} \leftrightarrow K_0$ and $\delta = \gamma^{-1} \leftrightarrow K_1$, it follows that $r(\gamma_{p/q}) = \gamma_{-p/q}$ represents the same element as $\gamma_{p/q}$ in $\pi_1(\mathcal{S})$.

We will show that if $p'/q' \neq \pm p/q + 2k$ for $k \in \mathbb{Z}$, then $\gamma_{p/q} \not\sim \gamma_{p'/q'}$ after computing traces; see Corollary 4.13. $\square$

4.4 Step 2: Computation of traces

Let $V_{p/q} \in \mathrm{SL}(2,\mathbb{C}) = \rho_x(\gamma_{p/q})$, where, since we want to compute $\operatorname{Tr} V_{p/q}$, we only need to consider $V_{p/q}$ up to cyclic permutation and inversion, and hence $\gamma_{p/q}$ only up to free homotopy. Rather than using the associated torus tree, we will work directly with a 4–holed sphere $\Sigma_{0,4}$ and the associated tree as described in [22]; see also [9]. Let $\alpha, \beta, \gamma, \delta$ denote loops round the four holes, oriented so that $\alpha\beta\gamma\delta = \mathrm{id}$. The fundamental group is identified with the free group F_3 with generators α, β, γ. A representation $\rho\colon F_3 \to \mathrm{SL}(2,\mathbb{C})$ is determined up to conjugation by its values on seven elements as follows (where we use $\widehat{w}$ in place of w in [22] etc to distinguish it from a variable w already in other use):

$$\operatorname{Tr}\rho(\alpha) = a; \quad \operatorname{Tr}\rho(\beta) = b; \quad \operatorname{Tr}\rho(\gamma) = c; \quad \operatorname{Tr}\rho(\delta) = d$$
$$\operatorname{Tr}\rho(\alpha\beta) = \widehat{x}; \quad \operatorname{Tr}\rho(\beta\gamma) = \widehat{y}; \quad \operatorname{Tr}\rho(\gamma\alpha) = \widehat{z}$$

related by the equation

$$\widehat{x}^2 + \widehat{y}^2 + \widehat{z}^2 + \widehat{x}\widehat{y}\widehat{z} = \widehat{p}\widehat{x} + \widehat{q}\widehat{y} + \widehat{r}\widehat{z} + \widehat{s}, \tag{4-1}$$

where

$$\widehat{p} = ab + cd, \quad \widehat{q} = bc + ad, \quad \widehat{r} = ac + bd, \quad \widehat{s} = 4 - a^2 - b^2 - c^2 - d^2 - abcd.$$

We identify our generators K_i as: $\alpha \leftrightarrow K_0$, $\beta \leftrightarrow K_1$, $\gamma \leftrightarrow K_2$, $\delta \leftrightarrow K_3$. Thus we find

$$a = b = c = d = -1, \qquad \begin{aligned} \widehat{x} &= \operatorname{Tr} K_0K_1 = x, \\ \widehat{y} &= \operatorname{Tr} K_1K_2 = 2, \\ \widehat{z} &= \operatorname{Tr} K_2K_0 = -x + 1. \end{aligned}$$

As a check, it is easy to verify that the trace identity (4-1) holds. Notice that none of the expressions $\widehat{p}, \dots, \widehat{z}$ depend on the sign choices made in Section 3.2.2.

The traces can be arranged in a trivalent tree in the usual way. As explained above, we have $\gamma_{0/1} \leftrightarrow K_0K_1$, $\gamma_{1/0} \leftrightarrow \mathrm{id}$, $\gamma_{1/1} \leftrightarrow K_0K_1^{-1}$. As explained in [22, Section 2.10], there are now 3 moves, depending on the values of $\widehat{p}, \widehat{q}, \widehat{r}$. In our case, $\widehat{p} = \widehat{q} = \widehat{r} = 2$, so the three moves described there coincide. Following [22], if u, v, w are labels round a vertex, with v, w labels adjacent along a common edge e, then the label at the vertex at the opposite end of e is $u' = 2 - vw - u$; compare Figure 3 in which $u' = vw - u$.

Clearly this procedure gives an algorithm for arranging curves and computing traces on a trivalent tree by analogy with that described in Section 2. Curves generated in this way inherit a natural labelling from the usual procedure of Farey addition as described in Section 2. Denote the curve which inherits the label p/q by $\delta_{p/q}$; we say this curve is in *Farey position* p/q on the tree. We shall refer to this tree together with its new rule for computing traces as the $\mathcal{S}$–tree, to distinguish it from the Markoff tree of Section 2.

We need to show that $\delta_{p/q}$ is the same as the curve $\gamma_{p/q}$ described in the previous section, namely the curve with $2q$ intersections with the meridian ∂m and a twist by p.

Lemma 4.10 *With the above notation,* $\delta_{p/q} = \gamma_{p/q}$.

Proof By definition we have $\delta_{p/q} = \gamma_{p/q}$ for $p, q \in \{0, 1\}$. With the notation above, these are the curves $\alpha\beta, \beta\gamma, \gamma\alpha$, each of which separates the punctures in pairs.

Call two essential simple nonperipheral curves on $\partial\mathcal{S}$ *neighbours* if they intersect exactly twice when in minimal position. Note that of the initial triple, each pair adjacent along an initial edge are neighbours, so that the triple round the initial vertex are neighbours in pairs. Note also that given a pair of neighbours δ, δ', there are exactly two other curves which are neighbours of both δ and δ'. If δ, δ' are adjacent along an edge of the tree, then these two further curves are exactly the remaining curves adjacent to the vertices at the ends of e.

These two further curves can be found by surgery, more precisely, by the Luo product defined in [21]. This works as follows. Arrange δ, δ' so as to have minimal intersection, cut them at their two intersection points and then make a consistent choice of the direction in which to turn to rejoin the resulting arcs. The Luo product rejoins the arcs by turning left at each intersection point (relative to a fixed orientation on the surface) as illustrated in Figure 10; equally we could rejoin by turning right at both intersection points. We denote the resulting curves by $\delta \cdot_L \delta'$ and $\delta \cdot_R \delta'$, respectively. It is not hard to see that $\delta \cdot_L \delta'$ is a neighbour of both δ and δ' and likewise for $\delta \cdot_R \delta'$. In particular, it is easy to check that $\delta_{0/1} \cdot_L \delta_{1/0} = \delta_{1/1}$ and $\delta_{0/1} \cdot_R \delta_{1/0} = \delta_{-1/1}$.

Now we show inductively that $\delta_{p/q} = \gamma_{p/q}$. As noted above, this is true for the initial values $0/1, 1/0, 1/1$ and $-1/1$. Suppose that it is true for neighbours $p/q, r/s$ where $|ps - rq| = 1$. By induction we may assume that $\delta_{p/q}, \delta_{r/s}$ are neighbours, hence adjacent along an edge e of the tree. By the above discussion we know that the additional curves at the two vertices of e are exactly $\delta_{p/q} \cdot_L \delta_{r/s}$ and $\delta_{p/q} \cdot_R \delta_{r/s}$. Moreover, by the inductive hypothesis, one of these curves must be $\delta_{r-p/s-q} = \gamma_{p-r/q-s}$ (or $\gamma_{r-p/s-q}$). Thus it remains only to show that the other curve is $\gamma_{p+r/q+s}$.

On each B_i, arrange $\delta_{p/q} = \gamma_{p/q}$ and $\delta_{r/s} = \gamma_{r/s}$ symmetrically with respect to the front and back of $\mathcal{S}$ as described above, then join the strands in the usual way.

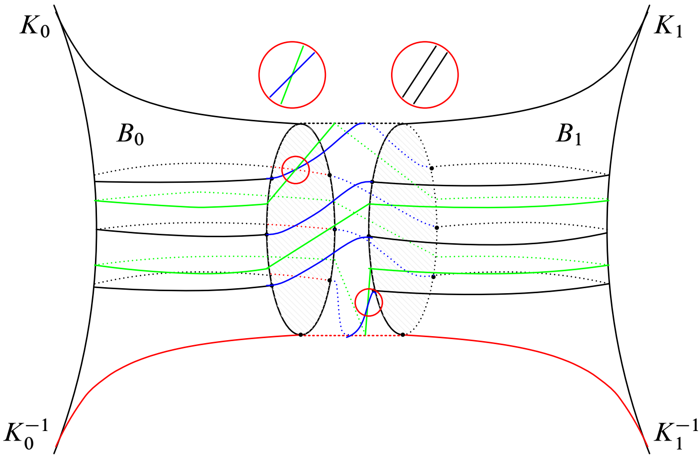


Figure 10: Here $\gamma_{1/3}$ and $\gamma_{1/2}$ are surgered to give $\gamma_{2/5}$; see the proof of Lemma 4.10. The inset circles show the direction of surgery.

With $\gamma_{p/q}$ in this position, its twist p is its intersection number with the geodesic α joining the two positive cone points given by the axes K_i, and likewise for $\gamma_{r/s}$. To take the Luo product, we have to cut $\gamma_{p/q}, \gamma_{r/s}$ at their intersection points and then make a consistent choice of which direction to rejoin the resulting arcs. Clearly the curve $\delta_{p/q} \cdot_L \delta_{r/s}$ with the "positive" surgery (see the inset circles in Figure 10) will have $2(q+s)$ intersection points with the meridian ∂m and intersection number $p+r$ with α, and hence must be $\gamma_{p+r/q+s}$. Since we already know the curve at one vertex of e is $\delta_{r-p/s-q} = \gamma_{p-r/q-s}$ (or $\gamma_{r-p/s-q}$), we must have $\delta_{p+r/q+s} = \gamma_{p+r/q+s}$. This completes the proof. □

In the following statement we make a particular, unimportant, choice of $\gamma_{p/q} \in \pi_1(\mathcal{S})$; see the beginning of Step 2 as above.

Proposition 4.11 *Let $V_{p/q}(x) = \rho_{\mathcal{S}}(x)(\gamma_{p/q})$. Then:*

(1) $\operatorname{Tr} V_{p/q} = \operatorname{Tr} V_{(p/q)+2} = \operatorname{Tr} V_{-p/q}$.

(2) $\operatorname{Tr} V_{p/q}(x) = \operatorname{Tr} V_{(p+q)/q}(1-x)$.

(3) $\operatorname{Tr} V_{p/q}$ *is a polynomial in x. If $0 \le p/q \le 1$, then its top two terms are* $(-1)^{p-q-1}(x^q - px^{q-1})$.

Remark 4.12 (3) should be compared to [15, Corollary 4.3] in which we showed that the leading term is of the form $(-1)^{p-q-1}cx^q$ for some $c > 0$; see also the remark following the corollary in that paper. Notice that by (1) and (2), it suffices to find the traces of curves the interval $0 \le p/q \le 1$.

Proof (1) This follows immediately from Proposition 4.9 and can also be proved easily by looking at the symmetries of the $\mathcal{S}$–tree.

(2) This results from the symmetry $x \mapsto 1-x$ which interchanges $\gamma_{0/1}, \gamma_{1/1}$.

(3) Note (1) holds for the three initial traces of $\gamma_{0/1}, \gamma_{1/0}, \gamma_{1/1}$. If curves $\gamma_{p/q}, \gamma_{r/s}$ are adjacent along an edge, then the two curves at the remaining vertices at the ends of the edge are $\gamma_{p\pm r/q\pm s}$. The result then follows easily by induction on the tree. □

Now we can prove the "only if" assertion of Proposition 4.9:

Corollary 4.13 *If $p'/q' \neq \pm p/q + 2k$ for $k \in \mathbb{Z}$, then $\gamma_{p/q} \not\sim \gamma_{p'/q'}$.*

Proof This follows directly by comparing the top two terms of $\operatorname{Tr} V_{p/q}, \operatorname{Tr} V_{p'/q'}$. □

4.5 Step 3: The exceptional Fuchsian case: computation of $\mathcal{P}_{0/1}$, $\mathcal{P}_{1/1}$

As above, let $\mathcal{P}_{p/q}$ denote the pleating ray of $\gamma_{p/q}$. The rays $\mathcal{P}_{0/1}, \mathcal{P}_{1/1}$ are exceptional. Since $\gamma_{0/1} \leftrightarrow K_0K_1$, $\gamma_{1/1} \leftrightarrow K_0K_1^{-1}$, we have $\operatorname{Tr} V_{0/1} = x$, $\operatorname{Tr} V_{1/1} = 1-x$. Thus the real locus for both trace polynomials is exactly the real axis, and on this locus, the group $G_{\mathcal{S}}(x)$, if discrete, is Fuchsian. This is exactly the situation discussed in [15, page 84].

In the ball model of $\mathbb{H}^3$, identify the extended real axis with the equatorial circle. Since the limit set is contained in $\widehat{\mathbb{R}}$, the convex core (the Nielsen region) of $G_{\mathcal{S}}(x)$ is contained in the equatorial plane. We can think that the convex core has been squashed flat and the bending lines are just the boundary of the Nielsen region, that is, the boundary of the convex core of the surface $\mathbb{H}^2/G_{\mathcal{S}}(x)$. Thus to find the bending lamination we just have to determine the boundary of $\mathbb{H}^2/G_{\mathcal{S}}(x)$.

Now if $x \in \mathbb{R}$, then either $\zeta \in \mathbb{R}$ and $x > 0$, or $\zeta \in i\mathbb{R}$ and $x < 0$. In both cases, we find a fundamental domain for $G_{\mathcal{S}}(x)$ as described in Section 4.1; see Figure 11. Thus regarded as a Fuchsian group acting on the upper half plane $\mathbb{H}$, $G_{\mathcal{S}}(x)$ represents a sphere with two order-3 cone points and one hole. However the cases $x < 0$ and $x > 0$ are slightly different, because of the relative directions of rotation of K_0 and K_1.

In both cases, the axis K_0 has fixed points $\pm i\sqrt{3}$ and its axis is oriented so that it is anticlockwise rotation about $i\sqrt{3}$. Thus $K_1 = PK_0P^{-1}$ rotates anticlockwise about $P(i\sqrt{3}) = -i\zeta^2/\sqrt{3}$. If $x < 0$ then $P(i\sqrt{3})$ is in the upper half plane $\mathbb{H}$, while if $x > 0$ then $P(i\sqrt{3})$ is in the lower half plane. Hence if $x < 0$ then K_0, K_1 rotate in the same sense about their fixed points in $\mathbb{H}$, while if $x > 0$ their rotation directions are opposite. This leads to the two different configurations shown in Figure 11.

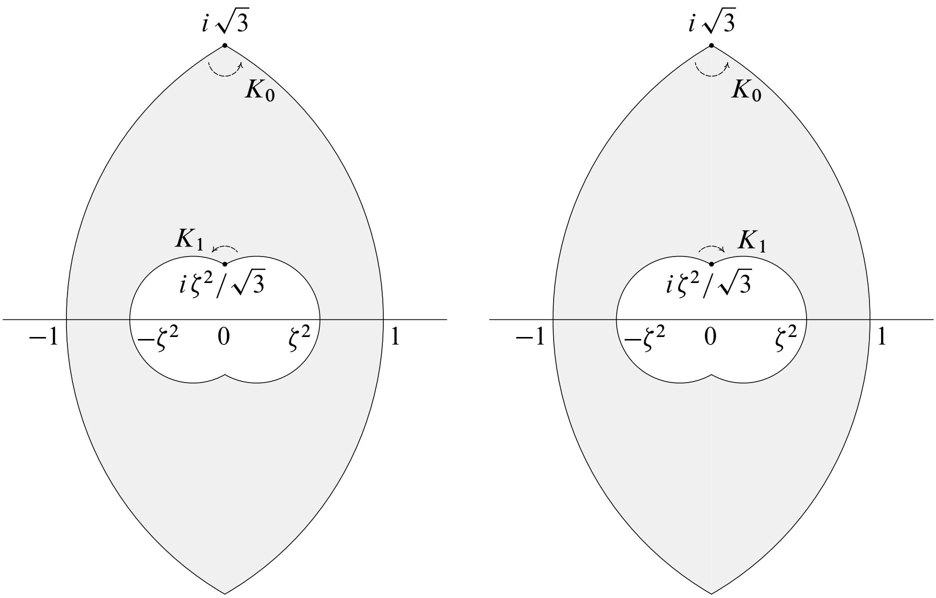


Figure 11: Configurations for $x \in \mathbb{R}$. Left: $\zeta \in i\mathbb{R}$, $x \le -2$; K_0 and K_1 rotate in the same directions $\mathbb{H}$ and the hole is $K_0 K_1$. Right: $\zeta \in \mathbb{R}$, $x \ge 3$; K_0 and K_1 rotate in opposite directions in $\mathbb{H}$ and the hole is $K_0 K_1^{-1}$.

As is easily checked, if $x > 0$ the boundary of the hole is thus $K_0 K_1^{-1}$ while if $x < 0$ the boundary of the hole is $K_0 K_1$. Since $K_0 K_1 \leftrightarrow \gamma_{0/1}$ and $K_0 K_1^{-1} \leftrightarrow \gamma_{1/1}$, combining this with information about the discreteness locus in the Fuchsian case from Section 4.1, we conclude that $\mathcal{P}_{0/1} = (-\infty, -2]$ and $\mathcal{P}_{1/1} = [3, \infty)$.

4.6 Step 4: Nonsingularity of pleating rays

This is the part of the argument which contains the deepest mathematics. Fortunately, the results needed have been proved elsewhere.

Theorem 4.14 [18; 15; 5] *Suppose that $\gamma \in \pi_1(\mathcal{S})$. Then $\mathcal{P}_\gamma$ is open and closed in the real trace locus $\mathbb{R}_\gamma$. Moreover, $\operatorname{Tr} \rho_x(\gamma)$ is a local coordinate for $\mathbb{C}$ in a neighbourhood of $\mathcal{P}_\gamma$, and is a global coordinate for $\mathcal{P}_\gamma$ on any nonempty connected component of $\mathcal{P}_\gamma$.*

Proof The statement that $\mathcal{P}_\gamma$ is open in $\mathbb{R}_\gamma$ is essentially [15, Proposition 3.1]; see also [18, Theorems 15 and 26]. The fact that $\operatorname{Tr} \rho_x(\gamma)$ is a local parameter is equivalent to the fact, also proved in both [15] and [14], that $\mathcal{P}_\gamma$ is a nonsingular 1–manifold. The openness and the final statement are actually a special case of Theorems B and C of [5] which state that for general hyperbolic manifolds, if the support of the bending lamination is rational (a union of closed curves), then the traces of these curves are local parameters for the deformation space in a neighbourhood of the corresponding pleating variety.

That $\mathcal{P}_\gamma$ is closed in $\mathbb{R}_\gamma$ can be proved as in [15, Theorem 3.7]. Here is a slightly more sophisticated version of the same idea. Suppose $x_n \to x_\infty$ with $x_n \in \mathcal{P}_\gamma$. The limit group $G_\mathcal{S}(x_\infty)$ is an algebraic limit of groups $G_\mathcal{S}(x_n)$ and hence the corresponding representation is discrete and faithful. Each of the two components of $(\partial\mathcal{C}/G_\mathcal{S}(x_n))\setminus\gamma$ is a flat surface corresponding to a conjugacy class of Fuchsian subgroup $F_j(x_n)$ for $j = 1, 2$ (the F–peripheral subgroups of [15]). Since the limit is algebraic, $F_j(x_n)$ limits on a Fuchsian subgroup $F_j(x_\infty)$, and similarly for all its conjugates in $G_\mathcal{S}(x_\infty)$.

The limit sets Λ_α of each of these subgroups F_α is spanned by a hyperbolic plane H_α in $\mathbb{H}^3$. The Nielsen regions of F_α in H_α fit together along the lifts of the bending line γ to $\mathbb{H}^3$, forming a pleated surface Π in $\mathbb{H}^3$. We claim that $\Pi = \partial\mathcal{C}(G_\mathcal{S}(x_\infty))$. This follows since the closure of the union of the Λ_α is the limit set of $G_\mathcal{S}(x_\infty)$; see also Proposition 7.2 in [17]. The result follows. □

Remark 4.15 The closedness of $\mathcal{P}_\gamma$ in $\mathbb{R}_\gamma$ is a simple case of both the "local limit theorem", Theorem 15 in [18] and the "lemme de fermeture" of [2]. These much more sophisticated results allow that the bending lines may be part of an irrational lamination. Our argument above, in which the bending lamination is supported on closed curves, is very close to that in the first part of the proof of Théorème 6 in [26].

Corollary 4.16 [14; 15; 5] *If $\mathcal{P}_\gamma \neq \varnothing$, then it is a union of connected nonsingular branches of the real trace locus $\mathbb{R}_\gamma$.*

Proof Suppose that $\mathcal{P}_\gamma \neq \varnothing$ and let $x \in \mathcal{P}_\gamma$, so that by Lemma 4.6, $x \in \mathbb{R}_\gamma$. By Theorem 4.14, $\mathcal{P}_\gamma$ is open and closed in $\mathbb{R}_\gamma$. Since $\operatorname{Tr}\rho_x(\gamma)$ is a local coordinate, in a neighbourhood of x the locus $\mathbb{R}_\gamma$ is a 1–manifold. □

Notice that the theorem says that $\operatorname{Tr}\rho_x(\gamma)$ is a local parameter even in a neighbourhood of a cusp where $\rho_x(\gamma)$ is parabolic [5, Theorem C]. Thus we have

Corollary 4.17 *Suppose that $x \in \mathcal{P}_\gamma$. Then there is a neighbourhood of x in $\mathbb{C}$ on which $x \in \mathbb{R}_\gamma$ implies that $x \in \mathcal{D}$.*

Corollary 4.18 *If $\mathcal{P}_\gamma \neq \varnothing$, then* $\operatorname{Tr}\rho_x(\gamma)$ *is unbounded on $\mathcal{P}_\gamma$.*

Proof Since $\operatorname{Tr}\rho_x(\gamma)$ is a local coordinate on connected components of $\mathcal{P}_\gamma$, this follows from the maximum principle on the branch; see [15, Theorem 4.1]. □

4.7 Step 5: Finding the nonempty pleating rays

Now we determine the pleating rays. As above, let $\mathcal{P}_{p/q}$ denote the ray corresponding to the curve $\gamma_{p/q}$ and write $\mathbb{R}_{p/q}$ for the real locus of $\operatorname{Tr} V_{p/q}$. From Proposition 4.9 we have $\mathcal{P}_{p/q} = \mathcal{P}_{(p+2q)/q} = \mathcal{P}_{-p/q}$.

By Section 4.6, $\mathcal{P}_{p/q}$ is a union of nonsingular branches of $\mathbb{R}_{p/q}$. We now find those $p/q \notin \{0, 1\}$ for which $\mathcal{P}_{p/q} \neq \varnothing$, at the same time resolving the connectivity issue. We follow the method of [15], using an inductive argument on position of the pleating rays and their asymptotic directions as $|x| \to \infty$, and at the same time correcting the second of the two errors referred to in Remark 4.7. We have:

Proposition 4.19 (cf [15, Theorem 4.1]) *The set $\mathcal{P}_{p/q}$ is the union of the two branches of $\mathbb{R}_{p/q}$ which are asymptotic to the half lines $\rho e^{\pm i\pi(p-q)/q}$ as $\rho \to \infty$.*

Proof Denote by $R(\theta)$ the ray $te^{i\theta}, t > 0$, in the x–plane. By Proposition 4.11, $\operatorname{Tr} V_{p/q}$ is a polynomial in x whose top term is $(-1)^{p-q-1}x^q$. Now $\operatorname{Tr} V_{p/q}$ takes real values on $\mathcal{P}_{p/q}$, moreover by Corollary 4.18 it is unbounded on $\mathcal{P}_{p/q}$. It follows that $\mathcal{P}_{p/q}$ must be asymptotic to one of the rays $R(k\pi/q)$ for some $k \in \mathbb{Z}$ as $|x| \to \infty$.

We have already identified $\mathcal{P}_{0/1}$ and $\mathcal{P}_{1/1}$ as the real intervals $(-\infty, -3]$ and $[2, \infty)$, respectively. It follows from Section 4.1 that the semicircular arc from -4 to 4 (say) in $\mathbb{H}$ is a continuous path in $\mathcal{D}$ from $\mathcal{P}_{0/1}$ to $\mathcal{P}_{1/1}$. Hence by the continuity theorem of [16], if $0 < p/q < 1$, there is a point on $\mathcal{P}_{p/q}$ in the upper half plane $\mathbb{H}$. Likewise, there is a point on $\mathcal{P}_{p/q}$ in the lower half plane. (This was missed in [15].) Since $\mathcal{P}_{0/1} \cup \mathcal{P}_{1/1}$ separates $\mathcal{D}$ into two connected components, this shows in particular that $\mathcal{P}_{p/q}$ must have at least two connected components.

Now we proceed by induction on the Farey tree. Suppose we have shown the result for two Farey neighbours $p/q, r/s$. Consider the locus $\mathcal{P}_{p+r/q+s}$. By the inductive hypothesis, $\mathbb{H}$ contains exactly one component of each of $\mathcal{P}_{p/q}, \mathcal{P}_{r/s}$, asymptotic to the rays $R(\pi(p-q)/q)$, $R(\pi(r-s)/s)$, respectively. Exactly as in [15], it is easy to check that there is exactly one integer $k \in \{0, 1, \dots, 2(q+s)-1\}$ for which $R(k\pi/(q+s))$ lies between $R(\pi(p-q)/q)$ and $R(\pi(r-s)/s)$, namely $k = (p+r)/(q+s)$. By the same continuity theorem as before, a path in this sector joining suitable points on $\mathcal{P}_{p/q}, \mathcal{P}_{r/s}$ must meet $\mathcal{P}_{p+r/q+s}$. Thus $\mathcal{P}_{p+r/q+s}$ has at least one connected component asymptotic to $R(\pi(p+r-q-s)/(q+s))$. A similar argument in the lower half plane gives another connected component asymptotic to $R(\pi(p+r+q+s)/(q+s))$. Since $\mathcal{P}_{p+r/q+s}$ has exactly two components by Proposition 4.20 below, the result follows. $\square$

The issue of connectivity of $\mathcal{P}_\gamma$ is a bit subtle. In the general theory (see [2; 5]), one shows that $\mathcal{P}_\gamma$ has one connected component. However this result holds in a space of manifolds which are consistently oriented throughout the space and all of whose convex cores have nonzero volume. In our case we have:

Proposition 4.20 *If $\gamma \neq \gamma_{0/1}, \gamma_{1/1}$ and $\mathcal{P}_\gamma \neq \varnothing$, then $\mathcal{P}_\gamma$ has exactly two connected components in $\mathcal{D}$.*

Proof The usual argument that the pleating ray of a rational lamination has one connected component goes as follows. Given a point on $\mathcal{P}_\gamma$, double the convex core along its boundary to obtain a cone manifold with a singular axis of angle $2(\pi-\theta)$ along γ, where θ is the bending angle along γ. (Notice that the convention on defining bending angles differs between papers by the first author and [2]. In our convention, a bending line contained in flat subsurface has bending angle 0 but cone angle 2π, whereas in [2], the bending angle along a line in a flat surface is defined to be π.) By [13], such a hyperbolic cone manifold is parametrised by its cone angle. One shows that one can continuously deform the cone angle to 0, at which point the curve whose axis is the bending line has to become parabolic. The doubled manifold is an oriented hyperbolic manifold with a rank-two cusp and finite volume. As long as we are working in a space in which all manifolds have consistent orientation, such a manifold is unique up to orientation-preserving isometry, from which one deduces that $\mathcal{P}_\gamma$ is connected.

In our case, the parameter space $\mathcal{D}$ is separated by two lines along which G is Fuchsian so that $\mathcal{C}(G)/G$ has zero volume and the above argument fails. Note however that, provided that G is not Fuchsian, $\mathcal{S}$ can be oriented by the triple consisting of the *oriented* axes of P, Q and the oriented line C from $\mathrm{Ax}\,K_0$ to $\mathrm{Ax}\,K_1$. The map $\zeta \to \overline{\zeta}$ reverses the relative orientations of $\mathrm{Ax}\,P, \mathrm{Ax}\,Q$ while fixing that of C. Thus $\mathcal{D}\setminus\mathbb{R}$ has two connected components in which $\mathcal{S}$ has naturally opposite orientations. The above argument shows that $\mathcal{P}_\gamma$ has at most one component in each component of $\mathcal{D}$. Since we have already shown in Proposition 4.19 that $\mathcal{P}_\gamma$ has at least one component in each of the upper and lower half planes, this completes the proof.

This result can alternatively be proved by the more ad hoc methods used in [15]. □

Remark 4.21 Proposition 4.19 shows that $\mathcal{P}_{p/q} \neq \varnothing$ for all $p/q \in \mathbb{Q}$. This can be viewed as a special case of the general result of [2, Theorem 1]; see also [5, Theorem 2.4]. We have to be careful to include the case, excluded in [2], that the group $G_{\mathcal{S}}(x)$ is Fuchsian so that $\mathcal{C}/G$ has zero volume. The conclusion is the following:

Proposition 4.22 *Let γ be an essential simple nonperipheral closed curve on $\partial\mathcal{S}$. Then $\mathcal{P}_\gamma \neq \varnothing$ if and only if γ is nontrivial in $\pi_1(\mathcal{S})$ and intersects the meridian disk $\gamma_{1/0}$ at least twice. If γ meets $\gamma_{1/0}$ exactly twice, then the bending angle is identically π and $G_{\mathcal{S}}(x)$ is Fuchsian.*

4.8 Step 6: Density of rational pleating rays

Finally, we justify the claim that the rational pleating rays are dense in $\mathcal{D}$:

Theorem 4.23 [14, Corollary 6.2; 15, Theorem 5.2] *Rational pleating rays are dense in $\mathcal{D}_{\mathcal{S}}$.*

Proof The proof of this result in any one-complex-dimensional parameter space is the same. Here is a quick sketch. Suppose that ν is an irrational lamination with corresponding pleating variety $\mathcal{P}_\nu$, and that $x \in \mathcal{P}_\nu$. Pick a sequence of rational measured laminations $\nu_n = c_n \delta_{\gamma_n}$ where $c_n \in \mathbb{R}^+$ so that $\nu_n \to \nu$ in the space of projective measured laminations on $\partial\mathcal{S}$, where δ_{γ_n} is the unit point mass on γ_n. Replace the traces of γ_n by complex length functions λ_n and scale to get complex analytic functions $c_n\lambda_n$. One shows that in a neighbourhood of $x \in \mathcal{P}_\nu$ these functions form a normal family which converges to a nonconstant analytic function [14, Theorem 6.9; 18, Theorem 20], whose real locus contains the pleating ray $\mathcal{P}_\nu$ [18, Theorem 23]. By Hurwitz's theorem, there are nearby points at which the approximating functions $c_n\lambda_n$ must take on real values. In a small enough neighbourhood of x, this is enough to force $y \in \mathcal{P}_{\gamma_n}$ [18, Theorem 31]. This gives density in $\operatorname{Int}\mathcal{D}$. By the result quoted in the introduction that $\mathcal{D} = \overline{\operatorname{Int}\mathcal{D}}$ we are done. □

Remark 4.24 $\mathcal{D}$ as defined above (see the beginning of Section 4) includes the parabolic cusp groups on $\partial\mathcal{D}$. In fact these groups are exactly the geometrically finite groups on $\partial\mathcal{D}$, and hence exactly the groups in $\mathcal{D}$ but not in $\mathcal{SCH}$ as defined in the introduction. Since there are only countably many such groups, and since $\operatorname{Int}\mathcal{D} = \mathcal{SCH}$, whether or not we include them in the parameter space does not materially affect our computations. See [23] for more on this and related issues.

4.9 The pleating rays for $\mathcal{H}$

By Corollary 4.3, $\mathcal{D}_{\mathcal{H}} = \mathcal{D}_{\mathcal{S}}$. Thus the rational rays for $\mathcal{D}_{\mathcal{S}}$ are also dense in $\mathcal{D}_{\mathcal{H}}$. However it is easy to see that a rational pleating laminations on $\partial\mathcal{H}(x)$ correspond exactly to those on $\mathcal{S}(x)$, and that although the actual bending curves differ, their traces are related by a simple formula.

Lemma 4.25 *Suppose that the bending lamination $\beta_{\mathcal{H}}(x)$ of $\mathcal{H}(x)$ is rational, so that its support λ is a union of disjoint simple closed curves on $\partial\mathcal{H}$. Let γ be a connected component of λ. Then either $\kappa(\gamma) = \gamma$ or the three curves $\gamma, \kappa(\gamma), \kappa^2(\gamma)$ are disjoint. The support of the bending lamination $\beta_{\mathcal{S}}(x)$ is exactly the projection of γ to $\mathcal{S}$, and all rational bending laminations of $\mathcal{S}$ arise in this way.*

Proof The limit set of $G_{\mathcal{H}}(x)$ and hence its convex core are invariant under the symmetry κ. Hence the support λ of $\beta_{\mathcal{H}}(x)$ is also κ–invariant. Let γ be a connected component of λ. Since connected components of λ are pairwise disjoint, either $\kappa(\gamma)=\gamma$ or the three curves $\gamma,\kappa(\gamma),\kappa^2(\gamma)$ are disjoint. In either case, γ cannot pass through a fixed point of κ: at the fixed point P the images of γ would meet at angles $\frac{2\pi}{3}$, so that $\gamma,\kappa(\gamma),\kappa^2(\gamma)$ would intersect at P, which is impossible.

Let π_κ be the projection $\mathcal{H}\to\mathcal{S}$. In a neighbourhood of a bending line π_κ is a covering map hence a local isometry. Since being a bending line can be characterised locally, $\beta_{\mathcal{S}}(x)$ is the projection of γ to $\mathcal{S}$.

Let γ be a simple closed curve on $\partial\mathcal{S}$. Clearly, by the same observation about local characterisation of bending lines, if γ is a bending line, then so is any connected component of its lift to $\partial\mathcal{H}$. This proves the converse. □

We remark that if p/q is congruent to $1/0$ or $0/1$ mod $\mathbb{Z}_2$, then the lift of $\gamma_{p/q}$ has three connected components which are permuted among themselves by κ, while if p/q is congruent to $1/1$, then its lift has one κ–invariant connected component. To see this, check by hand for the curves $\gamma_{1/0},\gamma_{0/1},\gamma_{1/1}$ and then note that the lifting property is invariant under the mapping class group of $\partial\mathcal{S}$ which at the same time acts transitively on p/q congruence classes mod $\mathbb{Z}_2$.

To actually compute the pleating rays for $\mathcal{D}_{\mathcal{S}}$, we computed the traces $\operatorname{Tr} V_{p/q}(x)$ corresponding to the curves $\gamma_{p/q}\in\pi_1(\mathcal{S})$. The above discussion shows that it is unnecessary to actually compute traces of lifted curves in $\pi_1(\mathcal{H})$. If for some reason one wanted to do this, either one could start again enumerating the curves on $\mathcal{H}$, or one could note that the complex length of a lift of $\gamma_{p/q}$ in $\mathcal{H}$ would be either the same as or three times that of the curve $\gamma_{p/q}$ in $\mathcal{S}$, depending on the $\mathbb{Z}_2$–parity of p/q.

5 Computing traces

To compute traces of the elements $V_{p/q}$, rather than use the $\mathcal{S}$–tree as in Section 4.4, we actually performed computations on the associated Markoff tree corresponding to the associated torus $\mathcal{T}$ of Section 3.2.3, referred to in this section as the $\mathcal{T}$–tree. To justify this, we need to compare the curves in Farey position p/q on the two trees to ensure that they do indeed correspond geometrically as expected. We also need to address the issue about lifting representations to $\mathrm{SL}(2,\mathbb{C})$ raised in Remark 3.2.

5.0.1 Correspondence of curves Homotopy classes of essential simple nonperipheral loops on $\partial\mathcal{T}$ are well known to be in bijective correspondence to unoriented lines

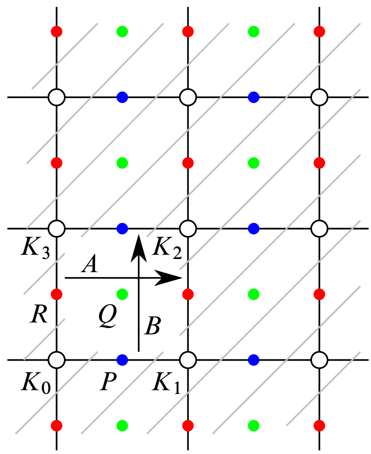


Figure 12: Lattice representation of a cover of $\partial\mathcal{S}$. The integer vertices (open circles) correspond to the endpoints of the order-3 axes on $\partial\mathcal{S}$; the endpoints of the order-2 elliptics P, Q and R are on horizontal segments (blue), in the middle of squares (green), and on vertical segments (red), respectively.

of rational slope in the plane; see for example [27; 14]. In fact the word $W_{p/q}$ generated by the concatenation process following the $\mathcal{T}$–tree described in Section 2 is the cutting sequence of a line of slope $p/q \in \widehat{\mathbb{Q}}$ across the lattice; see [27].

The key point here is that the plane with a cone singularity of angle $\frac{4\pi}{3}$ at integer lattice points (see Figure 12), is an intermediate covering between the universal cover $\mathbb{H}$ of $\partial\mathcal{T}$ and $\partial\mathcal{T}$ itself. As described in for example [15], the same lattice can also be viewed as an intermediate covering between $\mathbb{H}$ and $\partial\mathcal{S}$: the rectangle with vertices at $0, 1, 2i, 2i+1$ can be viewed as a fundamental domain for the lattice action corresponding to $\partial\mathcal{S}$ which projects, bijectively on its interior, to $\partial\mathcal{S}$. Likewise, the rectangle with vertices $0, 1, \frac{1}{2}i, 1+\frac{1}{2}i$ projects in a similar way to $\partial\mathcal{U}$ and the unit square projects to the torus $\partial\mathcal{T}$. The lattice points correspond to the cone points belonging to K_i for $i = 1, \dots, 4$ arranged as shown. Thus there is also a bijective correspondence between lines of rational slope in the punctured plane and simple essential nonperipheral curves on $\partial\mathcal{S}$. In this way, one can easily relate the words $W_{p/q}$ (on $\partial\mathcal{T}$) and $V_{p/q}$ (on $\partial\mathcal{S}$); this is explained in detail in [15].

In this picture, the meridian loop ∂m of Section 4.3 is identified as the “vertical” line of slope $1/0$. One sees easily that the line of slope p/q in the plane projects to a curve on $\partial\mathcal{S}$ which has exactly $2q$ intersections with ∂m and a twist of p as described in Section 4.3. It follows from Lemma 4.10 that the labelling of curves by lines of rational slope p/q exactly corresponds to the Farey labelling of curves by their position on the $\mathcal{S}$–tree.

As above, the curve represented in Farey position p/q on the $\mathcal{S}$–tree is denoted by $\gamma_{p/q}$, corresponding to the word $V_{p/q}$; while the curve represented in Farey position p/q on the $\mathcal{T}$–tree is denoted by $\omega_{p/q}$, corresponding to the word $W_{p/q}$. Now $\mathcal{S}$ projects to $\mathcal{U}$ by a four-fold cover and $\mathcal{T}$ projects to $\mathcal{U}$ by a two-fold cover.

Proposition 5.1 *The complex translation length of the geodesic representative of $\gamma_{p/q}$ is twice that of $\omega_{p/q}$; hence $\operatorname{Tr} V_{p/q}(\zeta) = \pm((\operatorname{Tr} W_{p/q}(\zeta))^2 - 2)$.*

Note that this allows for an ambiguity in the signs of the traces since the two lifts of $\pi_1(\mathcal{T})$ and $\pi_1(\mathcal{S})$ to $\mathrm{SL}(2,\mathbb{C})$ are not (indeed cannot be) chosen consistently.

Corollary 5.2 *Up to sign, the trace of the image $V_{p/q}$ of $\gamma_{p/q} \in \pi_1(\mathcal{U})$ may be computed using the formula of Proposition 5.1 and the $\mathcal{T}$–tree.*

Since we are aiming to compute pleating rays which are a geometrical construct and hence only depend on a $\mathrm{PSL}(2,\mathbb{C})$ representation, this would be sufficient for our purposes. However it is more satisfying to prove the following more precise result which shows that working with the $\mathrm{SL}(2,\mathbb{C})$ lift of the representation of $\pi_1(\mathcal{T})$ described in Section 3.2.3, we can fix the choice of sign.

Proposition 5.3 *With $W_{p/q}, V_{p/q}$ as above, let $f_{p/q}(\zeta) = \operatorname{Tr} V_{p/q}(\zeta)$ and $g_{p/q}(\zeta) = \operatorname{Tr} W_{p/q}(\zeta)$. Then $-f_{p/q}(\zeta) = (g_{p/q}(\zeta))^2 - 2$ for all $p/q \in \widehat{\mathbb{Q}}$.*

Proof It is easy to check that this is correct for $p/q = 0/1, 1/0, 1/1$. In detail, (recalling that as above $\gamma \leftrightarrow Z$ means that $\gamma \in \pi_1(\mathcal{S})$ or $\pi_1(\mathcal{T})$ is represented by $Z \in \mathrm{SL}(2,\mathbb{C})$):

- $\omega_{0/1} \leftrightarrow A$ and $\gamma_{0/1} \leftrightarrow K_0K_1$, and we have shown that $A^2 = -K_0K_1$. Thus $f_{0/1}(\zeta) = x$ and $(g_{0/1}(\zeta))^2 - 2 = (-x+2) - 2 = -x$.
- $\omega_{1/0} \leftrightarrow B$, $\gamma_{1/0} \leftrightarrow \mathrm{id}$ and $B^2 = -\mathrm{id}$. So $f_{1/0}(\zeta) = 2$ and $(g_{1/0}(\zeta))^2 - 2 = -2$.
- $\omega_{1/1} \leftrightarrow AB$ and $\gamma_{1/1} \leftrightarrow K_0K_1^{-1}$. So $f_{1/1}(\zeta) = 1-x$ and $(g_{1/1}(\zeta))^2 - 2 = x-1$.

Now we do an inductive proof. Suppose that in the $\mathcal{S}$–tree labels u, v are adjacent along an edge e with w the remaining label at one of the two vertices at the ends of e. By the formula in Section 4.4 the label at the other vertex is $2 - uv - w$.

Suppose that the corresponding labels on the $\mathcal{T}$–tree are u', v', w'. Then the remaining label at the vertex at the other end of e is $u'v' - w'$. Replace these labels by the negatives of the traces of the doubled curves to get labels $2-u'^2, 2-v'^2, 2-w'^2, 2-(u'v'-w')^2$ around the same 4 vertices. If we can show that

$$2 - (2-u'^2)(2-v'^2) - (2-w'^2) = 2 - (u'v' - w')^2,$$

we will be done. This is easily checked by multiplying out, noting that the trace identity (4-1) round a vertex of the $\mathcal{T}$–tree gives

$$u'^2 + v'^2 + w'^2 = u'v'w' + \operatorname{Tr}[A, B] + 2 = u'v'w' + 3. \qquad \square$$

5.0.2 The actual computations The above discussion justifies the method we actually used to perform computations involving traces on $\mathcal{S}$. Instead of computing on the $\mathcal{S}$–tree with initial traces $\operatorname{Tr}\rho_{\mathcal{S}}(\gamma_{0/1}) = \operatorname{Tr} K_0K_1 = x$, $\operatorname{Tr}\rho_{\mathcal{S}}(\gamma_{1/0}) = \operatorname{Tr}\mathrm{id} = 2$, $\operatorname{Tr}\rho_{\mathcal{S}}(\gamma_{1/1}) = \operatorname{Tr} K_0K_1^{-1} = 1-x$, we used the $\mathcal{T}$–tree with initial triple $\operatorname{Tr} A = \pm i(3/(2\zeta) - \zeta/2)$, $\operatorname{Tr} B = 0$ and $\operatorname{Tr} AB = \pm(3/(2\zeta) + \zeta/2)$ corresponding to the generators A, B of $G_{\mathcal{T}}$. As in Section 3.2.3, $A^2 = -K_0K_1$, so that $\operatorname{Tr} A^2 = -x$. Since $\operatorname{Tr} B = 0$, we can find $\operatorname{Tr} AB$ from the identity $(\operatorname{Tr} A)^2 + (\operatorname{Tr} AB)^2 = \operatorname{Tr}[A, B] + 2 = 3$. Thus setting $(a, b, c) = (\operatorname{Tr} A, \operatorname{Tr} B, \operatorname{Tr} AB)$ we have

$$a^2 - 2 = -x, \quad c^2 = 1 + x.$$

It is easily checked that this is in accord with (3-2). Thus associated to $G_{\mathcal{T}}(x)$ we have the torus tree $(a, b, c) = (\sqrt{-x+2}, 0, \sqrt{x+1})$. This is the method we actually used to compute the pleating rays shown in Figure 1.

Remark 5.4 The sign of the square roots in the above can be uniquely determined by the formulae for traces in terms of ζ. What we actually did was to make an arbitrary choice and plot rays corresponding to curves in the range $0 \le p/q \le 1$, thus making a picture in the upper half plane which we could then reflect. As can be seen from Figure 4, the signs of the square roots in fact alternate periodically with period 4 rather than period 2, so that, for example, $\operatorname{Tr}\rho_{\mathcal{T}}(\gamma_3) = -\operatorname{Tr}\rho_{\mathcal{T}}(\gamma_1)$.

5.0.3 Computations for the Riley slice The traces needed to find the pleating rays for the Riley slice in the lower frame of Figure 8 were computed by a method similar to that described above. Our parameter x can be related to the parameter ρ of [15] by comparing the traces of the word in Farey position $0/1$: these are K_0K_1 in our case and XY in the notation of [15]. Thus we find that x corresponds to $\rho + 2$. For the Riley group a similar computation to the one above with $\operatorname{Tr}[A, B] = -2$ gives immediately $(\operatorname{Tr} A)^2 = -(\rho + 2)$ and $(\operatorname{Tr} AB)^2 = \rho + 2$. Thus writing in terms of the x–coordinate we find the initial triple $(\sqrt{-x}, 0, \sqrt{x})$.

5.0.4 Comparison of Bowditch sets It is interesting to compare the Bowditch sets associated to the two initial triples (x, x, x) and $(\sqrt{-x+2}, 0, \sqrt{x+1})$. In the latter case, one needs to modify the definition of the Bowditch set: since $\phi(U) = 0$ for some $U \in \Omega$, there is a trace-preserving $\mathbb{Z}$–action on the associated tree $\mathbb{T}_{(\sqrt{-x+2}, 0, \sqrt{x+1})}$ corresponding to the action of a subgroup of $\operatorname{Aut}(F_2)$ generated by a parabolic; see,

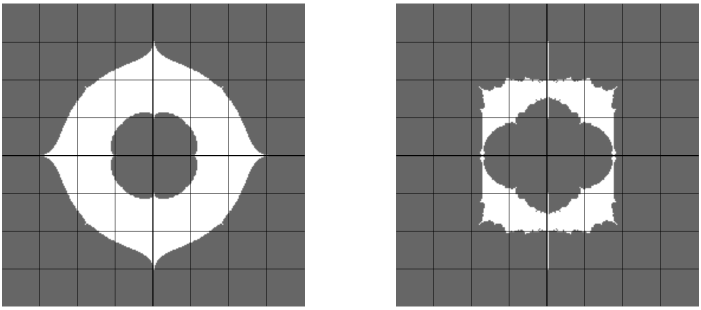


Figure 13: Bowditch sets (grey) plotted in the ζ–plane with range $[-4,4]\times[-4i,4i]$. Left: initial triple $(\sqrt{-x+2},0,\sqrt{x+1})$ corresponding to the torus group $G_{\mathcal{T}}(x)$. Right: initial triple (x,x,x) corresponding to the handlebody group $G_{\mathcal{H}}(x)$. The two regions are clearly distinct: the grey region on the right contains that on the left.

for example, [29, Theorem 1.6]. The Bowditch condition should actually be specified on $\Omega\setminus\{U\}/\sim$, where $\sim$ is the equivalence coming from this symmetry.

The results, plotted in the ζ–plane, are shown in Figure 13. On the right, the initial triple is (x,x,x) (with x related to ζ as in (3-1)) corresponding to the handlebody group $G_{\mathcal{H}}(x)$. On the left, the initial triple is $(\sqrt{-x+2},0,\sqrt{x+1})$ corresponding to the torus group $G_{\mathcal{T}}(x)$. The two regions are clearly distinct: the grey region on the right contains that on the left. Conjecturally, the left-hand grey region is also the discreteness locus for the groups $G_{\mathcal{T}}(x)$; see Figure 8 for the parametrisation in terms of x.

Note the various symmetries as discussed in Section 3.2.6, in particular note how Figure 5 looses the left-right reflectional symmetry seen in Figure 13. The coloured region in Figure 5 is the same region as the right frame of Figure 13, drawn in the x–plane.

References

[1] **H Akiyoshi**, **M Sakuma**, **M Wada**, **Y Yamashita**, *Punctured torus groups and 2–bridge knot groups, I*, Lecture Notes in Mathematics 1909, Springer (2007) MR

[2] **F Bonahon**, **J-P Otal**, *Laminations measurées de plissage des variétés hyperboliques de dimension 3*, Ann. of Math. 160 (2004) 1013–1055 MR

[3] **B H Bowditch**, *Markoff triples and quasi-Fuchsian groups*, Proc. London Math. Soc. 77 (1998) 697–736 MR

[4] **R D Canary**, *Pushing the boundary*, from "In the tradition of Ahlfors and Bers, III" (W Abikoff, A Haas, editors), Contemp. Math. 355, Amer. Math. Soc., Providence, RI (2004) 109–121 MR

[5] **Y-E Choi**, **C Series**, *Lengths are coordinates for convex structures*, J. Differential Geom. 73 (2006) 75–117 MR

[6] **M Culler**, *Lifting representations to covering groups*, Adv. in Math. 59 (1986) 64–70 MR

[7] **D B A Epstein**, **A Marden**, *Convex hulls in hyperbolic space, a theorem of Sullivan, and measured pleated surfaces*, from "Analytical and geometric aspects of hyperbolic space" (D B A Epstein, editor), London Math. Soc. Lec. Note Ser. 111, Cambridge Univ. Press (1987) 113–253 MR

[8] **W Fenchel**, *Elementary geometry in hyperbolic space*, De Gruyter Studies in Mathematics 11, De Gruyter, Berlin (1989) MR

[9] **W M Goldman**, *The modular group action on real* $\mathrm{SL}(2)$*–characters of a one-holed torus*, Geom. Topol. 7 (2003) 443–486 MR

[10] **W M Goldman**, *Trace coordinates on Fricke spaces of some simple hyperbolic surfaces*, from "Handbook of Teichmüller theory, Vol II" (A Papadopoulos, editor), IRMA Lect. Math. Theor. Phys. 13, Eur. Math. Soc., Zürich (2009) 611–684 MR

[11] **W Goldman**, **G McShane**, **G Stantchev**, **S P Tan**, *Automorphisms of two-generator free groups and spaces of isometric actions on the hyperbolic plane*, preprint (2015) arXiv

[12] **J Hempel**, 3*–Manifolds*, Ann. of Math. Studies 86, Princeton Univ. Press (1976) MR

[13] **C D Hodgson**, **S P Kerckhoff**, *Rigidity of hyperbolic cone-manifolds and hyperbolic Dehn surgery*, J. Differential Geom. 48 (1998) 1–59 MR

[14] **L Keen**, **C Series**, *Pleating coordinates for the Maskit embedding of the Teichmüller space of punctured tori*, Topology 32 (1993) 719–749 MR

[15] **L Keen**, **C Series**, *The Riley slice of Schottky space*, Proc. London Math. Soc. 69 (1994) 72–90 MR

[16] **L Keen**, **C Series**, *Continuity of convex hull boundaries*, Pacific J. Math. 168 (1995) 183–206 MR

[17] **L Keen**, **C Series**, *How to bend pairs of punctured tori*, from "Lipa's legacy" (J Dodziuk, L Keen, editors), Contemp. Math. 211, Amer. Math. Soc., Providence, RI (1997) 359–387 MR

[18] **L Keen**, **C Series**, *Pleating invariants for punctured torus groups*, Topology 43 (2004) 447–491 MR

[19] **Y Komori**, **C Series**, *The Riley slice revisited*, from "The Epstein birthday schrift" (I Rivin, C Rourke, C Series, editors), Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry (1998) 303–316 MR

[20] **I Kra**, *On lifting Kleinian groups to* $\mathrm{SL}(2,\mathbf{C})$, from "Differential geometry and complex analysis" (I Chavel, H M Farkas, editors), Springer (1985) 181–193 MR

[21] **F Luo**, *Simple loops on surfaces and their intersection numbers*, Math. Res. Lett. 5 (1998) 47–56 MR

[22] **S Maloni**, **F Palesi**, **S P Tan**, *On the character variety of the four-holed sphere*, Groups Geom. Dyn. 9 (2015) 737–782 MR

[23] **A Marden**, *Hyperbolic manifolds: an introduction in 2 and 3 dimensions*, 2nd edition, Cambridge Univ. Press (2016)

[24] **B Maskit**, *Kleinian groups*, Grundl. Math. Wissen. 287, Springer (1988) MR

[25] **S P K Ng**, **S P Tan**, *The complement of the Bowditch space in the $\mathrm{SL}(2,\mathbb{C})$ character variety*, Osaka J. Math. 44 (2007) 247–254 MR

[26] **J-P Otal**, *Sur le coeur convexe d'une variété hyperbolique de dimension* 3, preprint (1994) `http://www.math.univ-toulouse.fr/~otal/`

[27] **C Series**, *The geometry of Markoff numbers*, Math. Intelligencer 7 (1985) 20–29 MR

[28] **C Series**, *An extension of Wolpert's derivative formula*, Pacific J. Math. 197 (2001) 223–239 MR

[29] **S P Tan**, **Y L Wong**, **Y Zhang**, *Necessary and sufficient conditions for McShane's identity and variations*, Geom. Dedicata 119 (2006) 199–217 MR

[30] **S P Tan**, **Y L Wong**, **Y Zhang**, *Generalized Markoff maps and McShane's identity*, Adv. Math. 217 (2008) 761–813 MR

[31] **M Wada**, *Software OPTi* `http://delta-mat.ist.osaka-u.ac.jp/OPTi/`

[32] **M Wada**, *OPTi's algorithm for discreteness determination*, Experiment. Math. 15 (2006) 61–66 MR

Mathematics Institute, University of Warwick
Leeman Building, Coventry, CV4 7AL, United Kingdom

Department of Mathematics, National University of Singapore
Block S17, 10 Lower Kent Ridge Rd., Singapore 119076

Department of Information and Computer Sciences, Nara Women's University
Kitauoyanishi-machi, Nara 630-8506, Japan

`c.m.series@warwick.ac.uk`, `mattansp@nus.edu.sg`,
`yamasita@ics.nara-wu.ac.jp`

`http://homepages.warwick.ac.uk/~masbb/`, `http://www.math.nus.edu.sg/~mattansp/`, `http://vivaldi.ics.nara-wu.ac.jp/~yamasita/`

Received: 17 April 2016 Revised: 17 October 2016

Algebraic & Geometric Topology 17 (2017) 2283–2306

Untwisting information from Heegaard Floer homology

KENAN INCE

The unknotting number of a knot is the minimum number of crossings one must change to turn that knot into the unknot. We work with a generalization of the unknotting number due to Mathieu–Domergue, which we call the untwisting number. The p–untwisting number is the minimum number (over all diagrams of a knot) of full twists on at most $2p$ strands of a knot, with half of the strands oriented in each direction, necessary to transform that knot into the unknot. In previous work, we showed that the unknotting and untwisting numbers can be arbitrarily different. In this paper, we show that a common route for obstructing low unknotting number, the Montesinos trick, does not generalize to the untwisting number. However, we use a different approach to get conditions on the Heegaard Floer correction terms of the branched double cover of a knot with untwisting number one. This allows us to obstruct several 10– and 11–crossing knots from being unknotted by a single positive or negative twist. We also use the Ozsváth–Szabó τ invariant and the Rasmussen s invariant to differentiate between the p– and q–untwisting numbers for certain $p,q>1$.

57M25, 57M27; 57R58

1 Introduction

It is a natural knot-theoretic question to seek to measure "how knotted up" a knot is. One such "knottiness" measure is given by the *unknotting number* $u(K)$, the minimum number of crossings, taken over all diagrams of K, one must change to turn K into the unknot. By a *crossing change* we shall mean one of the two local moves on a knot diagram given in Figure 1.

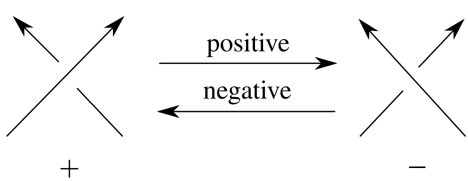


Figure 1: A positive and negative crossing change

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2283

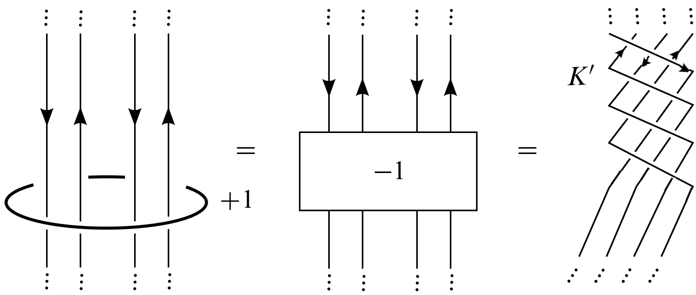


Figure 2: A left-handed, or positive, generalized crossing change

This invariant is quite simple to define but has proven itself very difficult to master. Fifty years ago, Milnor conjectured that the unknotting number for the (p,q)–torus knot was $\frac{1}{2}(p-1)(q-1)$; only in 1993, in two celebrated papers [6; 7], did Kronheimer and Mrowka prove this conjecture true. Hence, it is desirable to look at variants of the unknotting number which may be more tractable. One natural variant (due to Murakami [12]) is the *algebraic unknotting number* $u_a(K)$, the minimum number of crossing changes necessary to turn a given knot into an Alexander polynomial-one knot. Alexander polynomial-one knots are significant because they "look like the unknot" to *classical invariants*, knot invariants derived from the Seifert matrix. It is obvious that $u_a(K) \leq u(K)$ for any knot K, and there exist knots such that $u_a(K) < u(K)$ (for instance, the Whitehead double of any nontrivial knot).

In [9], Mathieu and Domergue defined another generalization of the unknotting number. In [8], Livingston worked with this definition. He described it as follows:

> "One can think of performing a crossing change as grabbing two parallel strands of a knot with opposite orientation and giving them one full twist. More generally, one can grab $2k$ parallel strands of K with k of the strands oriented in each direction and give them one full twist."

Following Livingston, we call such a twist a *generalized crossing change*. We describe in [4] how a crossing change may be encoded as a ± 1–surgery on a nullhomologous unknot $U \subset S^3 - K$ bounding a disk D such that $D \cap K = 2$ points. From this perspective, a generalized crossing change is a relaxing of the previous definition to allow $D \cap K = 2k$ points for any k, provided $\mathrm{lk}(K,U) = 0$; see Figure 2. In particular, any knot can be unknotted by a finite sequence of generalized crossing changes.

One may then naturally define the *untwisting number* $\mathrm{tu}(K)$ to be the minimum length, taken over all diagrams of K, of a sequence of generalized crossing changes beginning at K and resulting in the unknot. By $\mathrm{tu}_p(K)$, we will denote the minimum number of generalized crossing changes on $2p$ or fewer strands, with p strands oriented in each direction, needed to unknot K. Notice that $\mathrm{tu}_1 = u$ and that

$$\mathrm{tu} \leq \cdots \leq \mathrm{tu}_{p+1} \leq \mathrm{tu}_p \leq \cdots \leq \mathrm{tu}_1 = u.$$

The *algebraic untwisting number* $\mathrm{tu}_a(K)$ is the minimum number of generalized crossing changes, taken over all diagrams of K, needed to transform K into an Alexander polynomial-one knot. It is clear that $\mathrm{tu}_a(K) \leq \mathrm{tu}(K)$ for all knots K. In [4], we showed that, in fact, $\mathrm{tu}_a(K) = u_a(K)$ for all knots K; hence the unknotting and untwisting numbers are "algebraically the same". However, we also showed that tu and u can be arbitrarily different in general: there exists a family of knots $\{S_p^q\}$ such that $(u - \mathrm{tu}_q)(S_p^q) \geq p - 1$ for all $p, q \geq 2$.

Since the family $\{S_p^q\}$ consists of $(p, 1)$–cables of (untwisted) Whitehead doubles, most members of this family have very high crossing number. In this paper, we compare the unknotting and untwisting numbers for several 10– and 11–crossing knots with signature 0. In order to do this, we will develop an obstruction to a knot with signature 0 having untwisting number 1. This will require the methods of Heegaard Floer homology, specifically the *d–invariants* or *Heegaard Floer correction terms* of a 3–manifold.

In [19], Ozsváth and Szabó develop an unknotting number-1 obstruction using d–invariants. This obstruction relies on the *Montesinos trick*, which allows them to construct a definite 4–manifold with boundary the branched double cover $\Sigma(K)$ of an unknotting number-1 knot K. In Section 3, we give an infinite family of knots which have untwisting number 1 but which do not satisfy the Montesinos trick, eliminating that route toward a d–invariant obstruction:

Theorem 1.1 *There exists an infinite family $\{K_n\}_{n>1}$ of knots such that $\mathrm{tu}(K_n) = 1$ for all n, but $\Sigma(K_n)$ is not a half-integer surgery on any knot in S^3 for any n.*

In Section 4, we get around the failure of the Montesinos trick for untwisting number-1 knots by porting the machinery used by Owens and Strle in [16] and Nagel and Owens in [13] as an obstruction to low untwisting number:

Theorem 1.2 *Let K be a knot with signature $\sigma(K)$ which can be unknotted by p positive and n negative generalized crossing changes. Then $Y = \Sigma(K)$, the branched double cover of K, bounds a smooth, definite 4–manifold W with $b_2(W) = 2n + 2p$ and signature $2n - 2p + \sigma(K)$. Moreover, $H_2(W; \mathbb{Z})$ contains n classes of self-intersection $+2$ and p classes of self-intersection -2 which span a primitive sublattice; in other words, the quotient of $H_2(W; \mathbb{Z})$ by this sublattice is torsion-free.*

Once we have constructed a definite 4–manifold W with $\partial W = \Sigma(K)$, the next step is to apply a result of Ozsváth and Szabó to get conditions that the d–invariants of $\Sigma(K)$ must satisfy. These invariants are easily computable for alternating K via the *Goeritz matrix* associated to K. These computations are discussed further in Section 4. We successfully obstruct several 10–crossing knots from being unknotted by a single positive and/or negative generalized crossing change, though these untwisting numbers cannot be computed using the methods available prior to the development of Heegaard Floer homology:

Theorem 1.3 *The knots 10_{68} and 10_{96} have untwisting number 2, the knots 10_{22}, 10_{34}, 10_{35}, 10_{87} and 10_{90} cannot be unknotted by a single positive generalized crossing change, and the knot 10_{48} cannot be unknotted by a single negative generalized crossing change.*

Similarly, we apply these obstructions to all 11–crossing knots with signature 0, algebraic unknotting number 1, and unknotting number 2 to get the following:

Theorem 1.4 *The knots $11a_{37}, 11a_{103}, 11a_{169}$, $11a_{214}$ and $11a_{278}$ have untwisting number 2.*

Finally, we showed in [4] that there can be arbitrarily large gaps between the p–untwisting number and the 1–untwisting number (which by definition equals the unknotting number) for several families of knots. However, we had not yet been able to distinguish between tu_p and tu_q for $p, q > 1$.

In Section 6, we use invariants coming from Heegaard Floer homology (the Ozsváth–Szabó τ invariant) and Khovanov homology (the Rasmussen s invariant) to give lower bounds on the p–untwisting number for arbitrary p via the following theorem. While visiting Mark Powell at the Max Planck Institute, he suggested this theorem and outlined a proof similar to the proof of Powell and coauthors T Cochran, S Harvey, and A Ray that the τ and s invariants give lower bounds for their bipolar metrics (to appear in a future paper). The referee suggested a simpler approach involving the 4–genus, detailed in Section 6.

Theorem 1.5 *Let K be a knot which can be converted to the unknot via n generalized crossing changes, where for every i, the i^{th} generalized crossing change is performed on $2p_i$ strands. Then*

$$|\tau(K)| \leq \sum_{i=1}^{n} p_i^2 \quad \text{and} \quad \tfrac{1}{2}|s(K)| \leq \sum_{i=1}^{n} p_i^2.$$

This allows us to show that there exist $p,q>1$ such that the difference between the p–and q–untwisting numbers of several families of knots can be made arbitrarily large:

Example 1.6 Let K_{p^3} denote the $(p^3,1)$–cable of a knot K with genus 1 and $u(K)=1=\tau(K)$ (one example of such a K is the right-handed trefoil knot). We know from [4, Section 5] that $\mathrm{tu}_{p^3}(K_{p^3})=1$. We may use Theorem 1.5 to show that

$$\mathrm{tu}_p(K_{p^3})-\mathrm{tu}_{p^3}(K_{p^3}) \xrightarrow{p\to\infty} \infty.$$

Convention In this paper, all manifolds are assumed to be smooth, compact, orientable and connected, and all surfaces in manifolds are assumed to be smoothly embedded. When homology groups are given without specifying coefficients, they are assumed to have coefficients in $\mathbb{Z}$.

Acknowledgements Thanks to Stefan Friedl, Maciej Borodzik, Peter Horn, Matthias Nagel, and Mark Powell for many enlightening conversations. Thanks also to Stefan Friedl, Matthias Nagel, Brendan Owens, and the referee for providing comments on this paper. A Maple program written by Brendan Owens and Sašo Strle has been very useful in the computations in Section 5. I would also like to acknowledge the results of Brendan Owens and Sašo Strle in [16], Matthias Nagel and Brendan Owens in [13], Brendan Owens in [14], and Tim Cochran and William Lickorish in [3], all of which greatly inspired this work.

2 Preliminaries

2.1 Dehn surgery

In this section, we will describe the operation of Dehn surgery on knots.

Definition 2.1 Let $K\subset S^3$ be an oriented knot, let N be a closed tubular neighborhood of K, and consider the preferred framing for N (see [20, Definition 2E8]) in which the longitude L is oriented in the same way as K and the meridian M has linking number $+1$ with K. We may write any simple closed curve $J\subset\partial N$ in terms of the homology basis $\{\lambda=[L],\ \mu=[M]\}$:

$$[J]=q\lambda+p\mu\in H_1(\partial N).$$

The *result of* (p/q)*–surgery on* K is the 3–manifold

$$S^3_{p/q}(K):=(S^3-\mathring{N})\cup_h(S^1\times D^2),$$

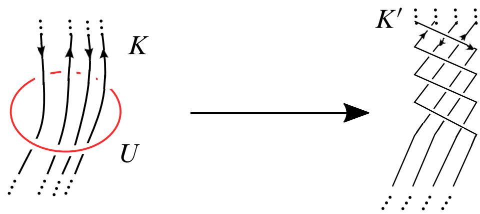


Figure 3: Performing $+1$–surgery on an unknot U gives the knot K a left-handed twist.

where $h\colon \partial(S^1 \times D^2) \to \partial N$ is a homeomorphism taking $* \times S^1$ onto a curve J of class $[J] = p\mu + q\lambda$ in $H_1(\partial N)$. By convention, we indicate that surgery is to be performed on K by writing the ratio p/q next to a diagram of K.

If $U \subset S^3 \setminus K$ is an unknot such that $\mathrm{lk}(K, U) = 0$, we define a *generalized crossing change diagram for* K to be a diagram of the link $K \sqcup U$ with the number ± 1 written next to U, indicating that U is to have ± 1–surgery performed on it.

There is an orientation-preserving homeomorphism Φ of the manifold $M := S^3_{\pm 1}(U)$ resulting from ± 1–surgery on U with S^3. However, $K' := \Phi(K) \subset S^3$ may have a different knot type than K. (Note that the knot type of K' does not depend on the choice of homeomorphism Φ since any two orientation-preserving homeomorphisms of S^3 are isotopic.) In particular, if D is a disk bounded by U such that $2p$ strands of K pass through D in straight segments, then each of the $2p$ straight pieces is replaced by a helix which screws through a neighborhood of D in the right- (respectively, left-) hand sense; see Figure 3.

The process of performing ± 1–surgery on an unknot U in a generalized crossing change diagram for a knot K, mapping the resulting manifold to S^3 via an orientation-preserving homeomorphism Φ, then erasing $\Phi(U)$ from the resulting diagram of $\Phi(K) \sqcup \Phi(U)$ is called a $\pm$ *generalized crossing change* on K. Now, it can be easily verified that performing a $-$ generalized crossing change on the knot K on the left side of Figure 4 transforms the crossing labeled $+$ into the crossing labeled $-$. The inverse process of introducing an unknot labeled with a $+1$ to the right side of Figure 4 and performing a $+$ generalized crossing change in the resulting generalized crossing change diagram transforms the crossing labeled $-$ into the crossing labeled $+$.

2.2 The untwisting number

In a generalized crossing change diagram for K consisting of a diagram of K and an unknot U, we have that K must pass through U an even number of times, for otherwise $\mathrm{lk}(K, U) \neq 0$. If at most $2p$ strands of K pass through an unknot U in

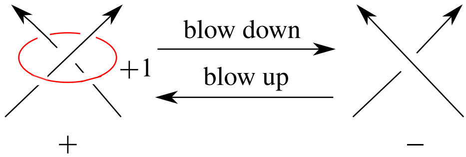


Figure 4: A crossing change is a 1–generalized crossing change.

a generalized crossing change diagram, we may call the associated $\pm$ generalized crossing change a *$\pm p$–generalized crossing change on K*.

The *untwisting number* $\mathrm{tu}(K)$ *of* K is the minimum length of a sequence of generalized crossing changes on K such that the result of the sequence is the unknot, where we allow ambient isotopy of the diagram in between generalized crossing changes. Note that by the reasoning on page 58 of [1], this definition is equivalent to taking the minimum length, over all diagrams of K, of a sequence of generalized crossing changes beginning with a fixed diagram of K such that the result of the sequence is the unknot, where we do not allow ambient isotopy of the diagram in between generalized crossing changes.

For $p = 1, 2, 3, \dots$, we define the *p–untwisting number* $\mathrm{tu}_p(K)$ to be the minimum length of a sequence of $\pm p$–generalized crossing changes on K resulting in the unknot, where we allow ambient isotopy of the diagram in between generalized crossing changes. It follows immediately that we have the chain of inequalities

$$\mathrm{tu}(K) \le \cdots \le \mathrm{tu}_{p+1}(K) \le \mathrm{tu}_p(K) \le \cdots \le \mathrm{tu}_2(K) \le \mathrm{tu}_1(K) = u(K). \tag{2-1}$$

2.3 Heegaard Floer homology

In this section, we will recall some properties of Heegaard Floer homology, a set of invariants of 3–manifolds defined by Ozsváth and Szabó. For details, refer to their papers, in particular [17; 18; 19].

Let Y be an oriented rational homology 3–sphere. Recall that one can associate to Y a set $\mathrm{Spin}^c(Y)$ of *spin^c structures on Y*. In the case where $|H^2(Y;\mathbb{Z})|$ is odd, there is a canonical bijection $H^2(Y;\mathbb{Z}) \leftrightarrow \mathrm{Spin}^c(Y)$ under which $0 \in H^2(Y;\mathbb{Z})$ is sent to the unique spin structure on Y. In this way, we may give $\mathrm{Spin}^c(Y)$ a group structure inherited from that of $H^2(Y;\mathbb{Z})$.

Fix a spin^c structure $\mathfrak{s}$ on Y. Then the *(plus flavor of) Heegaard Floer homology* $\mathrm{HF}^+(Y,\mathfrak{s})$ is a $\mathbb{Q}$–graded abelian group with a $\mathbb{Z}[U]$–action, where multiplication by U lowers the grading by 2. Associated to $\mathfrak{s}$ is a *d–invariant* $d(Y,\mathfrak{s}) \in \mathbb{Q}$ which satisfies the symmetry condition $d(Y,\mathfrak{s}) = -d(-Y,\mathfrak{s})$. The correction terms are useful for obstructing the existence of a 4–manifold with boundary Y:

Theorem 2.2 (Ozsváth and Szabó [17]) *Let X be a negative-definite 4–manifold with boundary Y and intersection form represented by a matrix Q, and let $\mathfrak{s}$ be any spin^c structure on X. Let $c_1(\mathfrak{s})$ denote the first Chern class associated to $\mathfrak{s}$. Then*

$$\begin{aligned} &\tfrac{1}{4}(c_1(\mathfrak{s})^2+b_2(X)) \le d(Y,\mathfrak{s}|_Y), \\ &\tfrac{1}{4}(c_1(\mathfrak{s})^2+b_2(X)) \equiv d(Y,\mathfrak{s}|_Y) \bmod 2. \end{aligned} \tag{2-2}$$

Following [17], we now show how to check this obstruction in practice. In addition to the assumptions of Theorem 2.2, suppose for simplicity that $\pi_1(X)=1$ and that $|H^2(Y;\mathbb{Z})|$ is odd. (This will always be true for the examples in this paper.) Let $r=b_2(X)$, the second Betti number of X. It is straightforward to see that $H_2(X;\mathbb{Z})$ is free of rank r in this case. Choose a basis $\{x_i\}_{i=1}^r$ for $H_2(X;\mathbb{Z})$ and let $Q=(Q_{ij})$ be a negative-definite $r\times r$ matrix representing the intersection pairing of X in this basis; then $\det Q=|H^2(Y;\mathbb{Z})|$. The dual basis $\{x^i\}_{i=1}^r$ for $H^2(X;\mathbb{Z})$ given by the universal coefficient theorem defines an isomorphism $H^2(X;\mathbb{Z})\cong\mathbb{Z}^r$. Under this isomorphism, the set $\{c_1(\mathfrak{s}) \mid \mathfrak{s}\in \mathrm{Spin}^c(X)\}\subset H^2(X;\mathbb{Z})$ of first Chern classes of spin^c structures on X is sent to the set of characteristic covectors $\mathrm{Char}(Q)$ for Q. (Recall that a *characteristic covector* for an $r\times r$ matrix Q is a vector $\xi=(\xi_1,\dots,\xi_r)\in\mathbb{Z}^r$ such that $\xi_i\equiv Q_{ii} \bmod 2$ for $i=1,\dots,r$.) In our basis, the square of the first Chern class of the spin^c structure corresponding to a characteristic covector ξ is given by $\xi^TQ^{-1}\xi$.

Define a function $m_Q\colon \mathbb{Z}^r/Q(\mathbb{Z}^r)\to\mathbb{Q}$ by

$$m_Q(g)=\max\left\{\tfrac{1}{4}(\xi^TQ^{-1}\xi+r) \mid \xi\in\mathrm{Char}(Q),\ [\xi]=g\right\},$$

where $[\xi]$ is the image of $\xi\in\mathbb{Z}^r$ under the projection to $\mathbb{Z}^r/Q(\mathbb{Z}^r)$. In computing m_Q, it is enough to consider characteristic covectors $\xi=(\xi_1,\dots,\xi_r)$ with $-Q_{ii}\ge\xi_i\ge Q_{ii}$; if, say, $\xi_i<Q_{ii}$, subtracting twice the i^{th} column of Q from ξ shows that $\xi^TQ^{-1}\xi$ is not maximal. Then we may simplify the conditions (2-2) as follows:

Theorem 2.3 (Ozsváth and Szabó) *Let Y be a rational homology 3–sphere which is the boundary of a simply connected, negative-definite 4–manifold X, with $|H^2(Y;\mathbb{Z})|$ odd. If the intersection pairing of X is represented in a basis by the matrix Q, then there is a group isomorphism*

$$\phi\colon \mathbb{Z}^r/Q(\mathbb{Z}^r)\to\mathrm{Spin}^c(Y)$$

such that for all $g\in\mathbb{Z}^r/Q(\mathbb{Z}^r)$,

$$\begin{aligned} &m_Q(g)\le d(Y,\phi(g)) \\ &m_Q(g)\equiv d(Y,\phi(g)) \bmod 2. \end{aligned} \tag{2-3}$$

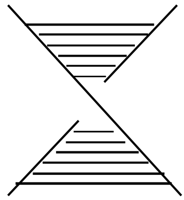


Figure 5: Crossing conventions for negative-definite Goeritz matrices of alternating knots

Under the assumptions of the theorem, we say that the 4–manifold X bounded by Y is *sharp* if equality holds in (2-3). In this case, we may compute the correction terms for Y using the values of m_Q. Moreover, if a rational homology sphere Y bounds a *positive*-definite 4–manifold X, we may compute the correction terms for Y by applying Theorem 2.3 to $-Y$.

If K is an alternating knot, we may compute the d–invariants of $\Sigma(K)$ using the negative-definite *Goeritz matrix* computed from an alternating diagram of K as follows. Consider a regular projection of K into a plane $\mathbb{R}^2 \subset \mathbb{R}^3 = S^3 \setminus \{\infty\}$. Color the regions of $\mathbb{R}^2 \setminus K$ alternately black and white so that the n white regions $X_1, \dots, X_n$ are separated by crossings of the type depicted in Figure 5.

For $0 \le i, j \le n$, where d is the number of double points incident to X_i and X_j, define

$$g_{ij} = \begin{cases} d, & i \neq j, \\ -\sum_{k \neq i} g_{ik}, & i = j. \end{cases}$$

Let $G' = (g_{ij})$. Then the *negative-definite Goeritz matrix* G associated to K is the $n \times n$ symmetric integer matrix obtained from G' by deleting the 0^{th} row and column of G'. It is shown in [19, Proposition 3.2] that G represents the intersection pairing of a sharp 4–manifold with boundary $\Sigma(K)$; thus, the correction terms for $\Sigma(K)$ are given by the values of m_G.

3 Failure of the Montesinos trick

The "Montesinos trick" relates crossing changes downstairs on K to surgery upstairs on $\Sigma(K)$, the branched double cover of K. We use the convention that the determinant of a knot is given by $|\Delta_K(-1)|$, where Δ_K is the Alexander polynomial for the knot K.

Theorem 3.1 [11] *If $u(K) = 1$, then $\Sigma(K) \cong S^3_{\pm D/2}(C)$ for some other knot $C \subset S^3$, where here D is the determinant of K.*

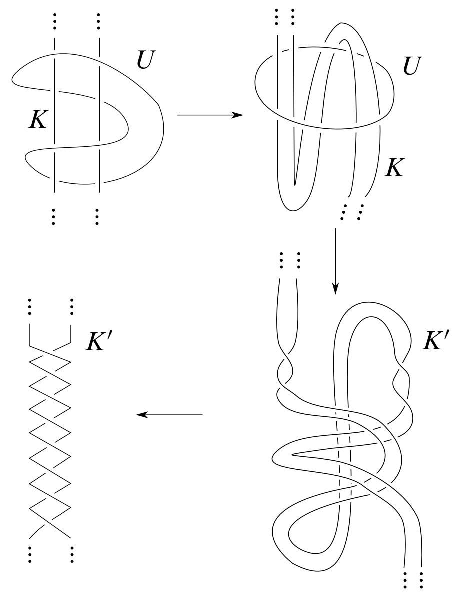


Figure 6: The (local) effect of performing a $+$ generalized crossing change on the unknot U

We show that this theorem does not generalize to untwisting number-1 knots:

Theorem 3.2 *There exists an infinite family* $\{K_n\}$ *of knots such that, for all* $n \geq 1$, $\mathrm{tu}(K_n) = 1$, *but* $\Sigma(K_n)$ *is not a half-integer surgery on any knot in* S^3.

In order to prove Theorem 3.2, we will need two main ingredients. The first is the following lemma:

Lemma 3.3 *The effect of performing a* $+$ *generalized crossing change on the unknot* U *in the local picture given in Figure 6 is to add* -4 *full twists to the knot* K.

Proof See Figure 6. The intermediate steps are left to the reader. □

Our second ingredient is the following theorem of McCoy [10]:

Theorem 3.4 *Let* K *be an alternating knot. Then the following are equivalent:*

1. $u(K) = 1$;
2. *the branched double cover* $\Sigma(K)$ *can be obtained by half-integer surgery on some knot in* S^3;
3. *in every minimal diagram of* K, *there exists a crossing which can be changed to unknot that diagram.*

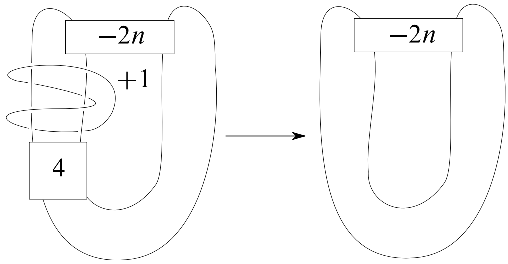


Figure 7: The knots K_n, together with the $+1$–generalized crossing change that unknots them. Here, positive (resp. negative) numbers in boxes denote right-handed (resp. left-handed) full twists.

Proof of Theorem 3.2 Fix an orientation on K_n. The generalized crossing change pictured in Figure 7 introduces -4–twists on the left side of K_n, which undo the 4–twists already present. Hence, $\mathrm{tu}(K_n) = 1$ for all n. Moreover, if $n > 1$, then K_n is a minimal diagram of an alternating knot. One can easily see that K_n cannot be unknotted by any single crossing change in this diagram. By Theorem 3.4, the branched double cover $\Sigma(K_n)$ cannot be obtained by half-integer surgery on any knot in S^3, and moreover, $u(K_n) > 1$. □

Note 3.5 The first knot in this family is $K_2 = 12a_{1166}$. The unknotting number of $12a_{1166}$ is listed as "not known" in the KnotInfo tables, but is either 1 or 2. By Theorem 3.4, we must have that $\mathrm{tu}(12a_{1166}) = 1 < 2 = u(12a_{1166})$.

Question 3.6 *Does there exist a knot K with $\mathrm{tu}(K) = 1$ such that $\Sigma(K)$ is not a surgery on* any *knot in S^3?*

4 Heegaard Floer-theoretic obstructions to untwisting number 1

Although the Montesinos trick does not hold for knots with untwisting number 1, we can still get obstructions to a knot K being unknotted by a single positive or negative generalized crossing change using techniques similar to those of Owens and Strle in [16] and Nagel and Owens in [13] together with Theorem 2.2.

In order to apply Theorem 2.2, we first compute a Goeritz matrix G for K and, from G, the function m_G as in Theorem 2.2. The image of $\mathbb{Z}^r/G(\mathbb{Z}^r)$ under m_G, where G is an $r \times r$ matrix, is the set of d–invariants for Y. We construct the 4–manifold W as in [13, Proposition 2.3] using the propositions below, then compute the m_Q and show that no isomorphism satisfying both conditions of (2-2) exists.

Proposition 4.1 *Let K be an oriented knot in S^3, and suppose that K can be unknotted by p positive and n negative generalized crossing changes. Then K bounds a disk Δ in a manifold $C \cong B^4 \#_n \mathbb{CP}^2 \#_p \overline{\mathbb{CP}^2}$ with $[\Delta] = 0 \in H_2(C, \partial C)$ and $\pi_1(C \setminus \Delta) = \mathbb{Z}$, generated by a meridian of K.*

Proof Suppose that K is an oriented knot in S^3 and that K can be unknotted by p positive and n negative generalized crossing changes. Then there is a sequence of knots

$$K := K_{p+n} \xrightarrow{\epsilon_{p+n}} K_{p+n-1} \xrightarrow{\epsilon_{p+n-1}} \cdots \xrightarrow{\epsilon_2} K_1 \xrightarrow{\epsilon_1} K_0 := U \tag{4-1}$$

for which K_i is obtained from K_{i+1} by a single generalized crossing change of sign $\epsilon_{i+1} \in \{\pm 1\}$ for $i = 1, \dots, p+n$, with precisely p of the ϵ_i equal to $+1$ and n of the ϵ_i equal to -1, and U is the unknot. Reversing our point of view, there is a sequence of knots

$$U := K_0 \xrightarrow{-\epsilon_1} K_1 \xrightarrow{-\epsilon_2} \cdots \xrightarrow{-\epsilon_{p+n-1}} K_{p+n-1} \xrightarrow{-\epsilon_{p+n}} K_{p+n} =: K \tag{4-2}$$

for which K_i is obtained from K_{i-1} by a single generalized crossing change of sign $-\epsilon_i$ for $i = 1, \dots, p+n$ and U is the unknot.

Consider U to be embedded in $\partial B^4 = S^3$. Since U is an unknot in S^3, it bounds an embedded disk $D \subset S^3$. We push D into B^4 to get a disk $\Delta_0 \subset B^4$ such that $\Delta_0 \cap \partial B^4 = U$ and $\pi_1(B^4 \setminus \Delta_0) = \mathbb{Z}$, where the latter is generated by a meridian of U.

Now, we build a 4–manifold C in which K bounds a disk Δ as follows. Let $C_0 := B^4$. We now build C from C_0 by sequentially thickening the boundary of C_0 and attaching 2–handles to the new boundary. First, we thicken the boundary $S_0 := \partial B^4$ to $S_0 \times [0,1]$, obtaining a new 4–manifold B_0. We denote the disk $\Delta_0 \cup (U \times I) \subset B_0$ by Δ_1. The first generalized crossing change can be realized via the attachment of a $-\epsilon_1$–framed 2–handle h_1 along an unknot $U_1 \subset S_0 \times \{1\}$ with $\operatorname{lk}(U \times \{1\}, U_1) = 0$. There is a unique orientation-preserving diffeomorphism from the new boundary S_1 resulting from this handle attachment to S^3, and after this diffeomorphism $U \times \{1\}$ is isotopic to K_1. We denote by C_1 the new 4–manifold resulting from this handle attachment. Since attaching a ± 1–framed 2–handle to the boundary of a 4–manifold along an unknot results in connect-summing a $\pm\mathbb{CP}^2$, we have that $C_1 \cong C_0 \# {-\epsilon_1}\mathbb{CP}^2 = B^4 \# {-\epsilon_1}\mathbb{CP}^2$ (here $\pm\mathbb{CP}^2$ denotes $\mathbb{CP}^2$ or $\overline{\mathbb{CP}^2}$, respectively). Note that Δ_1 is still a disk in C_1 and that $\partial\Delta_1 = K_1$.

Attaching a 2–handle generally adds a relation to the fundamental group of the resulting manifold, where the relation is given by the attaching map. Since the attaching circle U_1 of h_1 is trivial in $H_1\big((S_0 \times \{1\}) \setminus (U \times \{1\})\big) \cong \mathbb{Z}\langle \mu_0 \rangle$, where μ_0 is a meridian of $U \times \{1\} \subset S_0 \times \{1\}$, it is also trivial in $\pi_1(B_0 \setminus \Delta_1) \cong \mathbb{Z}\langle \mu_0 \rangle$. Thus, we get that $\pi_1(C_1 \setminus \Delta_1) \cong \mathbb{Z}$ as well, generated by a meridian μ_1 of K_1.

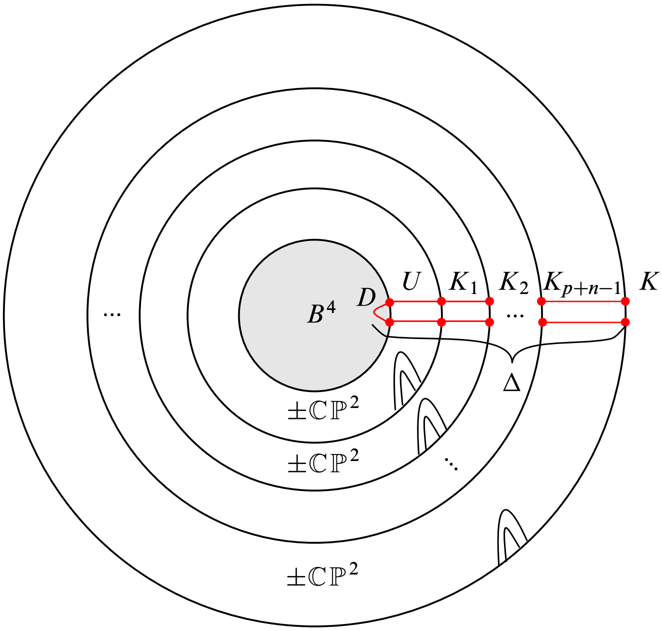


Figure 8: The construction of a manifold C in which K bounds a disk Δ

We continue in this way to iteratively get 4–manifolds $C_1, \dots, C_{p+n}$ so that C_{i+1} is obtained from C_i by adding a collar $\partial C_i \times [i, i+1]$ to ∂C_i and attaching a $-\epsilon_{i+1}$–framed 2–handle h_{i+1} to $\partial C_i \times \{i+1\}$. At each stage, the attaching circle $U_{i+1} \subset S_i \times \{i+1\}$ of h_{i+1} is trivial in

$$H_1\big((S_i \times \{i+1\}) \setminus (K_i \times \{i+1\})\big) \cong \mathbb{Z}\langle \mu_i \rangle,$$

where μ_i is a meridian of $K_i \times \{i+1\}$. Hence, U_{i+1} is trivial in $\pi_1(B_i \setminus \Delta_{i+1}) \cong \mathbb{Z}\langle \mu_i \rangle$. The end result of this process is a 4–manifold $C := C_{p+n} \cong B^4 \#_n \mathbb{CP}^2 \#_p \overline{\mathbb{CP}^2}$ in which $K := K_{p+n}$ bounds a disk $\Delta := \Delta_{p+n}$ such that $\pi_1(C \setminus \Delta) \cong \mathbb{Z}$, generated by a meridian μ_{p+n} of $K = K_{p+n}$.

We now consider the nondegenerate intersection form $H_2(C, \partial C) \times H_2(C) \to \mathbb{Z}$ in order to show that $[\Delta] = 0 \in H_2(C, \partial C)$. Since $H_2(C) \cong \mathbb{Z}^{p+n}$ is generated by the $\mathbb{CP}^1$ factors $\mathbb{CP}^1_1, \dots, \mathbb{CP}^1_{p+n}$, where $\mathbb{CP}^1_i$ is a generator of the second homology of the i^{th} connect-summed copy of $\pm\mathbb{CP}^2$, we know that an element $a \in H_2(C, \partial C)$ is 0 if and only if $a \cdot [\mathbb{CP}^1_i] = 0$ for all $i = 1, \dots, p+n$.

Let d_i denote the disk bounded by the unknot U_i, and let D_i denote the second D^2 factor in the i^{th} 2–handle attached to C. Then $\mathbb{CP}^1_i$ is homologous to

$$\big(d_{i-1} \times \{i - \tfrac{1}{2}\}\big) \cup \big(U_i \times [i - \tfrac{1}{2}, i]\big) \cup (* \times D_i).$$

The only intersections of Δ with $\mathbb{CP}^1_i$ come from the intersections of K_{i-1} with d_i. Since $\mathrm{lk}(K_{i-1}, U_i) = 0$ for all i, we have that $[K_{i-1}] \cdot [d_i] = 0$ for all i. Therefore, $[\Delta] = 0 \in H_2(C, \partial C)$. This completes the proof of the proposition. □

Next, we prove a generalization of [13, Proposition 2.3]:

Proposition 4.2 *Let K be a knot in $S^3 = \partial B^4$, and suppose K bounds a properly embedded disk Δ in $C := B^4 \#_n \mathbb{CP}^2 \#_p \overline{\mathbb{CP}^2}$ such that $[\Delta] = 0 \in H_2(C, \partial C)$ and $\pi_1(C \setminus \Delta) = \mathbb{Z}$, generated by a meridian of K. Then there exists an oriented 4–manifold W with boundary $\partial W = \Sigma(K)$, the branched double cover of K, such that*

(1) *W is simply connected;*

(2) *$H_2(W; \mathbb{Z}) \cong \mathbb{Z}^{2(p+n)}$;*

(3) *the signature of W is $\sigma(W) = 2(n-p) + \sigma(K)$;*

(4) *there exist $p+n$ pairwise disjoint classes in $H_2(W; \mathbb{Z})$ represented by p surfaces of self-intersection -2 and n surfaces of self-intersection $+2$ which span a primitive sublattice; in other words, the quotient of $H_2(W; \mathbb{Z})$ by this sublattice is torsion-free.*

Proof Since $\pi_1(C \setminus \Delta) = \mathbb{Z}$ with generator the meridian of K, we may take the double cover $W = \Sigma_2(C, \Delta)$ of C branched along Δ, and by definition, we have $\partial W = \Sigma_2(K)$.

(1) Let $p\colon (\widetilde{C \setminus N(\Delta)}) \to C \setminus N(\Delta)$ denote the two-fold, unbranched cover of the complement of an open tubular neighborhood of Δ in C. Since $\pi_1(C \setminus \Delta) \cong \mathbb{Z}$, we have that $\pi_1(\widetilde{C \setminus \Delta}) \cong \mathbb{Z}$ as well. The branched cover W may be recovered from $\widetilde{C \setminus N(\Delta)}$ by gluing back a closed neighborhood $\overline{N(\Delta)} \cong D^2 \times \Delta$ along $p^{-1}(\partial \overline{N(\Delta)}) \cong S^1 \times \Delta$. A straightforward application of the Seifert–van Kampen theorem to $W = \widetilde{C \setminus \Delta} \cup_{p^{-1}(\partial \overline{N(\Delta)})} \overline{N(\Delta)}$ shows that $\pi_1(W) = 1$.

(2) We will need the following claim.

Claim The Euler characteristic of W is $\chi(W) = 2(p+n) + 1$.

Proof of claim By a standard Mayer–Vietoris argument, we may show that

$$H_i(C) = \begin{cases} \mathbb{Z}, & i = 0, \\ 0, & i = 1, 3, \\ \mathbb{Z}^{p+n}, & i = 2, \\ 0, & i = 4, \end{cases}$$

where $H_4(C) = 0$ because C is a manifold with boundary. Thus, $\chi(C) = 1 + p + n$. We have that

$$\chi(C) = \chi(C \setminus \Delta) + \chi(\Delta) = \chi(C \setminus \Delta) + 1.$$

Therefore, the double cover $\widetilde{C \setminus \Delta}$ of $C \setminus \Delta$ has Euler characteristic $2(\chi(C) - 1)$. Since $W = \widetilde{C \setminus \Delta} \cup_{p^{-1}(\partial \overline{N(\Delta)})} \overline{N(\Delta)}$ as above, we have that

$$\chi(W) = 2(\chi(C) - 1) + 1 = 2(p + n) + 1. \qquad \square$$

Now, since $H_1(W; \mathbb{Z}) = 0$, the universal coefficient theorem together with the long exact cohomology sequence for $(X, \partial X)$ implies that $H^1(W, \partial W; \mathbb{Z}) = 0$ as well. By Poincaré–Lefschetz duality, we have that $H_3(W; \mathbb{Z}) = 0$ as well. Note that $H_4(W; \mathbb{Z}) = 0$ since W is a manifold with boundary. Now the Euler characteristic of W is

$$2p + 2n + 1 = \chi(W) = 1 + b_2(W).$$

Therefore, $b_2(W) = 2(p + n)$, and $H_2(W; \mathbb{Z})$ is free abelian of rank $2(p + n)$.

(3) Our proof follows the proof of [3, Theorem 3.7]. Let F_{-K} be a connected Seifert surface of the knot $-K$ with interior pushed into $-B^4$. Then the manifold $(\widehat{C}, F) := (C, \Delta) \cup_{(S^3, K)} (-B^4, F_{-K})$ is closed. Let $\widehat{W}$ denote the double cover $\Sigma_2(\widehat{C}, F)$ of $\widehat{C}$ branched over $F := \Delta \cup_K F_{-K}$. Then $\widehat{W} = W \cup_{\Sigma_2(K)} X_K$, where X_K is the double cover $\Sigma_2(F_{-K})$ of $-B^4$ branched along F_{-K}. By [21; 5], the signature of X_K is $-\sigma(K)$. Applying Novikov additivity, we get that

$$\sigma(\widehat{W}) = \sigma(W) + \sigma(X_K).$$

Furthermore, the G–signature theorem [2, Lemma 2.1] tells us that

$$\sigma(\widehat{W}) = 2\sigma(\widehat{C}) - \tfrac{1}{2}([F] \cdot [F]).$$

Since in this case $[\Delta] = 0 \in H_2(C, \partial C)$, we have that $[F] \cdot [F] = 0$ so

$$\sigma(W) = 2\sigma(C) + \sigma(K).$$

Since $\sigma(C) = n - p$, we get that $\sigma(W) = 2(n - p) + \sigma(K)$.

(4) We let S_i be a smoothly embedded surface representing the generator of $H_2(-\epsilon_i \mathbb{CP}^2_i)$, the i^{th} summand of C. We define x_i to be the homology class of the two-fold cover $\widehat{S}_i \subset W$ of S_i branched over $\Delta \cap S_i$, which is a subset of W. Since the S_i are pairwise disjoint downstairs, the $\widehat{S}_i$ are also pairwise disjoint. We show that the x_i have self-intersection $-2\epsilon_i$.

Let S_i^+ be a push-off of S_i. Then $S_i \cdot S_i^+ = -\epsilon_i$. We make the disk Δ disjoint from the (codimension-2) intersection points $S_i \cap S_i^+$. In the branched cover, denote the

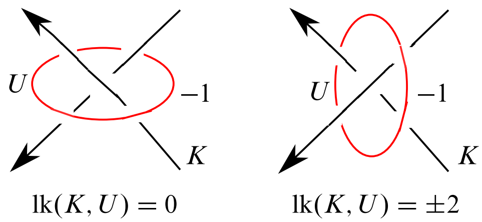


Figure 9: No matter the sign of the crossing to be changed, Nagel and Owens [13] may perform only -1–generalized crossing changes in order to do so.

preimage of S_i by T_i and the preimage of S_i^+ by T_i^+. Then T_i^+ is also a push-off of T_i. The intersection points of T_i and T_i^+ are the preimages of the intersection points of S_i and S_i^+; since the points of $S_i \cap S_i^+$ are disjoint from the branch set, there are geometrically two intersection points of T_i and T_i^+. Furthermore, the orientations upstairs give the same signs of intersection as downstairs. Therefore, $T_i \cdot T_i^+ = -2\epsilon_i$.

The proof of [13, Proposition 2.3] applies to our case to show that these classes span a primitive sublattice. This completes the proof of the proposition. □

Remark 4.3 The proof of Proposition 4.2 is very similar to the proof of [13, Proposition 2.3], with the caveat that Nagel and Owens use only -1–generalized crossing changes in order to unknot K, no matter the signs of the crossings of K that need to be changed (see Figure 9). The diagram on the right side of Figure 9 is not a generalized crossing change diagram, since $\mathrm{lk}(K,U) \neq 0$. Therefore, we must assume that K can be unknotted only by positive generalized crossing changes.

From Propositions 4.1 and 4.2, we derive a theorem analogous to [13, Theorem 1], but requiring the additional condition that the signature of the knot K is 0:

Theorem 4.4 *Let $K \subset S^3$ be an oriented knot with signature 0 which can be unknotted by p generalized crossing changes, all of sign $+1$. Then the double cover $Y := \Sigma(K)$ of S^3 branched along K bounds a smooth, simply connected, negative-definite 4–manifold W with $H_2(W;\mathbb{Z}) \cong \mathbb{Z}^{2p}$. Moreover, $H_2(W;\mathbb{Z})$ contains p pairwise disjoint homology classes of self-intersection -2 which span a primitive sublattice.*

Proof By Proposition 4.1, K bounds a disk Δ in a manifold $C \cong B^4 \#_p \mathbb{CP}^2$ such that $[\Delta] = 0 \in H_2(C, \partial C)$ and $\pi_1(C \setminus \Delta) = \mathbb{Z}\langle\mu\rangle$, where μ is a meridian of K. By Proposition 4.2, the double cover $W := \Sigma_2(C, \Delta)$ of C branched over Δ is simply connected, has $H_2(W;\mathbb{Z}) \cong \mathbb{Z}^{2p}$, and contains p pairwise disjoint homology classes of self-intersection -2 which span a primitive sublattice. Moreover, the signature of W is $\sigma(W) = -2p + \sigma(K) = -2p$, so W is negative definite. □

Note 4.5 If instead K can be unknotted using n generalized crossing changes, all of sign -1, Theorem 4.4 applied to $-K$ shows that the double cover $-Y$ of S^3 branched along $-K$ bounds a smooth negative-definite 4–manifold W with $b_1(W)=0$, $b_2(W)=2n$, and such that $H_2(W;\mathbb{Z})$ contains n pairwise disjoint surface classes of self-intersection -2 which span a primitive sublattice.

In the rest of this paper, we will say $\mathrm{tu}(K)=\pm 1$ if K can be unknotted by a single $\pm$ generalized crossing change. If $\sigma(K)=0$ and $\mathrm{tu}(K)=\pm 1$, we can always get a negative-definite 4–manifold W bounding $\pm\Sigma(K)$: if K can be unknotted by a positive generalized crossing change, then we get a negative-definite W bounding $+\Sigma(K)$, and if K can be unknotted by a negative generalized crossing change, then we get a negative-definite W bounding $-\Sigma(K)$. Moreover, the intersection form on W is represented by a definite 2×2 matrix Q. For an $n\times n$ matrix M, we denote by Γ_M the group $\mathbb{Z}^n/M(\mathbb{Z}^n)$. With this terminology established, we may state the following corollary of Proposition 4.1, which simplifies our computations:

Corollary 4.6 *Let K be an alternating knot such that $\mathrm{tu}(K)=\pm 1$ and $\sigma(K)=0$. We use the convention that $\det K=|\Delta_K(-1)|>0$. Let G be the negative-definite Goeritz matrix obtained from an alternating diagram for $\pm K$. Then there exists a negative-definite matrix of the form*

$$Q=\begin{pmatrix}-\frac{1}{2}(\det K+1) & 1\\ 1 & -2\end{pmatrix}$$

such that $\pm Y=\pm\Sigma(K)$ bounds a negative-definite 4–manifold with intersection form Q. Moreover, there is an isomorphism $\phi\colon \Gamma_Q\to\Gamma_G$ such that, for all $g\in\Gamma_Q$,

$$m_Q(g)\le m_G(\phi(g)), \tag{4-3}$$

$$m_Q(g)\equiv m_G(\phi(g)) \bmod 2. \tag{4-4}$$

Proof By Theorem 4.4, $\pm Y$ bounds a negative-definite 4–manifold with intersection form represented by

$$P=\begin{pmatrix}a & b\\ b & -2\end{pmatrix}$$

for some $a,b\in\mathbb{Z}$. By Theorem 2.2, there must exist isomorphisms

$$\Gamma_P \xrightarrow[\cong]{\phi} \mathrm{Spin}^c(\pm Y)\cong H^2(Y;\mathbb{Z}) \xrightarrow[\cong]{\mathrm{PD}} H_1(Y;\mathbb{Z}),$$

where the isomorphism labeled "PD" is from Poincaré duality and the order of $H_1(Y;\mathbb{Z})$ is equal to $\det K$. The matrix P presents the group $\mathbb{Z}/(\det P)\mathbb{Z}$. Therefore, we must have $\det P = \pm \det K$. Since $\det K$ is odd, we have that

$$b^2 \equiv -2a - b^2 = \det P \equiv \det K \equiv 1 \mod 2,$$

and hence b is odd. Therefore, we can use simultaneous row and column operations to change P into a matrix of form

$$Q = \begin{pmatrix} a & 1 \\ 1 & -2 \end{pmatrix}.$$

Since Q is negative definite, $\det Q \geq 0$, so we must have $\det Q = +\det K$. Therefore, $a = -\frac{1}{2}(\det K + 1)$. It follows from Theorem 3.1 that $m_Q(g) \leq m_G(g)$ and that the two are congruent modulo 2. The corollary follows. □

Note 4.7 Ozsváth and Szabó used a similar process to obstruct knots from having unknotting number 1 in [19], although their isomorphisms ϕ were also required to satisfy a "symmetry" condition which is not necessarily satisfied in our case. In [19, Corollary 1.3], Ozsváth and Szabó computed the m_Q and m_G for various knots to determine whether there exist isomorphisms ϕ of the type given in Corollary 4.6. The only knot with signature 0 which had its unknotting number determined by Ozsváth and Szabó for which the untwisting number was unknown and for which the "symmetry" condition was not necessary is 10_{68}. In this way, we get from their computations that $\mathrm{tu}(10_{68}) = 2 = u(10_{68})$, even though $u_a(10_{68}) = 1$.

5 Examples

In this section, we will prove Theorems 1.3 and 1.4 using Corollary 4.6. Following Ozsváth and Szabó in [19], we will refer to an isomorphism ϕ satisfying (4-3) as a *positive matching* and an isomorphism ϕ satisfying (4-4) as an *even matching*. We obstruct the existence of positive, even matchings for each of the cases listed in Theorem 1.3. We illustrate the proof that $\mathrm{tu}(10_{68}) = 2$; the remaining knots are obstructed from having untwisting number $+1$ and/or -1 similarly.

Example 5.1 Although Ozsváth and Szabó have already verified in [19] that $\Sigma(10_{68})$ cannot bound a 4–manifold with intersection form

$$Q = \begin{pmatrix} -29 & 1 \\ 1 & -2 \end{pmatrix},$$

as it would have to if $\mathrm{tu}(10_{68}) = 1$, we replicate the computation below. The knot 10_{68} has $\sigma(10_{68}) = 0$, $\det 10_{68} = 57$ and Goeritz matrix

$$G = \begin{pmatrix} -4 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -3 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -3 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & -2 \end{pmatrix}.$$

The values of m_G mod 2 are

0	98/57	50/57	28/19	86/57	56/57	36/19	14/57	2/57	24/19
110/57	2/57	30/19	32/57	56/57	16/19	8/57	50/57	20/19	2/3
98/57	4/19	8/57	86/57	6/19	32/57	14/57	26/19	110/57	110/57
26/19	14/57	32/57	6/19	86/57	8/57	4/19	98/57	2/3	20/19
50/57	8/57	16/19	56/57	32/57	30/19	2/57	110/57	24/19	2/57
14/57	36/19	56/57	86/57	28/19	50/57	98/57.			

If $\Sigma(10_{68})$ bounded a 4–manifold W as in Corollary 4.6, the matrix

$$Q = \begin{pmatrix} a & 1 \\ 1 & -2 \end{pmatrix}$$

representing the intersection form on W would have determinant equal to $-2a-1 = \det(10_{68}) = 57$, so that $a = -29$ and

$$Q = \begin{pmatrix} -29 & 1 \\ 1 & -2 \end{pmatrix}.$$

In this case, the values of m_Q mod 2 are

0	112/57	106/57	32/19	82/57	64/57	14/19	16/57	100/57	22/19
28/57	100/57	18/19	4/57	64/57	2/19	58/57	106/57	12/19	4/3
112/57	10/19	58/57	82/57	34/19	4/57	16/57	8/19	28/57	28/57
8/19	16/57	4/57	34/19	82/57	58/57	10/19	112/57	4/3	12/19
106/57	58/57	2/19	64/57	4/57	18/19	100/57	28/57	22/19	100/57
16/57	14/19	64/57	82/57	32/19	106/57	112/57.			

These lists are not identical (in particular, there is a $112/57$ in the m_Q list but not in the m_G list), so there are no even matchings here and $\mathrm{tu}(10_{68}) \neq +1$.

The Goeritz matrix for -10_{68} is

$$G' = \begin{pmatrix} -3 & 1 & 0 \\ 1 & -5 & 3 \\ 0 & 3 & -6 \end{pmatrix};$$

the values of $m_{G'}$ are

0	4/57	16/57	12/19	-50/57	-14/57	10/19	-32/57	28/57	-6/19
-56/57	28/57	2/19	-8/57	-14/57	-4/19	-2/57	16/57	-24/19	-2/3
4/57	-20/19	-2/57	-50/57	-30/19	-8/57	-32/57	-16/19	-56/57	-56/57
-16/19	-32/57	-8/57	-30/19	-50/57	-2/57	-20/19	4/57	-2/3	-24/19
16/57	-2/57	-4/19	-14/57	-8/57	2/19	28/57	-56/57	-6/19	28/57
-32/57	10/19	-14/57	-50/57	12/19	16/57	4/57.			

Using a Python program, we check all possible isomorphisms ϕ and find that there are no positive, even matchings between the values of m_Q and the values of $m_{G'}$. Therefore, $\mathrm{tu}(10_{68}) \neq -1$. Since $u(10_{68}) = 2$, we must have that $\mathrm{tu}(10_{68}) = 2$ as well.

6 Ozsváth–Szabó τ invariant and Rasmussen s invariant obstructions to p–untwisting number

In this section, we investigate p–generalized crossing changes for fixed p in order to prove Theorem 1.5.

Every p–generalized crossing change consists of $p(p-1)+p^2 = p(2p-1)$ standard crossing changes. Thus, for every positive integer p and every knot $K \subset S^3$, if $\mathrm{tu}_p(K) \leq n$, then there is an unknotting sequence consisting of $pn(2p-1)$ crossing changes such that

$$u(K) \leq p(2p-1)\,\mathrm{tu}_p(K),$$

whence

$$|\tau(K)| \leq u(K) \leq p(2p-1)\,\mathrm{tu}_p(K).$$

Thus, it is possible to use the τ invariant to get lower bounds on tu_p for all p. These bounds may be useful in distinguishing tu_p from tu_q for $p \neq q$. However, we may obtain a stronger bound using the smooth 4–genus as follows. While visiting Mark Powell at the Max Planck Institute, he suggested this theorem and outlined a somewhat more complicated proof. It is similar to the proof of Powell and coauthors T Cochran, S Harvey, and A Ray that the τ and s invariants give lower bounds for their bipolar metrics (to appear in a future paper). The following, simpler proof involving the 4–genus was suggested by the referee.

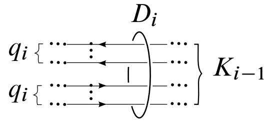


Figure 10: The result of the isotopy on D_i and the strands of K_{i-1}. We call the strands on the top *left-oriented* and those on the bottom *right-oriented*.

Theorem 6.1 *If K can be unknotted by k generalized crossing changes, where the i^{th} change is performed on $2q_i$ strands, then*

$$g_4(K) \leq \sum_{i=1}^{k} q_i^2.$$

Proof Suppose that K may be unknotted via k generalized crossing changes. Then there is a sequence of k generalized crossing changes taking K to U,

$$K = K_0 \xrightarrow{q_1\text{–gcc}} K_1 \xrightarrow{q_2\text{–gcc}} \cdots \xrightarrow{q_{k-1}\text{–gcc}} K_{k-1} \xrightarrow{q_k\text{–gcc}} K_k = U,$$

for which K_i is obtained from K_{i-1} by a single q_i–generalized crossing change for $i = 1, \dots, k$. Let D_i be the disk bounded by the unknot U_i on which the i^{th} q_i–generalized crossing change is performed.

First, note that we can isotope D_i so that the strands of K_{i-1} pass through it as in Figure 10. The strands passing through D_i are oriented in two different ways; we separate the q_i strands of each orientation as in the figure. Let us arbitrarily call one group of q_i strands (say, the ones on the top of the figure) "left-oriented" and the other group "right-oriented". Hence, we may assume without loss of generality that we have a local picture as in Figure 10.

A q_i–generalized crossing change can be undone by changing $q_i(2q_i - 1)$ crossings; one changes precisely one crossing between the i^{th} and j^{th} strands (s_i and s_j) for each $1 \leq i < j \leq 2q_i$. Since q_i of the strands are oriented in one direction and q_i in the other, q_i^2 of these crossing changes occur between strands oriented in opposite directions and $q_i(q_i - 1)$ occur between strands oriented in the same direction (see Figure 11 for an illustration in the case of a 4–generalized crossing change). Thus, q_i^2 of the crossing changes have one sign, and $q_i^2 - q_i$ have the other sign. Therefore, K can be unknotted by changing P positive crossings and N negative crossings, where

$$\max\{P, N\} \leq \sum_{i=1}^{k} q_i^2.$$

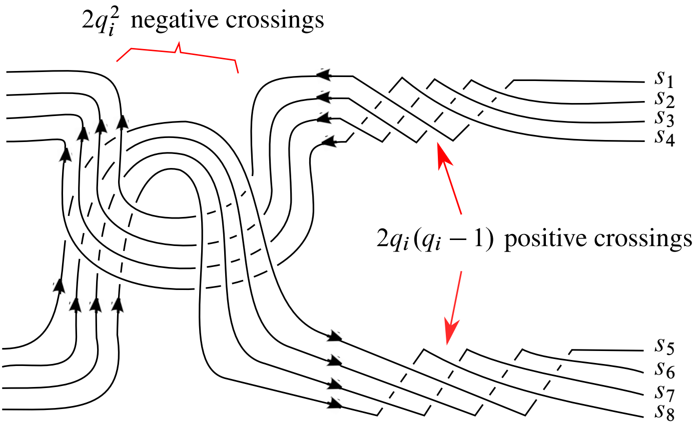


Figure 11: Two sets of four strands twisted around each other at a positive 4–generalized crossing change

However, it is well known, for instance by the argument in the third paragraph of the introduction of [15], that if K can be unknotted by changing P positive crossings and N negative crossings, then $g_4(K) \leq \max\{P, N\}$. □

Since the Ozsváth–Szabó τ invariant and Rasmussen s invariant give lower bounds on the slice genus of any knot, we immediately get the following:

Corollary 6.2 *Let K be a knot which can be converted to the unknot via k generalized crossing changes, where the i^{th} generalized crossing change is performed on $2q_i$ strands for $i = 1, \dots, k$. Then*

$$|\tau(K)| \leq \sum_{i=1}^{k} q_i^2 \quad \text{and} \quad \tfrac{1}{2}|s(K)| \leq \sum_{i=1}^{k} q_i^2.$$

This corollary gives rise to a method for distinguishing $\mathrm{tu}_q(K)$ from $\mathrm{tu}_p(K)$ for some $p, q > 1$. Suppose that $\mathrm{tu}_q(K) \leq n$. Then there exists an untwisting sequence for K consisting of n generalized crossing changes on $2p_i$ strands each, where $i = 1, \dots, n$ and $p_i \leq q$ for all i. Applying the corollary, we get that

$$|\tau(K)| \leq \sum_{i=1}^{n} p_i^2 \leq \sum_{i=1}^{n} q^2 = nq^2,$$

so we must have

$$n \geq \frac{|\tau(K)|}{q^2},$$

and similarly for $\frac{1}{2}|s(K)|$ in place of $|\tau(K)|$. We thus obtain the following obstruction to $\mathrm{tu}_q(K) = n$:

Corollary 6.3 *For all integers $q \geq 1$ and all knots $K \subset S^3$,*

$$\mathrm{tu}_q(K) \geq \frac{|\tau(K)|}{q^2} \quad \textit{and} \quad \mathrm{tu}_q(K) \geq \frac{|s(K)|}{2q^2}.$$

Note 6.4 The above obstruction shows that $|\tau(K)| \leq p^2 \cdot \mathrm{tu}_p(K)$ for all K, which is stronger than the obstruction $|\tau(K)| \leq p(2p-1)\,\mathrm{tu}_p(K)$ given by representing a p–generalized crossing change as $p(2p-1)$ standard crossing changes.

Example 6.5 Let K_{p^3} denote the $(p^3, 1)$–cable of a knot K with $u(K) = 1 = \tau(K) = g(K)$ (one example is the right-handed trefoil knot). We know from [4, Section 5.1] that $\mathrm{tu}_{p^3}(K_{p^3}) = 1$ and that $\tau(K_{p^3}) = p^3$. However, the above result shows that

$$\mathrm{tu}_p(K_{p^3}) \geq \frac{|\tau(K_{p^3})|}{p^2} = p$$

for all integers $p \geq 1$. Hence

$$\mathrm{tu}_p(K_{p^3}) - \mathrm{tu}_{p^3}(K_{p^3}) = \mathrm{tu}_p(K_{p^3}) - 1 \geq p - 1 \xrightarrow{p\to\infty} \infty.$$

References

[1] **C C Adams**, *The knot book*, W. H. Freeman, New York (1994) MR

[2] **A J Casson**, **C M Gordon**, *On slice knots in dimension three*, from "Algebraic and geometric topology, Part 2" (R J Milgram, editor), Proc. Sympos. Pure Math. XXXII, Amer. Math. Soc., Providence, RI (1978) 39–53 MR

[3] **T D Cochran**, **W B R Lickorish**, *Unknotting information from 4–manifolds*, Trans. Amer. Math. Soc. 297 (1986) 125–142 MR

[4] **K Ince**, *The untwisting number of a knot*, Pacific J. Math. 283 (2016) 139–156 MR

[5] **L H Kauffman**, **L R Taylor**, *Signature of links*, Trans. Amer. Math. Soc. 216 (1976) 351–365 MR

[6] **P B Kronheimer**, **T S Mrowka**, *Gauge theory for embedded surfaces, I*, Topology 32 (1993) 773–826 MR

[7] **P B Kronheimer**, **T S Mrowka**, *Gauge theory for embedded surfaces, II*, Topology 34 (1995) 37–97 MR

[8] **C Livingston**, *The slicing number of a knot*, Algebr. Geom. Topol. 2 (2002) 1051–1060 MR

[9] **Y Mathieu**, **M Domergue**, *Chirurgies de Dehn de pente ±1 sur certains nœuds dans les 3–variétés*, Math. Ann. 280 (1988) 501–508 MR

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

[11] **J M Montesinos**, *Three-manifolds as 3–fold branched covers of S^3*, Quart. J. Math. Oxford Ser. 27 (1976) 85–94 MR

[12] **H Murakami**, *Algebraic unknotting operation*, from "Proceedings of the Second Soviet-Japan Joint Symposium of Topology", volume 8 (1990) 283–292 MR

[13] **M Nagel**, **B Owens**, *Unlinking information from 4–manifolds*, Bull. Lond. Math. Soc. 47 (2015) 964–979 MR

[14] **B Owens**, *Unknotting information from Heegaard Floer homology*, Adv. Math. 217 (2008) 2353–2376 MR

[15] **B Owens**, *On slicing invariants of knots*, Trans. Amer. Math. Soc. 362 (2010) 3095–3106 MR

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

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

[18] **P Ozsváth**, **Z Szabó**, *On the Floer homology of plumbed three-manifolds*, Geom. Topol. 7 (2003) 185–224 MR

[19] **P Ozsváth**, **Z Szabó**, *Knots with unknotting number one and Heegaard Floer homology*, Topology 44 (2005) 705–745 MR

[20] **D Rolfsen**, *Knots and links*, Mathematics Lecture Series 7, Publish or Perish, Berkeley, CA (1976) MR

[21] **O J Viro**, *Branched coverings of manifolds with boundary, and invariants of links, I*, Izv. Akad. Nauk SSSR Ser. Mat. 37 (1973) 1241–1258 MR In Russian; translated in Math. USSR-Izvestiya 7 (1973) 1239–1256

Department of Mathematics, Westminster College
1840 South 1300 East, Salt Lake City, UT 84105, United States

kince@westminstercollege.edu

Received: 20 April 2016 Revised: 7 November 2016

Algebraic & Geometric Topology 17 (2017) 2307–2356

Cyclotomic structure in the topological Hochschild homology of DX

CARY MALKIEWICH

Let X be a finite CW complex, and let DX be its dual in the category of spectra. We demonstrate that the Poincaré/Koszul duality between $\mathrm{THH}(DX)$ and the free loop space $\Sigma^\infty_+ LX$ is in fact a genuinely S^1–equivariant duality that preserves the C_n–fixed points. Our proof uses an elementary but surprisingly useful rigidity theorem for the geometric fixed point functor Φ^G of orthogonal G–spectra.

19D55, 55P43; 55P25, 55P91

1 Introduction

Topological Hochschild homology (THH) is a powerful and computable invariant of rings and ring spectra. Like ordinary Hochschild homology, it is built by a cyclic bar construction on the ring R, but with the tensor products of abelian groups $R \otimes_Z R$ replaced by smash products of spectra $R \wedge_{\mathbb{S}} R$.

This construction was originally developed by Bökstedt [10], using ideas of Goodwillie and Waldhausen. The result is a spectrum $\mathrm{THH}(R)$ with a circle action. Out of its fixed points one builds topological cyclic homology $\mathrm{TC}(R)$, a very close approximation to the algebraic K–theory spectrum $K(R)$. This machinery has been tremendously successful at advancing our understanding of $K(R)$ when R is a discrete ring, and Waldhausen's functor $A(X) = K(\Sigma^\infty_+ \Omega X)$ for any space X, to say nothing of the K–theory of other ring spectra. The THH construction is also of intrinsic interest when one studies topological field theories, and TC appears to be an analogue of "crystalline cohomology" from algebraic geometry.

In this paper we use THH to study the ring spectrum DX, the Spanier–Whitehead dual of a finite CW complex. We are motivated by classical work on the Hochschild homology of the cochains $C^*(X)$. Jones [18] proved that when X is simply connected there is an isomorphism

$$HH_*(C^*(X)) \cong H^*(LX),$$

where LX is the space of free loops in X, and all homology is taken with field coefficients. We investigate a lift of this theorem to spectra. Namely, the functional

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2307

dual of $\mathrm{THH}(DX)$ is equivalent to $\Sigma^\infty_+ LX$ when X is finite and simply connected:

$$D(\mathrm{THH}(DX)) \simeq \Sigma^\infty_+ LX \simeq \mathrm{THH}(\Sigma^\infty_+ \Omega X). \tag{1}$$

This was observed by Cohen in the course of some string-topology calculations. Kuhn proved a more general statement for the tensor of the commutative ring DX with any unbased finite complex K, not just the circle S^1; see Kuhn [20]. The THH duality (1) was also extended by Campbell from $(\Sigma^\infty_+ \Omega X, DX)$ to other pairs of Koszul-dual ring spectra; see Campbell [13]. These generalizations can also be seen as special cases of the Poincaré/Koszul duality theorem of Ayala and Francis [4].

If $X = M$ is a closed smooth manifold, we refer to (1) as Atiyah duality for the infinite-dimensional manifold LM. Classical Atiyah duality is an equivalence of ring spectra $M^{-TM} \simeq DM$, where M^{-TM} has the intersection product described by Cohen and Jones [14]. If K is a finite set, the K–fold multiplicative norm of M^{-TM} is

$$N^K(M^{-TM}) = \wedge^k(M^{-TM}) = (M^k)^{-TM^{\oplus k}} = \mathrm{Map}(K, M)^{-T\,\mathrm{Map}(K,M)}.$$

By analogy, we define the "Thom spectrum" of the infinite-dimensional virtual bundle $-TLM$ over LM to be the multiplicative S^1–norm of M^{-TM}:

$$LM^{-TLM} = \mathrm{Map}(S^1, M)^{-T\,\mathrm{Map}(S^1,M)} = N^{S^1}(M^{-TM}).$$

By Angeltveit, Blumberg, Gerhardt, Hill, Lawson and Mandell [2], the THH of a commutative ring spectrum is a model for this multiplicative S^1–norm, so the duality (1) may be interpreted as

$$D(LM^{-TLM}) \simeq \Sigma^\infty_+ LM.$$

Previous work on the duality (1) has left open the question of whether it actually preserves any of the fixed points under the circle action. We address this with the following theorem:

Theorem 1.1 *When X is finite and simply connected, the map of (1) is an equivalence of cyclotomic spectra. It therefore induces equivalences of fixed point spectra*

$$\Phi^{C_n} D(\mathrm{THH}(DX)) \simeq \Phi^{C_n} \Sigma^\infty_+ LX,$$
$$[D(\mathrm{THH}(DX))]^{C_n} \simeq [\Sigma^\infty_+ LX]^{C_n},$$

for all finite subgroups $C_n \leq S^1$.

These notions of fixed points are recalled in Section 3.1. Cyclotomic spectra are recalled in Section 5.1; the main examples are $\mathrm{THH}(R)$ and $\Sigma^\infty_+ LX$, and this is the structure which allows us to compute $\mathrm{TC}(R)$ and $\mathrm{TC}(X)$.

Implicit in the above theorem is the construction of a cyclotomic structure on the dual $D(\mathrm{THH}(DX))$. In fact we show that for any associative ring spectrum R, the functional dual $D(\mathrm{THH}(R))$ comes with a natural *precyclotomic* structure, and in the case of $R = DX$ with X finite and simply connected, this becomes a cyclotomic structure.

We believe that this theorem suggests deeper connections between Waldhausen's functor $A(X)$ and the algebraic K–theory of DX. We will attempt to explore this idea further in future work.

Our work on $\mathrm{THH}(DX)$ builds on very recent results of Angeltveit, Blumberg, Gerhardt, Hill, Lawson and Mandell [1; 2], along with the thesis of Martin Stolz [29]. They establish that the cyclic bar construction, in orthogonal spectra, has the same equivariant behavior as Bökstedt's original construction [10] of topological Hochschild homology. But in many respects, this cyclic bar construction is much simpler. This leads to simplifications in the theory of THH, as well as new results, including those outlined above.

Our proofs also have consequences for the general theory of cyclotomic spectra and G–spectra. Let G be a compact Lie group. We prove a rigidity result for the smash powers and geometric fixed points of orthogonal spectra, which appears to be new and of independent interest. Let Φ be the functor from k–tuples of orthogonal G–spectra to orthogonal spectra

$$\Phi(X_1, \dots, X_k) = \Phi^G X_1 \wedge \cdots \wedge \Phi^G X_k,$$

where Φ^G is the monoidal geometric fixed point functor of Mandell and May [26].

Theorem 1.2 *Suppose $\eta\colon \Phi \to F$ is a natural transformation, and η is an isomorphism on every k–tuple of free G–spectra. Then there are only two natural transformations from Φ to F: the given transformation η, and zero.*

We emphasize that this theorem applies to point-set functors of orthogonal spectra, not to functors defined on the homotopy category. It is designed to prove that certain point-set constructions strictly agree, thereby eliminating the need to construct coherence homotopies between them.

The rigidity theorem has a host of technical corollaries. Here are two of them.

Corollary 1.3 *For G a finite group, the Hill–Hopkins–Ravenel diagonal map*

$$\Phi^H X \xrightarrow{\Delta} \Phi^G N_H^G X$$

is the only nonzero natural transformation from $\Phi^H X$ to $\Phi^G N_H^G X$.

This also applies to the subcategory of cofibrant spectra, giving an easy proof that the diagonal isomorphism constructed by Brun, Dundas and Stolz [12] agrees with the one constructed by Hill, Hopkins and Ravenel [17].

Corollary 1.4 *For G a compact Lie group, the commutation map*

$$\Phi^G X \wedge \Phi^G Y \xrightarrow{\alpha} \Phi^G (X \wedge Y)$$

is the only such natural transformation that is nonzero.

The rigidity theorem gives a useful framework for understanding how multiplicative structure interacts with cyclotomic structure in orthogonal spectra. Motivated by Kaledin's ICM address [19], we use Theorem 1.2 to place certain tensors and internal homs into the model category of cyclotomic spectra; see Blumberg and Mandell [8]. In particular, we get

Corollary 1.5 *The homotopy category of cyclotomic spectra is tensor triangulated.*

Barwick and Glasman [5] have recently extended this program further.

The paper is organized as follows. In Section 2 we review the theory of cyclic spaces and spectra. In Section 3 we review orthogonal G–spectra, and prove Theorem 1.2. In Section 4 we combine the previous two sections and develop the norm model of THH following [2]. In Section 5 we study the interaction of multiplicative structure and cyclotomic structure, proving Theorem 1.1.

The author is grateful to acknowledge Andrew Blumberg, Jon Campbell, Ralph Cohen and Randy McCarthy for several helpful and inspiring conversations throughout this project. He thanks Nick Kuhn for insightful comments on the first version of the paper, and the anonymous referee for a very close reading that substantially improved the exposition throughout. This paper represents a part of the author's PhD thesis, written under the direction of Ralph Cohen at Stanford University.

2 Review of cyclic spaces

A *cyclic set* is a simplicial set with extra structure, which allows the geometric realization to carry a natural S^1–action [15]. Similarly one may define *cyclic spaces* and *cyclic spectra*. In this section we collect together the main results of the theory of cyclic spaces, and their extensions to cocyclic spaces. We also describe (co)cyclic orthogonal spectra, though we defer the study of their equivariant behavior to Section 4. This section is all standard material from [16; 18; 11; 23] or a straightforward generalization thereof, but we make an effort to be definite and explicit in areas where our later proofs require it. We will also be brief; the reader seeking more complete proofs is referred to the author's thesis [24].

2.1 The category $\mathbf{\Lambda}$ and the natural circle action

Recall that $\mathbf{\Delta}$ is a category with one object $[n] = \{0, 1, \ldots, n\}$ for each $n \geq 0$. The morphisms $\mathbf{\Delta}([m],[n])$ are the functions $f\colon [m] \to [n]$ which preserve the total ordering. It is generated by the *coface maps* and *codegeneracy maps*

$$\left.\begin{aligned} d^i\colon [n-1] \to [n], \quad & j \mapsto \begin{cases} j & \text{if } j < i, \\ j+1 & \text{if } j \geq i, \end{cases} \\ s^i\colon [n+1] \to [n], \quad & j \mapsto \begin{cases} j & \text{if } j \leq i, \\ j-1 & \text{if } j > i, \end{cases} \end{aligned}\right. \qquad \text{for } 0 \leq i \leq n.$$

A simplicial object of $\boldsymbol{C}$ is a contravariant functor $X_\bullet\colon \mathbf{\Delta}^{\mathrm{op}} \to \boldsymbol{C}$. We are interested in the case where $\boldsymbol{C}$ is based spaces or orthogonal spectra. Any simplicial object $X_\bullet$ has a canonical presentation

$$\bigvee_{m,n} \mathbf{\Delta}(\bullet,[m])_+ \wedge \mathbf{\Delta}([m],[n])_+ \wedge X_n \rightrightarrows \bigvee_n \mathbf{\Delta}(\bullet,[n])_+ \wedge X_n \to X_\bullet.$$

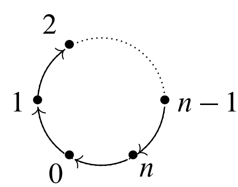
There is a geometric realization functor $|-|$ taking simplicial spaces to spaces. It is the unique colimit-preserving functor that takes $\Delta[n]$ to Δ^n, the convex hull of the standard basis vectors in $\mathbb{R}^{n+1}$. It turns out that for simplicial based spaces $X_\bullet$, the realization $|X_\bullet|$ is given by either of the two coequalizers

$$\coprod_{m,n} \Delta^m \times \mathbf{\Delta}([m],[n]) \times X_n \rightrightarrows \coprod_n \Delta^n \times X_n \to |X_\bullet|,$$

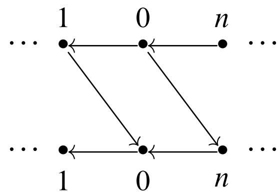
$$\bigvee_{m,n} \Delta^m_+ \wedge \mathbf{\Delta}([m],[n])_+ \wedge X_n \rightrightarrows \bigvee_n \Delta^n_+ \wedge X_n \to |X_\bullet|.$$

When $X_\bullet$ is a simplicial orthogonal spectrum we define $|X_\bullet|$ by the latter of these two formulas.

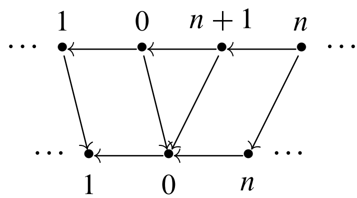
Connes's cyclic category $\mathbf{\Lambda}$ has the same objects as $\mathbf{\Delta}$, but more morphisms. Let $[n]$ denote the free category on the arrows:



The geometric realization $|N_\bullet[n]|$ of the nerve of the category $[n]$ is homotopy equivalent to the circle. The set $\mathbf{\Lambda}([m],[n])$ consists of those functors $[m] \to [n]$ which give a degree 1 map on the geometric realizations. This is generated by maps in $\mathbf{\Delta}$ plus a *cycle map* $\tau_n\colon [n] \to [n]$ for each $n \geq 0$:



We may also generate $\mathbf{\Lambda}$ by $\mathbf{\Delta}$ and an *extra degeneracy map* $s^{n+1}\colon [n+1] \to [n]$ for each $n \geq 0$, corresponding to the functor $[n+1] \to [n]$ pictured below:



We note that a morphism $f \in \mathbf{\Lambda}([m],[n])$ is determined by the underlying map of sets $\mathbb{Z}/(m+1) \to \mathbb{Z}/(n+1)$, unless this map of sets is constant, in which case f is determined by which arrow in $[m]$ is sent to a nontrivial arrow in $[n]$.

Definition 2.1 A *cyclic based space* is a functor $X_\bullet\colon \mathbf{\Lambda}^{\mathrm{op}} \to \mathbf{Top}_*$. The *geometric realization* $|X_\bullet|$ is defined by restricting $X_\bullet$ to $\mathbf{\Delta}^{\mathrm{op}}$ and taking the geometric realization of the resulting simplicial space.

Theorem 2.2 (eg [16]) *The geometric realization $|X_\bullet|$ of a cyclic based space X carries a natural based S^1–action.*

Proof The cyclic space $X_\bullet$ is a colimit of representable cyclic sets

$$\Lambda[n] = \mathbf{\Lambda}(-,[n]).$$

So, it suffices to prove that the space

$$\Lambda^n := |\Lambda[n]|$$

has an S^1 action for all n, commuting with the action of the category $\mathbf{\Lambda}$. By a combinatorial argument, we have homeomorphisms $\Lambda^n \cong S^1 \times \Delta^n$, and we define an S^1 action by translation on the first coordinate. These actions commute with the action of $\mathbf{\Lambda}$, and so they pass to the realization. We draw a few special cases of Λ^n and how it compares to the simplicial circle times Δ^n in Table 1. □

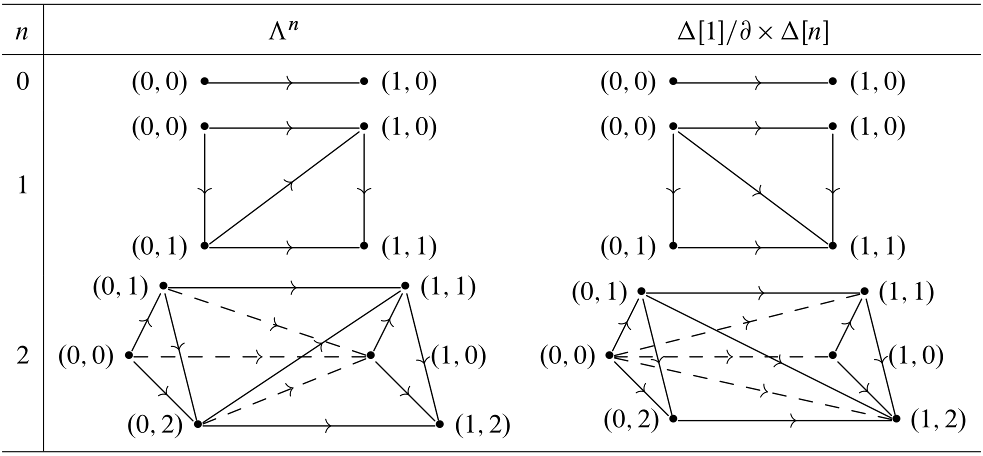


Table 1

2.2 Skeleta and latching objects

When $X_\bullet$ is a simplicial space, the n^{th} *skeleton* $\mathrm{Sk}_n X_\bullet$ is obtained by restricting $X_\bullet$ to the subcategory of $\mathbf{\Delta}^{\mathrm{op}}$ on the objects $0, \dots, n$ and then taking a left Kan extension back. The geometric realization of each skeleton is obtained from the previous one by a pushout square:

$$\begin{tikzcd} L_n X \times \Delta^n \cup_{L_n X \times \partial\Delta^n} X_n \times \partial\Delta^n \arrow[r] \arrow[d] & X_n \times \Delta^n \arrow[d] \\ |\mathrm{Sk}_{n-1} X_\bullet| \arrow[r] & |\mathrm{Sk}_n X_\bullet| \end{tikzcd} \tag{2}$$

Here $L_n X$ is the n^{th} *latching object*, the subspace of X_n consisting of all points in the images of some degeneracy map $s_i\colon X_{n-1} \to X_n$ for $0 \le i \le n-1$. Alternatively, to each proper subset $S \subseteq \{0, 1, \dots, n\}$ that contains 0, we define a map of totally ordered sets $[n] \to S$ by rounding down to the nearest element of S. This makes X_S into a subspace of X_n, and the colimit of these subspaces under inclusions $S \subset T$ gives the subspace $L_n X$.

Definition 2.3 $X_\bullet$ is *Reedy q–cofibrant* if each $L_n X \to X_n$ is a cofibration in the Quillen model structure on based spaces. $X_\bullet$ is *Reedy h–cofibrant* if each $L_n X \to X_n$ is a classical cofibration, ie a map satisfying the unbased homotopy extension property.

We have stated these definitions for based spaces, but they also apply to orthogonal spectra. There is a standard cofibrantly generated model structure that provides the q–cofibrations, while the h–cofibrations are defined as maps having the homotopy extension property with respect to the cylinders $X \wedge I_+$ [27]. So the following standard theorem applies to both spaces and spectra, with either notion of "cofibration":

Proposition 2.4 *If $X_\bullet$ is Reedy cofibrant then $|X_\bullet|$ is cofibrant. If both $X_\bullet$ and $Y_\bullet$ are Reedy cofibrant, then any map $X_\bullet \xrightarrow{\sim} Y_\bullet$ that is an equivalence on each simplicial level induces an equivalence $|X_\bullet| \xrightarrow{\sim} |Y_\bullet|$.*

Proof For simplicial spaces, the proof is an induction up the cube-shaped diagram defining $L_n X$, using the usual pushout and pushout-product properties for cofibrations. The use of unbased h–cofibrations was critical — the theorem is not true with based h–cofibrations, unless all the spaces are well-based.

For orthogonal spectra and q–cofibrations the proof is largely the same. For h–cofibrations of orthogonal spectra, the theorem is a little surprising since we do not assume any of the spectra involved are well-based. The hardest piece of the proof is the statement that if $f\colon K \to L$ is a relative CW complex and $g\colon A \to X$ is an h–cofibration of orthogonal spectra, the pushout-product $f \square g$ is an h–cofibration. This follows from the formal pairing result of Schwänzl and Vogt [28, Corollary 2.9]. $\square$

When $X_\bullet$ is a cyclic space, the simplicial skeleton $|\mathrm{Sk}_n X_\bullet|$ is of limited utility because it is not closed under the circle action. So we draw motivation from [6] and make the following definitions. Since it is important, we remark that here and elsewhere we work in the category of compactly generated, weak Hausdorff spaces.

Definition 2.5 For $n \geq 0$ we define the n^{th} *cyclic skeleton* $\mathrm{Sk}_n^{\text{cyc}} X$ by restricting $X_\bullet$ to the subcategory of $\mathbf{\Lambda}^{\text{op}}$ on the objects $0, \dots, n$ and then taking a left Kan extension back. This may be reexpressed as the coequalizer

$$\bigvee_{k,\ell \leq n} \mathbf{\Lambda}(\bullet, [k])_+ \wedge \mathbf{\Lambda}([k], [\ell])_+ \wedge X_\ell \rightrightarrows \bigvee_{k \leq n} \mathbf{\Lambda}(\bullet, [k])_+ \wedge X_k \to \mathrm{Sk}_n^{\text{cyc}} X_\bullet.$$

We take the $(-1)^{\text{st}}$ cyclic skeleton to be the space X_{-1}, defined as the equalizer of the degeneracy and extra degeneracy maps:

$$\mathrm{Sk}_{-1}^{\text{cyc}} X = X_{-1} \to X_0 \rightrightarrows X_1.$$

Definition 2.6 The n^{th} *cyclic latching object* $L_n^{\text{cyc}} X \subset X_n$ is the closed subspace consisting of all points lying in the image of some degeneracy map

$$s_i\colon X_{n-1} \to X_n, \quad 0 \leq i \leq n.$$

The 0^{th} latching object is also taken to be $\mathrm{Sk}_{-1}^{\text{cyc}} X \subset X_0$ rather than being empty.

The only difference between $L_n X$ and $L_n^{\mathrm{cyc}} X$ is that the *extra* degeneracy is included in $L_n^{\mathrm{cyc}} X$. Equivalently, $L_n^{\mathrm{cyc}} X$ is the closure of $L_n X$ under the action of the cycle map t_n. It follows that $|\mathrm{Sk}_n^{\mathrm{cyc}} X_\bullet|$ is the closure of $|\mathrm{Sk}_n X_\bullet|$ under the circle action.

We briefly prove an equivalent characterization of $L_n^{\mathrm{cyc}} X$. Let $[n]$ denote the cycle category with $n+1$ objects from the definition of $\mathbf{\Lambda}$. Each inclusion of a nonempty subset $S \subset \{0, \dots, n\}$ gives a degree 1 functor $[n] \to [|S|-1]$ which rounds down to the nearest element of S. By the cyclic structure of X, this gives a map $X_S := X_{|S|-1} \to X_n$. If S is empty then we define $X_S = X_{-1}$, and define $X_S \to X_n$ by including into X_0 and applying any composition of degeneracy maps $X_0 \to X_n$.

Proposition 2.7 *This forms a cube-shaped diagram of subspaces of X_n, indexed by the subsets of $\{0, \dots, n\}$ and inclusions. Restricting to the proper subsets, the colimit of this diagram is $L_n^{\mathrm{cyc}} X$.*

Proof If $n < 1$ then this is easy, so we assume $n \geq 1$. It is straightforward to check that our rule respects inclusions of subsets. Each edge of the cube is a standard degeneracy map, which is split by some face map. Since we are working in weak Hausdorff spaces, this implies that each X_S is a closed subspace of X_n. To prove that their colimit is equal to their union, it suffices to check $X_S \cap X_T = X_{S \cap T}$. This reduces to the following claim: For each $0 \leq i \leq n$, let $D_i \colon X_n \to X_n$ be the map induced by the functor $[n] \to [n]$ that sends i to $i-1$ and fixes all other points. Then X_S is precisely the subspace that is fixed by D_i for every i in the complement of S.

To prove this when S is nonempty, note there is a natural projection map $d_S \colon X_n \to X_S$ induced by the inclusion of S into $[n]$. Thinking of this as a map $X_n \to X_n$, the subspace of fixed points is precisely X_S. On the other hand, we may write the complement of S as some cyclically ordered set $\{m_1, \dots, m_k\}$, arranged so that $m_k + 1 \in S$, and then we have the identity

$$d_S = d_{\{m_k\}^c} \cdots d_{\{m_1\}^c} = D_{m_k} \cdots D_{m_1}.$$

Therefore, being in X_S is equivalent to being fixed by D_i for all $i \in S^c$.

If S is empty, then $X_\varnothing = X_{-1}$ is contained in every X_0 and so is fixed by all the projections D_i. Conversely, anything fixed by all the projections is in every subspace of the form $X_{\{s\}} \cong X_0$. In particular it lies in $X_{\{0\}}$ and $X_{\{1\}}$. This gives two points $x_0, x_1 \in X_0$ whose images under the two degeneracy maps are the same point $x \in X_1$. But each face map splits both degeneracy maps, so $x_0 = x_1$ and this point of X_0 lies in the subspace X_{-1}. □

Now we give the analogue of the standard pushout square (2). We expect this is known, but have not found a reference.

Proposition 2.8 *For each $n \geq 0$, there is a natural pushout square of S^1–spaces*

$$\begin{tikzcd} L_n^{\mathrm{cyc}} X \times_{C_{n+1}} \Lambda^n \cup_{L_n^{\mathrm{cyc}} X \times \partial\Lambda^n} X_n \times_{C_{n+1}} \partial\Lambda^n \arrow[r] \arrow[d] & X_n \times_{C_{n+1}} \Lambda^n \arrow[d] \\ |\mathrm{Sk}_{n-1}^{\mathrm{cyc}} X_\bullet| \arrow[r] & |\mathrm{Sk}_n^{\mathrm{cyc}} X_\bullet| \end{tikzcd} \tag{3}$$

for each unbased cyclic space $X_\bullet$, and the obvious variant with smash products when $X_\bullet$ is a based cyclic space.

Proof The square is clearly defined and natural, and the top horizontal map is the inclusion of a subspace. We treat the case $n = 0$ separately, where the square becomes

$$\begin{tikzcd} (L_0^{\mathrm{cyc}} X \times S^1) \amalg \varnothing \arrow[r] \arrow[d] & X_0 \times S^1 \arrow[d] \\ L_0^{\mathrm{cyc}} X \arrow[r] & |\mathrm{Sk}_0^{\mathrm{cyc}} X_\bullet| \end{tikzcd}$$

which is easily checked to be a pushout. For $n \geq 1$, it suffices to check that it is a pushout when $X_\bullet = \Lambda(\bullet, [m])$ is the standard cyclic m–simplex. The square may be rewritten as:

$$\begin{tikzcd} (L_n^{\mathrm{cyc}} \Lambda[m] \times_{C_{n+1}} \Lambda^n) \amalg (\Lambda_n[m] - L_n^{\mathrm{cyc}} \Lambda[m]) \times_{C_{n+1}} \partial\Lambda^n \arrow[r] \arrow[d] & \Lambda_n[m] \times_{C_{n+1}} \Lambda^n \arrow[d] \\ |\mathrm{Sk}_{n-1}^{\mathrm{cyc}} \Lambda[m]| \arrow[r] & |\mathrm{Sk}_n^{\mathrm{cyc}} \Lambda[m]| \end{tikzcd}$$

The top map is a disjoint union of some isomorphisms and some nontrivial inclusions. We strike out the isomorphisms without changing whether the square is a pushout:

$$\begin{tikzcd} (\Lambda_n[m] - L_n^{\mathrm{cyc}} \Lambda[m]) \times_{C_{n+1}} \partial\Lambda^n \arrow[r] \arrow[d] & (\Lambda_n[m] - L_n^{\mathrm{cyc}} \Lambda[m]) \times_{C_{n+1}} \Lambda^n \arrow[d] \\ |\mathrm{Sk}_{n-1}^{\mathrm{cyc}} \Lambda[m]| \arrow[r] & |\mathrm{Sk}_n^{\mathrm{cyc}} \Lambda[m]| \end{tikzcd}$$

The complement of the latching object $L_n^{\mathrm{cyc}} \Lambda[m]$ consists of maps in $\mathbf{\Lambda}([n], [m])$ for which the $n+1$ points $0, \dots, n$ go to distinct points in $0, \dots, m$. The C_{n+1}–action on these maps is free and each orbit has a unique representative that comes from $\mathbf{\Delta}([n], [m])$, so we can again simplify the square to:

$$\begin{tikzcd}
(\Delta_n[m] - L_n\Delta[m]) \times \partial\Lambda^n \arrow[r] \arrow[d] & (\Delta_n[m] - L_n\Delta[m]) \times \Lambda^n \arrow[d] \\
|\mathrm{Sk}^{\mathrm{cyc}}_{n-1}\Lambda[m]| \arrow[r] & |\mathrm{Sk}^{\mathrm{cyc}}_{n}\Lambda[m]|
\end{tikzcd}$$

Now one may identify this square as the standard simplicial pushout square for $\Delta[m]$, multiplied by the identity map on S^1. Alternatively, one can enumerate the cells of $|\mathrm{Sk}^{\mathrm{cyc}}_n\Lambda[m]|$ missing from $|\mathrm{Sk}^{\mathrm{cyc}}_{n-1}\Lambda[m]|$ and check that the above map attaches precisely those cells. So the square is a pushout and the proof is complete. □

As a result, Proposition 2.4 applies to cyclic spectra whose cyclic latching maps are cofibrations, including the "$(-1)^{\mathrm{st}}$ latching map" $* \to X_{-1}$. One can even check that being Reedy cofibrant in the cyclic sense is stronger than being Reedy cofibrant in the ordinary sense.

We will need to know when $|X_\bullet|$ is a cofibrant as a space with an S^1 action:

Definition 2.9 If G is a topological group, a map $X \to Y$ of based G–spaces is a *cofibration* if it is a retract of a relative cell complex built out of cells of the form

$$(G/H \times \partial D^n)_+ \hookrightarrow (G/H \times D^n)_+$$

with $n \geq 0$ and $H \leq G$ any closed subgroup.

Proposition 2.10 *If $X_\bullet$ is a cyclic space, X_{-1} is a cofibrant space and each cyclic latching map $L^{\mathrm{cyc}}_n X \to X_n$ is a cofibration of C_{n+1}–spaces, then $|X_\bullet|$ is a cofibrant S^1–space.*

Proof It suffices to show that each map of cyclic skeleta

$$|\mathrm{Sk}^{\mathrm{cyc}}_{n-1}X| \to |\mathrm{Sk}^{\mathrm{cyc}}_{n}X|$$

is an S^1–cofibration. The (-1)–skeleton is already assumed to be cofibrant, and it has trivial S^1–action, so it is also S^1–cofibrant. For the induction we use the square from Proposition 2.8:

$$\begin{tikzcd}
L^{\mathrm{cyc}}_n X \times_{C_{n+1}} \Lambda^n \cup_{L^{\mathrm{cyc}}_n X \times \partial\Lambda^n} X_n \times_{C_{n+1}} \partial\Lambda^n \arrow[r] \arrow[d] & X_n \times_{C_{n+1}} \Lambda^n \arrow[d] \\
|\mathrm{Sk}^{\mathrm{cyc}}_{n-1}X_\bullet| \arrow[r] & |\mathrm{Sk}^{\mathrm{cyc}}_{n}X_\bullet|
\end{tikzcd}$$

It suffices to prove that the top horizontal is an S^1–cofibration. Since $L^{\mathrm{cyc}}_n X \to X$ is a C_{n+1}–cofibration and $\partial\Lambda^n \to \Lambda^n$ is a free S^1–cofibration, this reduces to proving that the C_{n+1} orbits of the simpler pushout-product

$$[(C_{n+1}/C_r \times \partial D^k \to C_{n+1}/C_r \times D^k)_+ \square (S^1 \times \partial D^\ell \to S^1 \times D^\ell)_+]_{C_{n+1}}$$

is an S^1–cofibration. By associativity of the pushout-product we rewrite this as

$$[(C_{n+1}/C_r \times S^1)_+ \wedge (\partial D^{k+\ell} \to D^{k+\ell})_+]_{C_{n+1}},$$

which simplifies to

$$(S^1/C_r)_+ \wedge (\partial D^{k+\ell} \to D^{k+\ell})_+,$$

and this is one of the generating S^1–cofibrations. □

2.3 Fixed points and subdivision

We turn our attention to the fixed points $|X_\bullet|^{C_r}$, where $C_r \leq S^1$ is the cyclic subgroup of order r. The C_r–fixed points have an action of S^1/C_r, which we usually regard as an S^1–action by pulling back along the group isomorphism

$$\rho_r \colon S^1 \xrightarrow{\cong} S^1/C_r.$$

We will recall the standard result that the C_r–fixed points of $|X_\bullet|$ are built from the spaces $X_{rk-1}^{C_r}$ for $k \geq 1$. One applies a subdivision functor to $X_\bullet$ to obtain a new simplicial space $\mathrm{sd}_r\, X_\bullet$, whose realization is homeomorphic to $|X_\bullet|$, but with simplicial C_r action, giving a homeomorphism

$$|X_\bullet|^{C_r} \cong |(\mathrm{sd}_r\, X_\bullet)^{C_r}|.$$

In fact, one may even put S^1 actions on everything in sight, and the relevant maps are all equivariant. We recall the precise definitions and theorems below.

Definition 2.11 [11] The *r–fold edgewise subdivision functor* is a map of categories $\mathbf{\Delta} \xrightarrow{\mathrm{sd}_r} \mathbf{\Delta}$ which takes $[k-1]$ to $[rk-1]$. Each order-preserving map $[m-1] \to [n-1]$ is repeated r times to give a map $[rm-1] \to [rn-1]$. Given a simplicial space X, we let the *r–fold edgewise subdivision* $\mathrm{sd}_r\, X$ denote the simplicial space obtained by composing with sd_r.

Definition 2.12 The *r–cyclic category* $\mathbf{\Lambda}_r$ is the subcategory of $\mathbf{\Lambda}$ on the objects of the form $[rk-1]$ for $k \geq 1$, generated by all maps in the image of $\mathrm{sd}_r \colon \mathbf{\Delta} \to \mathbf{\Delta}$ in addition to the cycle maps. When working in $\mathbf{\Lambda}_r$ we relabel the object $[rk-1]$ as $[k-1]$. Equivalently, $\mathbf{\Lambda}_r([k-1],[n-1])$ consists of all nondecreasing functions $f \colon \mathbb{Z} \to \mathbb{Z}$ such that $f(x+k) = f(x)+n$, up to the equivalence relation $f \sim f + rn$.

Proposition 2.13 *If $X_\bullet$ is a cyclic space, its r–fold subdivision $\mathrm{sd}_r\, X_\bullet$ is naturally an r–cyclic object in C_r–spaces. The C_r–action is generated by t_{rn-1}^n at simplicial level $n-1$.*

Proposition 2.14 [11, 1.1] *There is a natural diagonal homeomorphism*

$$|\mathrm{sd}_r\, X_\bullet| \xrightarrow{D_r} |X_\bullet|$$

which sends each $(k-1)$–simplex in X_{rk-1} to the corresponding $(rk-1)$–simplex in X_{rk-1} by the diagonal

$$(u_0,\dots,u_{k-1}) \mapsto \Big(\frac{1}{r}u_0,\dots,\frac{1}{r}u_{k-1},\frac{1}{r}u_0,\dots,\frac{1}{r}u_{k-1},\dots\Big).$$

Theorem 2.15 [11, 1.6–1.8, 1.11] *The realization of any r–cyclic space carries a natural S^1–action. The generator of the subgroup $C_r \le S^1$ acts by the simplicial map t^n_{rn-1}. If $X_\bullet$ is a cyclic space, the diagonal homeomorphism D_r is S^1–equivariant.*

Now that we can freely replace $|X_\bullet|$ with $|\mathrm{sd}_r\, X_\bullet|$ as an S^1 space, we see that the C_r–fixed points can be built from the levelwise fixed points $(\mathrm{sd}_r\, X_\bullet)^{C_r}$. These levelwise fixed points are a priori an r–cyclic space, but they are actually a cyclic space because they factor through the following quotient functor:

Definition 2.16 The quotient functor

$$P_r\colon \mathbf{\Lambda}_r([m-1],[n-1]) \to \mathbf{\Lambda}([m-1],[n-1])$$

takes a function $f\colon \mathbb{Z}\to\mathbb{Z}$ up to $f\sim f+rn$ and mods out by the stronger equivalence relation $f\sim f+n$.

We always consider $(\mathrm{sd}_r\, X_\bullet)^{C_r}$ to be a cyclic space, reserving the notation $P_r(\mathrm{sd}_r\, X_\bullet)^{C_r}$ for the corresponding r–cyclic space. With these conventions, the isomorphism between $|X_\bullet|^{C_r}$ and $|(\mathrm{sd}_r\, X_\bullet)^{C_r}|$ is S^1–equivariant:

Proposition 2.17 [11, 1.10–1.12] *The passage between cyclic and r–cyclic structures on $\mathrm{sd}_r\, X_\bullet$ and $(\mathrm{sd}_r\, X_\bullet)^{C_r}$, together with the diagonal of Proposition 2.14, give natural S^1–equivariant homeomorphisms*

$$|(\mathrm{sd}_r\, X_\bullet)^{C_r}| \cong \rho_r^*|P_r(\mathrm{sd}_r\, X_\bullet^{C_r})| \cong \rho_r^*(|\mathrm{sd}_r\, X_\bullet|^{C_r}) \xrightarrow{D_r} \rho_r^*(|X_\bullet|^{C_r})$$

making the following triangle commute:

$$\begin{tikzcd}
|(\mathrm{sd}_{rs}\, X_\bullet)^{C_{rs}}| \arrow[d, "\cong"] \arrow[dr, "\cong"] & \\
\rho_r^*|(\mathrm{sd}_s\, X_\bullet)^{C_s}|^{C_r} \arrow[r, "\cong"] & \rho_{rs}^*|X_\bullet|^{C_{rs}}
\end{tikzcd}$$

2.4 Cocyclic spaces

The previous section dualizes easily. Recall that a cosimplicial object is a covariant functor $X^\bullet\colon \mathbf{\Delta} \to \boldsymbol{C}$. This can be canonically expressed as an equalizer

$$X^\bullet \to \prod_n \mathrm{Map}(\Delta_\bullet[n], X^n) \rightrightarrows \prod_{m,n} \mathrm{Map}(\Delta_\bullet[m] \times \mathbf{\Delta}(m,n), X^n)$$

and so a right adjoint out of cocyclic spaces is determined by what it does to the cosimplicial space $\mathrm{Map}(\Delta_\bullet[n], X^n)$. The totalization is the unique limit-preserving functor to spaces which takes $\mathrm{Map}(\Delta_\bullet[n], A)$ to $\mathrm{Map}(\Delta^n, A)$. It is given by the equalizer

$$\mathrm{Tot}(X^\bullet) \to \prod_n \mathrm{Map}(\Delta^n, X^n) \rightrightarrows \prod_{m,n} \mathrm{Map}(\Delta^m \times \mathbf{\Delta}(m,n), X^n).$$

The totalization of a cosimplicial orthogonal spectrum is given by the same formula.

If $X^\bullet$ is not just cosimplicial, but cocyclic, then its totalization is the equalizer

$$\mathrm{Tot}(X^\bullet) \to \prod_n \mathrm{Map}(\Lambda^n, X^n) \rightrightarrows \prod_{m,n} \mathrm{Map}(\Lambda^m \times \mathbf{\Lambda}(m,n), X^n),$$

which is enough to prove:

Proposition 2.18 *The totalization of a cocyclic space $X^\bullet$ carries a natural S^1–action. Similarly, the totalization of an r–cocyclic space $Y^\bullet$ carries a natural S^1–action, in which the action of $C_r \leq S^1$ is the totalization of a cosimplicial map.*

In the special case of $X^\bullet = \mathrm{Map}(E_\bullet, X)$, where $E_\bullet$ is a cyclic space, the canonical homeomorphism

$$\mathrm{Tot}(X^\bullet) \cong \mathrm{Map}(|E_\bullet|, X)$$

is S^1–equivariant. A useful example to keep in mind is $\mathrm{Map}(S^1_\bullet, X)$, the standard cosimplicial model for the free loop space LX.

Next we recall Reedy fibrancy, which we will only need for cosimplicial spectra (as opposed to spaces). We recall that the construction of the latching map $L_n X \to X_n$ for simplicial spectra dualizes to that of the matching map $X^n \to M_n X$ for cosimplicial spectra. We say that $X^\bullet$ is *Reedy fibrant* if these matching maps are fibrations in the stable model structure on orthogonal spectra. The standard analogue of Proposition 2.4 is:

Proposition 2.19 *A weak equivalence of Reedy fibrant cosimplicial spectra induces a weak equivalence on the totalizations.*

As expected, one can always replace a cosimplicial spectrum by a Reedy fibrant one that is equivalent on every cosimplicial level. In this paper, we will only use Reedy fibrant cosimplicial spectra of the form $F(X_\bullet, Y)$, where $F(-,-)$ is the internal hom in orthogonal spectra, Y is a fibrant spectrum, and $X_\bullet$ is a Reedy q–cofibrant simplicial spectrum. It is straightforward to verify from the properties of the model structure in [27] that such an $F(X_\bullet, Y)$ is always Reedy fibrant.

Finally, any cocyclic space $X^\bullet$ may be composed with sd_r to give an r–cocyclic space $\mathrm{sd}_r\, X^\bullet$. As before, the fixed points of $\mathrm{Tot}(X^\bullet)$ can be recovered as $\mathrm{Tot}((\mathrm{sd}_r\, X^\bullet)^{C_r})$:

Proposition 2.20 *If $X^\bullet$ is a cosimplicial space, there is a natural diagonal homeomorphism*

$$\mathrm{Tot}(X^\bullet) \xrightarrow{D_r} \mathrm{Tot}(\mathrm{sd}_r\, X^\bullet).$$

If $X^\bullet$ is cocyclic, D_r is S^1–equivariant.

Proposition 2.21 *If $X^\bullet$ is a cocyclic space, then $(\mathrm{sd}_r\, X^\bullet)^{C_r}$ may be regarded as a cocyclic space, and there are natural S^1–equivariant homeomorphisms*

$$\mathrm{Tot}((\mathrm{sd}_r\, X^\bullet)^{C_r}) \cong \rho_r^*\mathrm{Tot}(\mathrm{sd}_r\, X^\bullet)^{C_r} \xleftarrow{D_r^{C_r}} \rho_r^*\mathrm{Tot}(X^\bullet)^{C_r}.$$

The proofs are easy dualizations or direct copies of the proofs for cyclic spaces.

2.5 The suspension spectrum of LX

We end this section with a more concrete example. If X is any unbased space, then $\mathrm{Map}(S^1_\bullet, X)$ is a cocyclic space. We add a disjoint basepoint, and smash every level with the sphere spectrum, yielding a cocyclic spectrum

$$\mathbb{S} \wedge \mathrm{Map}(S^1_\bullet, X)_+ = \Sigma^\infty_+ X^{\bullet+1}.$$

It is not hard to check that there is a natural map

$$\Sigma^\infty_+ LX = \Sigma^\infty_+ \mathrm{Tot}(X^{\bullet+1}) \to \mathrm{Tot}(\Sigma^\infty_+ X^{\bullet+1}) \tag{4}$$

given by the interchange

$$\mathbb{S} \wedge \prod_k \mathrm{Map}_*(\Delta^k_+, X^k_+) \to \prod_k \mathbb{S} \wedge \mathrm{Map}_*(\Delta^k_+, X^k_+) \to \prod_k F(\Delta^k_+, \mathbb{S} \wedge X^k_+), \tag{5}$$

where $F(A, E)$ denotes the mapping spectrum or cotensor of a space A with an orthogonal spectrum E. On each spectrum level, the map (4) is a bijection on the underlying sets, but it is likely not a homeomorphism, because assembly maps of the form $A \wedge \mathrm{Map}_*(B, C) \to \mathrm{Map}_*(B, A \wedge C)$ fail to be closed inclusions [21, Appendix A, 8.6]. It does not really matter, because the cocyclic spectrum $\Sigma^\infty_+ X^{\bullet+1}$ is not Reedy fibrant, and so it must be replaced if the totalization is to be homotopically meaningful. Taking a Reedy fibrant replacement $R\Sigma^\infty_+ X^{\bullet+1}$ and totalizing gives a derived version of the interchange map

$$\Sigma^\infty_+ LX \to \mathrm{Tot}(\Sigma^\infty_+ X^{\bullet+1}) \to \mathrm{Tot}(R\Sigma^\infty_+ X^{\bullet+1}).$$

Proposition 2.22 *This composite is a stable equivalence when X is simply connected.*

Proof We first recall that the case where X is finite follows from [20, 6.6], with $K = S^1$ and $Z = X$. To see why, we observe that the cyclic bar construction on the dual DX can be made into a Reedy cofibrant cyclic spectrum (see Section 4). Applying $F(-, f\mathbb{S})$, where $f\mathbb{S}$ is a fibrant replacement of the sphere spectrum, gives a Reedy fibrant cosimplicial spectrum replacing $\Sigma^\infty_+ X^{\bullet+1}$. One then checks that the map of Kuhn's theorem lines up with the interchange we described above.

To get the general case, it suffices to show that both sides of the interchange map commute with filtered homotopy colimits of simply connected spaces. Using the "cube of retracts" terminology from [25], we identify the fibers of the coskeletal filtration of $\mathrm{Tot}(R\Sigma^\infty_+ X^{\bullet+1})$ as

$$F(\Delta^n/\partial\Delta^n, \Sigma^\infty_+ X \wedge \Sigma^\infty X^{\wedge n}) \simeq \Omega^n \Sigma^\infty X^{\wedge n} \vee \Omega^n \Sigma^\infty X^{\wedge(n+1)}.$$

The connectivity of these fibers tends to infinity when X is simply connected, and it follows easily that the limit of the tower commutes with such filtered homotopy colimits.

An alternative argument uses the "cyclic coskeletal filtration" for the right-hand side, whose fibers are

$$F^{C_{n+1}}(\Lambda^n/\partial\Lambda^n, \Sigma^\infty X^{\wedge(n+1)}) \simeq \Omega^n \Sigma^\infty X^{\wedge(n+1)} \vee \Omega^{n+1} \Sigma^\infty X^{\wedge(n+1)}.$$

Along the interchange map, this filtration can be shown to agree with Arone's model of the Taylor tower for $\Sigma^\infty_+ LX$ from [3]. □

3 Orthogonal G–spectra, equivariant smash powers and rigidity

We will now review the theory of orthogonal G–spectra and prove our rigidity theorem for the geometric fixed point functor Φ^G. This result is a technical linchpin that underlies the rest of our treatment of cyclotomic spectra and the cyclic bar construction. It allows us to cleanly reconstruct and extend the model of THH presented in [2].

3.1 Basic definitions, model structures and fixed points

We take these definitions from [26; 17].

Definition 3.1 If G is a fixed compact Lie group, an *orthogonal G–spectrum* is a sequence of based spaces $\{X_n\}_{n=0}^{\infty}$ equipped with

- a continuous action of $G \times O(n)$ on X_n for each n,
- a G–equivariant structure map $\Sigma X_n \to X_{n+1}$ for each n,

such that the composite

$$S^p \wedge X_n \to \cdots \to S^1 \wedge X_{(p-1)+n} \to X_{p+n}$$

is $O(p) \times O(n)$–equivariant. A map of orthogonal G–spectra $X \to Y$ is a collection of maps $X_n \to Y_n$ commuting with all the structure, including the G–actions.

Definition 3.2 Let U be a complete G–universe as in [26]. The category $\mathscr{J}_G$ has objects the finite-dimensional G–representations $V \subset U$, or any orthogonal G–representation isomorphic to such a subspace. The mapping spaces $\mathscr{J}_G(V, W)$ are the Thom spaces $O(V, W)^{W-V}$, consisting of linear isometries $f\colon V \to W$ with choices of point in the orthogonal complement $W - f(V)$. The group G acts on $O(V, W)^{W-V}$ by conjugating the map and acting on the point in $W - f(V)$.

Definition 3.3 A *$\mathscr{J}_G$–space* is an equivariant functor $\mathscr{J}_G$ into based G–spaces and nonequivariant maps. That is, each V is assigned to a based space $X(V)$, and for each pair V, W the map

$$\mathscr{J}_G(V, W) \to \mathrm{Map}_*(X(V), X(W))$$

is equivariant. A map of $\mathscr{J}_G$–spaces is a collection of G–equivariant maps

$$X(V) \to Y(V)$$

commuting with the action of $\mathscr{J}_G$.

Proposition 3.4 *Every $\mathscr{J}_G$–space gives an orthogonal G–spectrum by restricting to $V = \mathbb{R}^n$; denote this functor by $\mathcal{I}_U^{\mathbb{R}^\infty}$. Conversely, given an orthogonal G–spectrum X one may define a $\mathscr{J}_G$–space by the rule*

$$X(V) = X_n \wedge_{O(n)} O(\mathbb{R}^n, V)_+, \quad n = \dim V,$$

with G acting diagonally on X_n and on $O(\mathbb{R}^n, V) = \mathscr{J}_G(\mathbb{R}^n, V)$. Denote this functor by $\mathcal{I}_{\mathbb{R}^\infty}^U$. Then $\mathcal{I}_{\mathbb{R}^\infty}^U$ and $\mathcal{I}_U^{\mathbb{R}^\infty}$ are inverse equivalences of categories.

Definition 3.5 Given a G–representation V and based G–space A, the *free spectrum* $F_V A$ is the $\mathscr{J}_G$–space

$$(F_V A)(W) := \mathscr{J}_G(V, W) \wedge A.$$

For fixed V, the functor $A \mapsto F_V A$ is the left adjoint to the functor that evaluates a $\mathscr{J}_G$–space at V.

Proposition 3.6 [26] *There is a cofibrantly generated model structure on the category of orthogonal G–spectra, in which the cofibrations are the retracts of the cell complexes built from*

$$F_V((G/H \times \partial D^k)_+) \hookrightarrow F_V((G/H \times D^k)_+), \quad k \geq 0,\ H \leq G,\ V \subset U,$$

and the weak equivalences are the maps inducing isomorphisms on the stable homotopy groups

$$\pi_k^H(X) = \begin{cases} \operatorname{colim}_{V \subset U} \pi_k\big(\mathrm{Map}_*^H(S^V, X(V))\big), & k \geq 0, \\ \operatorname{colim}_{V \subset U} \pi_0\big(\mathrm{Map}_*^H(S^{V-\mathbb{R}^{|k|}}, X(V))\big), & k < 0,\ \mathbb{R}^{|k|} \subset V, \end{cases}$$

where $\mathrm{Map}_*^H(-,-)$ *denotes the space of H–equivariant maps.*

Proposition 3.7 [26] *The category $\mathscr{J}$ is symmetric monoidal, using the direct sum of representations. The Day convolution along $\mathscr{J}$ defines a smash product on the category of orthogonal G–spectra, which makes it into a closed symmetric monoidal category. This smash product is a left Quillen bifunctor with respect to the above model structure.*

When working with $G = S^1$, it is common to consider a broader class of weak equivalences that see only the finite subgroups $C_n \leq S^1$.

Definition 3.8 A map of S^1–spectra is an $\mathcal{F}$*–equivalence* if it is an equivalence as a map of C_n–spectra for all $n \geq 1$; equivalently it is an isomorphism on the homotopy groups $\pi_k^{C_n}(X)$ for all $n \geq 1$.

Next we recall the definitions of genuine and geometric fixed points. If X is a G–space and $H \leq G$ is a subgroup, the fixed point subspace X^H has a natural action by only the normalizer $NH \leq G$. Of course H acts trivially and so we are left with a natural action by the *Weyl group*

$$WH = NH/H \cong \mathrm{Aut}_G(G/H).$$

When X is a G–spectrum there are two natural notions of H–fixed points, each of which gives a WH–spectrum:

Definition 3.9 For a $\mathscr{J}_G$–space X and a subgroup $H \le G$, the $\mathscr{J}_{WH}$–space of *categorical fixed points* X^H is defined on each H–fixed G–representation $V \subset U^H \subset U$ as just the fixed points $X(V)^H$. More simply, if X is an orthogonal G–spectrum then X^H is obtained by taking H–fixed points levelwise.

Proposition 3.10 *The categorical fixed points are a Quillen right adjoint from G–spectra to WH–spectra. Their right-derived functor is called the spectrum of genuine fixed points.*

Definition 3.11 If X is a $\mathscr{J}_G$–space and $H \le G$ then the *geometric fixed points* $\Phi^H X$ are defined as the coequalizer

$$\bigvee_{V,W} F_{W^H} S^0 \wedge \mathscr{J}_G^H(V,W) \wedge X(V)^H \rightrightarrows \bigvee_V F_{V^H} S^0 \wedge X(V)^H \to \Phi^H X.$$

These are naturally $\mathscr{J}_{WH}$–spaces on the complete WH–universe U^H.

Theorem 3.12 *The geometric fixed points Φ^H satisfy these technical properties:*

(1) *There is a natural isomorphism of WH–spectra*
$$\Phi^H F_V A \cong F_{V^H} A^H.$$

(2) *Φ^H commutes with all coproducts, pushouts along a levelwise closed inclusion, and filtered colimits along levelwise closed inclusions.*

(3) *Φ^H preserves all cofibrations, acyclic cofibrations and weak equivalences between cofibrant objects.*

(4) *If $H \le K \le G$ then Φ^H commutes with the change-of-groups from G down to K.*

(5) *There is a canonical commutation map*
$$\Phi^G(X \wedge Y) \xrightarrow{\alpha} \Phi^G X \wedge \Phi^G Y,$$
which is an isomorphism when X or Y is cofibrant [8, A.1].

Remark 3.13 The geometric fixed point functor Φ^H is not a left adjoint, since it does not commute with all colimits. A simple counterexample with $G = \mathbb{Z}/2$ is given by the suspension spectra of the diagram of spaces:

$$\begin{tikzcd} (\mathbb{Z}/2)_+ \arrow[r] \arrow[d] & (*)_+ \\ (*)_+ & \end{tikzcd}$$

Therefore Φ^H is not a Quillen left adjoint. However, since it still preserves weak equivalences between cofibrant orthogonal G–spectra, we define a *left-derived geometric fixed point functor* $X \rightsquigarrow \Phi^H(cX)$ by composing Φ^H with a cofibrant replacement functor c in the above model structure.

It will be important for us that these derived geometric fixed points measure the weak equivalences of G–spectra. This is standard; for instance we can deduce it from [17, 2.52; 26, 3.5(vi), 4.12].

Proposition 3.14 *A map $X \to Y$ of orthogonal G–spectra is a weak equivalence if and only if the induced map of derived geometric fixed points $\Phi^H(cX) \to \Phi^H(cY)$ is an equivalence of spectra for all $H \leq G$.*

Consequently, a map of S^1–spectra $X \to Y$ is an $\mathcal{F}$–equivalence if and only if $\Phi^{C_n}(cX) \to \Phi^{C_n}(cY)$ is an equivalence for all $n \geq 1$.

Finally, though it does not seem to appear in the literature, the iterated fixed points map of [8] easily generalizes:

Proposition 3.15 *If $H \leq K \leq NH \leq G$ then there is a natural iterated fixed points map*

$$\Phi^K X \xrightarrow{\text{it}} \Phi^{K/H} \Phi^H X,$$

which is an isomorphism when $X = F_V A$, and therefore an isomorphism on all cofibrant spectra. When H and K are normal, this is a map of G/K–spectra.

3.2 The Hill–Hopkins–Ravenel norm isomorphism

When X is an orthogonal spectrum, the smash product $X^{\wedge n}$ has an action of $C_n \cong \mathbb{Z}/n$ which rotates the factors. This makes $X^{\wedge n}$ into an orthogonal C_n–spectrum. It is natural to guess that the geometric fixed points of this C_n–action should be X itself, and in fact there is natural diagonal map

$$X \xrightarrow{\Delta} \Phi^{C_n} X^{\wedge n}.$$

When X is cofibrant, this map is an *isomorphism*. More generally, if G is a finite group, $H \leq G$ and X is an orthogonal H–spectrum, we can define a smash product of copies of X indexed by G,

$$N_H^G X := \bigwedge\nolimits_{g_i H \in G/H} (g_i H)_+ \wedge_H X \cong \bigwedge\nolimits^{|G/H|} X.$$

This construction is the *multiplicative norm* defined by Hill, Hopkins and Ravenel. This can be given a G–action, which depends on some fixed choice of representatives $g_i H$ for each left coset of H (see [9; 17]). Changing the choice of representatives changes this action, but only up to natural isomorphism. We therefore implicitly assume that such representatives have been chosen. The general form of the above observation about $X^{\wedge n}$ is then:

Theorem 3.16 [17, B.209] *There is a natural "diagonal" map of orthogonal spectra*

$$\Phi^H X \xrightarrow{\Delta} \Phi^G N_H^G X.$$

When X is cofibrant, Δ is an isomorphism.

The full proof now appears in [17], but for the reader's convenience we also summarize the proof below:

Proof If A is just a based H–space, the indexed smash product of A over G/H has fixed points A^H:

$$A^H \xrightarrow{\cong} (N_H^G A)^G \cong \left(\bigwedge^{|G/H|} A\right)^G.$$

Here the map from left to right is the diagonal,

$$a \in A^H \mapsto (a, \dots, a).$$

Now suppose X is an orthogonal H–spectrum. We start by taking its coequalizer presentation

$$\bigvee_{V,W} F_W S^0 \wedge \mathscr{J}_H(V,W) \wedge X(V) \rightrightarrows \bigvee_V F_V S^0 \wedge X(V) \to X$$

and taking $\Phi^G N_H^G$ of everything in sight. Since $\Phi^G N_H^G$ commutes with wedges and smashes up to isomorphism, this gives

$$\bigvee_{V,W} \Phi^G N_H^G F_W S^0 \wedge (N_H^G \mathscr{J}_H(V,W))^G \wedge (N_H^G X(V))^G \\ \rightrightarrows \bigvee_V \Phi^G N_H^G F_V S^0 \wedge (N_H^G X(V))^G \to \Phi^G N_H^G X,$$

which simplifies to

$$\bigvee_{V,W} \Phi^G N_H^G F_W S^0 \wedge \mathscr{J}_H^H(V,W) \wedge X(V)^H \rightrightarrows \bigvee_V \Phi^G N_H^G F_V S^0 \wedge X(V)^H \to \Phi^G N_H^G X.$$

As a diagram, this is no longer guaranteed to be a coequalizer system, but it still commutes. We can simplify using the string of isomorphisms

$$\Phi^G N_H^G F_V A \cong \Phi^G F_{\mathrm{Ind}_H^G V}(N_H^G A) \cong F_{(\mathrm{Ind}_H^G V)^G}(N_H^G A)^G \cong F_{V^H} A^H$$

for any based H–space A and H–representation V. This gives

$$\bigvee_{V,W} F_{W^H} S^0 \wedge \mathscr{J}_H^H(V,W) \wedge X(V)^H \rightrightarrows \bigvee_V F_{V^H} S^0 \wedge X(V)^H \to \Phi^G N_H^G X$$

and the coequalizer of the first two terms is exactly $\Phi^H X$. The universal property of the coequalizer then gives us a map

$$\Phi^H X \to \Phi^G N_H^G X$$

and we take this as the definition of the diagonal map.

Now consider the special case when $X = F_V A$. The inclusion of the term

$$F_{V^H} S^0 \wedge A^H$$

into the above coequalizer system maps forward isomorphically to $\Phi^H X$, and so we can evaluate the diagonal map by just examining this term. But back at the top of our proof, the inclusion of the term

$$\Phi^G N_H^G F_V S^0 \wedge (N_H^G A)^G$$

also maps forward isomorphically to $\Phi^G N_H^G X$. Therefore, up to isomorphism, the diagonal map becomes the string of maps we used to connect $F_{V^H} S^0 \wedge A^H$ to $\Phi^G N_H^G F_V S^0 \wedge (N_H^G A)^G$, but these maps were all isomorphisms. Therefore the diagonal is an isomorphism when $X = F_V A$. It is straightforward to verify that both sides preserve coproducts, pushouts along h–cofibrations and sequential colimits along h–cofibrations, so, by induction, the diagonal is an isomorphism for all cofibrant X. □

3.3 A rigidity theorem for geometric fixed points

Let G be a compact Lie group. We will prove that the geometric fixed point functor Φ^G is *rigid*, in the sense that it admits very few point-set level natural transformations into other functors. Let $G\mathbf{Sp}^O$ denote the category of orthogonal G–spectra and G–equivariant maps between them. Let **Free** be the full subcategory on the free spectra $F_V A$ for all G–representations V and based G–spaces A. Let

$$\wedge \circ (\Phi^G, \ldots, \Phi^G) \colon \prod^k \mathbf{Free} \to \mathbf{Sp}^O$$

denote the composite of the geometric fixed points and the k–fold smash product, with $k \geq 1$.

Proposition 3.17 *The only endomorphisms of $\wedge \circ (\Phi^G)^k$ are zero and the identity.*

Proof A natural transformation $T \colon \wedge \circ (\Phi^G)^k \to \wedge \circ (\Phi^G)^k$ assigns to a k–tuple $(F_0 S^0, F_0 S^0, \ldots, F_0 S^0)$ a map of spectra

$$F_0 S^0 \to F_0 S^0,$$

which is determined at level 0 by a choice of point in S^0. So there are only two such maps, the identity and zero.

Assume that T is the identity on this object. Then consider T on the k–tuple $(F_{V_1} S^0, F_{V_2} S^0, \ldots, F_{V_k} S^0)$:

$$F_{V_1^G \oplus V_2^G \oplus \ldots \oplus V_k^G} S^0 \to F_{V_1^G \oplus V_2^G \oplus \ldots \oplus V_k^G} S^0.$$

Let $m_i := \dim V_i^G$ and fix an isomorphism between $\mathbb{R}^{m_i}$ and V_i^G. The above map is determined by what it does at level $m_1 + \cdots + m_k$:

$$O(m_1 + \cdots + m_k)_+ \to O(m_1 + \cdots + m_k)_+.$$

This map, in turn, is determined by the image of the identity point, which is some element $P \in O(m_1+\cdots+m_k)_+$. Now for any point $(t_1, \dots, t_k) \in S^{m_1} \wedge \cdots \wedge S^{m_k}$ we can choose maps of spectra $F_{V_i} S^0 \to F_0 S^0$ which at level V_i send the nonbasepoint of S^0 to the point $t_i \in S^{m_i} \cong (S^{V_i})^G$. Since T is a natural transformation, this square commutes for all choices of $(t_1, \dots, t_k)$:

$$\begin{tikzcd}
O(m_1 + \cdots + m_k)_+ \arrow[r, "\cdot P"] \arrow[d, "\mathrm{ev}_{(t_1,\dots,t_k)}"] & O(m_1 + \cdots + m_k)_+ \arrow[d, "\mathrm{ev}_{(t_1,\dots,t_k)}"] \\
S^{m_1+\cdots+m_k} \arrow[r, "\mathrm{id}"] & S^{m_1+\cdots+m_k}
\end{tikzcd}$$

Since $O(m_1 + \cdots + m_k)$ acts faithfully on the sphere $S^{m_1+\cdots+m_k}$, we must have $P = \mathrm{id}$. Therefore, our natural transformation T acts as the identity on the k–tuple of spectra $(F_{V_1} S^0, F_{V_2} S^0, \dots, F_{V_k} S^0)$.

Finally, let $A_1, \dots, A_k$ be a sequence of G–spaces, and consider T on the k–tuple $(F_{V_1} A_1, \dots, F_{V_k} A_k)$. Each collection of choices of point $a_i \in A_i^G$ gives a sequence of maps $F_{V_i} S^0 \to F_{V_i} A_i$, and applying T to this sequence of maps gives a commuting square:

$$\begin{tikzcd}
F_{V_1^G \oplus \dots \oplus V_k^G} S^0 \wedge \cdots \wedge S^0 \arrow[r, "\mathrm{id}"] \arrow[d, "F_{\dots}(a_1,\dots,a_k)"] & F_{V_1^G \oplus \dots \oplus V_k^G} S^0 \wedge \cdots \wedge S^0 \arrow[d, "F_{\dots}(a_1,\dots,a_k)"] \\
F_{V_1^G \oplus \dots \oplus V_k^G} A_1^G \wedge \cdots \wedge A_k^G \arrow[r, "T"] & F_{V_1^G \oplus \dots \oplus V_k^G} A_1^G \wedge \cdots \wedge A_k^G
\end{tikzcd}$$

From inspection of level $m_1 + \cdots + m_k$, the bottom map must be the identity on the point $\mathrm{id} \wedge (a_1, \dots, a_k)$. But this is true for all $(a_1, \dots, a_k)$ and so the bottom map is the identity. Therefore, T is the identity on $(F_{V_1} A_1, \dots, F_{V_k} A_k)$, so it is the identity on every object in $\prod^k \mathbf{Free}$.

For the second case, we assume T is zero on $(F_0 S^0, \dots, F_0 S^0)$ and follow the same steps as before, concluding that T is zero on $(F_{V_1} S^0, \dots, F_{V_k} S^0)$ and then it is zero on $(F_{V_1} A_1, \dots, F_{V_k} A_k)$. □

To derive corollaries, we say that a functor $\phi\colon \prod^k G\mathbf{Sp}^O \to \mathbf{Sp}^O$ is *rigid* if restricting to the subcategory $\prod^k \mathbf{Free}$ gives an injective map on natural transformations out of ϕ. In other words, a natural transformation out of ϕ is determined by its behavior on the subcategory $\mathbf{Free}$.

Corollary 3.18 *If ϕ_1 and ϕ_2 are functors $\prod^k G\mathbf{Sp}^O \to \mathbf{Sp}^O$ which when restricted to the subcategory $\prod^k$ **Free** are separately isomorphic to $\wedge \circ (\Phi^G)^k$, and ϕ_1 is rigid, then there is at most one nonzero natural transformation $\phi_1 \to \phi_2$.*

The example we are interested in is the smash product of geometric fixed points.

Proposition 3.19 *The functor $\wedge \circ (\Phi^G, \dots, \Phi^G)$ is rigid.*

Proof For any orthogonal G–spectrum X, let ξ denote the map

$$\xi\colon \bigvee_{V \subset U} F_V X(V) \to X$$

whose V^{th} summand is adjoint to the identity map of $X(V)$. It suffices to show that $\wedge \circ (\Phi^G, \dots, \Phi^G)$ takes $(\xi, \dots, \xi)$ to a map of orthogonal spectra that is surjective on every spectrum level. We will describe this in detail in the case of $k = 2$, that is, $(X, Y) \rightsquigarrow \Phi^G X \wedge \Phi^G Y$, but the other cases are similar.

The smash product commutes with colimits in each variable, and this gives a definition of $\Phi^G X \wedge \Phi^G Y$ as a colimit of a diagram with four terms. We rearrange this into a single coequalizer diagram and conclude that there is a natural levelwise surjection of spectra

$$\bigvee_{V', W' \subset U} F_{V'^G} X(V')^G \wedge F_{W'^G} Y(W')^G \to \Phi^G X \wedge \Phi^G Y$$

for all orthogonal G–spectra X and Y. Applying this construction to (ξ, ξ) gives a commuting square

$$\begin{tikzcd}
\bigvee_{V,W \subset U} \Phi^G F_V X(V) \wedge \Phi^G F_W Y(W) \arrow[r, "\Phi^G(\xi) \wedge \Phi^G(\xi)"] & \Phi^G X \wedge \Phi^G Y \\
Z \arrow[u, two heads] \arrow[r] & \bigvee_{V', W' \subset U} F_{V'^G} X(V')^G \wedge F_{W'^G} Y(W')^G \arrow[u, two heads]
\end{tikzcd}$$

where

$$Z = \bigvee_{V', W', V, W \subset U} F_{V'^G} [\mathscr{J}_G(V, V') \wedge X(V)]^G \wedge F_{W'^G} [\mathscr{J}_G(W, W') \wedge Y(W)]^G,$$

in which the vertical maps are levelwise surjections. We wish to show $\Phi^G(\xi) \wedge \Phi^G(\xi)$ is surjective, and for this it suffices to show that the bottom horizontal map is surjective. This follows by examining the summands where $V = V'$ and $W = W'$, and noting that the action map $O(V)_+ \wedge X(V) \to X(V)$ is surjective on the G–fixed points. (Alternatively, one can show that the top horizontal and right vertical maps may be identified by a homeomorphism.) □

As a result, we get new rigidity statements for the maps relating geometric fixed points and smash powers:

Theorem 3.20 *Let X and Y denote arbitrary G–spectra. Then the commutation map*

$$\Phi^G X \wedge \Phi^G Y \xrightarrow{\alpha} \Phi^G(X \wedge Y)$$

is the only nonzero natural transformation from $\Phi^G X \wedge \Phi^G Y$ to $\Phi^G(X \wedge Y)$.

Remark 3.21 If X and Y are G–spectra and $H \le G$, then there is more than one natural map

$$\Phi^H X \wedge \Phi^H Y \to \Phi^H(X \wedge Y).$$

Indeed, we could take any element g in the center $Z(G)$, and postcompose α_H with the map $\mathcal{I}^U_{\mathbb{R}^\infty} g$ that acts on the trivial-representation levels by the action of g. However, α_H is the only natural transformation that respects the forgetful functor to H–spectra. In other words, it is the only one that is natural with respect to all of the H–equivariant maps of spectra, and not just the G–equivariant ones. Similar considerations apply to the iterated fixed points map below.

Theorem 3.22 *Let G be a finite group and let X denote an arbitrary H–spectrum with $H \le G$. Then the Hill–Hopkins–Ravenel diagonal map*

$$\Phi^H X \xrightarrow{\Delta} \Phi^G N_H^G X$$

is the only such map that is both natural and nonzero.

Theorem 3.23 *If X is a G–spectrum and $N \le G$ is a normal subgroup, then the iterated fixed points map*

$$\Phi^G X \xrightarrow{\mathrm{it}} \Phi^{G/N} \Phi^N X$$

is characterized by the property that it is natural in X and nonzero.

We end with five more corollaries, which served as the motivation for the rigidity result. The first corollary is the most important for our work on tensors and duals of cyclotomic spectra.

Proposition 3.24 *If X and Y are a G–spectra and $N \le G$ is a normal subgroup, then the following rectangle commutes:*

$$\begin{tikzcd}
\Phi^G X \wedge \Phi^G Y \arrow[rr, "\alpha_G"] \arrow[d, "\mathrm{it} \wedge \mathrm{it}"] & & \Phi^G(X \wedge Y) \arrow[d, "\mathrm{it}"] \\
\Phi^{G/N}\Phi^N X \wedge \Phi^{G/N}\Phi^N Y \arrow[r, "\alpha_{G/N}"] & \Phi^{G/N}(\Phi^N X \wedge \Phi^N Y) \arrow[r, "\Phi^{G/N}\alpha_N"] & \Phi^{G/N}\Phi^N(X \wedge Y)
\end{tikzcd}$$

The next two corollaries help us simplify and clarify the theory of cyclic orthogonal spectra.

Proposition 3.25 (see also [2, Lemma 4.5]) *If X is a G–spectrum and $g \in Z(G)$, then multiplication by g on the trivial representation levels gives a map of $\mathscr{J}_G$–spaces*

$$X \xrightarrow{\mathcal{I}_{\mathbb{R}^\infty}^U g} X$$

which on fixed points,

$$\Phi^G X \xrightarrow{\Phi^G \mathcal{I}_{\mathbb{R}^\infty}^U g} \Phi^G X,$$

is the identity map.

Proposition 3.26 *If X and Y are orthogonal spectra, then the self-map of orthogonal C_r–spectra*

$$f\colon N^{C_r}(X \wedge Y) \cong X^{\wedge r} \wedge Y^{\wedge r} \to X^{\wedge r} \wedge Y^{\wedge r}$$

which rotates only the Y factors but not the X factors fits into a commuting triangle:

$$\begin{tikzcd} & \Phi^{C_r}(X^{\wedge r} \wedge Y^{\wedge r}) \arrow[d, "\Phi^{C_r} \mathcal{I}_{\mathbb{R}^\infty}^U f"] \\ X \wedge Y \arrow[ur, "\Delta"] \arrow[dr, "\Delta"'] & \\ & \Phi^{C_r}(X^{\wedge r} \wedge Y^{\wedge r}) \end{tikzcd}$$

The next corollary requires more explanation. Let X be an orthogonal spectrum, and consider the diagonal map

$$X^{\wedge m} \xrightarrow{\Delta_n} \Phi^{C_n}(X^{\wedge m})^{\wedge n}.$$

If we write $(X^{\wedge m})^{\wedge n}$ in lexicographical order

$$(X^{\wedge m}) \wedge (X^{\wedge m}) \wedge \cdots \wedge (X^{\wedge m}),$$

then there is an obvious C_{mn}–action which rotates the terms. This commutes with the action of the subgroup C_n, so it passes to a C_{mn}–action on the geometric fixed points. By Proposition 3.25, the subgroup C_n acts trivially, giving a C_m–action on the fixed points.

Proposition 3.27 *Under these conventions, Δ_n is C_m–equivariant.*

Proof Let g denote the generator of C_m and h the generator of C_{mn}. Since the diagonal is natural, Δ_n is equivariant with respect to the action of g, but with g acting on $(X^{\wedge m})^{\wedge n}$ by rotating each $X^{\wedge m}$ separately. If we apply g and then the inverse of h, the composite matches the description of the map f of Proposition 3.26. Therefore, $f^{-1} \circ \Delta_n = \Delta_n$, so

$$\Delta_n \circ g = g \circ \Delta_n = g \circ f^{-1} \circ \Delta_n = h \circ \Delta_n.$$

Therefore, Δ_n is C_m–equivariant. □

Remark 3.28 This argument generalizes: the diagonal map Δ_n commutes with any automorphism of $X^{\wedge mn}$ coming from a self-map of the C_n–set $C_m \times C_n$ that gives the identity on the quotient set C_m. In particular, C_{mn} may be identified with $C_m \times C_n$ as C_n–sets with quotient C_m.

Our final corollary will be the key ingredient for showing that the cyclotomic structure maps on the cyclic bar construction are compatible with each other.

Proposition 3.29 *If X is an ordinary spectrum and $m, n \geq 0$ then the following square commutes:*

$$\begin{tikzcd} X \arrow[r, "\Delta_{C_{mn}}"] \arrow[d, "\Delta_{C_m}"'] & \Phi^{C_{mn}} X^{\wedge mn} \arrow[d, "\mathrm{it}"] \\ \Phi^{C_m} X^{\wedge m} \arrow[r, "\Phi^{C_m}(\Delta_n)"] & \Phi^{C_m} \Phi^{C_n} X^{\wedge mn} \end{tikzcd}$$

Remark 3.30 It is reasonable to expect that Δ_n coincides with the generalized HHR diagonal

$$N^{C_{mn}/C_n} X \xrightarrow{\Delta_*} \Phi^{C_n} N^{C_{mn}} X$$

of [2, Proposition 2.19]. Of course the above proposition is true for Δ_* as well.

4 Cyclic orthogonal spectra and the cyclic bar construction

Now we will integrate the modern technology from Section 3 into the classical theory from Section 2. We prove a few more properties of cyclic and cocyclic orthogonal spectra that concern the genuinely equivariant structure. Then we describe the construction and properties of the cyclic bar construction in orthogonal spectra, expanding on the treatment in [2].

4.1 Equivariant properties of cyclic and cocyclic spectra

Let $X_\bullet$ be a cyclic orthogonal spectrum. Then $\mathrm{sd}_r X_\bullet$ is an r–cyclic orthogonal spectrum. At each simplicial level, $(\mathrm{sd}_r X)_{n-1}$ is an orthogonal spectrum with C_r–action generated by the n^{th} power of the cycle map t_{rn-1}^n. This commutes with all the face, degeneracy and cycle maps, making $\mathrm{sd}_r X_\bullet$ an r–cyclic object in orthogonal C_r–spectra. So we may take the geometric fixed points on each level separately.

Proposition 4.1 *If $X_\bullet$ is a cyclic spectrum then $\Phi^{C_r} \mathrm{sd}_r X_\bullet$ is naturally a cyclic spectrum, and there is a natural S^1–equivariant isomorphism*

$$|\Phi^{C_r} \mathrm{sd}_r X_\bullet| \cong \rho_r^* \Phi^{C_r} |X_\bullet|.$$

Proof Since geometric fixed points is a functor, we know that $\Phi^{C_r}\,\mathrm{sd}_r\,X_\bullet$ is at least an r–cyclic orthogonal spectrum. By Proposition 3.25, the n^{th} power of the cycle map t_{rn-1}^n acts trivially on the geometric fixed points. Therefore $\Phi^{C_r}\,\mathrm{sd}_r\,X_\bullet$ is actually a cyclic spectrum, ie it factors in a canonical way through the quotient functor $P_r\colon \mathbf{\Lambda}_r \to \mathbf{\Lambda}$.

Using $P_r\Phi^{C_r}\,\mathrm{sd}_r\,X_\bullet$ to denote $\Phi^{C_r}\,\mathrm{sd}_r\,X_\bullet$ as an r–cyclic spectrum, we have the equivariant isomorphisms

$$|\Phi^{C_r}\,\mathrm{sd}_r\,X_\bullet| \cong \rho_r^*|P_r\Phi^{C_r}\,\mathrm{sd}_r\,X_\bullet| \cong \rho_r^*\Phi^{C_r}|\mathrm{sd}_r\,X_\bullet| \cong \rho_r^*\Phi^{C_r}|X_\bullet|,$$

where the middle map is the canonical commutation of Φ^{C_r} with geometric realization. These are obtained from the maps of Proposition 2.17 applied to the term $F_{V^{C_r}}S^0 \wedge X(V)^{C_r}$ in the coequalizer system for $\Phi^{C_r}X$. They pass to the coequalizer because ρ_r^*, P_r, sd_r and geometric realization all commute with colimits. □

We already know (Proposition 2.4) that the realization functor $|X_\bullet|$ preserves weak equivalences when $X_\bullet$ is Reedy cofibrant. We will also need to know when $|X_\bullet|$ is cofibrant.

Proposition 4.2 *If $X_\bullet$ is a cyclic spectrum, X_{-1} is a cofibrant spectrum and each cyclic latching map $L_n^{\mathrm{cyc}}X \to X_n$ is a cofibration of C_{n+1}–spectra, then $|X_\bullet|$ is a cofibrant S^1–spectrum.*

Proof As in Proposition 2.10, we reduce to checking that the C_{n+1} orbits of a pushout-product of a C_{n+1}–cell of spectra and a free S^1–cell of spaces is an S^1–cofibration,

$$[(F_V(C_{n+1}/C_r\times\partial D^k)_+ \to F_V(C_{n+1}/C_r\times D^k)_+)\,\square\,(S^1\times\partial D^\ell \to S^1\times D^\ell)_+]_{C_{n+1}}.$$

Here V is any finite-dimensional C_r–representation. This simplifies to

$$[F_V(C_{n+1}/C_r)_+ \wedge_{C_{n+1}} S^1_+] \wedge (\partial D^{k+\ell} \to D^{k+\ell})_+.$$

It suffices to show the left-hand term is cofibrant as an S^1–spectrum, but it is obtained by applying the left Quillen functor $-\wedge_{C_{n+1}} S^1_+$ to the C_{n+1}–cofibrant object $F_V(C_{n+1}/C_r)_+$, so it is cofibrant. □

Next, let $X^\bullet$ be a cocyclic orthogonal spectrum. Then $\mathrm{sd}_r\,X^\bullet$ is an r–cocyclic orthogonal spectrum, and, by the same argument as above, $\Phi^{C_r}\,\mathrm{sd}_r\,X^\bullet$ is naturally a cocyclic orthogonal spectrum. As before, we get the string of equivariant maps

$$\mathrm{Tot}(\Phi^{C_r}\,\mathrm{sd}_r\,X^\bullet) \cong \rho_r^*\mathrm{Tot}(P_r\Phi^{C_r}\,\mathrm{sd}_r\,X^\bullet) \leftarrow \rho_r^*\Phi^{C_r}\mathrm{Tot}(\mathrm{sd}_r\,X^\bullet) \cong \rho_r^*\Phi^{C_r}\mathrm{Tot}(X^\bullet).$$

The middle map is the canonical commutation of Φ^{C_r} with totalization, but as one might expect, it is not an isomorphism.

Proposition 4.3 *There is a natural interchange map*

$$\Phi^{C_r}\operatorname{Tot}(Z^\bullet) \to \operatorname{Tot}(\Phi^{C_r} Z^\bullet)$$

for cosimplicial spectra with C_r–actions.

Proof The interchange map is given canonically by universal properties, using the shorthand diagram:

$$\begin{tikzcd}
 & \operatorname{Tot}\Phi^{C_r} \arrow[rr] & & \prod_k \Phi^{C_r} \arrow[rr, shift left] \arrow[rr, shift right] & & \prod_{k,\ell} \Phi^{C_r} \\
\Phi^{C_r}\operatorname{Tot} \arrow[ur, dashed] & & & & & \\
 & & & \prod_k \bigvee_V \arrow[uu] \arrow[rr, shift left] \arrow[rr, shift right] & & \prod_{k,\ell} \bigvee_V \arrow[uu] \\
\bigvee_V \operatorname{Tot} \arrow[uu] \arrow[rr] & & \bigvee_V \prod_k \arrow[ur] \arrow[rr, shift left] \arrow[rr, shift right] & & \bigvee_V \prod_{k,\ell} \arrow[ur] & \\
 & & & \prod_k \bigvee_{V,W} \arrow[uu, shift left] \arrow[uu, shift right] & & \\
\bigvee_{V,W} \operatorname{Tot} \arrow[uu, shift left] \arrow[uu, shift right] \arrow[rr] & & \bigvee_{V,W} \prod_k \arrow[uu, shift left] \arrow[uu, shift right] \arrow[ur] & & &
\end{tikzcd}$$

A diagram chase shows this is natural with respect to maps of cosimplicial spectra $Z^\bullet \to \tilde{Z}^\bullet$. □

Corollary 4.4 *If $X^\bullet$ is a cocyclic spectrum then $\Phi^{C_r}\operatorname{sd}_r X^\bullet$ is naturally a cocyclic spectrum, and there is a natural S^1–equivariant map*

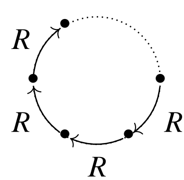
$$\rho_r^*\Phi^{C_r}\operatorname{Tot}(X^\bullet) \to \operatorname{Tot}(\Phi^{C_r}\operatorname{sd}_r X^\bullet).$$

4.2 The cyclic bar construction

Let R be an orthogonal ring spectrum. The *cyclic bar construction on R* is the cyclic spectrum $N_\bullet^{\mathrm{cyc}} R$ with

$$N_n^{\mathrm{cyc}} R = R^{\wedge(n+1)} = R^{\wedge n} \wedge \underline{R}.$$

We underline the last copy of R since in the simplicial structure it plays a special role. The action of $\mathbf{\Lambda}$ is best visualized by taking the category $[n]$ and labeling the arrows with copies of R:



Each map $[k] \to [n]$ induces a map $R^{\wedge(n+1)} \to R^{\wedge(k+1)}$ as follows. Each arrow $i \to i+1$ in $[k]$ is sent to some composition $j \to \cdots \to j+\ell$ in $[n]$, which corresponds to ℓ copies of R in $R^{\wedge(n+1)}$. We send this smash product $R^{\wedge\ell}$ to the copy of R in slot i of $R^{\wedge(k+1)}$, using the multiplication on R. When $\ell = 0$, we interpret this as the unit map $\mathbb{S} \to R$.

More generally, if $\boldsymbol{C}$ is a category enriched in orthogonal spectra, the cyclic nerve on $\boldsymbol{C}$ is defined as

$$N_n^{\mathrm{cyc}}\boldsymbol{C} = \bigvee_{c_0,\ldots,c_n \in \mathrm{ob}\,\boldsymbol{C}} \boldsymbol{C}(c_0,c_1) \wedge \boldsymbol{C}(c_1,c_2) \wedge \cdots \wedge \boldsymbol{C}(c_{n-1},c_n) \wedge \underline{\boldsymbol{C}(c_n,c_0)}.$$

One may think of these objects loosely as "functors" from $[k]$ into $\boldsymbol{C}$, where ordinary products have been substituted by smash products, and this suggests the correct face, degeneracy and cycle maps. In particular, as indicated below, the 0^{th} face map $d_0\colon N_n^{\mathrm{cyc}}\boldsymbol{C} \to N_{n-1}^{\mathrm{cyc}}\boldsymbol{C}$ switches the first term $\boldsymbol{C}(c_0,c_1)$ past the others and composes it into $\boldsymbol{C}(c_n,c_0)$. The extra degeneracy map $s_{n+1}\colon N_n^{\mathrm{cyc}}\boldsymbol{C} \to N_{n+1}^{\mathrm{cyc}}\boldsymbol{C}$ inserts a unit $\mathbb{S} \to \boldsymbol{C}(c_0,c_0)$ into the underlined factor in the smash product. The cycle map $t_n\colon N_n^{\mathrm{cyc}}\boldsymbol{C} \to N_n^{\mathrm{cyc}}\boldsymbol{C}$ rotates the factors towards the right:

$$\begin{aligned}
d_0\colon \boldsymbol{C}(c_0,c_1) \wedge \boldsymbol{C}(c_1,c_2) \wedge \cdots \wedge \underline{\boldsymbol{C}(c_n,c_0)} &\to \boldsymbol{C}(c_1,c_2) \wedge \cdots \wedge \underline{\boldsymbol{C}(c_n,c_1)},\\
s_{n+1}\colon \cdots \wedge \boldsymbol{C}(c_{n-1},c_n) \wedge \underline{\boldsymbol{C}(c_n,c_0)} \wedge \mathbb{S} &\to \cdots \wedge \boldsymbol{C}(c_{n-1},c_n) \wedge \boldsymbol{C}(c_n,c_0) \wedge \underline{\boldsymbol{C}(c_0,c_0)},\\
t_n\colon \boldsymbol{C}(c_0,c_1) \wedge \cdots \wedge \boldsymbol{C}(c_{n-1},c_n) \wedge \underline{\boldsymbol{C}(c_n,c_0)} &\\
&\to \boldsymbol{C}(c_n,c_0) \wedge \boldsymbol{C}(c_0,c_1) \wedge \cdots \wedge \underline{\boldsymbol{C}(c_{n-1},c_n)}.
\end{aligned}$$

If $\boldsymbol{C}$ has a single object, we recover the definition of $N^{\mathrm{cyc}}R$ we gave above.

Definition 4.5 The *topological Hochschild homology* of $\boldsymbol{C}$ is the geometric realization of the cyclic nerve

$$\mathrm{THH}(\boldsymbol{C}) := |N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|.$$

The cyclic bar construction of orthogonal spectra is remarkable because its geometric fixed points are isomorphic to the original spectrum.

Theorem 4.6 *If $\boldsymbol{C}$ is a spectral category then there are natural maps of S^1–spectra, for $r \geq 0$,*

$$\gamma_r\colon \mathrm{THH}(\boldsymbol{C}) \to \rho_r^*\Phi^{C_r}\mathrm{THH}(\boldsymbol{C}).$$

They are compatible in the following sense: if $T = \mathrm{THH}(\boldsymbol{C})$ then the square

$$\begin{tikzcd}
T \arrow[r, "\gamma_{mn}"] \arrow[d, "\gamma_m"'] & \rho_{mn}^*\Phi^{C_{mn}}T \arrow[d, "\mathrm{it}"] \\
\rho_m^*\Phi^{C_m}T \arrow[r, "\rho_m^*\Phi^{C_m}\gamma_n"] & \rho_m^*\Phi^{C_m}\rho_n^*\Phi^{C_n}T
\end{tikzcd}$$

strictly commutes. Furthermore, if every $\boldsymbol{C}(c_i, c_j)$ is a cofibrant orthogonal spectrum, then every γ_r is an isomorphism.

Remark 4.7 This extends one of the main results of [2] from ring spectra to spectral categories. This turns out to not be so difficult. However the treatment in [2] does not prove the above compatibility square, which seems to be harder. Our rigidity theorem allows us to check the compatibility easily.

Proof In essence, we need to understand the geometric fixed points of $\mathrm{THH}(\boldsymbol{C})$. We start with the isomorphism of S^1–spectra from Proposition 4.1:

$$|\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C}| \xrightarrow{\cong} \rho_r^*\Phi^{C_r}|N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|.$$

It thus suffices to understand the geometric fixed points of the subdivision $\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C}$. This is an r–cyclic spectrum. At simplicial level $n-1$ it is a wedge of smash products

$$\bigvee_{c_0,\dots,c_{rn-1}\in\mathrm{ob}\,\boldsymbol{C}} \boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge \underline{\boldsymbol{C}(c_{rn-1},c_0)}$$

and the C_r–action is by t_{rn-1}^n, which rotates this rn–fold smash product by n slots. In particular, the generator $\alpha\in C_r$ sends the summand A indexed by $c_0,\dots,c_{rn-1}$ to the summand $\alpha(A)$ indexed by

$$c_{(r-1)n},\dots,c_{rn-1},c_0,\dots,c_{(r-1)n-1}$$

by a homeomorphism. The summands A and $\alpha(A)$ coincide precisely when the list $c_0,\dots,c_{rn-1}$ repeats with period n:

$$c_0,c_1,\dots,c_{n-1},c_0,c_1,\dots,c_{n-1},c_0,c_1,\dots,c_{n-1}.$$

If this is not the case, then the C_r–closure $\widetilde{A}$ of A does not have any levelwise C_r–fixed points: $\widetilde{A}(V)^{C_r}=*$. This is because any fixed point would have in its C_r–orbit a point $x\in A$, but then x must be in the intersection $A\cap\alpha(A)=*$.

It is therefore a good idea to write $Y=\mathrm{sd}_r\,N_{n-1}^{\mathrm{cyc}}\boldsymbol{C}$ as the wedge of two spectra $X\vee X'$, where X is the wedge of those summands A such that $A=\alpha(A)$, and X' contains the remaining summands. Since the levelwise fixed point functor $(-)(V)^{C_r}$ preserves wedge sums, we immediately conclude that the inclusion $X\to Y$ induces a homeomorphism on each level $X(V)^{C_r}\cong Y(V)^{C_r}$. Recalling the definition of Φ^{C_r} (Definition 3.11), we conclude that the inclusion also induces an isomorphism on the geometric fixed points $\Phi^{C_r}X\cong\Phi^{C_r}Y$.

In conclusion, the geometric fixed points of the subdivision can be rewritten as

$$\begin{aligned}\Phi^{C_r}\,\mathrm{sd}_r\,N_{n-1}^{\mathrm{cyc}}\boldsymbol{C} &\cong \Phi^{C_r}\big(\textstyle\bigvee_{c_0,\dots,c_{n-1}}(\boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge\boldsymbol{C}(c_{n-1},c_0))^{\wedge r}\big)\\ &\cong \textstyle\bigvee_{c_0,\dots,c_{n-1}}\Phi^{C_r}(\boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge\boldsymbol{C}(c_{n-1},c_0))^{\wedge r}.\end{aligned}$$

It remains to compare this last term to $N^{\mathrm{cyc}}_{n-1}\boldsymbol{C}$ using Hill–Hopkins–Ravenel norm diagonal

$$\boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge\boldsymbol{C}(c_{n-1},c_0)\xrightarrow{\Delta}\Phi^{C_r}(\boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge\boldsymbol{C}(c_{n-1},c_0))^{\wedge r}.$$

We want to show that these diagonal maps for each $n\geq 1$ assemble into a map of cyclic spectra

$$N^{\mathrm{cyc}}_\bullet\boldsymbol{C}\xrightarrow{\Delta}\Phi^{C_r}\operatorname{sd}_r N^{\mathrm{cyc}}_\bullet\boldsymbol{C}$$

(see [2, Definition 4.6]). It easily commutes with most of the face and degeneracy maps because the diagonal is natural. One runs into issues with d_0 and t_{rn-1}, but these are fixed by the argument we used in Proposition 3.27. In brief, the r–fold smash $(d_0)^{\wedge r}$ of d_0 from the cyclic structure is not the same map as d_0 in the r–cyclic structure, but they differ by the map

$$f\colon(\boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge\boldsymbol{C}(c_{n-1},c_0))^{\wedge r}\to(\boldsymbol{C}(c_0,c_1)\wedge\cdots\wedge\boldsymbol{C}(c_{n-1},c_0))^{\wedge r}$$

that takes the factors $\boldsymbol{C}(c_{n-1},c_0)$ and cycles them while leaving all the other terms fixed. It suffices to show that f commutes with Δ, but we did that in Proposition 3.26. A similar argument works for t_{rn-1}.

This proves that the Hill–Hopkins–Ravenel diagonal gives a map of cyclic spectra. We define γ_r to be its geometric realization, combined with the S^1–equivariant isomorphism of Proposition 4.1:

$$|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}|\xrightarrow{|\Delta_r|}|\Phi^{C_r}\operatorname{sd}_r N^{\mathrm{cyc}}_\bullet\boldsymbol{C}|\xrightarrow{\cong}\rho_r^*\Phi^{C_r}|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}|.$$

When all the $\boldsymbol{C}(c_i,c_{i+1})$ are cofibrant, γ_r is a realization of isomorphisms at each level, so γ_r is an isomorphism.

Now we check compatibility. The compatibility square may be expanded and subdivided:

$$\begin{tikzcd}
|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[r, "\Delta_{mn}"] \arrow[d, "\Delta_m"] & \Phi^{C_{mn}}|\operatorname{sd}_{mn}N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[r, "\Phi^{C_{mn}}D_{mn}", "\cong"'] \arrow[d, "\mathrm{it}"] & \Phi^{C_{mn}}|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[d, "\mathrm{it}"] \\
\Phi^{C_m}|\operatorname{sd}_m N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[d, "\Phi^{C_m}D_m", "\cong"'] & \Phi^{C_m}\Phi^{C_n}|\operatorname{sd}_{mn}N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[r, "\Phi^{C_m}\Phi^{C_n}D_{mn}", "\cong"'] \arrow[d, "\Phi^{C_m}\Phi^{C_n}D_m", "\cong"'] & \Phi^{C_m}\Phi^{C_n}|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[d, equal] \\
\Phi^{C_m}|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[r, "\Phi^{C_m}\Delta_n"] & \Phi^{C_m}\Phi^{C_n}|\operatorname{sd}_n N^{\mathrm{cyc}}_\bullet\boldsymbol{C}| \arrow[r, "\Phi^{C_m}\Phi^{C_n}D_n", "\cong"'] & \Phi^{C_m}\Phi^{C_n}|N^{\mathrm{cyc}}_\bullet\boldsymbol{C}|
\end{tikzcd}$$

The top-right square commutes by naturality of the iterated fixed points map, and the bottom-right commutes by Proposition 2.17. The left-hand rectangle is subtle, so we expand and subdivide it once more:

$$\begin{tikzcd}
{|N^{\mathrm{cyc}}_\bullet \boldsymbol{C}|} \arrow[rrr, "\Delta_{mn}"] \arrow[d, "\Delta_m"] & & & \Phi^{C_{mn}}|\mathrm{sd}_m\,\mathrm{sd}_n\,N^{\mathrm{cyc}}_\bullet \boldsymbol{C}| \arrow[d, "\mathrm{it}"] \\
\Phi^{C_m}|\mathrm{sd}_m\,N^{\mathrm{cyc}}_\bullet \boldsymbol{C}| \arrow[r, "\Phi^{C_m}\mathrm{sd}_m\Delta_n"] \arrow[d, "\cong"', "\Phi^{C_m}D_m"] & \Phi^{C_m}|\mathrm{sd}_m\,\Phi^{C_n}\,\mathrm{sd}_n\,N^{\mathrm{cyc}}_\bullet \boldsymbol{C}| \arrow[rr, "\mathrm{int}", "\cong"'] \arrow[d, "\cong"', "\Phi^{C_m}D_m"] & & \Phi^{C_m}\Phi^{C_n}|\mathrm{sd}_m\,\mathrm{sd}_n\,N^{\mathrm{cyc}}_\bullet \boldsymbol{C}| \arrow[d, "\cong"', "\Phi^{C_m}\Phi^{C_n}D_m"] \\
\Phi^{C_m}|N^{\mathrm{cyc}}_\bullet \boldsymbol{C}| \arrow[r, "\Phi^{C_m}\Delta_n"] & \Phi^{C_m}\Phi^{C_n}|\mathrm{sd}_n\,N^{\mathrm{cyc}}_\bullet \boldsymbol{C}| \arrow[rr, equal] & & \Phi^{C_m}\Phi^{C_n}|\mathrm{sd}_n\,N^{\mathrm{cyc}}_\bullet \boldsymbol{C}|
\end{tikzcd}$$

The bottom-left square inside commutes by naturality of D_m. The interchange map "int" is the obvious identification of the two cyclic spectra, which at simplicial level $k-1$ are both given by $\Phi^{C_m}\Phi^{C_n}N^{\mathrm{cyc}}_{mnk-1}\boldsymbol{C}$. The lower-right square then easily commutes, and the remaining rectangle commutes by Proposition 3.29. □

In order to do homotopy theory, we need to know which maps $\boldsymbol{C} \to \boldsymbol{D}$ are sent to weak equivalences $\mathrm{THH}(\boldsymbol{C}) \to \mathrm{THH}(\boldsymbol{D})$, and we need conditions guaranteeing that $\mathrm{THH}(\boldsymbol{C})$ will be cofibrant. By our work above, this reduces to a calculation of the latching maps and cyclic latching maps. Let $\boldsymbol{S}$ denote the initial spectrally enriched category on the objects of $\boldsymbol{C}$:

$$\boldsymbol{S}(c_i, c_j) = \begin{cases} \mathbb{S}, & c_i = c_j, \\ *, & c_i \neq c_j. \end{cases}$$

The latching maps of the cyclic bar construction can be described concisely in terms of the canonical functor $\boldsymbol{S} \to \boldsymbol{C}$.

Proposition 4.8 *For every $n \geq 0$ the latching map $L_n N^{\mathrm{cyc}}\boldsymbol{C} \to N^{\mathrm{cyc}}_n\boldsymbol{C}$ is the wedge of pushout-products*

$$\bigvee_{c_0,\dots,c_n \in \mathrm{ob}\,\boldsymbol{C}} (\boldsymbol{S}(c_0,c_1) \to \boldsymbol{C}(c_0,c_1)) \,\square \cdots \square\, (\boldsymbol{S}(c_{n-1},c_n) \to \boldsymbol{C}(c_{n-1},c_n)) \\ \square\, (* \to \underline{\boldsymbol{C}(c_n,c_0)})$$

and the cyclic latching map $L^{\mathrm{cyc}}_n N^{\mathrm{cyc}}_\bullet\boldsymbol{C} \to N^{\mathrm{cyc}}_n\boldsymbol{C}$ is the wedge of pushout-products

$$\bigvee_{c_0,\dots,c_n \in \mathrm{ob}\,\boldsymbol{C}} (\boldsymbol{S}(c_0,c_1) \to \boldsymbol{C}(c_0,c_1)) \,\square \cdots \square\, (\underline{\boldsymbol{S}(c_n,c_0)} \to \underline{\boldsymbol{C}(c_n,c_0)}).$$

Proof One proves by induction that the pushout-product of $n+1$ different maps $f_0\colon A_0 \to X_0, \dots, f_n\colon A_n \to X_n$ comes from a cube-shaped diagram indexed by the subsets $S \subseteq \{0,\dots,n\}$ and inclusions. Each S is assigned to the smash product of those A_i for $i \notin S$ and X_i for $i \in S$. The pushout-product $f_0 \,\square \cdots \square\, f_n$ is then the map that includes into the final vertex the colimit of the remaining vertices.

Therefore it suffices to identify the cube for the pushout-product with the cube from Proposition 2.7 for the n^{th} cyclic latching object L_n^{cyc}. Each cube sends $S \subseteq \{0, \dots, n\}$ to a smash product in which the smash summand for (c_{i-1}, c_i) is $\boldsymbol{C}(c_{i-1}, c_i)$ if $i \in S$ and $\boldsymbol{S}(c_{i-1}, c_i)$ if $i \notin S$. In the pushout-product cube, the map induced by the inclusion $S \subseteq T$ is a smash product of $\boldsymbol{S}(c_{i-1}, c_i) \to \boldsymbol{C}(c_{i-1}, c_i)$ for each $i \in T - S$, together with the identity map on $\boldsymbol{S}(c_{i-1}, c_i)$ for $i \notin T$ and $\boldsymbol{C}(c_{i-1}, c_i)$ for $i \in S$. But this is the same as the map in the cyclic latching cube, because the rounding down map $T \to S$ preserves every arrow which ends in S and squashes the rest, so in the cyclic structure this induces a map that includes the unit for every arrow not ending in S and preserves the rest. Therefore the two cubes coincide. Restricting attention to subsets S containing 0 gives the cube for the simplicial latching object, giving a pushout-product in which the last factor is always $\underline{\boldsymbol{C}(c_n, c_0)}$. □

Remark 4.9 We have claimed that the 0^{th} cyclic latching map is the wedge of unit maps $\iota\colon \boldsymbol{S}(c,c) \to \boldsymbol{C}(c,c)$. In general, this is not quite correct — it is actually the wedge of inclusions of the images of these unit maps. However the inclusion of the image of ι is still a pushout of ι, so it does not matter which one we use in the latching square from Proposition 2.8.

The previous proposition suggests that we need a very weak cofibrancy assumption on $\boldsymbol{C}$ to guarantee that $\mathrm{THH}(\boldsymbol{C})$ is well behaved.

Definition 4.10 $\boldsymbol{C}$ is *cofibrant* if every map $\boldsymbol{S}(c_i, c_j) \to \boldsymbol{C}(c_i, c_j)$ is a cofibration of orthogonal spectra. Equivalently, every $\boldsymbol{C}(c_i, c_j)$ is a cofibrant orthogonal spectrum.

Proposition 4.11 *If $\boldsymbol{C}$ is cofibrant then $|N_\bullet^{\text{cyc}} \boldsymbol{C}|$ is a cofibrant S^1–spectrum. Moreover the inclusion of each cyclic skeleton into the next is a cofibration of S^1–spectra.*

Proof By Proposition 4.2, it suffices to show that the cyclic latching map from Proposition 4.8

$$\bigvee_{c_0,\dots,c_{n-1} \in \mathrm{ob}\,\boldsymbol{C}} (\boldsymbol{S}(c_0, c_1) \to \boldsymbol{C}(c_0, c_1)) \,\square \cdots \square\, (\underline{\boldsymbol{S}(c_{n-1}, c_0)} \to \underline{\boldsymbol{C}(c_{n-1}, c_0)})$$

is a C_n–cofibration of spectra. We restrict to one wedge summand at a time and consider its C_n–orbit. If there is no periodicity in the objects $c_0, \dots, c_{n-1}$ then the orbit is of the form $(C_n)_+$ smashed with a pushout-product of cofibrations, so it is automatically a C_n–cofibration. When there is r–fold periodicity, the problem instead reduces to showing that an r–fold pushout-product of a single cofibration f of orthogonal spectra becomes a C_r–cofibration $f^{\square r}$. Since $\square$ preserves retracts, it suffices to show that if f is a cell complex of orthogonal spectra then $f^{\square r}$ is a cell complex of orthogonal C_r–spectra.

In fact, it is a cell complex of orthogonal Σ_r–spectra. The argument for this is tedious but very formal. It holds because the categories of orthogonal G–spectra with varying G satisfy the following assumptions: the domains of our G–cells are small with respect to relative cell complexes; $\wedge$ commutes with colimits in each variable; a pushout-product of an H–cell with a K–cell is a coproduct of $H \times K$–cells; the operation $G \wedge_H -$ takes H–cells to G–cells; restriction of group actions takes G–cells to H–cell complexes; and the n–fold pushout-product of a single cell is a Σ_n–cell complex. This last assumption can be observed for orthogonal spectra by combining the space-level argument (eg [25, 3.4]) with the fact that an n–fold smash power of a free spectrum $F_{\mathbb{R}^m} A$ is isomorphic to $F_{\oplus^n \mathbb{R}^m} A^{\wedge n}$ as a Σ_n–spectrum. □

Proposition 4.12 *If $\boldsymbol{C}$ and $\boldsymbol{D}$ are cofibrant, and $\boldsymbol{C} \to \boldsymbol{D}$ is a pointwise weak equivalence which is the identity on objects, then it induces an $\mathcal{F}$–equivalence of S^1–spectra $|N_\bullet^{\mathrm{cyc}} \boldsymbol{C}| \to |N_\bullet^{\mathrm{cyc}} \boldsymbol{D}|$* (see Definition 3.8).

Proof It is easy to check that $N_\bullet^{\mathrm{cyc}} \boldsymbol{C} \to N_\bullet^{\mathrm{cyc}} \boldsymbol{D}$ is a levelwise stable equivalence. By Proposition 4.8, both simplicial spectra are Reedy cofibrant, so the map of realizations is an equivalence of nonequivariant spectra. By Proposition 4.11, both of these realizations are cofibrant S^1–spectra, and by Theorem 4.6 each one is naturally equivalent its own geometric fixed points. It follows that the map of left-derived geometric fixed points

$$\Phi^{C_n} |N_\bullet^{\mathrm{cyc}} \boldsymbol{C}| \to \Phi^{C_n} |N_\bullet^{\mathrm{cyc}} \boldsymbol{D}|$$

is an equivalence for all $n \geq 1$. By Proposition 3.14, the map $|N_\bullet^{\mathrm{cyc}} \boldsymbol{C}| \to |N_\bullet^{\mathrm{cyc}} \boldsymbol{D}|$ is therefore an $\mathcal{F}$–equivalence. □

5 Tensors and duals of cyclotomic spectra

In this final section, we discuss how to tensor and dualize cyclotomic structures, and use this to prove Theorem 1.1.

5.1 A general framework for dualizing cyclotomic structures

Recall that a *cyclotomic spectrum* is an orthogonal S^1–spectrum T with compatible maps of S^1–spectra, for all $n \geq 1$,

$$c_n \colon \rho_n^* \Phi^{C_n} T \to T$$

for which the composite map

$$\rho_n^* \Phi^{C_n} (cT) \to \rho_n^* \Phi^{C_n} T \to T \tag{6}$$

is an $\mathcal{F}$–equivalence of S^1–spectra (Definition 3.8). Here c refers to cofibrant replacement in the stable model structure on orthogonal S^1–spectra; see Proposition 3.6. To be more specific about the compatibility, we require that for all $m, n \geq 1$ the square

$$\begin{tikzcd}
\rho_{mn}^* \Phi^{C_{mn}} X \arrow[r, "c_{mn}"] \arrow[d, "\mathrm{it}"] & X \\
\rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} X \arrow[r, "\rho_m^* \Phi^{C_m} c_n"] & \rho_m^* \Phi^{C_m} X \arrow[u, "c_m"']
\end{tikzcd}$$

commutes. The left vertical is the canonical iterated fixed points map described in [8, Proposition 2.4], and it is an isomorphism when X is cofibrant as an S^1–spectrum.

A *precyclotomic spectrum* has all the same structure except that the map (6) need not be an equivalence. An *op-precyclotomic spectrum* has the above structure, but every map has the opposite direction, except for the iterated fixed points map.

In contrast to this, we give a more restrictive definition:

Definition 5.1 A *tight cyclotomic spectrum* is a cofibrant S^1–spectrum with isomorphisms $\gamma_n\colon T \xrightarrow{\cong} \rho_n^* \Phi^{C_n} T$ of S^1–spectra for all $n \geq 0$ compatible in the following way:

$$\begin{tikzcd}
T \arrow[r, "\gamma_{mn}", "\cong"'] \arrow[d, "\gamma_m", "\cong"'] & \rho_{mn}^* \Phi^{C_{mn}} T \arrow[d, "\mathrm{it}", "\cong"'] \\
\rho_m^* \Phi^{C_m} T \arrow[r, "\rho_m^* \Phi^{C_m} \gamma_n", "\cong"'] & \rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} T
\end{tikzcd}$$

Here "cofibrant" means in the stable model structure of Proposition 3.6. This implies that the geometric fixed points compute the left-derived geometric fixed points, ie the first map of (6) is always an equivalence. So a tight cyclotomic spectrum may be regarded as a cyclotomic spectrum by taking $c_n = \gamma_n^{-1}$ and forgetting that it is an isomorphism. We can summarize most of the previous section in a single theorem:

Theorem 5.2 *If R is an orthogonal ring spectrum which is cofibrant as an orthogonal spectrum, then $\mathrm{THH}(R)$ is a tight cyclotomic spectrum. If $\boldsymbol{C}$ is a cofibrant spectral category, then $\mathrm{THH}(\boldsymbol{C})$ is a tight cyclotomic spectrum.*

The point of these definitions is to dualize cyclotomic structures. Our first result is:

Proposition 5.3 *If T is a tight cyclotomic spectrum and T' is precyclotomic then the function spectrum $F(T, T')$ has a natural precyclotomic structure.*

Corollary 5.4 *If T is a tight cyclotomic spectrum then the functional dual $DT = F(T, \mathbb{S})$ is precyclotomic.*

Proof We define the structure map c_r as the composite

$$\rho_r^* \Phi^{C_r} F(T,T') \xrightarrow{\overline{\alpha}} F(\rho_r^* \Phi^{C_r} T, \rho_r^* \Phi^{C_r} T') \xrightarrow{F(\gamma_r, c_r)} F(T,T'),$$

where $\overline{\alpha}$ is the "restriction" map adjoint to

$$\rho_r^* \Phi^{C_r} F(T,T') \wedge \rho_r^* \Phi^{C_r} T \xrightarrow{\alpha} \rho_r^* \Phi^{C_r} (F(T,T') \wedge T) \to \rho_r^* \Phi^{C_r} T'$$

and α is the usual commutation of Φ^{C_r} with smash products. By the usual rules for equivariant adjunctions, c_r is automatically S^1–equivariant. We verify that these maps are compatible. Clearly they are natural in T and T', so in the diagram

$$\begin{tikzcd}
\rho_{mn}^* \Phi^{C_{mn}} F(T,T') \arrow[r, "\overline{\alpha}"] \arrow[d, "\mathrm{it}"] & F\left(\begin{array}{c}\rho_{mn}^* \Phi^{C_{mn}} T, \\ \rho_{mn}^* \Phi^{C_{mn}} T'\end{array}\right) \arrow[r, "F(\mathrm{id},\mathrm{it})"] & F\left(\begin{array}{c}\rho_{mn}^* \Phi^{C_{mn}} T, \\ \rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} T'\end{array}\right) \\
\rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} F(T,T') \arrow[r, "\Phi^{C_m}\overline{\alpha}"] & \rho_m^* \Phi^{C_m} F\left(\begin{array}{c}\rho_n^* \Phi^{C_n} T, \\ \rho_n^* \Phi^{C_n} T'\end{array}\right) \arrow[r, "\overline{\alpha}"] \arrow[d, "F(\Phi^{C_m}\gamma_n, \Phi^{C_m} c_n)"'] & F\left(\begin{array}{c}\rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} T, \\ \rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} T'\end{array}\right) \arrow[u, "F(\mathrm{it},\mathrm{id})", "\cong"'] \arrow[d, "F(\Phi_{C_m}\gamma_n, \Phi^{C_m} c_n)"'] \\
& \rho_m^* \Phi^{C_m} F(T,T') \arrow[r, "\overline{\alpha}"] & F(\rho_m^* \Phi^{C_m} T, \rho_m^* \Phi^{C_m} T') \arrow[d, "F(\gamma_m, c_m)"'] \\
& & F(T,T')
\end{tikzcd}$$

the small square automatically commutes. The left-most and right-most paths compose to give the two maps we are trying to compare. So, we just need to show that the big rectangle at the top commutes. It is adjoint to:

$$\begin{tikzcd}
\begin{array}{r}\rho_{mn}^* \Phi^{C_{mn}} F(T,T') \\ \wedge \rho_{mn}^* \Phi^{C_{mn}} T\end{array} \arrow[r, "\alpha"] \arrow[d, "\mathrm{it} \wedge \mathrm{it}"] & \rho_{mn}^* \Phi^{C_{mn}} (F(T,T') \wedge T) \arrow[r] \arrow[d, "\mathrm{it}"] & \rho_{mn}^* \Phi^{C_{mn}} T' \arrow[d, "\mathrm{it}"] \\
\begin{array}{r}\rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} F(T,T') \\ \wedge \rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} T\end{array} \arrow[r, "\alpha \circ \alpha"] & \rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} (F(T,T') \wedge T) \arrow[r] & \rho_m^* \Phi^{C_m} \rho_n^* \Phi^{C_n} T'
\end{tikzcd}$$

The right square is by naturality of the iterated fixed points map, and the left square is by Proposition 3.24. □

We have chosen to state these results for tight cyclotomic spectra, because then every object we work with has precyclotomic structure, as opposed to a mix of objects with precyclotomic and op-precyclotomic structure. If we freely allow ourselves to use both structures then the we get the following more general conclusion. It tells us that we have something close to, but not quite, a closed symmetric monoidal category of spectra with these structures.

Proposition 5.5 *If X and Y are op-precyclotomic spectra and Z is a precyclotomic spectrum then $X \wedge Y$ is op-precyclotomic, $F(Y,Z)$ is precyclotomic and the adjunction*

$$F(X \wedge Y, Z) \cong F(X, F(Y,Z))$$

respects the precyclotomic structure.

Proof The above proof generalizes to show that $F(Y,Z)$ is precyclotomic, since we only used the maps γ_n for Y and c_n for Z. For $X \wedge Y$ we define the op-cyclotomic structure by

$$X \wedge Y \xrightarrow{\gamma_n \wedge \gamma_n} \Phi^{C_n} X \wedge \Phi^{C_n} Y \xrightarrow{\alpha} \Phi^{C_n}(X \wedge Y),$$

where the ρ_n^* are suppressed. By an easy diagram chase, the compatibility reduces again to Proposition 3.24. When we check that the adjunction preserves the cyclotomic structures, we reduce to the claim that the interchange map α has an associativity property. This can be proven from the definitions with a little bit of work, but it also follows effortlessly from the rigidity theorem. □

This analysis does not quite apply to the categories of precyclotomic or cyclotomic spectra, because we get zigzags when we try to define a cyclotomic structure on their tensor product. However this problem goes away if we restrict attention to cofibrant objects, so we can draw a conclusion about the homotopy category:

Proposition 5.6 *The homotopy categories of precyclotomic spectra and of cyclotomic spectra from [8] have a tensor triangulated structure.*

Proof For simplicity we suppress ρ_n^*. If X and Y are cofibrant (pre)cyclotomic spectra, we make $X \wedge Y$ into a (pre)cyclotomic spectrum using the structure maps

$$\Phi^{C_n}(X \wedge Y) \underset{\cong}{\xleftarrow{\alpha}} \Phi^{C_n} X \wedge \Phi^{C_n} Y \xrightarrow{c_n \wedge c_n} X \wedge Y.$$

The relevant compatibility square is:

$$\begin{tikzcd}
\Phi^{C_{mn}}(X\wedge Y) \arrow[dd, "\mathrm{it}"] & \Phi^{C_{mn}}X\wedge\Phi^{C_{mn}}Y \arrow[l, "\alpha"', "\cong"] \arrow[d, "\mathrm{it}\wedge\mathrm{it}"] \arrow[r, "c_{mn}\wedge c_{mn}"] & X\wedge Y \\
& \Phi^{C_m}\Phi^{C_n}X\wedge\Phi^{C_m}\Phi^{C_n}Y \arrow[d, "\alpha"] \arrow[r, "\Phi^{C_m}c_n\wedge\Phi^{C_m}c_n"] & \Phi^{C_m}X\wedge\Phi^{C_m}Y \arrow[u, "c_m\wedge c_m"'] \arrow[d, "\alpha"', "\cong"] \\
\Phi^{C_m}\Phi^{C_n}(X\wedge Y) & \Phi^{C_m}(\Phi^{C_n}X\wedge\Phi^{C_n}Y) \arrow[l, "\Phi^{C_m}\alpha"', "\cong"] \arrow[r, "\Phi^{C_m}(c_n\wedge c_n)"] & \Phi^{C_m}(X\wedge Y)
\end{tikzcd}$$

Again Proposition 3.24 gives us the left-hand rectangle, the top-right square is the smash product of the compatibility squares for X and Y, and the bottom-right commutes by naturality of α. It is straightforward to check that this smash product preserves colimits and cofibers of (pre)cyclotomic spectra, so this gives the desired tensor triangulated structure on the homotopy category. □

Remark 5.7 The analogue of this theorem for p–precyclotomic spectra and p–cyclotomic spectra is also true, and it is much easier.

Returning to the precyclotomic structure on $F(T,T')$, our main example of interest will be when $T=|N^{\mathrm{cyc}}\boldsymbol{C}|$ is the cyclic nerve of a ring or category. We have just proven that $F(|N^{\mathrm{cyc}}\boldsymbol{C}|,T')$ is a precyclotomic spectrum. It is also the totalization of the cocyclic S^1–spectrum

$$Y^k=F(N_k^{\mathrm{cyc}}\boldsymbol{C},T').$$

To be precise, the S^1 is acting only on the T', and $\boldsymbol{\Lambda}$ is acting by the dual of the $\boldsymbol{\Lambda}^{\mathrm{op}}$ action on $N_k^{\mathrm{cyc}}\boldsymbol{C}$. This puts two commuting S^1–actions on the totalization, but we restrict attention to the diagonal S^1–action, because this is the action that agrees with the precyclotomic structure we just defined.

In order to compare this to the cocyclic spectrum $\Sigma^\infty_+X^{\bullet+1}$, we will need to describe our cyclotomic structure maps using only the cocyclic structure on $F(|N^{\mathrm{cyc}}\boldsymbol{C}|,T')$:

Proposition 5.8 *The cyclotomic structure map on* $\mathrm{Tot}(Y^\bullet)\cong F(|N^{\mathrm{cyc}}\boldsymbol{C}|,T')$ *is equal to the composite of* S^1*–equivariant maps*

$$\begin{aligned}
\rho_r^*\Phi^{C_r}\mathrm{Tot}(F(N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T'))&\xrightarrow{D_r}\rho_r^*\Phi^{C_r}\mathrm{Tot}(F(\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T'))\\
&\to\rho_r^*\mathrm{Tot}(\Phi^{C_r}F(\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T'))\xrightarrow{\bar\alpha}\rho_r^*\mathrm{Tot}(F(P_r\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},\Phi^{C_r}T'))\\
&\xrightarrow{\cong}\mathrm{Tot}(F(\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},\rho_r^*\Phi^{C_r}T'))\xrightarrow{F(\Delta,c_r)}\mathrm{Tot}(F(N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T')),
\end{aligned}$$

where the undecorated map is the interchange of Proposition 4.3.

Proof We compare to the structure map we defined above:

$$
\begin{tikzcd}
\rho_r^*\Phi^{C_r}F(|N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,T') \arrow[rr, "\cong"] \arrow[d, "\bar\alpha"] \arrow[dr, "D_r", "\cong"'] & & \rho_r^*\Phi^{C_r}\mathrm{Tot}(F(N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T')) \arrow[d, "D_r"] \\
F\left(\substack{\rho_r^*\Phi^{C_r}|N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,\\ \rho_r^*\Phi^{C_r}T'}\right) \arrow[dr, "D_r", "\cong"'] & \rho_r^*\Phi^{C_r}F(|\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,T') \arrow[r, "\cong"] \arrow[d, "\bar\alpha"] & \rho_r^*\Phi^{C_r}\mathrm{Tot}(F(\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T')) \arrow[d] \\
& F\left(\substack{\rho_r^*\Phi^{C_r}|\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,\\ \rho_r^*\Phi^{C_r}T'}\right) \arrow[d, "\cong"] & \rho_r^*\mathrm{Tot}(\Phi^{C_r}F(\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},T')) \arrow[d, "\bar\alpha"] \\
& F\left(\substack{\rho_r^*|P_r\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,\\ \rho_r^*\Phi^{C_r}T'}\right) \arrow[r, "\cong"] \arrow[d, "\cong"] & \rho_r^*\mathrm{Tot}\left(F\left(\substack{P_r\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},\\ \Phi^{C_r}T'}\right)\right) \arrow[d, "\cong"] \\
& F\left(\substack{|\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,\\ \rho_r^*\Phi^{C_r}T'}\right) \arrow[r, "\cong"] \arrow[d, "F(\Delta{,}c_r)"] & \mathrm{Tot}\left(F\left(\substack{\Phi^{C_r}\,\mathrm{sd}_r\,N_\bullet^{\mathrm{cyc}}\boldsymbol{C},\\ \rho_r^*\Phi^{C_r}T'}\right)\right) \arrow[d, "F(\Delta{,}c_r)"] \\
& F(|N_\bullet^{\mathrm{cyc}}\boldsymbol{C}|,T') \arrow[r, "\cong"] & \mathrm{Tot}(F(N_\bullet^{\mathrm{cyc}}\boldsymbol{C},\Phi^{C_r}T'))
\end{tikzcd}
$$

Most of these squares commute easily. The nontrivial one in the middle can be simplified to the following: if $X_\bullet$ is a simplicial C_r–spectrum and T is a C_r–spectrum then the middle rectangle of

$$
\begin{tikzcd}
F_{W^{C_r}}S^0\wedge \mathrm{Map}_*^{C_r}(|X_\bullet|,\mathrm{sh}^W T) \arrow[d] & \\
\Phi^{C_r}F(|X_\bullet|,T) \arrow[r, "\cong"] \arrow[d, "\bar\alpha"] & \Phi^{C_r}\mathrm{Tot}(F(X_\bullet,T)) \arrow[d] \\
F(\Phi^{C_r}|X_\bullet|,\Phi^{C_r}T) \arrow[d, "\cong"] & \mathrm{Tot}(\Phi^{C_r}F(X_\bullet,T)) \arrow[d, "\bar\alpha"] \\
F(|\Phi^{C_r}X_\bullet|,\Phi^{C_r}T) \arrow[r, "\cong"] & \mathrm{Tot}(F(\Phi^{C_r}X_\bullet,\Phi^{C_r}T)) \arrow[d] \\
& F(F_{V^{C_r}}S^0\wedge\Delta^k_+\wedge X_k(V)^{C_r},\Phi^{C_r}T)
\end{tikzcd}
$$

commutes. Here $\mathrm{sh}^W T$ is shorthand for the mapping spectrum $F(F_W S^0, T)$, which is used in the standard formula for level W of a mapping spectrum

$$F(|X_\bullet|, T)(W) \cong F(|X_\bullet|, \mathrm{sh}^W T)(0) \cong \mathrm{Map}_*(|X_\bullet|, \mathrm{sh}^W T).$$

Now, it suffices to show that the composites from the top to the bottom of our rectangle are identical for any C_r–representations V and W and any integer $k \geq 0$. For the left-hand branch it is easy to check this is adjoint to the composite:

$$\begin{array}{c}
F_{V^{C_r}} S^0 \wedge \Delta^k_+ \wedge X_k(V)^{C_r} \wedge F_{W^{C_r}} S^0 \wedge \mathrm{Map}_*^{C_r}(|X_\bullet|, \mathrm{sh}^W T) \\
\Big\downarrow \text{include into } \Phi^{C_r} \\
\Phi^{C_r}(\Delta^k_+ \wedge X_k) \wedge \Phi^{C_r} F(|X_\bullet|, T) \\
\Big\downarrow \text{include into } |X_\bullet| \\
\Phi^{C_r}|X_\bullet| \wedge \Phi^{C_r} F(|X_\bullet|, T) \\
\Big\downarrow \alpha \\
\Phi^{C_r}(|X_\bullet| \wedge F(|X_\bullet|, T)) \\
\Big\downarrow \Phi^{C_r}(\mathrm{ev}) \\
\Phi^{C_r} T
\end{array}$$

A careful trace through the diagram in Proposition 4.3 shows that the right-hand branch is the composite

$$\begin{aligned}
&F_{W^{C_r}} S^0 \wedge \mathrm{Map}_*^{C_r}(|X_\bullet|, \mathrm{sh}^W T) \xrightarrow{\text{restrict}} F_{W^{C_r}} S^0 \wedge \mathrm{Map}_*^{C_r}(\Delta^k_+ \wedge X_k, \mathrm{sh}^W T) \\
&\xrightarrow{\text{assembly}} F(\Delta^k_+, F_{W^{C_r}} S^0 \wedge \mathrm{Map}_*^{C_r}(X_k, \mathrm{sh}^W T)) \xrightarrow{\text{include}} F(\Delta^k_+, \Phi^{C_r} F(X_k, T)) \\
&\xrightarrow{\bar{\alpha}} F(\Delta^k_+, F(\Phi^{C_r} X_k, \Phi^{C_r} T)) \xrightarrow{\text{include}} F(\Delta^k_+, F(F_{V^{C_r}} S^0 \wedge X_k(V)^{C_r}, \Phi^{C_r} T)).
\end{aligned}$$

The adjoint of this map does indeed agree with the first, by a very long diagram chase. The essential ingredients are functoriality of Φ^{C_r}, naturality of α and ev, and associativity of α. □

Now we know that $F(T, T')$ has a precyclotomic structure. This won't be very useful unless we can make cofibrant and fibrant replacements of T and T', respectively, while preserving that structure. For this task, we use the model structure on cyclotomic and precyclotomic spectra defined in [8]. It has following attractive property:

Lemma 5.9 *If T is cofibrant or fibrant in the model* category on (pre)cyclotomic spectra, then it is also cofibrant or fibrant, respectively, as an orthogonal S^1–spectrum in the $\mathcal{F}$–model structure.*

Proof The fibrant part is true by definition. For the cofibrant part it suffices to check that the monad

$$\mathbb{C}X = \bigvee_{n\geq 1}\rho_n^*\Phi^{C_n}X$$

preserves cofibrant objects in the $\mathcal{F}$–model structure. This is true because wedge sums, geometric fixed points and change of groups all preserve cofibrations. □

In light of this fact, we can replace T' with a fibrant cyclotomic spectrum fT', resulting in the precyclotomic spectrum $F(T, fT')$, whose underlying S^1–spectrum has the homotopy type of the derived mapping spectrum from T to T' (ie the first input is cofibrant and the second input is fibrant). Specializing to $T' = \mathbb{S}$ gives a precyclotomic structure on the dual $F(T, f\mathbb{S})$.

Remark 5.10 If T is finite as a genuine S^1 spectrum, then $F(T, f\mathbb{S})$ is actually cyclotomic, not just precyclotomic. In general, however, this is not true. One can check that $T = \Sigma^\infty_+\mathbb{R}\mathbb{P}^\infty$ gives a counterexample. In the next section we will consider an example where T is infinite, but $F(T, f\mathbb{S})$ is still cyclotomic, mainly for reasons of connectivity.

5.2 The equivariant duality between THH(DX) and $\Sigma^\infty_+ LX$

Let X be a finite based CW complex and let $DX = F(X_+, \mathbb{S})$ denote its Spanier–Whitehead dual. Though $\mathbb{S}$ is not fibrant, X is compact, so DX has the correct homotopy type. It is also finite, of course, but it is no longer compact, and this slightly complicates our proof below.

DX is a commutative ring with multiplication given by the dual of the diagonal map of X. Likewise, the spectrum $\widetilde{D}X = F(X, \mathbb{S})$ has a commutative multiplication given by the dual of the smash diagonal $X \to X \wedge X$. It does not have a unit, but we can make $\mathbb{S} \vee \widetilde{D}X$ into a ring spectrum by having $\mathbb{S}$ act as the unit. The levelwise fiber sequence of spectra

$$F(X, \mathbb{S}) \to F(X_+, \mathbb{S}) \to \mathbb{S}$$

preserves the multiplications, and this allows us to form an equivalence of ring spectra

$$\mathbb{S} \vee \widetilde{D}X \xrightarrow{\sim} DX.$$

Let $c\widetilde{D}X$ denote cofibrant replacement of DX as a unitless ring, so that

$$cDX := \mathbb{S} \vee c\widetilde{D}X \to \mathbb{S} \vee \widetilde{D}X$$

is a particularly nice cofibrant replacement of ring spectra. We'll take as our example of a tight cyclotomic spectrum

$$T = \mathrm{THH}(cDX).$$

We recall that this cofibrant replacement ensures that $\mathrm{THH}(cD(-))$ is homotopy invariant (Proposition 4.12). Our starting point is the following consequence of Proposition 2.22. In its statement, we assume implicitly that cofibrant replacements are taken before each application of D or THH.

Theorem 5.11 [20; 13] *When X is a finite simply connected CW complex, there is an equivalence of spectra with an S^1–action*

$$D(\mathrm{THH}(DX)) \simeq \mathrm{THH}(\Sigma^\infty_+ \Omega X) \simeq \Sigma^\infty_+ LX$$

in which $LX = \mathrm{Map}(S^1, X)$ is the free loop space.

Remark 5.12 If M is a manifold then $DM \simeq M^{-TM}$ is a Thom spectrum. But the analysis of [7] does not apply, because the multiplication on M^{-TM} does not arise from the normal bundle $M \to BO$ being a loop map.

We will spend the rest of this section proving a more highly structured version of that result:

Theorem 5.13 *Let $f\mathbb{S}$ be a fibrant replacement of $\mathbb{S}$ as a cyclotomic spectrum. Then for every unbased space X there is a natural map of precyclotomic spectra*

$$\Sigma^\infty_+ LX \to F(\mathrm{THH}(cDX), f\mathbb{S}).$$

The left-hand side is always cyclotomic. When X is a finite simply connected CW complex, the right-hand side is cyclotomic and the map is an $\mathcal{F}$–equivalence.

Corollary 5.14 *When X is a finite simply connected CW complex, the equivalence between $\mathrm{THH}(\Sigma^\infty_+ \Omega X)$ and the functional dual of $\mathrm{THH}(DX)$ is an equivalence of cyclotomic spectra.*

Proof We will describe explicitly the map of Theorem 5.11 and check that it respects the precyclotomic structures. Then we will use connectivity arguments to argue that these precyclotomic spectra are actually cyclotomic when X is finite.

As above, let $Y^\bullet$ denote the cocyclic S^1–spectrum

$$Y^k = F(N^{\mathrm{cyc}}_k cDX, f\mathbb{S}) = F((cDX)^{\wedge(k+1)}, f\mathbb{S}).$$

The totalization of $Y^\bullet$ is isomorphic to $F(|N^{\mathrm{cyc}} cDX|, f\mathbb{S})$, and Proposition 5.8 gives us a recipe for the precyclotomic structure. Furthermore, $Y^\bullet$ is the dual of a Reedy cofibrant simplicial spectrum, and is therefore Reedy fibrant.

We will construct a map $\Sigma^\infty_+ LX \to \mathrm{Tot}(Y^\bullet)$ by going through an intermediary $\mathrm{Tot}(Z^\bullet)$. Let $Z^\bullet$ be the cocyclic spectrum $\Sigma^\infty_+ \mathrm{Map}(S^1_\bullet, X)$, so that

$$Z^k = \Sigma^\infty_+ \mathrm{Map}(\mathbf{\Lambda}([k],[0]), X) \cong \Sigma^\infty_+ X^{k+1}$$

with $\mathbf{\Lambda}$ action given by applying Σ^∞_+ to the usual $\mathbf{\Lambda}^{\mathrm{op}}$ action on the $\mathbf{\Lambda}(-,[0])$ term. The interchange of Proposition 2.22 gives a map of spectra

$$\Sigma^\infty_+ LX \to \mathrm{Tot}(Z^\bullet).$$

Next we construct a map of cocyclic spectra $Z^\bullet \to Y^\bullet$. The evaluation map composed with the product in $\mathbb{S}$ and fibrant replacement

$$(\Sigma^\infty_+ X)^{\wedge(k+1)} \wedge c(DX)^{\wedge(k+1)} \to (\Sigma^\infty_+ X)^{\wedge(k+1)} \wedge (DX)^{\wedge(k+1)} \to (\mathbb{S})^{\wedge(k+1)} \to \mathbb{S} \to f\mathbb{S}$$

is adjoint to a map

$$Z^k = \Sigma^\infty_+ X^{k+1} \to F((cDX)^{\wedge(k+1)}, f\mathbb{S}) = Y^k.$$

Of course, this map is actually an equivalence when X is finite. The map clearly commutes with the S^1–action on each level coming from $f\mathbb{S}$. We check that it commutes with the cocyclic structure: for each $\gamma \in \mathbf{\Lambda}([k],[\ell])$ we have the square

$$\begin{tikzcd}
\mathrm{Map}(\Lambda[0]_k, X) \cong X^{k+1} \arrow[r] \arrow[d, "\gamma"] & F((cDX)^{\wedge k+1}, f\mathbb{S}) \arrow[d, "\gamma"] \\
\mathrm{Map}(\Lambda[0]_\ell, X) \cong X^{\ell+1} \arrow[r] & F((cDX)^{\wedge \ell+1}, f\mathbb{S})
\end{tikzcd}$$

which commutes if this one commutes:

$$\begin{tikzcd}
X^{k+1} \wedge (cDX)^{\ell+1} \arrow[r, "\gamma\wedge\mathrm{id}"] \arrow[d, "\mathrm{id}\wedge\gamma"] & X^{\ell+1} \wedge (cDX)^{\ell+1} \arrow[d] \\
X^{k+1} \wedge (cDX)^{k+1} \arrow[r] & \mathbb{S}
\end{tikzcd}$$

Both branches have the same description: γ gives a map from a necklace with $k+1$ beads and every segment labeled by X to a necklace with $\ell+1$ beads and every segment labeled by DX. Each copy of X is sent by γ to a string of a copies of DX; we apply the diagonal to $X_+ \xrightarrow{\Delta} \left(\prod^a X\right)_+$ and pair with those a copies of DX.

Therefore we have a map of cocyclic S^1–spectra $Z^\bullet \to Y^\bullet$, with S^1 acting trivially on each cosimplicial level of $Z^\bullet$. Composing with the interchange map of Proposition 2.22 gives an S^1–equivariant map

$$\Sigma^\infty_+ LX \to \mathrm{Tot}(Z^\bullet) \to \mathrm{Tot}(Y^\bullet). \tag{7}$$

When X is finite, this is the equivalence of Theorem 5.11. In fact, when X is finite the map $Z^\bullet \to Y^\bullet$ is an equivalence on each cosimplicial level, and we may therefore consider $Y^\bullet$ to be a Reedy fibrant replacement of $Z^\bullet$, so (7) is also a model for the derived interchange map of Proposition 2.22.

Our next task is to check that the map (7) respects the precyclotomic structures on the two ends. The recipe in Proposition 5.8 actually defines a precyclotomic structure on $\mathrm{Tot}(Z^\bullet)$ as well, so our problem breaks up into two steps:

$$\begin{tikzcd}
\Phi^{C_r}\Sigma^\infty_+ LX \arrow[r, "\cong"] \arrow[dddd, "\cong"] & \rho_r^*\Phi^{C_r}\mathrm{Tot}(Z^\bullet) \arrow[r] \arrow[d, "\cong"', "D_r"] & \rho_r^*\Phi^{C_r}\mathrm{Tot}(Y^\bullet) \arrow[d, "\cong"', "D_r"] \\
& \rho_r^*\Phi^{C_r}\mathrm{Tot}(\mathrm{sd}_r Z^\bullet) \arrow[r] \arrow[d] & \rho_r^*\Phi^{C_r}\mathrm{Tot}(\mathrm{sd}_r Y^\bullet) \arrow[d] \\
& \rho_r^*\mathrm{Tot}(\Phi^{C_r}\mathrm{sd}_r Z^\bullet) \arrow[r] & \rho_r^*\mathrm{Tot}(\Phi^{C_r}\mathrm{sd}_r Y^\bullet) \arrow[d, "F(\Delta,c_r)\circ\overline{\alpha}"] \\
& & \\
\Sigma^\infty_+ LX \arrow[r, "\cong"] & \mathrm{Tot}(Z^\bullet) \arrow[r] \arrow[uu, "\cong", "\Delta"'] & \mathrm{Tot}(Y^\bullet)
\end{tikzcd} \tag{8}$$

We start with the left-hand rectangle of (8), where everything is a suspension spectrum and so all maps are completely determined by what they do at spectrum level 0. The horizontal homeomorphisms may be computed by observing that $\Lambda[0]_k = \mathbf{\Lambda}([k],[0])$ has $k+1$ points $f_0,\dots,f_k$, where $f_i\colon \mathbb{Z}\to\mathbb{Z}$ sends 0 through $i-1$ to 0 and i through k to 1 (or if $i=0$ it sends 0 through k to 0). Using our choice of homeomorphism $|\Lambda[0]| \cong \mathbb{R}/\mathbb{Z}$ from Section 2, the k–simplex given by f_i maps down to the circle $\mathbb{R}/\mathbb{Z}$ by the formula

$$(t_0,\dots,t_k) \mapsto (t_i+\cdots+t_k) \sim (1-(t_0+\cdots+t_{i-1})).$$

Negating the circle and reparametrizing $\Delta^k \subset \mathbb{R}^k$ as points $(x_1,\dots,x_k)$ for which $0 \le x_1 \le x_2 \le \cdots \le x_k \le 1$ according to the rule $x_i = t_0+\cdots+t_{i-1}$, we arrive at the simple rule

$$(f_i, x_1,\dots,x_k) \mapsto x_i, \qquad x_0 := 0.$$

So now the map $LX \to \mathrm{Tot}(X^{\bullet+1})$ can be expressed by the formula

$$\Delta^{k-1}\times LX \to X^k, \quad (r_1,\dots,r_{k-1},\gamma) \mapsto (\gamma(0),\gamma(r_1),\dots,\gamma(r_{k-1})),$$

as in [14]. Under this change of coordinates, both branches give

$$\gamma(-) \to (r_1,\dots,r_{k-1}) \mapsto \left(\gamma(0), \gamma\left(\frac{1}{r}r_1\right), \gamma\left(\frac{1}{r}r_2\right), \dots, \gamma\left(\frac{1}{r}r_{k-1}\right), \gamma(0), \gamma\left(\frac{1}{r}r_1\right), \dots\right)$$

and so the square commutes.

Returning to (8), the top and middle squares of the right-hand row automatically commute by the naturality of the cosimplicial diagonal and the interchange map with geometric fixed points. The final square is then:

$$\begin{tikzcd}
\mathrm{Tot}(\Phi^{C_r}\,\mathrm{sd}_r\, Z^\bullet) \arrow[r] & \rho_r^*\mathrm{Tot}(\Phi^{C_r}\,\mathrm{sd}_r\, Y^\bullet) \arrow[d, "F(\Delta{,}c_r)\circ\overline{\alpha}"] \\
\mathrm{Tot}(Z^\bullet) \arrow[r] \arrow[u, "\cong", "\Delta"'] & \mathrm{Tot}(Y^\bullet)
\end{tikzcd}$$

The map Δ is the cocyclic map

$$\Phi^{C_r}\Sigma^\infty_+ X^{rk} \xleftarrow{\cong} \Sigma^\infty_+ X^k$$

given by the Hill–Hopkins–Ravenel diagonal; this is almost tautologically cosimplicial. The map $F(\Delta,c_r)\circ\overline{\alpha}$ is also cocyclic, so to check that this square commutes it suffices to check level $k-1$. This boils down to this rectangle:

$$\begin{tikzcd}
\Phi^{C_r}\Sigma^\infty_+ X^{rk}\wedge\Phi^{C_r}(cDX)^{\wedge rk} \arrow[r, "\alpha"] & \Phi^{C_r}(\Sigma^\infty_+ X^{rk}\wedge(cDX)^{\wedge rk}) \arrow[r] & \Phi^{C_r}\mathbb{S} \arrow[d, "\cong"] \\
\Sigma^\infty_+ X^k\wedge(cDX)^{\wedge k} \arrow[u, "\Delta\wedge\Delta"] \arrow[ur, "\Delta"'] \arrow[rr] & & \mathbb{S}
\end{tikzcd}$$

The top triangle commutes because the norm diagonal commutes with smash products. The trapezoid commutes because the inverse of the right-hand isomorphism is the norm diagonal on $\mathbb{S}$ (in fact there is only one isomorphism $\mathbb{S}\to\mathbb{S}$), and the norm diagonal is natural. This finishes the proof that $\Sigma^\infty_+ LX \to \mathrm{Tot}(Y^\bullet)$ is a map of precyclotomic spectra.

For the second phase of the proof, we assume that X is finite and 1–connected, and we check that $\mathrm{Tot}(Y^\bullet)$ is actually cyclotomic; in other words, the map

$$\Phi^{C_r}\mathrm{Tot}(Y^\bullet) \to \mathrm{Tot}(Y^\bullet)$$

is nonequivariantly an equivalence when Φ^{C_r} is left-derived. For simplicity, we may forget the S^1–actions and remember only the cosimplicial C_r–action on $\mathrm{sd}_r\, Y^\bullet$, making it a cosimplicial C_r–spectrum. Then our structure maps respect the restriction to the k–skeleton for each $k\geq 0$:

$$
\begin{tikzcd}
\Phi^{C_r} cF(|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \arrow[r] \arrow[d, "\bar\alpha"] & \Phi^{C_r} cF(|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \arrow[d, "\bar\alpha"] \\
F(\Phi^{C_r}|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, \Phi^{C_r} f\mathbb{S}) \arrow[r] \arrow[d, "F(\Delta, c_r)"] & F(\Phi^{C_r}|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, \Phi^{C_r} f\mathbb{S}) \arrow[d, "F(\Delta, c_r)"] \\
F(|N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \arrow[r] & F(|\mathrm{Sk}_k N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})
\end{tikzcd}
\tag{9}
$$

We first argue that the right vertical composite is an equivalence for each value of k. The skeleta $|\mathrm{Sk}_k N^{\mathrm{cyc}}_\bullet cDX|$ and $|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|$ all have the homotopy type of a finite spectrum, so by [22, III.1.9] the interchange map $\bar\alpha$ is an equivalence. Of course, the diagonal isomorphism Δ from the proof of Theorem 4.6 is an isomorphism of simplicial objects, so it also gives an isomorphism of skeleta. This is enough to conclude that the map $F(\Delta, c_r)$ on the right-hand column is an equivalence.

Thus we get two equivalent towers of spectra underneath $\Phi^{C_r} cF(|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})$ and $F(|N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})$, giving an equivalence of homotopy inverse limits:

$$
\begin{tikzcd}
\Phi^{C_r} cF(|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \arrow[r] \arrow[d] & \mathrm{holim}_k\, \Phi^{C_r} cF(|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \arrow[d, "\sim"] \\
F(|N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \arrow[r, "\sim"] & \mathrm{holim}_k\, F(|\mathrm{Sk}_k N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})
\end{tikzcd}
$$

To finish proving that the left vertical map is an equivalence, it remains to show that on the top, the derived geometric fixed points $\Phi^{C_r}(c-)$ commute with the homotopy inverse limit. This will require us to look more closely at the homotopy fibers of the maps in the homotopy limit system.

Although $\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX$ and $N^{\mathrm{cyc}}_\bullet cDX^{\wedge r}$ are not isomorphic as simplicial objects, they have the same degeneracy maps and therefore have isomorphic latching maps. The cofiber of this latching map

$$(\mathbb{S} \to (cDX)^{\wedge r})^{\square k} \,\square\, (* \to (cDX)^{\wedge r})$$

is the smash product of k copies of the C_r–equivariant cofiber of $\mathbb{S} \to (cDX)^{\wedge r}$ and one copy of $(cDX)^{\wedge r}$. The C_r–equivariant dual of this is a smash product of k copies of $\Sigma^\infty X^r$ and one copy of $\Sigma^\infty_+ X^r$.

To evaluate the homotopy fiber of the map of our homotopy limit system

$$F(|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \to F(|\mathrm{Sk}_{k-1}\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}),$$

we observe that it is the dual of the cofiber of the inclusion of skeleta. By the usual latching square, this cofiber is the k–fold suspension of the cofiber of the latching map.

Therefore our desired homotopy fiber is equivalent as a C_r–spectrum to

$$\Omega^k \Sigma^\infty (X^r)^{\wedge k} \wedge X^r_+.$$

Since X is 1–connected, we can arrange so that its lowest nonbasepoint cell is in dimension 2. This leads to a C_r–equivariant cell structure on X^r in which the lowest nonbasepoint cell is in the diagonal, and is also dimension 2, so $(X^r)^{\wedge k} \wedge X^r_+$ has lowest nonbasepoint cell in dimension $2k$. By induction on these cells, the genuine fixed points

$$(f \Omega^k \Sigma^\infty (X^r)^{\wedge k} \wedge X^r_+)^H$$

are at least $(k-1)$–connected, for each subgroup $H \le C_r$. Since genuine fixed points commute with homotopy limits, we conclude that the fiber of the map from the homotopy limit to the k^{th} term in the homotopy limit system

$$F(|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \to F(|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})$$

has k–connected genuine fixed points for all $H \le C_r$.

The derived geometric fixed points of this fiber are also k–connected. To see this, we use an equivalent definition for the derived geometric fixed points of E, as the genuine fixed points of $\widetilde{E}P \wedge E$ for a certain complex $\widetilde{E}P$ [17, B.10.1]. Our claim then follows by induction on the cells of $\widetilde{E}P$, using the identifications

$$(f(\Sigma^n G/H_+ \wedge E))^G \simeq \Sigma^n F(G/H_+, fE)^G \simeq \Sigma^n (fE)^H.$$

In fact, this proves that for any finite G, a G–spectrum E with k–connected genuine fixed points $(fE)^H$ for all $H \le G$ will also have k–connected geometric fixed points $\Phi^H cE$ for all $H \le G$.

Finally, since derived geometric fixed points commute with fiber sequences, we conclude that the map of derived geometric fixed points

$$\Phi^{C_r} cF(|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \to \Phi^{C_r} cF(|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})$$

is $(k+1)$–connected. Therefore the map to the homotopy limit is an equivalence:

$$\Phi^{C_r} cF(|\mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}) \xrightarrow{\sim} \operatorname{holim}_k \Phi^{C_r} cF(|\mathrm{Sk}_k\, \mathrm{sd}_r\, N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S}).$$

This finishes the proof that $\mathrm{Tot}(Y^\bullet) = F(|N^{\mathrm{cyc}}_\bullet cDX|, f\mathbb{S})$ is cyclotomic.

In conclusion, our map $\Sigma^\infty_+ LX \to \mathrm{Tot}(Y^\bullet)$ is a map of cyclotomic spectra. We already know that it is a stable equivalence if we ignore the circle action. But any such equivalence of cyclotomic spectra is automatically an $\mathcal{F}$–equivalence of S^1 spectra, so we are done. □

Remark 5.15 One may similarly check that this duality preserves multiplications and Adams operations. As a result, when $n \geq 1$, the homology of $\mathrm{THH}(DS^{2n+1})$ is a tensor of a divided power algebra and an exterior algebra

$$H_*(\mathrm{THH}(DS^{2n+1})) \cong H^{-*}(LS^{2n+1}) \cong \Gamma[\alpha] \otimes \Lambda[\beta],$$

where $|\alpha| = -2n$ and $|\beta| = -(2n+1)$. The Adams operations ψ^n are given by

$$\psi^n(\alpha_i \beta^j) = n^i \alpha_i \beta^j.$$

References

[1] **V Angeltveit**, **A J Blumberg**, **T Gerhardt**, **M A Hill**, **T Lawson**, *Interpreting the Bökstedt smash product as the norm*, Proc. Amer. Math. Soc. 144 (2016) 5419–5433 MR

[2] **V Angeltveit**, **A Blumberg**, **T Gerhardt**, **M Hill**, **T Lawson**, **M Mandell**, *Topological cyclic homology via the norm*, preprint (2014) arXiv

[3] **G Arone**, *A generalization of Snaith-type filtration*, Trans. Amer. Math. Soc. 351 (1999) 1123–1150 MR

[4] **D Ayala**, **J Francis**, *Poincaré/Koszul duality*, preprint (2014) arXiv

[5] **C Barwick**, **S Glasman**, *Cyclonic spectra, cyclotomic spectra, and a conjecture of Kaledin*, preprint (2016) arXiv

[6] **C Berger**, **I Moerdijk**, *On an extension of the notion of Reedy category*, Math. Z. 269 (2011) 977–1004 MR

[7] **A J Blumberg**, **R L Cohen**, **C Schlichtkrull**, *Topological Hochschild homology of Thom spectra and the free loop space*, preprint (2008) arXiv

[8] **A J Blumberg**, **M A Mandell**, *The homotopy theory of cyclotomic spectra*, preprint (2013) arXiv

[9] **A M Bohmann**, *A comparison of norm maps*, Proc. Amer. Math. Soc. 142 (2014) 1413–1423 MR

[10] **M Bökstedt**, *Topological Hochschild homology*, preprint, Bielefeld (1985)

[11] **M Bökstedt**, **W C Hsiang**, **I Madsen**, *The cyclotomic trace and algebraic K–theory of spaces*, Invent. Math. 111 (1993) 465–539 MR

[12] **M Brun**, **B I Dundas**, **M Stolz**, *Equivariant structure on smash powers*, preprint (2016) arXiv

[13] **J A Campbell**, *Derived Koszul duality and topological Hochschild homology*, preprint (2014) arXiv

[14] **R L Cohen**, **J D S Jones**, *A homotopy theoretic realization of string topology*, Math. Ann. 324 (2002) 773–798 MR

[15] **A Connes**, *Cohomologie cyclique et foncteurs* Ext^n, C. R. Acad. Sci. Paris Sér. I Math. 296 (1983) 953–958 MR

[16] **W G Dwyer**, **M J Hopkins**, **D M Kan**, *The homotopy theory of cyclic sets*, Trans. Amer. Math. Soc. 291 (1985) 281–289 MR

[17] **M A Hill**, **M J Hopkins**, **D C Ravenel**, *On the nonexistence of elements of Kervaire invariant one*, Ann. of Math. 184 (2016) 1–262 MR

[18] **J D S Jones**, *Cyclic homology and equivariant homology*, Invent. Math. 87 (1987) 403–423 MR

[19] **D Kaledin**, *Motivic structures in non-commutative geometry*, from "Proceedings of the International Congress of Mathematicians, II" (R Bhatia, A Pal, G Rangarajan, V Srinivas, M Vanninathan, editors), Hindustan Book Agency, New Delhi (2010) 461–496 MR

[20] **N J Kuhn**, *The McCord model for the tensor product of a space and a commutative ring spectrum*, from "Categorical decomposition techniques in algebraic topology" (G Arone, J Hubbuck, R Levi, M Weiss, editors), Progr. Math. 215, Birkhäuser, Basel (2004) 213–236 MR

[21] **L G J Lewis**, *The stable category and generalized Thom spectra*, PhD thesis, The University of Chicago (1978) MR Available at `http://search.proquest.com/docview/251798616`

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

[23] **I Madsen**, *Algebraic K–theory and traces*, from "Current developments in mathematics, 1995" (R Bott, M Hopkins, A Jaffe, I Singer, D Stroock, S-T Yau, editors), Int. Press, Cambridge, MA (1994) 191–321 MR

[24] **C Malkiewich**, *Duality and linear approximations in Hochschild homology, K–theory, and string topology*, PhD thesis, Stanford University (2014) Available at `http://math.stanford.edu/~carym/thesis.pdf`

[25] **C Malkiewich**, *A tower connecting gauge groups to string topology*, J. Topol. 8 (2015) 529–570 MR

[26] **M A Mandell**, **J P May**, *Equivariant orthogonal spectra and S–modules*, Mem. Amer. Math. Soc. 755, Amer. Math. Soc., Providence, RI (2002) MR

[27] **M A Mandell**, **J P May**, **S Schwede**, **B Shipley**, *Model categories of diagram spectra*, Proc. London Math. Soc. 82 (2001) 441–512 MR

[28] **R Schwänzl**, **R M Vogt**, *Strong cofibrations and fibrations in enriched categories*, Arch. Math. (Basel) 79 (2002) 449–462 MR

[29] **M Stolz**, *Equivariant structure on smash powers of commutative ring spectra*, PhD thesis, University of Bergen (2011) Available at `https://web.math.rochester.edu/people/faculty/doug/otherpapers/mstolz.pdf`

Department of Mathematics, University of Illinois at Urbana-Champaign
Urbana, IL 61801, United States

`cmalkiew@illinois.edu`

`http://math.uiuc.edu/~cmalkiew/`

Received: 17 May 2016 Revised: 21 January 2017

Algebraic & Geometric Topology 17 (2017) 2357–2412

Spectral sequences in smooth generalized cohomology

DANIEL GRADY
HISHAM SATI

We consider spectral sequences in smooth generalized cohomology theories, including differential generalized cohomology theories. The main differential spectral sequences will be of the Atiyah–Hirzebruch (AHSS) type, where we provide a filtration by the Čech resolution of smooth manifolds. This allows for systematic study of torsion in differential cohomology. We apply this in detail to smooth Deligne cohomology, differential topological complex K-theory and to a smooth extension of integral Morava K-theory that we introduce. In each case, we explicitly identify the differentials in the corresponding spectral sequences, which exhibit an interesting and systematic interplay between (refinements of) classical cohomology operations, operations involving differential forms and operations on cohomology with $U(1)$ coefficients.

55N15, 55T10, 55T25; 53C05, 55S05, 55S35

1 Introduction

Spectral sequences are very useful algebraic tools that often allow for efficient computations that would otherwise require brute force; see McCleary [54] for a broad survey. The Atiyah–Hirzebruch spectral sequence (henceforth AHSS) for K-theory and any generalized cohomology theory, in the topological sense, was introduced by Atiyah and Hirzebruch in [3]. An excellent introduction to the generalized cohomology AHSS can also be found in Hilton [38] and Adams [1, Section III.7]. Other useful references on the subject include Switzer [67] (Section 15, from a homology point of view, including the Gysin sequence from AHSS), and interesting remarks in relation to spectra are given in Rudyak [59]: Theorem 3.45 (homology), Remark 4.24 (sheaves and Čech), Remark 4.34 (Postnikov) and Corollary 7.12. A description with an eye for applications is given in Husemöller, Joachim, Jurčo and Schottenloher [42, Chapter 21].

The goal of this paper is to systematically study the spectral sequence in the context of smooth or differential cohomology; see Cheeger and Simons [20], Freed [27], Hopkins and Singer [41], Simons and Sullivan [66], Bunke [13], Bunke and Schick [17] and Schreiber [63]. Existence and interesting aspects of the AHSS in twisted forms of

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2357

such differential cohomology theories have been considered briefly by Bunke and Nikolaus [15], where the main interest was the effect of the geometric part of the twist on the spectral sequence. In this paper, we take a step back and consider untwisted differential generalized cohomology to systematically study the corresponding AHSS in generality and determine the differentials explicitly as cohomology operations. From the geometric point of view, one might expect on general grounds that the geometric information carried by the differential cohomology theory should somehow manifest itself within the spectral sequence. On the other hand, from an algebraic point of view, one might a priori expect not much of that information to be retained, or even expect it to be totally stripped out while running through the homological algebra machine. We will show that the answer lies somewhat in between, and both intuitions are to some extent correct: the differentials in the spectral sequence will be essentially refinements of classical ones, but with additional operations on differential forms. We recently characterized such operations in [33], and so this paper is a natural continuation of that work.

Just as generalized cohomology theories are represented by spectra, differential cohomology theories are represented by certain sheaves of smooth spectra called *differential function spectra*. The original definition of differential function spectra was due to Hopkins and Singer [41], generalized by Bunke, Nikolaus and Völkl [16], and reformulated in terms of cohesion by Schreiber [63]. The terms *smooth cohomology* and *differential cohomology* seem to be used interchangeably in some of the literature; see eg Bunke and Schick [18]. However, we will find it useful for us to provide a specific and precise usage, where the first is viewed as being more general than the second. We also present most of our ∞–categories as combinatorial, simplicial model categories, rather than quasicategories. We believe that this way, nice objects are more easily and explicitly identifiable, which is desirable when dealing with differential cohomology. Indeed, our discussion will be very explicit, and the results will be readily utilizable.

Ordinary cohomology has smooth extension with various different realizations, including those of Cheeger and Simons [20], Gajer [30], Brylinski [11], Dupont and Ljungmann [23], Hopkins and Singer [41] and Bunke, Kreck and Schick [14]. All these realizations are in fact isomorphic [66; 18]. A description of K-theory with coefficients that combines vector bundles, connections and differential forms into a topological context was initiated by Karoubi [45]. Using Karoubi's description, Lott introduced $\mathbb{R}/\mathbb{Z}$–valued K-theory [49] as well as differential flat K-theory [50]. Currently, there are various geometric models of differential K-theory; see Lott [49], Bunke and Schick [17], Simons and Sullivan [66], Freed and Lott [28], and Tradler, Wilson and Zeinalian [69; 70]. As in the case of ordinary differential cohomology, these models should be equivalent. Indeed, explicit isomorphisms between various models have been demonstrated: for instance, between the differential K-theory group of Hopkins and Singer [41] and that of Freed and Lott [28] in Klonoff [46], between Lott's $\mathbb{R}/\mathbb{Z}$ K-theory and Lott–Freed differential K-theory in [28], between Bunke–Schick differential K-theory and Lott(–Freed) differential K-theory in Ho [40], and between Simons–Sullivan [66] and Freed–Lott [28] in Ho [39].

The group structure of differential K-theory splits into odd and even-degree parts; thus the refinement preserves the grading. However, the odd part turns out to be more delicate than the even part. In particular, while any two differential extensions of even K-theory are isomorphic by the uniqueness results in [18], odd K-theory requires extra data in order to obtain uniqueness. There are various concrete models in the odd case: using smooth maps to the unitary group [69], via loop bundles (see Hekmati, Murray, Schlegel and Vozzo [37]) and via Hilbert bundles (see Gorokhovsky and Lott [31]). Our results in both even and odd K-theory will, of course, not depend on the particular model chosen.

Suppose $\mathcal{E}$ is a spectrum and X is a space of the homotopy type of a CW-complex. Then there is a half-plane spectral sequence (AHSS)

$$E_2^{p,q} \cong H^p(X; \mathcal{E}^q(*)),$$

converging conditionally to $\mathcal{E}^*(X)$. An immediate matter that we encounter in setting up the spectral sequence which calculates the generalized differential cohomology of a smooth manifold X is how to deal with filtrations. Classically, Maunder [52] gave two approaches to any generalized cohomology theory. The first is by filtering over the q–skeletons X^q of the topological space X, and the second by filtering over the Postnikov systems of spaces Y_q, which are the layers of an Ω–spectrum associated to the cohomology theory. Maunder also gives an isomorphism between the two approaches. While we expect this to be the case in the differential setting, the proof might require considerable work. Hence we leave this as an open problem. Maunder sets up his construction in the simplicial complex setting, which is equivalent to doing so in the CW-complex setting, as the geometric realization of a simplicial set is a CW-complex. Simplicial and Čech spectral sequences are discussed by May and Sigurdsson [53, Chapter 22].

We will prefer the filtration of the spaces/manifolds rather than of the corresponding spectra, as this will naturally bring out the geometry desired in the smooth setting. We first would like to replace a topological space with skeletal filtration by a smooth manifold and then view this manifold as a stack. Hence, in doing this, we need an analogue of a skeleton in stacks. This will be done via Čech resolution of smooth spaces, and the replacement of skeletons of a space X will be the various intersections of open sets covering the smooth manifold X.

We will use $\mathrm{diff}(\Sigma^n \mathcal{E}, \mathrm{ch})$ to denote the differential refinement in degree n of a cohomology theory $\mathcal{E}$. This was the notation used in [41] and carries more data than other notation, such as $\mathcal{E}(n)$. It also avoids possible confusion with other notations, eg when dealing with Morava K-theory $K(n)$ at chromatic level n. The axiomatic approach is very useful for characterizing a smooth cohomology theory, but one still needs the model of [41] for actually constructing examples of such smooth spectra. We will be using features of two main approaches at once, namely from [41] with $I\colon \mathrm{diff}(\Sigma^n \mathcal{E}, \mathrm{ch}) \to \underline{\mathcal{E}}$ and from [16; 63] with $I\colon \mathcal{E} \to \Pi\mathcal{E}$. Note that $\underline{\mathcal{E}}$ is not discrete while $\Pi\mathcal{E}$ is, but both are equivalent as smooth spectra: $\underline{\mathcal{E}} \simeq \Pi\mathcal{E}$. This essentially boils down to the fact that since $\Pi\mathcal{E}$ is locally constant, the underlying theory satisfies $\Pi\mathcal{E}^*(U) = \Pi\mathcal{E}^*(*)$ on contractible open sets. On the other hand, the homotopy invariance of the theory $\underline{\mathcal{E}}$ implies the same thing: namely, $\underline{\mathcal{E}}(U) \simeq \underline{\mathcal{E}}(*)$ for a contractible U. These relationships are discussed in further detail in [16].

We will be interested in how the differentials look in our spectral sequences. One might a priori suspect that the differentials in the refined theories should at least loosely be connected to the differentials of the underlying topological theory. We will make this precise below, and so it seems appropriate to understand the form and structure of the differentials in the topological case. To illustrate the point, we will focus on what might perhaps be the most prominent example, namely the first differential $d_3\colon H^*(X, K^0(*)) \to H^*(X, K^0(*))$ in complex topological K-theory $K(X)$ of a topological space X. This is given by $\mathrm{Sq}^3_{\mathbb{Z}}$; see Atiyah and Hirzebruch [3; 4]. There are exactly two stable cohomology operations $H^*(X;\mathbb{Z}) \to H^{*+3}(X;\mathbb{Z})$, since $H^{n+3}(K(\mathbb{Z},n)) = \mathbb{Z}/2$ for n sufficiently large. One of these is zero and the other is $\beta \circ \mathrm{Sq}^2 \circ \rho_2$, where β is the Bockstein associated to the sequence $\mathbb{Z} \xrightarrow{\times 2} \mathbb{Z} \xrightarrow{\rho_2} \mathbb{Z}_2$ with ρ_2 denoting both the mod 2 reduction and its effect on cohomology with these as coefficients, ie $\rho_2\colon H^i(X;\mathbb{Z}) \to H^i(X;\mathbb{Z}/2)$.

The above class, which is a priori in mod 2 cohomology, turned out to be a class in integral cohomology. One could work at any prime [4] by noting the following; see eg Fomenko, Fuchs and Gutenmacher [26] or Hatcher [36]. For any class $x \in H^n(X;\mathbb{Z}/p)$, and with β_p the Bockstein associated with the sequence $\mathbb{Z}_p \xrightarrow{\times p} \mathbb{Z}_{p^2} \xrightarrow{\rho_p} \mathbb{Z}_p$, the element $\beta_p(x)$ is an integral class in $H^{n+1}(X;\mathbb{Z}/p)$; ie it belongs to the image of the reduction homomorphism $\rho_p\colon H^{n+1}(X;\mathbb{Z}) \to H^{n+1}(X;\mathbb{Z}/p)$. This can be used to prove the integrality of the class $d \in H^3(K(\mathbb{Z}/p,2);\mathbb{Z}/p)$ as follows; see [26]. The cohomology Serre spectral sequence for the path-loop fibration $\Omega K(\mathbb{Z},2) \to PK(\mathbb{Z},3) \to K(\mathbb{Z},3)$ gives that $H^*(K(\mathbb{Z},3);\mathbb{Z}/p)$ has a single additive generator $\bar{d}$ in dimension $\le 2p$. Now we have a map $\beta\colon K(\mathbb{Z}/p,2) \to K(\mathbb{Z},3)$ such that $\beta^*(\bar{d}) = d \in H^3(K(\mathbb{Z}/p,2);\mathbb{Z}/p)$, constructed via the Serre spectral sequence of the path-loop fibration $K(\mathbb{Z}/p,1) \to PK(\mathbb{Z}/p,2) \to K(\mathbb{Z}/p,2)$. The map β induces a map of loop spaces which are also Serre fibrations:

$$\begin{tikzcd} K(\mathbb{Z}/p,1) \arrow[r] & PK(\mathbb{Z}/p,2) \arrow[d] & PK(\mathbb{Z},3) \arrow[d] & K(\mathbb{Z}/p,2) \arrow[l] \\ & K(\mathbb{Z}/p,2) \arrow[r] & K(\mathbb{Z},3) & \end{tikzcd}$$

The induced homomorphism on the special sequences sends $\bar{d}$ to d by the construction of β. Now we have $H^3(K(\mathbb{Z}/p,2);\mathbb{Z}/p)=\mathbb{Z}/p$; hence d is contained in the image of the homomorphism $\rho_p\colon H^3(K(\mathbb{Z}/p,2);\mathbb{Z})\to H^3(K(\mathbb{Z}/p,2);\mathbb{Z}/p)$. Therefore, d is an integral class. This is attractive as it makes it readily amenable to differential refinement.

Such statements, and generalizations to other primes and to other generalized cohomology theories, can be made at the level of spectra; see eg Schwede [64]. The first nontrivial k–invariant of connective complex K-theory spectrum ku is a morphism $\boldsymbol{k}_2(ku)\in H^2(H\mathbb{Z},\mathbb{Z})$, which is equal to $\beta\circ\mathrm{Sq}^2$, where $\beta\colon H\mathbb{Z}/2\to\Sigma(H\mathbb{Z})$ is the Bockstein operator associated to the extension $\mathbb{Z}\xrightarrow{\times 2}\mathbb{Z}\to\mathbb{Z}/2$, and $\mathrm{Sq}^2_{\mathbb{Z}}$ is the pullback of the Steenrod operation $\mathrm{Sq}^2\in H^2(H\mathbb{Z}/2,\mathbb{Z}/2)$ along the projection morphism $\rho_2\colon H\mathbb{Z}\to H\mathbb{Z}/2$ given by mod 2 reduction. Since ku is a symmetric ring spectrum, then by [64, Proposition 8.8], the k–invariants are derivations. The only derivations (up to units) in the mod p Steenrod algebra $\mathcal{A}_p$ are the Milnor primitives $Q_n\in H^{2p^n-1}(H\mathbb{Z}/p,\mathbb{Z}/p)$. At the lowest level, we have $Q_0=\beta_p$, the mod p Bockstein, and the others are realized as k–invariants of symmetric spectra, the connective Morava K-theory spectra $k(n)$. That is, we have $Q_n=\boldsymbol{k}_{2p^n-2}(k(n))$. We will consider refinements of integral lifts of these.

The classical AHSS collapses already at the first page if the generalized cohomology theory is rational. In fact, it can be shown that for any reasonably behaved spectrum like all the ones we consider, all the differentials in the AHSS are torsion, ie zero when rationalized; see [59, Corollary 7.12]. The differentials in the AHSS in the topological case are analyzed by systematically by Arlettaz [2]. Using the structure of the integral homology of the Eilenberg–Mac Lane spectra, it is proved there that for any connected space X, there are integers R_r such that $R_r d_r^{s,t}=0$ for all $r\geq 2$ and for all s, t. Some aspects of this general feature will continue to hold in the differential setting. From a homotopy point of view, there is not much difference between the localizations at $\mathbb{R}$ and at $\mathbb{Q}$. However, from a geometric point of view there is a considerable difference. Nevertheless, we will still use the term "rationalize" when we discuss localization at $\mathbb{R}$, as is customary in the homotopy theory literature. We stress that the distinction is needed in certain geometric settings (see Griffiths and Morgan [35]), but it will not be an issue for us in this paper.

Note that although the differential cohomology diamond, ie the diagram that characterizes such theories (see Remark 12), certainly detects torsion classes in the flat part of the theory, it does not distinguish between torsion at various primes. As a by-product, our analysis can be seen as a systematic method for addressing p–primary torsion in differential theories. In [33], we found that the Deligne–Beilinson squaring operation admits lower-degree operations refining the Steenrod squares. We have the familiar pattern

$$\mathrm{DD},\ \widehat{\mathrm{Sq}}^1,\ \widehat{\mathrm{Sq}}^2,\ \widehat{\mathrm{Sq}}^3,\ \ldots,\ \mathrm{DD}^2,\ \ldots,$$

where DD is the Dixmier–Douady class: a nontorsion differential cohomology operation. The refined squares $\widehat{\mathrm{Sq}}^{2k+1}$, as the classical squares Sq^{2k}, are operations that are 2–torsion. In this paper, we get $\widehat{\mathrm{Sq}}^{2k+1}$ as we expect, but also differentials d_{2m} at lowest degree for every m:

$$d_{2m}\colon \prod_k \Omega^{2k}(M) \to H^{2m}(M;U(1)). \tag{1-1}$$

We consider this as a cohomology operation, which can be viewed as first projecting on to the homogeneous component ch_{2m} of the Chern character. A $U(1)$–valued Čech cocycle is obtained by restricting to $2m$–fold intersections of an open cover, pairing with an appropriate simplex of degree $2m$ and exponentiating; this will be spelled out in detail in Section 4. If indeed the form ch_{2m} arises as the curvature of a bundle, it must represent a closed form with *integral* periods. The differential d_{2m} can therefore be understood as the obstruction to this condition. Similar results hold for the odd part, ie for differentially refined K^1–theory, where the refined Steenrod square takes the same form as in differential K^0–theory, while the differentials arising from forms — the analogues of those in (1-1) — are now of odd degrees.

The paper is organized as follows. In Section 2, we start by carefully setting up the background in smooth and differential cohomology, preparing the scene for our constructions. In particular, in Section 2.1, we adapt abstract general results on stacks (or simplicial sheaves) to our context and spell out specific definitions and constructions that will be useful for us in later sections; more general and comprehensive accounts can be found in Jardine [43], Lurie [51] and Schreiber [63]. Then in Section 2.2, we take the approach to differential cohomology that allows for a direct generalization. Our main constructions will be in Section 3. In particular, in Section 3.1, we provide the filtration via Čech resolutions; then we construct the AHSS for smooth spectra in Section 3.2 and compare to the AHSS of the underlying topological theory. This refinement will depend on whether the degree is positive, negative or zero. Then we explore the compatibility of the differentials with the product structure in Section 3.3.

Having given the main construction, our main applications of the general spectral sequence to various differential cohomology theories will be presented in Section 4.

The construction is general enough to apply to any structured cohomology theory whose coefficients are known. We will explicitly emphasize three main examples: ordinary differential cohomology, differential K-theory and a differential version of integral Morava K-theory that we introduce. As a test of our method, in Section 4.1, we recover the usual hypercohomology spectral sequence for the Deligne complex (see [11] and Esnault and Viehweg [24, Appendix]), and we do so for manifolds, then products of these, and then more generally for smooth fiber bundles. Then the AHSS for K-theory is generalized in Section 4.2 to differential K-theory, where the differentials involve refinements of Steenrod squares, in the sense of [33], as well as operations on forms, as indicated above around expression (1-1). We also show that the odd case, ie smooth extension of K^1, leads to a similar construction, but with the differentials now involving odd forms. Then in Section 4.3, we first introduce a refinement of the integral form of Morava K-theory, discussed in Kriz and Sati [47], Sati [60] and Sati and Westerland [62], and then characterize the corresponding differentials, which turn out to have a similar pattern as in K-theory, where the operation that gets refined is the Milnor primitive Q_n encountered above. We end with an application to an example from M-theory and string theory.

Notation We have the following morphism that we will use repeatedly throughout. Denote by $\rho_p\colon \mathbb{Z} \to \mathbb{Z}/p$ the mod p reduction on coefficients with corresponding morphism using the same notation on the cohomology groups with these as coefficients. We will denote by β, β_p and $\widetilde{\beta}$ the Bockstein homomorphisms associated with the coefficient sequences

$$0 \to \mathbb{Z} \to \mathbb{R} \xrightarrow{\exp} U(1) \to 0,$$

$$0 \to \mathbb{Z}/p \xrightarrow{\times p} \mathbb{Z}/p^2 \xrightarrow{\rho_p} \mathbb{Z}/p \to 0,$$

$$0 \to \mathbb{Z} \xrightarrow{\times p} \mathbb{Z} \xrightarrow{\rho_p} \mathbb{Z}/p \to 0,$$

respectively. We will let $\Gamma_2\colon \mathbb{Z}/2 \hookrightarrow U(1)$ denote the representation as the square roots of unity, and also the induced map $\Gamma_2\colon H^n(-;\mathbb{Z}/2) \to H^n(-;U(1))$ on cohomology. We will also use more refined Bockstein homomorphisms associated with spectra, and these will be defined as we need them.

2 Smooth cohomology

2.1 Smooth cohomology and the stable category of smooth stacks

In this section, we adapt abstract general results on stacks (or simplicial sheaves) to our context and spell out specific definitions and constructions that will be useful for us in later sections. The interested reader can find more general and comprehensive accounts in [43; 51; 63]. For the reader who is more interested in the applications to differential cohomology theories, this section can be skipped. However, we would like to emphasize that although the language used in this section is rather abstract, the generality gained from this formalism is far reaching and allows this machinery to be used for a wide variety theories beyond just differential cohomology theories.

Essentially, the axioms characterizing a smooth cohomology theory are not much different from the axioms characterizing usual cohomology theories. The big difference is where the theory takes place. More precisely, we want to consider homotopical functors on the category of pointed *smooth stacks* $\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+$ with $\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}$ the category of Cartesian spaces, rather than the category of pointed topological spaces $\mathcal{T}\mathrm{op}_+$. Let $\mathcal{A}\mathrm{b}_{\mathrm{gr}}$ be the category of graded abelian groups.

Definition 1 (smooth cohomology) Let $\mathcal{E}^*\colon \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+^{\mathrm{op}} \to \mathcal{A}\mathrm{b}_{\mathrm{gr}}$ be a functor satisfying the following axioms:

(1) **Invariance** $\mathcal{E}^*$ sends equivalences to isomorphisms.

(2) **Additivity** For small coproducts (ie ones forming sets) of pointed stacks, $\bigvee_\alpha X_\alpha$, we have
$$\mathcal{E}^*\big(\textstyle\bigvee_\alpha X_\alpha\big) = \prod_\alpha \mathcal{E}^*(X_\alpha).$$

(3) **Mayer–Vietoris** For any homotopy pushout of pointed stacks
$$\begin{tikzcd} Z \arrow[r] \arrow[d] & Y \arrow[d] \\ X \arrow[r] & X \cup_Z Y \end{tikzcd}$$
the induced sequence
$$\mathcal{E}^*(X \cup_Z Y) \to \mathcal{E}^*(X) \oplus \mathcal{E}^*(Y) \to \mathcal{E}^*(Z)$$
is exact.

(4) **Suspension** For any stack X, there is an isomorphism $\mathcal{E}^{n+1}(\Sigma X) \simeq \mathcal{E}^n(X)$.

Then we call $\mathcal{E}^*$ a *smooth cohomology theory*.

Remark 2 Note that the Mayer–Vietoris axiom implies the usual Mayer–Vietoris sequence. Indeed, let M be a manifold and let V be a local chart of M. Let U be an open set such that $\{U, V\}$ is a cover of M. Then the strict pushout

$$\begin{tikzcd} U \cap V \arrow[r] \arrow[d] & V \arrow[d] \\ U \arrow[r] & U \cup V \end{tikzcd}$$

is actually a homotopy pushout. We can equivalently write this diagram as a homotopy coequalizer

$$U \cap V \rightrightarrows U \sqcup V \to U \cup V$$

in which the homotopy cofiber of the second map can be identified with $\Sigma U \cap V$. By iterating this argument and applying $\mathcal{E}^*$ to the resulting diagram, one obtains the long exact sequence

$$\cdots \to \mathcal{E}^*(U \cap V) \to \mathcal{E}^*(M) \to \mathcal{E}^*(U) \oplus \mathcal{E}^*(V) \to \mathcal{E}^{*+1}(U \cap V) \to \cdots,$$

which is the familiar Mayer–Vietoris sequence.

The above axioms can be taken as a generalization of the Eilenberg–Steenrod axioms (see [1; 38]), where the Mayer–Vietoris axiom subsumes both the excision axiom and the long exact sequence axiom. It is interesting to note that the axioms *do not* require homotopy invariance. Namely, if two manifolds M and N are *homotopic*, they may fail to be equivalent as stacks. In fact, an equivalence of stacks requires, in particular, that for every sheaf F (embedded as a stack), we have an isomorphism

$$F(N) \simeq \pi_0 \operatorname{Map}(N, F) \simeq \pi_0 \operatorname{Map}(M, F) \simeq F(M).$$

In particular, we can take the sheaf of smooth $\mathbb{R}$–valued functions on a manifold. Then if every homotopy equivalence $f: M \to N$ induced an equivalence of stacks, we would have an induced isomorphism

$$f^*: C^\infty(N; \mathbb{R}) \to C^\infty(M; \mathbb{R}).$$

Taking $N = *$ and $M = \mathbb{R}^n$ immediately gives a contradiction. On the other hand, every equivalence of stacks does produce a weak homotopy equivalence of geometric realizations. To see this, simply note that the geometric realization functor

$$\Pi: \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}) \to s\mathcal{S}\mathrm{et},$$

being a Quillen functor, has a derived functor by Ken Brown's lemma [10]. It therefore preserves weak equivalences between fibrant objects. But these objects are exactly those that satisfy descent, namely stacks (eg manifolds) [63; 22].

Remark 3 Given a smooth cohomology theory $\mathcal{E}^*$, we always get a presheaf of graded abelian groups on the site $\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}$ by precomposing with the Yoneda embedding:

$$\mathcal{E}^*\colon \mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p} \xhookrightarrow{Y} \mathrm{Sh}(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}) \xhookrightarrow{\mathrm{sk}_0} \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}) \xrightarrow{+} \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+ \xrightarrow{\mathcal{E}^*} \mathcal{A}\mathrm{b}_{\mathrm{gr}},$$

where sk_0 embeds a sheaf as a discrete simplicial sheaf. We will use this fact later in the construction of the spectral sequence in Theorem 25.

Just as all cohomology theories are representable by Ω–spectra, via Brown representability, all smooth cohomology theories are representable by smooth spectra. This follows from the version of Brown representability formulated by Jardine in [43] applied to the stable homotopy category of smooth stacks. We will quickly review the basic properties of this category (see [51; 44]) to establish where our objects of interest live.

We first recall some operations on stacks that are counterparts to standard operations on topological spaces. Let X and Y be two pointed stacks.

(i) The wedge product $X \vee Y$ is defined via the pushout diagram:

$$\begin{tikzcd} Y \arrow[r] & Y \vee X \\ * \arrow[u] \arrow[r] & X \arrow[u] \end{tikzcd}$$

(ii) The smash product $X \wedge Y$ is defined as the quotient $X \wedge Y := X \times Y / X \vee Y$ of the Cartesian product by the wedge product.

(iii) The suspension ΣX is defined via the homotopy pushout diagram:

$$\begin{tikzcd} X \arrow[r] \arrow[d] & * \arrow[d] \\ * \arrow[r] & \Sigma X \end{tikzcd}$$

(iv) The looping, ie loop space, ΩX is defined via the homotopy pullback:

$$\begin{tikzcd} \Omega X \arrow[r] \arrow[d] & * \arrow[d] \\ * \arrow[r] & X \end{tikzcd}$$

Definition 4 We define the *stabilization* $\mathrm{Stab}(\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+)$ *of smooth pointed stacks* to be the following category:

- The objects of $\mathrm{Stab}(\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+)$ are sequences of pointed stacks
$$\{\mathcal{E}_n\} \subset \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+, \quad n \in \mathbb{Z},$$
equipped with maps $\sigma_n \colon \Sigma\mathcal{E}_n \to \mathcal{E}_{n+1}$.
- The morphisms between $\mathcal{E}$ and $\mathcal{F}$ are defined to be the levelwise morphisms $\mathcal{E}_n \to \mathcal{F}_n$ commuting with the σ_n.

This category carries a stable model structure given by first taking the projective model structure on sequences of stacks and then performing Bousfield localization with respect to stable weak equivalences in the usual way. This process is described in detail in [43; 51; 44], and we summarize the relevant results found there. The category $\mathrm{Stab}(\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+)$ admits a stable, closed, simplicial model structure with the following properties:

- The weak equivalences are *stable* weak equivalences. That is, a morphism of smooth spectra $f \colon \mathcal{E}_\bullet \to \mathcal{F}_\bullet$ is a weak equivalence if and only if it induces a weak equivalence
$$Q(f)\colon \lim_{i\to\infty} \Omega^i\mathcal{E}_{n+i} \to \lim_{j\to\infty} \Omega^j\mathcal{F}_{n+j}.$$
- The fibrant objects are precisely the smooth Ω–Spectra, that is, the sequence of stacks $X_\bullet$ whose structure maps
$$\sigma_n\colon \Sigma\mathcal{E}_n \to \mathcal{E}_{n+1}$$
induce equivalences $\mathcal{E}_n \xrightarrow{\simeq} \Omega\mathcal{E}_{n+1}$.

Remark 5 We will refer to the stable model category $\mathrm{Stab}(\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+)$ as the category of *smooth spectra* and denote it by
$$\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{S}\mathrm{p}) := \mathrm{Stab}(\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+).$$

Example 6 Let $M \in \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})_+$ be a manifold, viewed a stack and equipped with a basepoint. We can define the smooth spectrum $\Sigma^\infty M$ in the usual way, as the sequence of suspensions of the manifold M. Given a smooth Ω–spectrum $\mathcal{E}$, we can define a smooth cohomology theory $\mathcal{E}^*$, by setting
$$\mathcal{E}^q(M) \simeq \pi_0\,\mathrm{Map}(\Sigma^{-q}\Sigma^\infty M, \mathcal{E}).$$

Differential cohomology theories are examples of the theories introduced above, although it may not be immediately apparent where the differential cohomology "diamond" diagram [66] fits into this context. In fact, it was observed by Bunke, Nikolaus and Völkl in [16] that the diamond provides a further characterization of *all* smooth cohomology theories in terms of refinement of topological theories. This characterization happens in addition to the Brown representability described above, and it happens only when the category of stacks exhibits so-called *cohesion*. We now review the properties of the cohesive structure on smooth stacks [63] that we need, along with the characterization of smooth cohomology theories described in [16]. It is shown in [63] that the category $\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})$ admits a quadruple ∞–categorical adjunction $(\Pi \dashv \mathrm{disc} \dashv \Gamma \dashv \mathrm{codisc})$

$$\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}) \begin{array}{c} \xrightarrow{\Pi} \\ \xleftarrow{\mathrm{disc}} \\ \xrightarrow{\Gamma} \\ \xleftarrow{\mathrm{codisc}} \end{array} s\mathcal{S}\mathrm{et}, \tag{2-1}$$

where Π preserves finite ∞–limits, and the functors disc and codisc are fully faithful.

One implication of this is that $s\mathcal{S}\mathrm{et}$ embeds into $\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})$ as an ∞–subcategory in two different ways: one reflective, the other reflective and coreflective. From the reflectors, one can produce two monads and one comonad defined as follows:

$$\Pi := \Pi \circ \mathrm{disc}, \quad \flat := \mathrm{disc} \circ \Gamma, \quad \sharp := \mathrm{codisc} \circ \Gamma.$$

These monads fit into a triple ∞–adjunction $(\Pi \dashv \flat \dashv \sharp)$ which is called a *cohesive* adjunction.

Remark 7 Each monad in the cohesive adjunction picks out a different part of the nature of a smooth stack. This nature is perhaps best exemplified by how the adjoints behave on smooth manifolds (viewed as stacks). More precisely, if M is a smooth manifold, then for instance:

(i) The comonad $\flat$ takes the underlying set of points of the manifold and then embeds this set back into stacks as a discrete object. This functor therefore misses the smooth structure of the manifold and treats it instead as a discrete object.

(ii) The monad Π essentially takes the singular nerve of the manifold using *smooth* paths and higher smooth simplices on the manifold. It therefore retains the geometry of the manifold and "knows" that the points of the manifold ought to be connected together in a smooth way.

The following observation on lifting from simplicial sets to spectra is known [63, Proposition 4.1.9], but we supply a proof for completeness.

Proposition 8 *The ∞–adjunction (2-1) lifts to an ∞–adjunction*

$$\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{S}\mathrm{p}) \begin{array}{c} \xrightarrow{\Pi^s} \\ \xleftarrow{\mathrm{disc}^s} \\ \xrightarrow{\Gamma^s} \\ \xleftarrow{\mathrm{codisc}^s} \end{array} \mathcal{S}\mathrm{p}$$

on the stable ∞–category of smooth spectra. Moreover, the adjoints satisfy the same condition as the ∞–adjunction (2-1) does.

Proof The category of smooth stacks is presented by the combinatorial simplicial model category

$$\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}) = [\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}, s\mathcal{S}\mathrm{et}]_{\mathrm{loc,proj}},$$

where loc denotes the Bousfield localized model structure at the maps out of Čech nerves. The quadruple adjunction is presented by Quillen adjoints $(\Pi \dashv \mathrm{disc} \dashv \Gamma \dashv \mathrm{codisc})$ [63]. We need to show that this adjunction holds on the stable model category of smooth spectra. The adjunction immediately gives an underlying categorical adjunction by simply applying the functors degreewise. In the projective model structure, the right adjoints are Quillen by definition, and the closed model axioms imply that the left adjoints are also Quillen.

Now the functors (in the global model structure on $\mathcal{S}\mathrm{p}$) disc and codisc both preserve homotopy limits. Hence for a local weak equivalence $f\colon \mathcal{E} \to \mathcal{F}$ of spectra, we have

$$\begin{aligned} \lim_{i\to\infty} \Omega^i \mathrm{disc}(\mathcal{E})_{n+i} &\simeq \mathrm{disc}\Big(\lim_{i\to\infty} \Omega^i \mathcal{F}_{n+i}\Big) \\ &\simeq \mathrm{disc}\Big(\lim_{j\to\infty} \Omega^j \mathcal{F}_{n+j}\Big) \\ &\simeq \lim_{j\to\infty} \Omega^j \mathrm{disc}(\mathcal{F})_{n+j}, \end{aligned}$$

and $\mathrm{disc}(f)$ induces a weak equivalence $Q(\mathrm{disc}(f))$. Hence $\mathrm{disc}(f)$ is a weak equivalence. In the same way, codisc preserves local weak equivalences. It follows by the basic properties of Bousfield localization that disc and codisc are right Quillen adjoints. Again, by the axioms of a closed model category, it follows that the entire adjunction holds as a Quillen adjunction of stable model categories. □

Remark 9 The proof of the previous proposition implies that both disc and codisc preserve Ω–spectra. However, Π and Γ need not take Ω–spectra to Ω–spectra. This problem can be remedied by taking Π^s (or Γ^s) to be the composition $R \circ \Pi$ (or $R \circ \Gamma$), where R is the fibrant replacement in spectra. Since R defines a left ∞–adjoint to the inclusion of fibrant objects (and preserves finite ∞–limits), we will still have an adjunction at the level of ∞–categories (although this is not presented by a Quillen adjunction).

As in the case of smooth stacks, the quadruple adjunction in Proposition 8 produces adjoint monads $(\Pi^s \dashv \flat^s \dashv \sharp^s)$ exhibiting *stable* cohesion. The main observation in [16], recast in the cohesive setting in [63], is the following. Let $j\colon \flat^s \to \mathrm{id}$ be the counit of the comonad $\flat^s$, and let $I\colon \mathrm{id} \to \Pi^s$ be the unit of the monad Π^s. Let $\mathcal{E} \in \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{S}\mathrm{p})$ be a smooth spectrum. Then $\mathcal{E}$ sits inside a hexagon diagram

$$
\begin{tikzcd}
 & \mathrm{fib}(\eta)(\mathcal{E}) \arrow[rr] \arrow[dr] & & \mathrm{cofib}(\epsilon)(\mathcal{E}) \arrow[dr] & \\
\Sigma^{-1}\Pi^s \,\mathrm{cofib}(\epsilon)(\mathcal{E}) \arrow[ur] \arrow[dr] & & \mathcal{E} \arrow[ur] \arrow[dr, "I"] & & \Pi^s \,\mathrm{cofib}(\epsilon)(\mathcal{E}) \\
 & \flat^s\mathcal{E} \arrow[rr] \arrow[ur, "j"] & & \Pi^s\mathcal{E} \arrow[ur] &
\end{tikzcd}
\tag{2-2}
$$

where the diagonals are fiber sequences (by definition), the top and bottom sequences are fiber sequences, and the two squares in the hexagon are homotopy Cartesian; ie both are homotopy pullback squares and hence homotopy pushouts (via the equivalence of the two in the stable setting). The latter property is key because it is a homotopy Cartesian square, as on the right of the hexagon, which Hopkins and Singer [41] took as the definition of differential cohomology (for a specific choice of the object of differential forms). Bunke, Nikolaus and Völkl [16] observed that by the hexagon, *every* smooth spectrum satisfies this kind of Hopkins–Singer definition, if one just allows more general objects of differential forms, which is the object $\mathrm{cofib}(\epsilon)(\mathcal{E})$ in our notation above.

It often happens in practice that the smooth spectra $\mathrm{fib}(\eta)(\mathcal{E})$ and $\mathrm{cofib}(\epsilon)(\mathcal{E})$ contain no information away from degree 0. In particular, it often happens that for $n > 0$,

$$\pi_n \,\mathrm{Map}(M, \mathrm{cofib}(\epsilon)(\mathcal{E})) \simeq 0, \tag{2-3}$$

$$\pi_{-n} \,\mathrm{Map}(M, \mathrm{fib}(\eta)(\mathcal{E})) \simeq 0. \tag{2-4}$$

In this case, the $\mathcal{E}$–cohomology of a manifold can be calculated as either the flat cohomology or the underlying topological cohomology in all degrees but 0. This is summarized as the following result.

Proposition 10 *Let $\mathcal{E}$ be a smooth spectrum such that (2-3) and (2-4) are satisfied. Then the $\mathcal{E}$–theory of a manifold M is given by*

$$\mathcal{E}^n(M) := \begin{cases} (\Pi^s\mathcal{E})^n(M), & n > 0, \\ (\flat^s\mathcal{E})^n(M), & n < 0, \end{cases}$$

where $\mathcal{E}(M)$ is already characterized in degree 0 by the diamond (2-2).

Proof Since the diagonals of the diamond are fiber sequences, they induce long exact sequences in cohomology. Let n be a positive integer. The sequence

$$\flat^s\mathcal{E} \to \mathcal{E} \to \mathrm{cofib}(\epsilon)(\mathcal{E})$$

gives the section of the long sequence

$$\pi_{n+1}\operatorname{Map}(M,\operatorname{cofib}(\epsilon)(\mathcal{E}))\to \flat^s\mathcal{E}^{-n}(M)\to \mathcal{E}^{-n}(M)\to \pi_n\operatorname{Map}(M,\operatorname{cofib}(\epsilon)(\mathcal{E})).$$

By assumption, the leftmost and rightmost groups are 0. Thus we have an isomorphism

$$(\flat^s\mathcal{E})^{-n}(M)\simeq \mathcal{E}^{-n}(M).$$

Similarly, the sequence

$$\operatorname{fib}(\eta)(\mathcal{E})\to\mathcal{E}\to\Pi^s\mathcal{E}$$

gives the long sequence

$$\pi_{-n}\operatorname{Map}(M,\operatorname{fib}(\eta)(\mathcal{E}))\to \mathcal{E}^n(M)\to (\Pi^s\mathcal{E})^n(M)\to \pi_{-n-1}\operatorname{Map}(M,\operatorname{fib}(\eta)(\mathcal{E})),$$

and again we get the desired isomorphism. □

2.2 Differential cohomology and differential function spectra

The main applications we have in mind, as we indicated in the introduction, concern *differential cohomology theories*. In this section, we review some of the concepts established in [13; 16; 63] (which generalize [66]), adapted to our context.

Definition 11 Let $\mathcal{E}^*$ be a cohomology theory. A *differential refinement* $\widehat{\mathcal{E}}^*$ of $\mathcal{E}^*$ consists of the following data:

(1) a functor $\widehat{\mathcal{E}}^*\colon \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}_+)^{\mathrm{op}}\to \mathcal{A}\mathrm{b}_{\mathrm{gr}}$;

(2) three natural transformations:
  - (a) **Integration** $I\colon \widehat{\mathcal{E}}^*\to\mathcal{E}^*$;
  - (b) **Curvature** $R\colon \widehat{\mathcal{E}}^*\to Z_*(\Omega^*\otimes\mathcal{E}^*(*))$;
  - (c) **Secondary Chern character** $a\colon \Omega^*\otimes\mathcal{E}^*(*)[1]/\operatorname{im}(d)\to\widehat{\mathcal{E}}^*$;

such that the following axioms hold:

- **Chern–Weil** We have a commutative diagram

$$\begin{tikzcd} \widehat{\mathcal{E}}^* \arrow[r, "R"] \arrow[d, "I"] & Z_*(\Omega^*\otimes\mathcal{E}^*(*)) \arrow[d, "q"] \\ \mathcal{E}^* \arrow[r, "\mathrm{ch}"] & H_*(\Omega^*\otimes\mathcal{E}^*(*)) \end{tikzcd}$$

  where ch is the Chern character map.

- **Secondary Chern–Weil** We have a commutative diagram

$$\begin{tikzcd} \Omega^* \otimes \mathcal{E}^*(*)[1]/\operatorname{im}(d) \arrow[rr, "d"] \arrow[dr, "a"'] & & Z_*(\Omega^* \otimes \mathcal{E}^*(*)) \\ & \widehat{\mathcal{E}}^* \arrow[ur, "R"'] & \end{tikzcd}$$

and an exact sequence

$$\cdots \to \mathcal{E}^*[1] \to \Omega^* \otimes \mathcal{E}^*(*)[1]/\operatorname{im}(d) \to \widehat{\mathcal{E}}^* \to \mathcal{E}^* \to \cdots .$$

Note that in the Chern–Weil axiom above, $H_*(\Omega^* \otimes \mathcal{E}^*(*))$ appears as the codomain of the Chern character. As explained in [16], this becomes a locally constant stack equivalent to just the locally constant stack on the rationalization of $\mathcal{E}^*$; ie ch is equivalent to ch: $\mathcal{E}^* \to \mathcal{E}^* \wedge H\mathbb{R}$ (or $M\mathbb{R}$).

Remark 12 The above characterization can ultimately be summarized by saying that differential cohomology fits into an exact diamond

$$\begin{tikzcd} & \Omega^* \otimes \mathcal{E}^*(*)[1]/\operatorname{im}(d) \arrow[rr, "d"] \arrow[dr, "a"] & & Z_*(\Omega^* \otimes \mathcal{E}^*(*)) \arrow[dr] & \\ \mathcal{E}^{*-1} \otimes \mathbb{R} \arrow[ur] \arrow[dr] & & \widehat{\mathcal{E}}^* \arrow[ur, "R"] \arrow[dr, "I"] & & \mathcal{E}^* \otimes \mathbb{R} \\ & \mathcal{E}^{*-1}_{\mathbb{R}/\mathbb{Z}} \arrow[ur] \arrow[rr, "\beta_{\mathcal{E}}"] & & \mathcal{E}^* \arrow[ur, "\mathrm{ch}"] & \end{tikzcd}$$

where the diagonal, top and bottom sequences are all part of long exact sequences. The bottom sequence is obtained by observing that the cofiber of the rationalization map is an $MU(1)$ (Eilenberg–Moore spectrum), where we identify $\mathbb{R}/\mathbb{Z}$ with $U(1)$ throughout. That is, we have a cofiber sequence involving the unit map from the sphere spectrum $\mathbb{S} = M\mathbb{Z}$:

$$\mathbb{S} \to M\mathbb{R} \to MU(1).$$

Smashing on the left with the theory $\mathcal{E}$, we obtain a "Bockstein sequence"

$$\mathcal{E} \to \mathcal{E} \wedge M\mathbb{R} \to \mathcal{E} \wedge MU(1) \xrightarrow{\beta_{\mathcal{E}}} \Sigma\mathcal{E}.$$

We define the flat theory as

$$\mathcal{E}_{U(1)} := \mathcal{E} \wedge MU(1)$$

and the rational theory as

$$\mathcal{E}_{\mathbb{R}} := \mathcal{E} \wedge M\mathbb{R}.$$

Remark 13 Differential cohomology theories are special cases of smooth cohomology theories, while differential function spectra are special cases of smooth spectra. Thus, this section can be viewed as describing a special case of the previous section.

Since differential cohomology theories will arise as certain homotopy pullbacks (in Definition 17 below), we will first need to establish the components of the pullback. We begin with the following lemma that can be found in [13, Lemma 6.10], which explains how we can transition from a topological cohomology theory to a smooth one, in a process whose direction is opposite to that of the map I.

Lemma 14 *Let $\mathcal{E}$ be a spectrum and define the smooth presheaf of spectra $\underline{\mathcal{E}}$ via the assignments*

$$\begin{aligned}\text{objects:}\quad & U \mapsto \mathrm{Map}(\Sigma^\infty U, \mathcal{E}),\\ \text{morphisms:}\quad & (f\colon U \to V) \mapsto \big(f^*\colon \mathrm{Map}(\Sigma^\infty V, \mathcal{E}) \to \mathrm{Map}(\Sigma^\infty U, \mathcal{E})\big).\end{aligned}$$

Then $\underline{\mathcal{E}}$ satisfies descent.

Proof Let $C^\bullet(\{U_\alpha\})$ denote the Čech nerve of a good open cover $\{U_\alpha\}$ of some manifold M. The Yoneda lemma and basic properties of the mapping space functor imply that we have the sequence of equivalences

$$\begin{aligned}\underline{\mathcal{E}}(M) &:= \mathrm{Map}(\Sigma^\infty M, \mathcal{E})\\ &\simeq \mathrm{Map}(\Sigma^\infty \operatorname{hocolim}_{\Delta^{\mathrm{op}}} C^\bullet(\{U_\alpha\}), \mathcal{E})\\ &\simeq \mathrm{Map}(\operatorname{hocolim}_{\Delta^{\mathrm{op}}} \Sigma^\infty C^\bullet(\{U_\alpha\}), \mathcal{E})\\ &\simeq \operatorname{holim}_{\Delta^{\mathrm{op}}} \mathrm{Map}(\Sigma^\infty C^\bullet(\{U_\alpha\}), \mathcal{E})\\ &\simeq \operatorname{holim}\Big\{\cdots \overset{\longleftarrow}{\underset{\longleftarrow}{\Longrightarrow\!\!\!\!\Longleftarrow}} \textstyle\prod_{\alpha\beta\gamma} \mathrm{Map}(\Sigma^\infty U_{\alpha\beta\gamma}, \mathcal{E}) \overset{\longleftarrow}{\underset{\longleftarrow}{\Longrightarrow}} \prod_{\alpha\beta} \mathrm{Map}(\Sigma^\infty U_{\alpha\beta}, \mathcal{E})\\ &\qquad\qquad \overset{\longrightarrow}{\underset{\longrightarrow}{\longleftarrow}} \textstyle\prod_\alpha \mathrm{Map}(\Sigma^\infty U_\alpha, \mathcal{E})\Big\}\\ &\simeq \operatorname{holim}\Big\{\cdots \overset{\longleftarrow}{\underset{\longleftarrow}{\Longrightarrow\!\!\!\!\Longleftarrow}} \textstyle\prod_{\alpha\beta\gamma} \underline{\mathcal{E}}(U_{\alpha\beta\gamma}) \overset{\longleftarrow}{\underset{\longleftarrow}{\Longrightarrow}} \prod_{\alpha\beta} \underline{\mathcal{E}}(U_{\alpha\beta}) \overset{\longrightarrow}{\underset{\longrightarrow}{\longleftarrow}} \prod_\alpha \underline{\mathcal{E}}(U_\alpha)\Big\},\end{aligned}$$

and so $\underline{\mathcal{E}}$ satisfies descent. □

The other components of the pullback we want to establish are presented by sheaves of chain complexes. There is a general functorial construction by which one can turn an unbounded chain complex into a spectrum, which we now describe; see [65] for details. This functor is called the *Eilenberg–Mac Lane* functor

$$H\colon \mathcal{C}\mathrm{h} \to \mathcal{S}\mathrm{p}, \tag{2-5}$$

and acts on objects as follows. Let $C_\bullet$ be an unbounded chain complex, and let Z_n denote the subgroup of cycles in degree n. The functor H takes $C_\bullet$ and forms the sequence $C_\bullet(\bullet)$ of truncated bounded chain complexes:

$$\begin{aligned}
C_\bullet(0) &= (\cdots \to C_n \to C_{n-1} \to \cdots \to C_1 \to Z_0),\\
C_\bullet(1) &= (\cdots \to C_n \to C_{n-1} \to \cdots \to C_0 \to Z_{-1}),\\
C_\bullet(2) &= (\cdots \to C_n \to C_{n-1} \to \cdots \to C_{-1} \to Z_{-2}),\\
&\ \ \vdots\\
C_\bullet(k) &= (\cdots \to C_n \to C_{n-1} \to \cdots \to C_{-k+1} \to Z_{-k}),\\
&\ \ \vdots
\end{aligned}$$

The reason for the group of cycles appearing in degree 0 comes from using the *right* adjoint to the inclusion $i\colon \mathcal{C}\mathrm{h}^+ \to \mathcal{C}\mathrm{h}$ (as opposed to the left). The left adjoint simply truncates the complex in degree 0, while the right adjoint truncates and then takes only the cycles in degree 0.

Continuing with our discussion, at each level in the sequence, H applies the Dold–Kan functor $\mathrm{DK}\colon \mathcal{C}\mathrm{h}^+ \to s\mathcal{S}\mathrm{et}$ to the bounded chain complex in that degree. This gives a sequence $\mathrm{DK}(C_\bullet(\bullet))$ of spaces

$$\mathrm{DK}(C_\bullet(0)),\ \mathrm{DK}(C_\bullet(1)),\ \mathrm{DK}(C_\bullet(2)),\ \ldots,\ \mathrm{DK}(C_\bullet(k)),\ \ldots.$$

Since DK preserves looping (being a right Quillen adjoint) and equivalences (being a Quillen equivalence of model categories), we get induced equivalences

$$\sigma_k\colon \mathrm{DK}(C_\bullet(k)) \to \Omega\,\mathrm{DK}(C_\bullet(k-1)),$$

which turns $\mathrm{DK}(C_\bullet(\bullet))$ into a spectrum.

Example 15 Consider the unbounded chain complex $\mathbb{Z}[0]$, with $\mathbb{Z}$ concentrated in degree 0. Then

$$H(\mathbb{Z}[0]) \simeq H\mathbb{Z},$$

where the right-hand side denotes the Eilenberg–Mac Lane spectrum.

Example 16 Fix a manifold M and consider the de Rham complex

$$\Omega^* := (\cdots \to 0 \to \Omega^0(M) \to \Omega^1(M) \to \cdots \to \Omega^k(M) \to \cdots),$$

where the nonzero terms are concentrated in negative degrees. Then H takes Ω^* to the spectrum:

$$H(\Omega^*(M)) = \begin{cases} \mathrm{DK}(\cdots \to 0 \to \Omega^0_{\mathrm{cl}}(M)), \\ \mathrm{DK}(\cdots \to 0 \to \Omega^0(M) \to \Omega^1_{\mathrm{cl}}(M) \to \dots), \\ \mathrm{DK}(\cdots \to 0 \to \Omega^0(M) \to \Omega^1(M) \to \Omega^2_{\mathrm{cl}}(M) \to \dots), \\ \quad\vdots \\ \mathrm{DK}(\cdots \to 0 \to \Omega^0(M) \to \Omega^1(M) \to \cdots \to \Omega^k_{\mathrm{cl}}(M) \to \dots), \\ \quad\vdots \end{cases}$$

By the basic properties of the Dold–Kan functor, the stable homotopy groups of this spectrum are computed as

$$\begin{aligned} \pi^s_n H(\Omega^*(M)) &\simeq \lim_{k\to\infty} \pi_{k+n}\, \mathrm{DK}(\cdots \to 0 \to \Omega^0(M) \to \Omega^1(M) \to \cdots \to \Omega^k_{\mathrm{cl}}(M)) \\ &\simeq \lim_{k\to\infty} H_{k+n}(\cdots \to 0 \to \Omega^0(M) \to \Omega^1(M) \to \cdots \to \Omega^k_{\mathrm{cl}}(M)). \end{aligned}$$

For $n > 0$, these groups are 0. For $n \le 0$, they are the n^{th} de Rham groups $H^n_{\mathrm{dR}}(M)$.

Now the functor H in (2-5) prolongs to a functor on prestacks

$$H\colon [\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}, \mathcal{C}\mathrm{h}] \to [\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}, \mathcal{S}\mathrm{p}].$$

In fact, using the properties of the Dold–Kan correspondence, it is fairly straightforward to show that this functor preserves local weak equivalences [10]. We therefore get a functor of smooth stacks

$$H\colon \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{C}\mathrm{h}) \to \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{S}\mathrm{p}). \tag{2-6}$$

Recall that for an Ω–spectrum $\mathcal{E}$, we always have a rational equivalence

$$\mathrm{r}\colon \mathcal{E} \wedge M\mathbb{R} \to H(\pi_*(\mathcal{E}) \otimes \mathbb{R}),$$

where $M\mathbb{R}$ denotes an Eilenberg–Moore spectrum. Now, since we are working over the site of Cartesian spaces, the Poincaré lemma implies that the inclusion $j\colon \mathbb{R}[0] \to \Omega^*$ induces an equivalence

$$\mathrm{id} \otimes j\colon \pi_*(\mathcal{E}) \otimes \mathbb{R}[0] \to \pi_*(\mathcal{E}) \otimes \Omega^*,$$

where $\pi_*(\mathcal{E}) = \mathcal{E}(*)$ (which follows from suspension).

Definition 17 Let $\mathcal{E}$ be a spectrum. For an unbounded chain complex $C_\bullet$, let $\tau_{\le 0} C_\bullet$ denote the truncated complex

$$\tau_{\le 0} C_\bullet = (\cdots \to 0 \to C_0 \to C_{-1} \to \cdots \to C_{-n} \to \cdots).$$

A differential function spectrum $\mathrm{diff}(\mathcal{E},\mathrm{ch})$ is a homotopy pullback

$$\begin{tikzcd}
\mathrm{diff}(\mathcal{E},\mathrm{ch}) \arrow[r] \arrow[d] & H(\tau_{\leq 0}\Omega^* \otimes \pi_*(\mathcal{E})) \arrow[d] \\
\mathcal{E} \arrow[r, "\mathrm{ch}"] & H(\Omega^* \otimes \pi_*(\mathcal{E}))
\end{tikzcd}$$

where $\mathrm{ch} = j \circ \mathrm{r}$, and j induces an equivalence $j\colon \pi_*(\mathcal{E}) \otimes \mathbb{R}[0] \xrightarrow{\simeq} \pi_*(\mathcal{E}) \otimes \Omega^*$.

Remark 18 In our definition, we have chosen the complex $\Omega^* \otimes \pi_*(\mathcal{E})$ as the de Rham complex modeling our rational theory. In general, the differential function spectrum depends on this choice and on the equivalence j [13]. For the purposes of clarity and utility, we will always choose this model, although other models can be treated analogously. We do, however, keep the dependence on the map ch explicit to emphasize this fact.

Example 19 (Deligne cohomology) Let $\mathcal{E} = H(\mathbb{Z}[n]) \simeq \Sigma^n H\mathbb{Z}$ be the n–fold suspension of the Eilenberg–Mac Lane spectrum. In unbounded chain complexes, we have a natural isomorphism

$$\underline{\mathbb{Z}}[n] \otimes \Omega^* \simeq \Omega^*[n],$$

where $\underline{\mathbb{Z}}[n]$ is the sheaf of locally constant integer-valued functions in degree n, and the complex on the right-hand side has been shifted up n units. That is, Ω^n is in degree 0, while Ω^0 is in degree n. Since $\Sigma^n H\mathbb{Z}$ is in the image of the Eilenberg–Mac Lane functor H, and H preserves homotopy pullbacks, the homotopy pullback

$$\begin{tikzcd}
\mathrm{diff}(\Sigma^n H\mathbb{Z},\mathrm{ch}) \arrow[r] \arrow[d] & H(\tau_{\leq 0}\Omega^*[n]) \arrow[d] \\
\Sigma^n H\underline{\mathbb{Z}} \arrow[r, "\mathrm{ch}"] & H(\Omega^*[n])
\end{tikzcd}$$

is presented by the homotopy pullback of unbounded chain complexes:

$$\begin{tikzcd}
\underline{\mathbb{Z}}[n] \times^h_{\Omega^*[n]} \tau_{\geq 0}\Omega^*[n] \arrow[r] \arrow[d] & \tau_{\leq 0}\Omega^*[n] \arrow[d] \\
\underline{\mathbb{Z}}[n] \arrow[r] & \Omega^*[n]
\end{tikzcd}$$

By stability, we can identify the homotopy pullback with the shifted mapping cone:

$$\underline{\mathbb{Z}}[n] \times^h_{\Omega^*[n]} \tau_{\leq 0}\Omega^*[n] \simeq \mathrm{cone}\big(\underline{\mathbb{Z}}[n] \oplus \tau_{\leq 0}\Omega^* \to \Omega^*[n]\big)[-1].$$

The right-hand side is precisely the Deligne complex $\mathbb{Z}_{\mathcal{D}}^{\infty}(n+1)$. We therefore have an equivalence

$$H(\mathbb{Z}_{\mathcal{D}}^{\infty}(n+1)) \simeq \mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch}).$$

The underlying theory this spectrum represents is precisely Deligne cohomology. In fact, by the Dold–Kan correspondence, we have an isomorphism of graded abelian groups

$$\pi_0 \hom_{\mathcal{C}\mathrm{h}}\big(N\big(C(\{U_i\}), \mathbb{Z}_{\mathcal{D}}^{\infty}(n+1)\big)\big) \simeq \pi_0 \operatorname{Map}(\Sigma^{\infty} M, \mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch})).$$

Here N denotes the normalized Moore complex (adjoint to the Dold–Kan functor DK) and $C(\{U_i\})$ denotes the Čech nerve of some good open cover of X. The right-hand side is simply the definition of $\mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch})^0(M)$, while the left-hand side is the shifted total complex of the Čech–Deligne double complex. It therefore computes the degree-n Deligne cohomology $H^n(M; \mathbb{Z}_{\mathcal{D}}^{\infty}(n+1))$.

The above example illustrates what exactly differential function spectra have to do with differential cohomology theories. The following definition can be found in [16].

Definition 20 Let $\mathcal{E}$ be a spectrum, and let

$$\mathrm{ch}\colon \mathcal{E} \to H(\tau_{\leq 0}\Omega^* \otimes \pi_*(\mathcal{E}))$$

be the Chern character map as in Definition 17. The *differential $\mathcal{E}$–cohomology of a manifold* is the smooth cohomology theory with degree-n component

$$\widehat{\mathcal{E}}^n(M) \simeq \mathrm{diff}(\Sigma^n \mathcal{E}, \mathrm{ch})^0(M).$$

Since $\mathrm{diff}(\Sigma^n \mathcal{E}, \mathrm{ch})$ is a smooth spectrum for each n, it fits into a diamond diagram of the form (2-2), as established in [16; 64]. In [16], it was shown that the form that this diamond takes is precisely the differential cohomology diamond in Remark 12. In particular, Proposition 10 allows us to calculate the $\mathrm{diff}(\Sigma^n \mathcal{E}, \mathrm{ch})$ cohomology in degrees away from 0 as

$$\mathrm{diff}(\Sigma^n \mathcal{E}, \mathrm{ch})^q(M) = \begin{cases} \mathcal{E}^{n+q}(M), & q > 0, \\ \mathcal{E}_{U(1)}^{n-1+q}(M), & q < 0. \end{cases}$$

3 The smooth Atiyah–Hirzebruch spectral sequence (AHSS)

In this section, we describe general machinery to construct an Atiyah–Hirzebruch spectral sequence (AHSS) from a smooth spectrum $\mathcal{E}$. We also describe how to compare this spectral sequence to the classical AHSS spectral sequence for the underlying theory $\Pi\mathcal{E}$, in nice cases.

3.1 Construction of the spectral sequence via Čech resolutions

The trick to describing the spectral sequence is to choose the right filtration on a fixed manifold. In the local (projective) model structure on smooth stacks, a natural choice arises: namely, the *Čech-type filtration on good open covers*. This is indeed the most natural choice since the maps which are weakly inverted in the local model structure are precisely those arising from taking the Čech nerve of a good open cover of a manifold. That is, we have a weak equivalence

$$w\colon \operatorname{hocolim}\Big\{ \cdots \coprod_{\alpha\beta\gamma} U_{\alpha\beta\gamma} \rightrightarrows \coprod_{\alpha\beta} U_{\alpha\beta} \rightrightarrows \coprod_{\alpha} U_{\alpha} \Big\} \to X.$$

We now explicitly describe a filtration on $C(\{U_i\})$. Recall that any simplicial diagram $J\colon \Delta^{\mathrm{op}} \to \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})$ can be filtrated by skeleta. More precisely, let $i\colon \Delta_{\le k} \hookrightarrow \Delta$ denote the embedding of the full subcategory of linearly ordered sets $[r]$ such that $r \le k$. Then i induces a restriction between functor categories (the k^{th} truncation)

$$\tau_{\le k}\colon [\Delta^{\mathrm{op}}, \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})] \to [\Delta^{\mathrm{op}}_{\le k}, \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})].$$

By general abstract nonsense (the existence of left and right Kan extensions), there are left and right adjoints $(\mathrm{sk}_k \dashv \tau_{\le k} \dashv \mathrm{cosk}_k)$

$$\begin{tikzcd}[\Delta^{\mathrm{op}}, \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})] \arrow[r, "\tau_{\le k}"] & {[\Delta^{\mathrm{op}}_{\le k}, \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})]} \arrow[l, bend right, "\mathrm{sk}_k"'] \arrow[l, bend left, "\mathrm{cosk}_k"]\end{tikzcd}.$$

Furthermore, by composing adjoints, we have an adjunction $(\mathrm{sk}_k \dashv \mathrm{cosk}_k)$

$$\begin{tikzcd}[\Delta^{\mathrm{op}}, \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})] \arrow[r, shift left, "\mathrm{sk}_k"] & {[\Delta^{\mathrm{op}}, \mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})]} \arrow[l, shift left, "\mathrm{cosk}_k"]\end{tikzcd}.$$

The functor sk_k freely fills in degenerate simplices above level k, while cosk_k probes a simplicial object with simplices only up to level k (the singular k–skeleton).

Proposition 21 *Let $Y_\bullet$ be a simplicial object in $\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})$. Then we can filter $Y_\bullet$ by skeleta*

$$\mathrm{sk}_0 Y_\bullet \to \mathrm{sk}_1 Y_\bullet \to \cdots \to \mathrm{sk}_k Y_\bullet \to \cdots \to Y_\bullet.$$

The homotopy colimit over $Y_\bullet$ is presented by the ordinary colimit

$$\operatorname*{hocolim}_{\Delta^{\mathrm{op}}}(Y_\bullet) \simeq \operatorname*{colim}_{k\to\infty} \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}}(\mathrm{sk}_k Y_\bullet),$$

where $\mathbb{L}\mathrm{colim}$ is the left derived functor of the colimit, hence computable upon suitable cofibrant replacement of the diagram.[1]

[1] We take this particular model of the homotopy colimit in order to ensure that taking the colimit of the resulting diagram makes sense. The claim will also hold for other presentations of the homotopy colimit.

Proof Since $\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})$ is presented by a combinatorial simplicial model category, the homotopy colimit over a filtered diagram is presented by the ordinary colimit, and the canonical map

$$\mathbb{L}\operatorname*{colim}_{k\to\infty} \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}}(\mathrm{sk}_k Y_\bullet) \to \operatorname*{colim}_{k\to\infty} \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}}(\mathrm{sk}_k Y_\bullet)$$

is an equivalence. Since homotopy colimits commute with homotopy colimits, we also have an equivalence

$$\mathbb{L}\operatorname*{colim}_{k\to\infty} \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}}(\mathrm{sk}_k Y_\bullet) \simeq \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}} \mathbb{L}\operatorname*{colim}_{k\to\infty}(\mathrm{sk}_k Y_\bullet).$$

Again, using the fact that the ordinary colimit over a filtered diagram presents the homotopy colimit, we have an equivalence

$$\mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}} \mathbb{L}\operatorname*{colim}_{k\to\infty}(\mathrm{sk}_k Y_\bullet) \to \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}} \operatorname*{colim}_{k\to\infty}(\mathrm{sk}_k Y_\bullet) \simeq \mathbb{L}\operatorname*{colim}_{\Delta^{\mathrm{op}}}(Y_\bullet). \qquad \square$$

Remark 22 The above proposition says that the homotopy colimit over the simplicial object is filtered by homotopy colimits of its skeleta. In particular, if M is a paracompact manifold, we can fix a good open cover on M and form the simplicial object given by its Čech nerve

$$C(\{U_i\}) := \cdots \begin{smallmatrix}\longrightarrow\\ \longleftarrow\\ \longrightarrow\\ \longleftarrow\\ \longrightarrow\\ \longleftarrow\\ \longrightarrow\end{smallmatrix} \coprod\nolimits_{\alpha\beta\gamma} U_{\alpha\beta\gamma} \begin{smallmatrix}\longrightarrow\\ \longleftarrow\\ \longrightarrow\\ \longleftarrow\\ \longrightarrow\end{smallmatrix} \coprod\nolimits_{\alpha\beta} U_{\alpha\beta} \begin{smallmatrix}\longrightarrow\\ \longleftarrow\\ \longrightarrow\end{smallmatrix} \coprod\nolimits_{\alpha} U_{\alpha}.$$

The homotopy colimit over this object is then filtered by its skeleta.

Let us see exactly what the skeleta look like in this case. To this end, we recall that in $\mathrm{Sh}_\infty(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p})$, the full homotopy colimit is presented by the local homotopy formula

$$\operatorname*{hocolim}_{\Delta^{\mathrm{op}}} C(\{U_i\}) = \int^{n\in\Delta} \coprod_{\alpha_0\cdots\alpha_n} U_{\alpha_0\cdots\alpha_n} \odot \Delta[n].$$

The filtration on this object is given by first truncating the Čech nerve and then freely filling in degenerate simplices. As a consequence, in degree k, we can forget about the simplices of dimension higher than k. The homotopy colimit over this skeleton is then given by a strict colimit over the diagram

$$\begin{aligned}\text{(3-1)}\quad \coprod\nolimits_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k} \odot \Delta[k] \cdots &\begin{smallmatrix}\longrightarrow\\ \longleftarrow\\ \longrightarrow\\ \longleftarrow\\ \longrightarrow\\ \longleftarrow\\ \longrightarrow\end{smallmatrix} \coprod\nolimits_{\alpha\beta\gamma} U_{\alpha\beta\gamma} \odot \Delta[2] \\ &\begin{smallmatrix}\longrightarrow\\ \longleftarrow\\ \longrightarrow\\ \longleftarrow\\ \longrightarrow\end{smallmatrix} \coprod\nolimits_{\alpha\beta} U_{\alpha\beta} \odot \Delta[1] \begin{smallmatrix}\longrightarrow\\ \longleftarrow\\ \longrightarrow\end{smallmatrix} \coprod\nolimits_{\alpha} U_{\alpha} \odot \Delta[0],\end{aligned}$$

where the face and degeneracy maps are induced by the face and degeneracy maps of $\Delta[k]$. Taking $k\to\infty$, we do indeed reproduce the coend representing the full homotopy colimit $C(\{U_i\})$.

We would like to eventually use this filtration to define a Mayer–Vietoris like spectral sequence for general cohomology theory $\mathcal{E}$. To get to this step, however, we will need to identify the successive quotients of the filtration. To simplify notation in what follows, we will fix a manifold M with Čech nerve $C(\{U_i\})$, and we set

$$X_k := \operatorname*{hocolim}_{\Delta^{\mathrm{op}}}\big(\mathrm{sk}_k C(\{U_i\})\big).$$

Then the quotient X_k/X_{k-1} can be identified from the previous discussion by quotienting out the face maps at level k described in diagram (3-1). Since the tensor of a simplicial set and a stack is given by the product of the stack with the discrete inclusion of the simplicial set, we can identify the quotient from the pushout of coends

$$\begin{array}{ccc} \displaystyle\int^{n<k} \coprod_{\alpha_0\cdots\alpha_n} U_{\alpha_0\cdots\alpha_n} \times \mathrm{disc}(\Delta[n]) & \longrightarrow & * \\ \Big\downarrow{\scriptstyle \partial} & & \\ \displaystyle\int^{m\le k} \coprod_{\alpha_0\cdots\alpha_m} U_{\alpha_0\cdots\alpha_m} \times \mathrm{disc}(\Delta[m]) & & \end{array}$$

where ∂ denotes the boundary inclusion. At the level of points (or elements), a simplex in $\int^{n<k} \coprod_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_n} \times \mathrm{disc}(\Delta[n])$ is given by a pair

$$(\rho,\sigma) \in \coprod_{\alpha_0\cdots\alpha_{k-1}} U_{\alpha_0\cdots\alpha_{k-1}} \times \mathrm{disc}(\Delta[k-1]),$$

which is glued to lower simplices via the face and degeneracy relations.

Let us identify where the boundary inclusion takes a generic simplex. Then the quotient X_k/X_{k-1} will be obtained by gluing these simplices together to a single point. Note that the face and degeneracy relations imply that simplices of the form $(\rho, s_{j+1}\sigma)$ are sent by d_j to $(d_j\rho,\sigma)$. Since simplices in the image of the face maps are precisely those which are collapsed to a point, we see that

$$(d_j\rho,\sigma) \sim * \quad \text{for every } \sigma.$$

We therefore see that each term of the coproduct $\coprod_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k}$ is joined to another by the inclusion into a lower intersection. These lower intersections are then collapsed to a point yielding the wedge product

$$\bigvee\nolimits_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k} \subset X_k/X_{k-1}.$$

Similarly, the simplex $(s_{j+1}\rho,\sigma)$ is sent to $(\rho, d_j\sigma)$ under d_j. We therefore identify the discrete simplicial sphere in the quotient

$$\mathrm{disc}\big(\Delta[k]/\partial\Delta[k]\big) \subset X_k/X_{k-1}.$$

Finally, the relations imposed by the coend imply that a simplex of the form $(s_j\rho, \sigma)$ is glued to $(\rho, d_j\sigma)$. The former are precisely those simplices in the simplicial sphere, while the latter are glued to the point. Similarly, $(\rho, s_j\sigma)$ is glued to the point. Thus we have the following.

Lemma 23 *We can identify the quotient with the smash product:*

$$X_k/X_{k-1} \simeq \mathrm{disc}(\Delta[k]/\partial\Delta[k]) \wedge \bigvee_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k} \simeq \Sigma^k\big(\bigvee_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k}\big).$$

Remark 24 (the filtration as a natural choice) Another way to think of our filtration above is the following. Let us form a Čech nerve of a manifold, then contract all the patches and intersections in that Čech nerve as points, so we obtain a simplicial set. Then *Borsuk's nerve theorem* (see [6] for a survey, [36, Corollary 4G.3] or [57, Theorem 3.21]) says that this simplicial set is equivalent — weak homotopy equivalent — to the singular simplicial complex of the manifold, hence to its homotopy type. Moreover, that singular simplicial complex (or rather its geometric realization) in turn gives a CW-complex realization of the original manifold. So with this in mind, one may view our filtration above as the natural smooth refinement of the filtration by CW-stages of the manifold. That is, in taking the Čech nerve *without* contracting all its patches to points, we retain exactly the smooth information that, via Borsuk's theorem, corresponds to each cell in the canonical CW-complex incarnation of the manifold. So in this sense, our refinement can be viewed as the canonical smooth refinement of the traditional filtering by CW-stages.

We are now ready to describe the spectral sequence.

Theorem 25 (AHSS for general smooth spectra) *Let M be a compact smooth manifold, and let $\mathcal{E}$ be a smooth spectrum. There is a spectral sequence with*

$$E_2^{p,q} = H^p(M, \mathcal{E}^q) \implies \mathcal{E}^{p+q}(M).$$

Here H^p denotes the p^{th} Čech cohomology with coefficients in the presheaf $\mathcal{E}^q$. Moreover, the differential on the E_1–page is given by the differential in Čech cohomology.

Proof The proof is almost immediate from the definitions. Recall that we have identified the quotients in Lemma 23. By the axioms for a smooth cohomology theory, we have that the $\mathcal{E}$–cohomology of the quotient is given by

$$\begin{aligned}\mathcal{E}^*(X_k/X_{k-1}) &\simeq \mathcal{E}^*\big(\Sigma^k\big(\textstyle\bigvee_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k}\big)\big)\\ &\simeq \mathcal{E}^{*-k}\big(\textstyle\bigvee_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k}\big)\\ &\simeq \bigoplus_{\alpha_0\cdots\alpha_k} \mathcal{E}^{*-k}(U_{\alpha_0\cdots\alpha_k}).\end{aligned}$$

Applying $\mathcal{E}^{p+q}$ to the cofiber sequence $X_p \hookrightarrow X_{p+1} \twoheadrightarrow X_{p+1}/X_p$ gives the long exact sequence in $\mathcal{E}$–cohomology

$$(3\text{-}2)\quad \cdots \to \mathcal{E}^{p+q}(X_{p+1}/X_p) \to \mathcal{E}^{p+q}(X_{p+1}) \to \mathcal{E}^{p+q}(X_p) \to \mathcal{E}^{p+q+1}(X_{p+1}/X_p) \to \cdots .$$

Forming the corresponding exact triangle, we get a spectral sequence with $E_1^{p,q}$–term

$$E_1^{p,q} = \bigoplus_{\alpha_0,\cdots,\alpha_p} \mathcal{E}^q(U_{\alpha_0\cdots\alpha_p}).$$

Now we want to show that the differential on this page is given by the Čech differential

$$\delta\colon E_1^{p,q} = \bigoplus_{\alpha_0\cdots\alpha_p} \mathcal{E}^q(U_{\alpha_0\cdots\alpha_p}) \to \bigoplus_{\alpha_0\cdots\alpha_{p+1}} \mathcal{E}^q(U_{\alpha_0\cdots\alpha_{p+1}}) = E^{p+1,q}.$$

To this end, note that differential on the E_1–page, by definition, comes from the exact sequence

$$\cdots \to \mathcal{E}^{p+q}(X_{p+1}/X_p) \xrightarrow{j} \mathcal{E}^{p+q}(X_{p+1}) \xrightarrow{i} \mathcal{E}^{p+q}(X_p) \xrightarrow{\partial} \mathcal{E}^{p+q+1}(X_{p+1}/X_p) \to \cdots .$$

We need to show that $\partial j = d_1 = \delta$ is the Čech differential. By naturality of the connecting homomorphism ∂, we have a commutative diagram

$$\begin{tikzcd}
\check{C}^{p-1}(M;\mathcal{E}^q) \arrow[rr, "d_1"] \arrow[d, "\simeq"'] & & \check{C}^p(M;\mathcal{E}^q) \arrow[d, "\simeq"] \\
\bigoplus_{\alpha_0\cdots\alpha_{p-1}} \mathcal{E}^q(U_{\alpha_0\cdots\alpha_{p-1}}) \arrow[rr] \arrow[d, "\simeq"'] & & \bigoplus_{\alpha_0\cdots\alpha_p} \mathcal{E}^q(U_{\alpha_0\cdots\alpha_p}) \arrow[d, "\simeq"] \\
\mathcal{E}^{p+q-1}(X_{p-1}/X_{p-2}) \arrow[r, "j"] \arrow[d] & \mathcal{E}^{p+q-1}(X_{p-1}) \arrow[r, "\partial"] \arrow[dd] & \mathcal{E}^{p+q}(X_p/X_{p-1}) \arrow[d] \\
\mathcal{E}^{p+q-1}(\partial\Delta[p]\times U_{\alpha_0\cdots\alpha_{p-1}}) \arrow[dr, "\mathrm{id}"] & & \mathcal{E}^{p+q}(\Delta[p]/\partial\Delta[p]\wedge U_{\alpha_0\cdots\alpha_p}) \\
& \mathcal{E}^{p+q-1}(\partial\Delta[p]\times U_{\alpha_0\cdots\alpha_{p-1}}) \arrow[ur, "\partial"] &
\end{tikzcd}$$

where the vertical bottom maps are induced from the inclusion of a factor

$$(3\text{-}3)\quad \begin{tikzcd}
\Delta[p]\times U_{\alpha_0\cdots\alpha_p} \arrow[r, hook] & X_p \\
\partial\Delta[p]\times U_{\alpha_0\cdots\alpha_{p-1}} \arrow[r, hook] \arrow[u] & X_{p-1} \arrow[u] \\
\varnothing \arrow[r] \arrow[u] & X_{p-2} \arrow[u]
\end{tikzcd}$$

into the p–level of the filtration. Comparing the top and bottom composite morphisms in the big diagram, we see that on $(p-1)$–fold intersections $U_{\alpha_0\cdots\alpha_{p-1}}$, the map d_1 is forced to map a section to the alternating sum of restrictions, as this is precisely the map induced by the boundary inclusion in (3-3).

All that remains is the convergence. To establish that, we simply note that compactness implies that, for large values of p, we have an equivalence $X_p \simeq X$. Moreover, there are only finitely many diagonal entries at each page of the sequence. With this assumption, the convergence to the corresponding graded complex

$$E_\infty^{p,q} = \frac{\ker\big(\mathcal{E}^{p+q}(X) \to \mathcal{E}^{p+q}(X_p)\big)}{\ker\big(\mathcal{E}^{p+q}(X) \to \mathcal{E}^{p+q}(X_{p+1})\big)} = \frac{F_p\mathcal{E}^{p+q}(X)}{F_{p+1}\mathcal{E}^{p+q}(X)}$$

follows exactly as in the classical case in [3]. □

Fiber bundles We can also construct a spectral sequence for a fiber bundle

$$F \to N \xrightarrow{p} M,$$

where each map is a smooth map of manifolds and M is compact. To that end, we note that for a fixed good open cover $\{U_i\}$ of M, the pullbacks $\{p^{-1}(U_i)\}$ define a good open cover of N. By local triviality, we have that each $p^{-1}(U_i) \simeq F \times U_i$. Then, using the filtration

$$X_k = \operatorname*{hocolim}_{\Delta^{\mathrm{op}}}\big(\mathrm{sk}_k C(\{p^{-1}(U_i)\})\big)$$

on the total space N, we identify the successive quotients

$$X_k/X_{k-1} \simeq \Sigma^k \textstyle\bigvee_{\alpha_0\cdots\alpha_k} U_{\alpha_0\cdots\alpha_k} \wedge F.$$

A similar argument as in the proof of Theorem 25 gives:

Theorem 26 (smooth AHSS for fiber bundles) *Let M, N and F be manifolds with M compact. Let $F \to N \xrightarrow{p} M$ be a fiber bundle. Let $\mathcal{E}$ be a sheaf of spectra. Then there is a spectral sequence*

$$E_2^{p,q} = H^p(M, \mathcal{E}^q(- \wedge F)) \implies \mathcal{E}^{p+q}(N).$$

Here H^p denotes the p^{th} Čech cohomology with coefficients in the presheaf $\mathcal{E}^{-q}(-\wedge F)$.

Remark 27 (unreduced theories) Note that the smooth spectral sequence works for reduced theories. One can treat unreduced theories similarly by setting

$$\mathcal{E}^q(M, *) := \widetilde{\mathcal{E}}^q(M_+),$$

where the tilde denotes the reduced theory and M_+ is the pointed stack with basepoint $*$. In this case, we have the slight modification on the second spectral sequence, which takes the form

$$E_2^{p,q} = H^p(M, \mathcal{E}^q(- \times F)) \implies \mathcal{E}^{p+q}(N).$$

3.2 Morphisms of smooth spectral sequences and refinement of the AHSS

Our next task will be to show that these spectral sequences do indeed refine the classical Atiyah–Hirzebruch spectral sequence (AHSS) [3]. Since any smooth theory $\mathcal{E}$ comes as a refinement of the underlying topological theory $\Pi\mathcal{E}$, we will immediately get a morphism of spectral sequences induced by the morphism of spectra

$$I\colon \mathcal{E} \to \Pi\mathcal{E}.$$

Unfortunately, this morphism does not allow us to compare the differentials of the spectral sequences in the way that we would ideally hope for. However, as we will progressively see, the situation can be remedied by constructing a slightly different morphism of spectral sequences. This morphism is related to the *boundary map* of spectral sequences which occurs when a morphism of spectra induces the 0 map on corresponding spectral sequences; see [55] for a discussion in the case of the Adams spectral sequence. We first discuss the morphism induced by I, then construct this "boundary-type" map, and prove that it indeed defines a morphism of spectral sequences.

Definition 28 Let $E_n^{p,q}$ and $F_n^{p,q}$ be spectral sequences, that is, a sequence of bigraded complexes $E_n^{p,q}$ and $F_n^{p,q}$, $n \in \mathbb{N}$. A *morphism of spectral sequences* is a morphism of bigraded complexes

$$f_n\colon E_n^{p,q} \to F_n^{p,q},$$

defined for all $n > N$, where N is some fixed positive integer. Furthermore, we require the map f_{n+1} to be the map on homology induced by f_n. We call the smallest integer N such that f_n are defined for $n > N$ the *rank* of the morphism.

We now apply this to the smooth AHSS. The next result should follow from general principles, but we emphasize it explicitly for clarity and for subsequent use.

Proposition 29 *Let $\mathcal{E}$ and $\mathcal{F}$ be smooth spectra. Then a map $f\colon \mathcal{E} \to \mathcal{F}$ induces a morphism of corresponding smooth AHSSs*

$$E_n^{p,q} \to F_n^{p,q}.$$

Proof Fix a manifold X and a good open cover $\{U_i\}$. Let X_p denote the p^{th} filtration of the Čech nerve as before. It is clear by naturality that a map of spectra $f\colon \mathcal{E} \to \mathcal{F}$ induces a morphism of long exact sequences (see (3-2))

$$\begin{tikzcd}
\cdots \mathcal{E}^{p+q}(X_{p+1}/X_p) \arrow[r] \arrow[d] & \mathcal{E}^{p+q}(X_{p+1}) \arrow[r] \arrow[d] & \mathcal{E}^{p+q}(X_p) \arrow[r] \arrow[d] & \mathcal{E}^{p+q+1}(X_{p+1}/X_p) \cdots \arrow[d] \\
\cdots \mathcal{F}^{p+q}(X_{p+1}/X_p) \arrow[r] & \mathcal{F}^{p+q}(X_{p+1}) \arrow[r] & \mathcal{F}^{p+q}(X_p) \arrow[r] & \mathcal{F}^{p+q+1}(X_{p+1}/X_p) \cdots
\end{tikzcd}$$

It follows immediately from the construction of the corresponding exact triangles that this morphism commutes with the differentials. □

This now allows us to compare the topological and the smooth theories.

Corollary 30 *Let $\mathcal{E}$ be a smooth spectrum and $\Pi\mathcal{E}$ the underlying topological theory. Let E_n and F_n denote the spectral sequences corresponding to $\mathcal{E}$ and $\Pi\mathcal{E}$, respectively. The natural map $I\colon \mathcal{E} \to \Pi\mathcal{E}$ induces a morphism of classical AHSSs*[2]

$$I\colon E_n^{p,q} \to F_n^{p,q}.$$

Remark 31 It is interesting that the smooth spectrum $\Pi\mathcal{E}$ is, by definition, locally constant. From the discussion around (2-1), this means that we have an isomorphism

$$\Pi E^q(U) \simeq \pi_{-q}\operatorname{Map}(U, \Pi\mathcal{E}) \simeq \pi_{-q}\operatorname{Map}(*, \Pi\mathcal{E}) \simeq \pi_{-q}\Pi\mathcal{E} \simeq \Pi\mathcal{E}^q(*)$$

for every element of a good open cover (or higher intersection) U. This connects, via Borsuk's theorem mentioned in Remark 24 above, the "smooth AHSS for locally constant coefficients" with the classical AHSS: the locally constant coefficients see each (contractible) patch as a point, and hence by Borsuk's theorem, they see our "Čech filtration" to be the classical CW-cell filtration.

From the construction of our smooth AHSS, it directly follows that the spectral sequence associated to the smooth spectrum is a refinement of the classical topological AHSS.

Corollary 32 *The spectral sequence $F_n^{p,q}$ is precisely the AHSS for the cohomology theory $\Pi\mathcal{E}$.*

[2]Here we have an unfortunate conflict of notation. We are using the same symbols for the pages in the spectral sequences for both the classical and the refined theories. We will aim to make the context explicit whenever a possible ambiguity arises.

We now would like to apply the above machinery to differential cohomology theories. In particular, we note that for a differential function spectrum $\mathrm{diff}(\mathcal{E},\mathrm{ch})$, the natural map

$$I\colon \mathrm{diff}(\mathcal{E},\mathrm{ch}) \to \underline{\mathcal{E}},$$

which strips the differential theory of the differential data and maps to the bare underlying theory, is precisely the map induced by the unit $I\colon \mathrm{id} \to \Pi$. In the above discussion, we observed that this map always induces a morphism of spectral sequences. Moreover, the target spectral sequence is exactly the AHSS for the underlying topological theory. One might hope to be able to use this map to compare the differentials in the refined theory with those differentials in the classical AHSS.

Unfortunately, this does not work in practice, as we will see when we discuss applications in Section 4. The core issue is that the spectral sequence for the refined theory usually ends up shifted with respect to the classical AHSS. As a consequence, the nonzero terms in each sequence are interlaced with respect to one another, and the map I ends up killing all the nonzero terms. This, in turn, stems from the appearance of the Bockstein map (which raises degree by 1) in the differential cohomology diagram.

However, there is often a different map between the *lower quadrants* of the two spectral sequences corresponding to $\mathrm{diff}(\mathcal{E},\mathrm{ch})$ and $\mathcal{E}$, which lowers the degree as to match the corresponding nonzero entries. This map is related to the so-called *boundary map* between spectral sequences studied in [55]. The next proposition concerns this map and will be essential for comparing the differentials in the refined theory to those of the classical theory.

Proposition 33 (i) *Let $\mathcal{E}$ be a spectrum such that $\pi_*(\mathcal{E})$ is concentrated in degrees which are a multiple of some integer $n \geq 2$ (eg K-theory, Morava K-theory). Suppose, moreover, that $\pi_*(\mathcal{E})$ is projective in those degrees. Then the sequence of spectra*

$$\mathcal{E} \to \mathcal{E} \wedge M\mathbb{R} \to \mathcal{E} \wedge MU(1) \xrightarrow{\beta_{\mathcal{E}}} \Sigma\mathcal{E}$$

induces a short exact sequence on coefficients

$$0 \to \pi_*(\mathcal{E}) \to \pi_*(\mathcal{E}) \otimes \mathbb{R} \to \pi_*(\mathcal{E}) \otimes U(1) \to 0. \tag{3-4}$$

(ii) *Let β denote the connecting homomorphism (ie the Bockstein) for the coefficient sequence (3-4). Let $E_n^{p,q}$ denote the spectral sequence corresponding to $\Sigma^{-1}\mathcal{E} \wedge MU(1)$ and let $F_n^{p,q}$ denote the spectral sequence corresponding to $\mathcal{E}$. Then*

$$\beta\colon E_n^{p,q} \to F_n^{p,q}$$

induces a morphism of spectral sequences of rank 2.

Proof Consider the long Bockstein sequence

$$\cdots \to \mathcal{E} \xrightarrow{r} \mathcal{E} \wedge M\mathbb{R} \xrightarrow{e} \mathcal{E} \wedge MU(1) \xrightarrow{\beta_{\mathcal{E}}} \Sigma\mathcal{E} \to \cdots$$

induced by the cofiber sequence

$$\mathbb{S} \to M\mathbb{R} \to MU(1).$$

Fix a manifold M and let X_p denote the p–level of the Čech filtration. Now each spectrum in the above sequence has a long exact sequence induced by the cofiber sequences

$$X_{p-1} \to X_p \to X_p/X_{p-1},$$

from which one builds the exact couple for the corresponding spectral sequence. Using the properties of $\pi_*(\mathcal{E})$ along with this sequence, we can fit the long exact sequences into a diagram

$$\begin{tikzcd}
\check{C}^p(X;\pi_{-q-1}(\mathcal{E})) \arrow[r, "q^*"] \arrow[d, "r"] & \mathcal{E}^{p+q-1}(X_p) \arrow[r, "i^*"] \arrow[d, "r"] & \mathcal{E}^{p+q-1}(X_{p-1}) \arrow[r, "\partial"] \arrow[d, "r"] & 0 \arrow[d] \\
\check{C}^p(X;\pi_{-q-1}(\mathcal{E}_{\mathbb{R}})) \arrow[r, "q^*"] \arrow[d, "e"] & \mathcal{E}_{\mathbb{R}}^{p+q-1}(X_p) \arrow[r, "i^*"] \arrow[d, "e"] & \mathcal{E}_{\mathbb{R}}^{p+q-1}(X_{p-1}) \arrow[r, "\partial"] \arrow[d, "e"] & 0 \arrow[d] \\
\check{C}^p(X;\pi_{-q-1}(\mathcal{E}_{U(1)})) \arrow[r, "q^*"] \arrow[d, "\beta_{\mathcal{E}}"] & \mathcal{E}_{U(1)}^{p+q}(X_p) \arrow[r, "i^*"] \arrow[d, "\beta_{\mathcal{E}}"] & \mathcal{E}_{U(1)}^{p+q}(X_{p-1}) \arrow[r, "\partial"] \arrow[d, "\beta_{\mathcal{E}}"] & 0 \arrow[d] \\
0 \arrow[r] & \mathcal{E}^{p+q}(X_{p+1}) \arrow[r] & \mathcal{E}^{p+q}(X_p) \arrow[r] & \check{C}^p(X;\pi_{-q+1}(\mathcal{E}))
\end{tikzcd}$$

where both the rows and columns are part of exact sequences, and $\check{C}^p(X;A)$ denotes the group of Čech p–cochains with coefficients in A. Since everything commutes, this induces a corresponding short exact sequence of E_1–pages. At each (p,q)–entry, this sequence is given by

$$0 \to C^p(X;\pi_{-q}(\mathcal{E})) \to C^p(X;\pi_{-q}(\mathcal{E})\otimes\mathbb{R}) \to C^p(X;\pi_{-q}(\mathcal{E})\otimes U(1)) \to 0.$$

Since the differentials on the E_1–page are precisely the Čech differentials, the construction of the Bockstein map in Čech cohomology will produce a map of E_2–pages

$$\beta\colon H^p(X;\pi_{-q}(\mathcal{E})\otimes U(1)) \to H^{p+1}(X;\pi_{-q}(\mathcal{E})).$$

We need to show that this map commutes with the differential. Choose a representative x of a class in $H^p(X;\pi_{-q}(\mathcal{E})\otimes U(1))$. By definition, $y=\beta(x)$ is a class such that $r(y)=\delta(\overline{x})$, where $\overline{x}$ is such that $e(\overline{x})=x$. Then

$$r(d_2 y) = d_2 r(y) = d_2\delta(\overline{x}).$$

We want to show that there is a lift z of d_2x such that $\delta(z) = d_2\delta(\overline{x})$. Indeed, if this is the case, then d_2y represents $\beta(d_2x)$ and we are done.

To construct z, recall that d_2x is defined by first pulling back by the quotient q, which lies in the image of the map induced by the inclusion $i\colon X_p \hookrightarrow X_{p+1}$, and then applying the boundary to an element of the preimage. Let w be such that

$$i^*(w) = q^*(x).$$

Chasing the diagram

$$\begin{tikzcd}
\check{C}^p(X;\pi_{-q-1}(\mathcal{E}_{\mathbb{R}})) \arrow[r, "q^*"] \arrow[d, "e"] & \mathcal{E}_{\mathbb{R}}^{p+q-1}(X_p) \arrow[r, "e"] \arrow[d, "e"] & \mathcal{E}_{\mathbb{R}}^{p+q-1}(X_{p-1}) \arrow[d, "e"] \\
\check{C}^p(X;\pi_{-q-1}(\mathcal{E}_{U(1)})) \arrow[r, "q^*"] \arrow[d, "\beta_{\mathcal{E}}"] & \mathcal{E}_{U(1)}^{p+q}(X_p) \arrow[r, "i^*"] \arrow[d, "\beta_{\mathcal{E}}"] & \mathcal{E}_{U(1)}^{p+q}(X_{p-1}) \arrow[d, "\beta_{\mathcal{E}}"] \\
0 \arrow[r] & \mathcal{E}^{p+q}(X_p) \arrow[r, "i^*"] & \mathcal{E}^{p+q}(X_{p-1})
\end{tikzcd}$$

we see that $0 = \beta_{\mathcal{E}}q^*(x) = \beta_{\mathcal{E}}i^*(w) = i^*(\beta_{\mathcal{E}}w)$. By exactness of the rows, this implies that $\beta_{\mathcal{E}}w = 0$. Therefore, there is a class $\overline{w} \in \mathcal{E}_{\mathbb{R}}^{p+q+1}(X_{p+1})$ such that $e(\overline{w}) = w$.

Now, by definition of the differential, we have

$$e(\partial\overline{w}) = \partial(e(\overline{w})) = \partial w = d_2x,$$

and $z := \partial\overline{w}$ is a lift of d_2x. Using the fact that $\delta = d_1 = \partial q^*$, we have

$$\delta(z) = \delta(\partial\overline{w}) = \partial(q^*\partial\overline{w}).$$

By exactness, we have

$$i^*(q^*\partial\overline{w}) = 0 = q^*\partial q^*(\overline{x}) = q^*(\delta(\overline{x})),$$

and it follows from the definition that $\delta(z) = d_2(\delta(\overline{x}))$.

To show that $H^*(\beta)$ commutes with the higher differentials, we proceed by induction. The above discussion proves the base case. Suppose β induces a map $H_n(\beta)$ on E_n which commutes with d_n. Then $H^n(\beta)$ induces a well-defined map $H^{n+1}(\beta)$ on the E_{n+1}–page. Let $x \in \bigcap_{i=1}^n \ker(d_{n+1})$ be a representative of a class on the E_n–page. Then by definition, $H^{n+1}(\beta)(x) = \beta(x)$, and the exact same argument as before (replacing d_2 with d_{n+1}) gives the result. $\square$

Having done the heavy lifting in the above proposition, we will now apply this to straightforwardly relate the differentials of the refined theory to those of the underlying topological theory. This will use an explicit alternative to the map I, along the lines of the discussion preceding Proposition 33.

Theorem 34 (refinement of differentials) *Let $\mathcal{E}$ be a spectrum satisfying the properties of Proposition 33, and let* $\mathrm{diff}(\mathcal{E},\mathrm{ch})$ *be a differential function spectrum refining $\mathcal{E}$. Let E_n and F_n denote the smooth AHSSs corresponding to* $\mathrm{diff}(\mathcal{E},\mathrm{ch})$ *and $\mathcal{E}$, respectively. Then the Bockstein β defines a rank-2 morphism of fourth quadrant spectral sequences*

$$\beta\colon E_n^{p,q}\to F_n^{p,q},\quad q<0.$$

Proof Recall that for $q<0$, Proposition 10 implies that $\mathrm{diff}(\mathcal{E},\mathrm{ch})^q(M)\simeq \mathcal{E}^{q-1}_{U(1)}(M)$. The claim then follows from the previous proposition. □

3.3 Product structure and the differentials

Let $\mathcal{E}$ be an E_∞ ring spectrum. Then the associative graded-commutative product on $\mathcal{E}^*$ induces a product (associative and graded-commutative) on the refinement $\mathrm{diff}(\Sigma^n\mathcal{E},\mathrm{ch})^*$, that is, a map

$$\cup\colon \mathrm{diff}(\Sigma^n\mathcal{E},\mathrm{ch})^k\otimes \mathrm{diff}(\Sigma^m\mathcal{E},\mathrm{ch})^j\to \mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{k+j} \tag{3-5}$$

(see [13; 71]). The goal of this section will be to establish the following very useful property, in analogy with the classical case.

Proposition 35 (compatibility with products) *The product*

$$\cup\colon \mathrm{diff}(\Sigma^n\mathcal{E},\mathrm{ch})^k\otimes \mathrm{diff}(\Sigma^m\mathcal{E},\mathrm{ch})^j\to \mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{k+j}$$

induces a morphism of spectral sequences

$$\cup\colon E_*(n)\times E_*(m)\to E_*(n+m).$$

Moreover, the differentials satisfy the Leibniz rule

$$d(xy)=d(x)y+(-1)^{p+q}xd(y).$$

Let us first work out what the cup product pairing is on the E_1–page. Recall from the construction of the spectral sequence that $E_1^{p,q}$ is given by

$$E_1^{p,q}=\bigoplus_{\alpha_0\cdots\alpha_p}\mathrm{diff}(\Sigma^n\mathcal{E},\mathrm{ch})^q(U_{\alpha_0\cdots\alpha_p})\simeq \check{C}^p(M;\mathrm{diff}(\Sigma^n\mathcal{E},\mathrm{ch})^q).$$

Using the product (3-5), we get a cross product map

$$\begin{aligned}\times\colon \bigoplus_{\alpha_0\cdots\alpha_p}\mathrm{diff}(\Sigma^n\mathcal{E},\mathrm{ch})^q(U_{\alpha_0\cdots\alpha_p})\times\bigoplus_{\alpha_0\cdots\alpha_r}\mathrm{diff}(\Sigma^m\mathcal{E},\mathrm{ch})^t(U_{\alpha_0\cdots\alpha_r})\\ \to\bigoplus_{\alpha_0\cdots\alpha_p}\bigoplus_{\alpha_0\cdots\alpha_r}\mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}(U_{\alpha_0\cdots\alpha_p}\times U_{\alpha_0\cdots\alpha_r}).\end{aligned} \tag{3-6}$$

We also have an isomorphism

$$\begin{aligned}\bigoplus_{\alpha_0\cdots\alpha_s} \mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}((U\times U)_{\alpha_0\cdots\alpha_s}) \\ &\simeq \mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}\big(\textstyle\bigvee_{\alpha_0\cdots\alpha_s}(U\times U)_{\alpha_0\cdots\alpha_s}\big) \\ &\simeq \mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}\big(\textstyle\bigvee_{\alpha_0\cdots\alpha_p}\bigvee_{\alpha_0\cdots\alpha_r}\bigvee_{p+r=s}(U_{\alpha_0\cdots\alpha_p}\times U_{\alpha_0\cdots\alpha_r})\big) \\ &\simeq \bigoplus_{\alpha_0\cdots\alpha_p}\bigoplus_{\alpha_0\cdots\alpha_r}\bigoplus_{p+r=s}\mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}(U_{\alpha_0\cdots\alpha_p}\times U_{\alpha_0\cdots\alpha_r})\end{aligned}$$

given by decomposing the product of the cover $\{U_\alpha\}$ with itself. Finally, we can pullback by the diagonal map

$$\begin{aligned}\Delta^*\colon \bigoplus_{\alpha_0\cdots\alpha_s} \mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}((U\times U)_{\alpha_0\cdots\alpha_s}) \\ \to \bigoplus_{\alpha_0\cdots\alpha_s}\mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}(U_{\alpha_0\cdots\alpha_s}) \simeq \check{C}^{p+r}(M;\mathrm{diff}(\Sigma^{n+m}\mathcal{E},\mathrm{ch})^{q+t}).\end{aligned}$$

The cup product on the E_1–page is defined by the composite map $\Delta^*\times$.

Lemma 36 *The differential d_1 on the E_1–page satisfies the Leibniz rule.*

Proof The construction of the cup product on the E_1–page is precisely the cup product structure for Čech cohomology. The Čech differential satisfies the Leibniz rule, and this is precisely d_1 by construction. □

We are now ready to prove Proposition 35.

Proof The proof follows by induction on the pages of the spectral sequence. The base case is satisfied by Lemma 36. Now suppose we have a cup product map

$$\cup\colon E(n)_k\times E(n)_k\to E(n+m)_k$$

such that d_k satisfies Leibniz. By definition, we have

$$E(n)^{p,q}_{k+1}=\frac{\ker\big(d_k\colon E(n)^{p,q}_k\to E(n)^{p+k,q+k-1}_k\big)}{\mathrm{im}\big(d_k\colon E(n)^{p-k,q-k+1}\to E(n)^{p,q}\big)},$$

and we define the cup product

$$\cup\colon E(n)^{p,q}_{k+1}\times E(m)^{r,s}_{k+1}\to E(n+m)^{p+r,q+s}_{k+1}$$

by restricting to elements in the kernel of d_k. The product is well defined since d_k satisfies the Leibniz rule. At this stage, the problem looks formally like the classical problem. Hence, analogously to the classical discussion in [36], it is tedious but straightforward to show that d_{k+1} also satisfies the Leibniz rule. □

4 Applications to differential cohomology theories

In this section, we would like to apply the spectral sequence constructed in the previous section to various differential cohomology theories. The construction is general enough to apply to any structured cohomology theory whose coefficients are known. We will explicitly emphasize three main examples. The first two are to known theories, namely ordinary differential cohomology and differential K-theory. We take this opportunity to explicitly develop the third theory, which is differential Morava K-theory, and then apply our smooth AHSS construction to it.

4.1 Ordinary differential cohomology theory

We begin by recovering the usual hypercohomology spectral sequence for the Deligne complex (see [11; 24, Appendix]) using our methods. We will first look at manifolds, then products of these, and then more generally to smooth fiber bundles.

Let us consider the smooth spectrum $\mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch})$ representing differential cohomology in degree n. We would like to see what our smooth AHSS gives in this case. We recall that $\mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch})$ is represented by Deligne cohomology of the sheaf of chain complexes $\mathbb{Z}_{\mathcal{D}}^{\infty}(n)$ via the Eilenberg–Mac Lane functor $H\colon \mathrm{Sh}_{\infty}(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{C}\mathrm{h}) \to \mathrm{Sh}_{\infty}(\mathcal{C}\mathrm{art}\mathcal{S}\mathrm{p}; \mathcal{S}\mathrm{p})$ (2-6). It follows from the general properties of this functor that the homotopy groups are given by

$$\pi_k \,\mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch}) \simeq H_k \mathbb{Z}_{\mathcal{D}}^{\infty}(n).$$

In this case, we have the immediate corollary to Theorem 25.

Corollary 37 *The spectral sequence for Deligne cohomology takes the form*

$$E_2^{p,q} = H^p(X; H_{-q}\mathbb{Z}_{\mathcal{D}}^{\infty}(n)) \implies H^{p+q}(X; \mathbb{Z}_{\mathcal{D}}^{\infty}(n)),$$

which is essentially the hypercohomology spectral sequence for the Deligne complex, but shifted as a fourth quadrant spectral sequence.

For the sake of completeness, we work out this spectral sequence and recover the differential cohomology diamond (2-2) from the sequence. This will help to illustrate how the general spectral sequence behaves and how it can be used to calculate general differential cohomology groups.

Now over the site of Cartesian spaces, the Poincaré lemma implies that we have an isomorphism of presheaves $d\colon \Omega^{n-1}/\,\mathrm{im(d)} \xrightarrow{\simeq} \Omega^n_{\mathrm{cl}}$. Since Ω^n_{cl} is a sheaf over the site of smooth manifolds, the gluing condition allows us to calculate the relevant terms on the E_2–page of the spectral sequence in Figure 1.

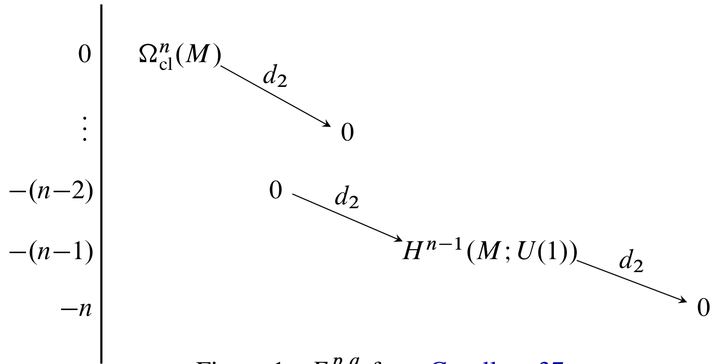


Figure 1: $E_2^{p,q}$ from Corollary 37

The term $H^{n-1}(M;U(1))$ will survive to the E_∞–page, and we have an isomorphism

$$H^{n-1}(M;U(1)) \simeq F_{n-1}\widehat{H}^n(M;\mathbb{Z}) / F_n\widehat{H}^n(M;\mathbb{Z}).$$

In fact, it is not hard to see that the definition of the filtration gives $F_n\widehat{H}^n(M;\mathbb{Z}) \simeq 0$, and we have an injection

$$H^{n-1}(M;U(1)) \simeq F_{n-1}\widehat{H}^n(M;\mathbb{Z}) \hookrightarrow \widehat{H}^n(M;\mathbb{Z}).$$

On the E_n–page, we get one possibly nonzero differential

$$d_n\colon \Omega^n(M)_{\mathrm{cl}} \to H^n(M;U(1)).$$

Proposition 38 *The differential d_n for the AHSS for Deligne cohomology can be identified with the composition*

$$\Omega^n_{\mathrm{cl}}(M) \to H^n_{\mathrm{dR}}(M) \xrightarrow{\int_{\Delta^n}} H^n(M;\mathbb{R}) \xrightarrow{\exp} H^n(M;U(1)),$$

and the kernel is precisely those forms which have integral periods.

Proof We will unpack the definition of the differential in the AHSS in detail. This in turn will require unpacking the connecting homomorphism in the Deligne model of ordinary differential cohomology; see [11]. Denote by X_p the Čech filtration, and let

$$\partial\colon \mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch})^q(X_p) \to \mathrm{diff}(\Sigma^n H\mathbb{Z}, \mathrm{ch})^{q+1}(X_{p+1}/X_p)$$

denote the connecting homomorphism in the long exact sequence associated to the cofiber sequence $X_p \hookrightarrow X_{p+1} \twoheadrightarrow X_{p+1}/X_p$ in the usual way. In what follows, we will denote Čech–Deligne cochains on the p–level of the filtration X_p as a p–tuple

$$(z_0, z_1, \dots, z_p) \in \widehat{C}^q(X_p),$$

where z_i is a $(q-i)$–form defined on i–fold intersections.

Now, by definition, $d_n\colon E_n^{0,0}\to E_n^{n,0}$ is given by $d_n=\partial(j^*)^{-1}$, where $(j^*)^{-1}$ denotes a choice of element in the preimage of the restriction j^* induced by $j\colon X_0\hookrightarrow X_{n-1}$.[3] Since we have $d_k=0$ for $k<n$, the differential d_n is defined on all elements $z\in\Omega^n_{\mathrm{cl}}(M)$. Let g_0 be a locally defined $(n-1)$–form trivializing z. Then we can choose $(j^*)^{-1}z$ to be the Čech–Deligne cocycle

$$(j^*)^{-1}z=\underbrace{(g_0,g_1,g_2,\dots,g_{n-2})}_{n-1}\in\widehat{C}^0(X_{n-1}),\tag{4-1}$$

where each g_k is a $(n-k-1)$–form that satisfies the cocycle condition $\delta(g_k)=(-1)^k dg_{k+1}$. To see where the boundary map takes this element, let y be a Čech–Deligne cochain given by

$$y=\underbrace{(g_0,g_1,g_2,\dots,g_{n-2},\exp(2\pi i g_{n-1}))}_{n}\in\widehat{C}^0(X_n),$$

where g_{n-1} is any smooth $\mathbb{R}$–valued function satisfying $d(g_{n-1})=(-1)^{n-1}\delta(g_{n-2})$.[4] Now y is not Čech–Deligne closed in general since

$$Dy=(d+(-1)^{n-1}\delta)y=(0,0,\dots,\exp((-1)^{n-1}2\pi i\cdot\delta(g_{n-1}))),$$

and g_{n-1} may not satisfy the cocycle condition $\delta(g^{n-1})=0$. However, by the Čech–de Rham isomorphism (see for example [7]), this element in the Čech– de Rham double complex is isomorphic to an $\mathbb{R}$–valued Čech cocycle on n–fold intersections. Explicitly, there is a constant $\mathbb{R}$–valued cocycle r_n such that $\delta(g^{n-1})=r_n$. It follows from the isomorphisms between the Čech, de Rham, and singular cohomologies that the class of r_n can be represented by the singular cocycle given by the pairing $\int_\sigma z$ for any cycle σ in M. Since the class $\int_\sigma z$ was just an unraveling of the boundary $\partial((j^*)^{-1}z)$, we have proved the claim. □

In the next section, we will need to make use of a differential refinement of the Chern character. To this end, we briefly discuss differential cohomology with rational coefficients $\widehat{H}^n(-;\mathbb{Q})$. These groups are obtained via the differential function spectra $\mathrm{diff}(\Sigma^n H\mathbb{Q},\mathrm{ch})$ which fit into the homotopy cartesian square:

$$\begin{tikzcd}
\mathrm{diff}(H\mathbb{Q},\mathrm{ch}) \arrow[r] \arrow[d] & H(\tau_{\le 0}\Omega^*[n]) \arrow[d] \\
\Sigma^n H\underline{\mathbb{Q}} \arrow[r] & H(\Omega^*[n])
\end{tikzcd}$$

[3] Note that the differential only takes this form at the $(0,0)$–entry. In general, the differential formed from the n^{th} derived couple will be more complicated.

[4] Note that this cocycle condition is necessary for y to be an lift of $(j^*)^{-1}z$ to the n–level of the filtration.

As a consequence of Proposition 10, the cohomology groups with values in this spectrum are calculated as

$$\mathrm{diff}(\Sigma^n H\mathbb{Q}, \mathrm{ch})^q(M) = \begin{cases} H^{n+q}(M), & q>0, \\ \widehat{H}^n(M;\mathbb{Q}), & q=0, \\ H^{n-1+q}(M;\mathbb{R}/\mathbb{Q}), & q<0. \end{cases}$$

The explicit calculation of the differential in Proposition 38 can be easily modified to get the following.[5]

Proposition 39 *The differential d_n on the E_n–page for the AHSS spectral sequence for $\mathrm{diff}(\Sigma^n H\mathbb{Q}, \mathrm{ch})$ is given by*

$$\Omega^n_{\mathrm{cl}}(M) \to H^n_{\mathrm{dR}}(M) \xrightarrow{\int_{\Delta^n}} H^n(M;\mathbb{R}) \to H^n(M;\mathbb{Q}/\mathbb{Z}),$$

and the kernel is precisely those forms which have rational periods.

We will make use of this result when we discuss the differentials in smooth K-theory in the next section. For now, from Proposition 38, we immediately get the following characterization of closed forms with integral periods and forms with rational periods using our smooth AHSS.

Corollary 40 (i) *The group of closed forms with integral periods on a manifold M is given by*

$$\Omega^n_{\mathrm{cl},\mathbb{Z}}(M) \simeq \widehat{H}^n(M;\mathbb{Z}) / F_1\widehat{H}^n(M;\mathbb{Z}).$$

(ii) *The group of closed forms with rational periods on a manifold M is given by*

$$\Omega^n_{\mathrm{cl},\mathbb{Q}}(M) \simeq \widehat{H}^n(M;\mathbb{Q}) / F_1\widehat{H}^n(M;\mathbb{Q}).$$

4.2 Differential K-theory

In this section, we examine the smooth AHSS for the differential function spectrum $\mathrm{diff}(K, \mathrm{ch})$, corresponding to complex K-theory. Proposition 10 allows us to calculate the cohomology groups on a paracompact manifold M as (see [49; 17; 66; 28])

$$\mathrm{diff}(K,\mathrm{ch})^q(M) = \begin{cases} K^q(M), & q>0, \\ \widehat{K}^0(M), & q=0, \\ K^q_{U(1)}(M), & q<0. \end{cases} \tag{4-2}$$

Both groups K and $K_{U(1)}$ are periodic. Indeed, $K_{U(1)}(M)$ fits into an exact sequence

$$\cdots \to K^{-1}(M)\otimes\mathbb{R} \to K^{-1}_{U(1)}(M) \to K(M) \to K(M)\otimes\mathbb{R} \to \cdots.$$

[5]The exact argument in the proof of Proposition 38 applies, with $\mathbb{R}/\mathbb{Q}$ in place of $\mathbb{R}/\mathbb{Z} \simeq U(1)$.

Consequently, the periodicity of both integral and rational K-theory, along with an application of the five lemma, imply that $K_{U(1)}$ is 2–periodic. In particular, we have

$$K_{U(1)}^{2q}(*) \simeq U(1) \quad \text{and} \quad K_{U(1)}^{2q+1}(*) \simeq 0, \quad q \in \mathbb{Z}.$$

Given (4-2), we see that for a contractible open set U, we have an isomorphism

$$\operatorname{diff}(K, \mathrm{ch})^{2q+1}(U) \simeq K_{U(1)}^{2q}(*) \simeq U(1)$$

for $q < 0$. For degree 0, the differential cohomology diamond in this case takes the form:

$$\begin{tikzcd}
 & \prod_{2k-1} \Omega^{2k-1}/\operatorname{im}(d) \arrow[rr, "d"] \arrow[dr, "a"] & & \prod_{2k} \Omega^{2k}_{\mathrm{cl}} \arrow[dr] & \\
K_{\mathbb{R}}^{-1} \arrow[ur] \arrow[dr] & & \widehat{K}^0 \arrow[ur, "R"] \arrow[dr, "I"] & & K_{\mathbb{R}}^0 \\
 & K_{U(1)}^{-1} \arrow[rr, "\beta_K"] \arrow[ur] & & K^0 \arrow[ur, "\mathrm{ch}"] &
\end{tikzcd}$$

This implies that for a contractible open set U, differential K-theory $\widehat{K}^0(U)$ fits into the short exact sequence

$$0 \to \prod_{2k-1} \Omega^{2k-1}/\operatorname{im}(d)(U) \to \widehat{K}^0(U) \to \mathbb{Z} \to 0.$$

Hence, over the site of Cartesian spaces, we have a naturally split short exact sequence of presheaves

$$0 \to \prod_{2k-1} \Omega^{2k-1}/\operatorname{im}(d) \to \widehat{K}^0 \to \underline{\mathbb{Z}} \to 0.$$

Over that site, the presheaf on the left-hand side is actually a sheaf and is naturally isomorphic (by the Poincaré lemma) to the sheaf $\prod_{2k} \Omega^{2k}_{\mathrm{cl}}$. We therefore make the identification

$$\widehat{K}^0 \simeq \prod_{2k} \Omega^{2k}_{\mathrm{cl}} \oplus \underline{\mathbb{Z}}. \tag{4-3}$$

Remark 41 It is important to note that the identification (4-3) is only true on the site of *Cartesian spaces*, which is to say that it holds only locally. On the site of smooth manifolds, this is of course not the case.

Next, since both $\Omega^{2k}_{\mathrm{cl}}$ and $\underline{\mathbb{Z}}$ are sheaves on the site of smooth manifolds, we can identify the degree-0 Čech cohomology with these coefficients with the value of this

$$
\begin{tikzcd}[column sep=small]
0 & \prod_{2k} \Omega^{2k}_{\mathrm{cl}}(M) \oplus \mathbb{Z} \arrow[drr, "d_2"] & & & & \\
-1 & & H^1(M;U(1)) \arrow[drr] & H^2(M;U(1)) \arrow[drr] & & \\
-2 & & & 0 \arrow[drr] & 0 & 0 \\
-3 & & & & & H^4(M;U(1))
\end{tikzcd}
$$

Figure 2: $E_2^{p,q}$ for even differential K-theory

sheaf on M. Isolating the terms on the E_2–page which converge to $\widehat{K}^0(M)$, we get Figure 2.

We see that all the differentials are zero except for the map labeled d_2 above. On the E_3–page, we get Figure 3.

The higher pages will fall into cases depending on the parity. We observe that for each even page E_{2m}, there is one nonzero differential given by d_{2m}. For the odd pages the differentials are given by an odd-degree $U(1)$–cohomology operation.

Note that, in the diagrams, we are interested in the case $p+q=0$, corresponding to diagonal entries. Now $p \geq 0$, as the Čech filtrations are of nonnegative degrees, which implies that $q \leq 0$. Hence the entries go down the diagonal. Our first goal will be to identify the even differentials d_{2m}. In order to do this, let us recall that there is a *differential Chern character* map (see [13; 63]) which is stably given by a morphism of smooth spectra

$$\widehat{\mathrm{ch}}\colon \mathrm{diff}(K, \mathrm{ch}) \to \prod_{2k} \mathrm{diff}(\Sigma^{2k} H\mathbb{Q}, \mathrm{ch}).$$

Postcomposing this map with the projection pr_{2m} onto the $2m$–component gives a map of smooth spectra

$$\mathrm{pr}_{2m}\,\widehat{\mathrm{ch}}\colon \mathrm{diff}(K, \mathrm{ch}) \to \mathrm{diff}(\Sigma^{2m} H\mathbb{Q}, \mathrm{ch}).$$

Using this map, we can prove the following analogue of Proposition 39.

Proposition 42 *The group of permanent cycles in bidegree* $(0,0)$ *in the AHSS for* $\mathrm{diff}(K, \mathrm{ch})$ *is a subgroup of even-degree closed forms with rational periods. That is, we have*

$$E_\infty^{0,0} \subset \prod_k \Omega^{2k}_{\mathrm{cl},\mathbb{Q}}(M) \oplus \mathbb{Z}.$$

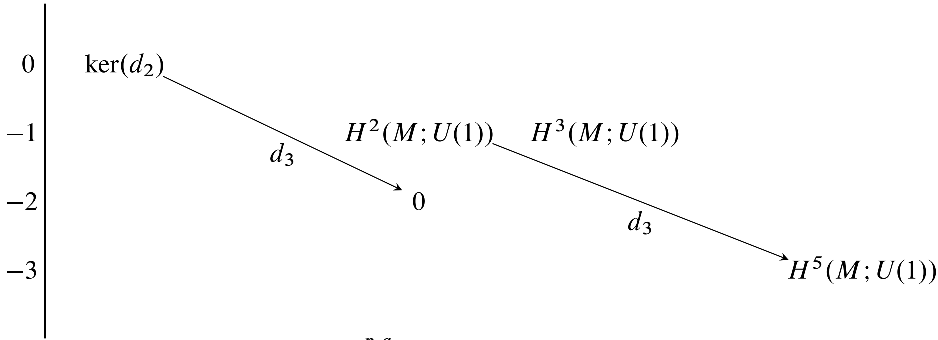


Figure 3: $E_3^{p,q}$ for even differential K-theory

Proof We prove by induction on the even pages[6] of the spectral sequence that, for all n, $E_{2n}^{0,0}$ must be a subgroup of

$$\prod_{2k\le 2n} \Omega^{2k}_{\mathrm{cl},\mathbb{Q}}(M) \oplus \prod_{2k>2n} \Omega^{2k}_{\mathrm{cl}}(M) \oplus \mathbb{Z}.$$

For the base case, observe that the map $\mathrm{pr}_2\, \widehat{\mathrm{ch}}$ induces a rank-1 morphism of AHSSs and therefore commutes with d_2. It is straightforward to check, using the definitions, that this leads to the following commutative diagram:

$$\begin{tikzcd}
\prod_{2k} \Omega^{2k}_{\mathrm{cl}}(M) \oplus \mathbb{Z} \arrow[r, "\mathrm{pr}_2"] \arrow[d, "d_2"] & \Omega^2_{\mathrm{cl}}(M) \arrow[d, "d_2'"] \\
H^2(M;\mathbb{R}/\mathbb{Z}) \arrow[r, "q"] & H^2(M;\mathbb{R}/\mathbb{Q})
\end{tikzcd}$$

We see that the kernel of d_2 must be a subgroup of $\Omega^2_{\mathrm{cl},\mathbb{Q}}(M) \oplus \prod_{2k>2} \Omega^{2k}(M) \oplus \mathbb{Z}$ by Proposition 38.

Now suppose the claim is true for d_{2n}. Again, we have that $\mathrm{pr}_{2n+2}\, \widehat{\mathrm{ch}}$ commutes with d_{2n+2}, and we have the following commutative diagram:

$$\begin{tikzcd}
\ker(d_{2n}) \arrow[r, "\mathrm{pr}_{2n+2}"] \arrow[d, "d_{2n+2}"] & \Omega^{2n+2}_{\mathrm{cl}}(M) \arrow[d, "d_{2n+2}'"] \\
H^{2n+2}(M;\mathbb{R}/\mathbb{Z}) \arrow[r, "q"] & H^{2n+2}(M;\mathbb{R}/\mathbb{Q})
\end{tikzcd}$$

[6]The differential is 0 for the odd pages, and so no generality is lost by restricting to the even pages.

By the induction hypothesis,

$$\ker(d_{2n}) \subset \prod_{2k\leq 2n} \Omega^{2k}_{\mathrm{cl},\mathbb{Q}}(M) \oplus \prod_{2k>n} \Omega^{2k}_{\mathrm{cl}}(M) \oplus \mathbb{Z},$$

and the kernel of d_{2n+2} is as claimed. □

We now turn to the first odd differential d_3. Recall that β and $\widetilde{\beta}$ denote the Bockstein homomorphisms corresponding to the sequences $0 \to \mathbb{Z} \to \mathbb{R} \xrightarrow{\exp} U(1) \to 0$ and $0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2 \to 0$, respectively. We still also denote by $\Gamma_2\colon H^n(-;\mathbb{Z}/2) \to H^n(-,U(1))$ the map induced by the representation of $\mathbb{Z}/2$ as the square roots of unity and $\rho_2\colon \mathbb{Z} \to \mathbb{Z}/2$ as the mod 2 reduction.

Proposition 43 (degree-3 differential) *The first odd-degree differential in the AHSS for differential K-theory is given by*

$$d_3 = \begin{cases} \widehat{\mathrm{Sq}}^3 := \Gamma_2 \mathrm{Sq}^2 \rho_2 \beta, & q<0, \\ \mathrm{Sq}^3_{\mathbb{Z}} := \widetilde{\beta}\mathrm{Sq}^2\rho_2, & q>0, \\ 0, & q=0. \end{cases}$$

Proof The case for $q=0$ is obvious. For $q>0$, this follows from the fact that the integration map defines an isomorphism $I\colon \mathrm{diff}(K,\mathrm{ch})^q(M) \xrightarrow{\simeq} K^q(M)$ for $q>0$. Since the differential d_3 for the classical AHSS is given by $\mathrm{Sq}^3_{\mathbb{Z}}$, and the integration map defines an isomorphism of corresponding first quadrant spectral sequences, the case $q>0$ is settled.

For $q<0$, Corollary 30 implies that the Bockstein β commutes with the differentials on the E_3–page. We therefore have

$$\beta d_3 = \mathrm{Sq}^3_{\mathbb{Z}}\beta = \widetilde{\beta}\mathrm{Sq}^3\rho_2\beta. \tag{4-4}$$

Rephrasing, we have the commuting diagram:

$$\begin{tikzcd} H^{n-1}(M;U(1)) \arrow[r, "d_3"] \arrow[d, "\beta"'] & H^{n+3-1}(M;U(1)) \arrow[d, "\beta"] \\ H^n(M;\mathbb{Z}) \arrow[r, "\mathrm{Sq}^3_{\mathbb{Z}}"] & H^{n+3}(M;\mathbb{Z}) \end{tikzcd}$$

We now claim that $\widetilde{\beta} = \beta \circ \Gamma_2$. Indeed, we have a morphism of short exact sequences:

$$\begin{tikzcd} \mathbb{Z} \arrow[r, "\times 2"] \arrow[d, "\mathrm{id}"] & \mathbb{Z} \arrow[r, "\rho_2"] \arrow[d, "\times\pi i"] & \mathbb{Z}/2 \arrow[d, "\Gamma_2"] \\ \mathbb{Z} \arrow[r, "\times 2\pi i"] & \mathbb{R} \arrow[r, "\exp"] & U(1) \end{tikzcd}$$

This morphism induces a morphism on the associated long exact sequences on cohomology. After delooping once to extend to the left, the homotopy commutativity of the resulting diagram

$$\begin{tikzcd} \mathbb{Z}/2 \arrow[r, "\widetilde{\beta}"] \arrow[d, "\Gamma_2"'] & B\mathbb{Z} \arrow[d, equal] \\ U(1) \arrow[r, "\beta"] & B\mathbb{Z} \end{tikzcd}$$

immediately establishes the claim.

Now it follows from (4-4) that $d_3 - \Gamma_2 \mathrm{Sq}^3 \rho_2 \beta$ is in the kernel of β. By exactness of the Bockstein, this implies that it must be in the image of the exponential map $\exp\colon H^*(-;\mathbb{R}) \to H^*(-;U(1))$. Hence there is an operation $\psi\colon H^*(-;U(1)) \to H^*(-;\mathbb{R})$ such that

$$\phi := \exp \circ \psi = \exp(\psi) = d_3 - \Gamma_2 \mathrm{Sq}^3 \rho_2 \beta.$$

Equivalently, we have a factorization:

$$\begin{tikzcd} H^*(-;U(1)) \arrow[rr, "\phi"] \arrow[dr, "\psi"'] & & H^*(-;U(1)) \\ & H^*(-;\mathbb{R}) \arrow[ur, "\exp"'] & \end{tikzcd}$$

We expect to have $\hom(H^*(-;U(1)), H^*(-;\mathbb{R})) = 0$ since $U(1)$ is almost completely torsion (and since the second argument is an $\mathbb{R}$–vector space). However, we need to be slightly careful here, since not all elements $a \in H^*(M;U(1))$ represent torsion classes. In fact, identifying $U(1) \simeq \mathbb{R}/\mathbb{Z}$, such an element will be torsion if and only if it represents an element in $H^*(M;\mathbb{Q}/\mathbb{Z}) \hookrightarrow H^*(M;\mathbb{R}/\mathbb{Z})$. To fix this issue, we observe that for any abelian group A, we have an isomorphism

$$r\colon \hom(H^*(M;\mathbb{R}), A) \to \hom(H^*(M;\mathbb{Q}), A)$$

given by restricting a map to the rationals $\mathbb{Q} \subset \mathbb{R}$. The inverse is given by restricting a map to the generators and extending with real coefficients. This implies, in turn, that we have an isomorphism

$$r\colon \hom(H^*(M;\mathbb{R}/\mathbb{Z}), A) \to \hom(H^*(M;\mathbb{Q}/\mathbb{Z}), A);$$

ie $\mathbb{R}/\mathbb{Z}$ and $\mathbb{Q}/\mathbb{Z}$ behave equivalently when taken as coefficients of cohomology inside the hom. Finally, since $\hom(H^*(M;\mathbb{Q}/\mathbb{Z}), H^*(M;\mathbb{R})) = 0$, we must have $\hom(H^*(M;\mathbb{R}/\mathbb{Z}), H^*(M;\mathbb{R})) = 0$, which forces $\psi = 0$. Consequently, $\exp \circ \psi = 0$, so that $\phi = 0$. Therefore, indeed we have

$$d_3 = \Gamma_2 \mathrm{Sq}^3 \rho_2 \beta. \qquad \square$$

Remark 44 The above proposition suggests that these operations are related to some sort of *differential* Steenrod squares. Indeed, this is the case, which has been investigated by the authors in [33], with $\widehat{\mathrm{Sq}}^3$ being one such operation.

Now that we have established the algebraic construction, we turn to investigating the convergence of the spectral sequence from a geometric point of view. In particular, we immediately observe that the only terms in the spectral sequence which contain information about differential forms are at $q=0$. These terms converge to elements in the filtered graded complex (since $q=0$)

$$\widehat{K}(M)/F_1\widehat{K}(M).$$

Since the filtration is given by the Čech-type filtration on M, we see that this quotient contains elements which have nontrivial data on all open sets, intersections and higher intersections. For the degrees $q<0$, the filtration quotients

$$F_p\widehat{K}(M)/F_{p+1}\widehat{K}(M)$$

have trivial data below p–intersections.

In fact, it is not too surprising that this occurs. There is a geometric model for reduced $\widehat{K}^0$ which is given by the moduli stack $\mathbf{B}U_{\mathrm{conn}}$ of unitary vector bundles, equipped with connection. Let Vect_∇ be the moduli stack of vector bundles with connections. It was shown in [16] that there is a cycle map

$$\mathrm{cycl}\colon \pi_0\,\mathrm{Map}(M,\mathrm{Vect}_\nabla)\to\widehat{K}^0(M),$$

which induces an isomorphism upon group completion. In our construction, this is equivalent to

$$\mathrm{cycl}\colon \pi_0\,\mathrm{Map}(M,\mathbf{B}U_{\mathrm{conn}})\to\widehat{K}^0(M).$$

Now the stack $\mathbf{B}U_{\mathrm{conn}}$ can be identified with the moduli stack obtained by taking the nerve of the action groupoid $C^\infty(-,U)//\Omega^1(-;\mathfrak{u})$ with the action given by gauge transformations, where $\mathfrak{u}$ is the Lie algebra of the unitary group. Let $\{U_\alpha\}$ be a good open cover of M. Then a map $M\to\mathbf{B}U_{\mathrm{conn}}$ is given by the following data:

- a choice of smooth $U(n)$–valued function $g_{\alpha\beta}$ on intersections $U_\alpha\cap U_\beta$,
- a choice of local connection 1–form $\mathcal{A}_{\alpha\beta}$ on open sets U_α.

This is precisely the data needed to define a unitary vector bundle on M.

Remark 45 More relevant to our needs, though, is the fact that the effects of the filtration become transparent when taking $\mathbf{B}U_{\mathrm{conn}}$ as a model for $\widehat{K}^0$. We now see that the $q=0$ terms converge to terms which involve the data of the connection, while the $q<0$ terms contain data about bundles with trivializable connections (in particular, flat connections).

Differential K^1–theory We now consider odd differential K-theory, K^1. In this case, the representing spectrum is the unitary group U itself. Viewing this as a classifying space, we can write $U = B\Omega U$. Of course we are interested in the corresponding stacks. Unfortunately, we do not have the analogue of the above group-loop group relation in stacks; ie $U_{\mathrm{conn}} \not\simeq \mathbf{B}\Omega U_{\mathrm{conn}}$. Nevertheless, the machinery that we set up will work equally well for differential K^1–theory, as far as the third differential goes; ie we still have $d_3 = \widehat{\mathrm{Sq}}^3$. However, the even differentials are now transgressed in degree by one, so that they are also of odd degree. This is expected as the Chern character in this case is a map to cohomology of odd degree.

The story for $\widehat{K}^1$ can be worked out similarly as we indicated above. Let us expand on this in more detail. In the odd case, the differential cohomology diamond takes the form

$$\begin{tikzcd}
 & \prod_{2k} \Omega^{2k}/\operatorname{im}(d) \arrow[rr, "d"] \arrow[dr, "a"] & & \prod_{2k+1} \Omega^{2k+1}_{\mathrm{cl}} \arrow[dr] & \\
K^0_{\mathbb{R}} \arrow[ur] \arrow[dr] & & \widehat{K}^1 \arrow[ur, "R"] \arrow[dr, "I"] & & K^1_{\mathbb{R}} \\
 & K^0_{U(1)} \arrow[ur] \arrow[rr, "\beta_K"] & & K^1 \arrow[ur, "\mathrm{ch}"] &
\end{tikzcd}$$

and we get a short exact sequence of presheaves (on the site of Cartesian spaces)

$$0 \to \underline{\mathbb{Z}} \to \prod_{2k} \Omega^{2k}/\operatorname{im}(d) \to \widehat{K}^1 \to 0.$$

It is straightforward to show that the map $\underline{\mathbb{Z}} \to \prod_{2k} \Omega^{2k}/\operatorname{im}(d)$ is zero. Consequently, we have the isomorphism

$$\widehat{K}^1 \simeq \prod_{2k} \Omega^{2k}/\operatorname{im}(d) \simeq \prod_{2k+1} \Omega^{2k+1}_{\mathrm{cl}}.$$

Using the same type of argument as in the even K-theory K^0, we likewise get a refinement of the differential of the underlying topological theory. More precisely, we see that the first nonzero differentials appear on the E_3–page as in Figure 4.

Proposition 46 *Proposition 43 holds for differential K^1–theory. That is, the degree-3 differential in $\widehat{K}^1$ is given by the refinement of the Steenrod square of dimension three.*

Also, using the same argument as in the proof of Proposition 42, we see that the permanent cycles in bidegree $(0,0)$ are a subgroup of odd-degree forms with rational periods.

$$
\begin{tikzcd}
0 & \prod_{2k+1}\Omega^{2k+1}_{\mathrm{cl}}(M) \arrow[ddrr, "d_3"] & & & \\
-1 & & & & \\
-2 & & H^2(M;U(1)) \arrow[ddrr, "\widehat{\mathrm{Sq}}^3"] & H^3(M;U(1)) & \\
-3 & & & & \\
-4 & & & & H^5(M;U(1))
\end{tikzcd}
$$

Figure 4: $E_2^{p,q}$ for odd differential K-theory

Proposition 47 *The group of permanent cycles in bidegree* $(0,0)$ *in the AHSS for* $\mathrm{diff}(\Sigma K, \mathrm{ch})$ *is a subgroup of odd degree closed forms with rational periods. That is, we have*

$$E_\infty^{0,0} \subset \prod_k \Omega^{2k-1}_{\mathrm{cl},\mathbb{Q}}(M) \oplus \mathbb{Z}.$$

Example 48 (fields in string theory and M-theory) In the string theory and M-theory literature, one encounters settings where cohomology classes are compared to K-theory elements, in the sense of asking when a cohomology class arises from or "lifts to" a K-theory class. This involves, in a sense, a physical modeling of the process of building the AHSS. One such obstruction is Sq^3, viewed as the first nontrivial differential d_3 in K-theory, so that the condition $\mathrm{Sq}^3 x = 0$ on a cohomology class x amounts to saying that the class lifts to K-theory. This is desirable in the study of the partition function of the fields in type-IIA string theory; see [21; 47]. On the other hand, it is desirable to have differential refinements for physical purposes. Therefore, now that we have the differential AHSS at our disposal, it is natural to consider expressions such as $d_3(\hat{x}) := \widehat{\mathrm{Sq}}^3\hat{x} = 0$ on the differential cohomology class $\hat{x}$ that refines the topological class x. This can be viewed as a condition on cohomology with $U(1)$–coefficients (or flat n–bundles), in order that they lift to flat elements in $\widehat{K}$.[7] If the degree of the class x is even, then we are in type-IIA string theory, and we lift to differential K^0–theory. On the other hand, being in type-IIB string theory means the degree of x is odd, and we are lifting to differential K^1–theory. The new differentials d_{2m} and d_{2m+1} arising from differential forms will correspond to even- and odd-degree closed differential forms as the particular forms representing the physical fields F_{2m} and F_{2m+1} via the Chern character.

[7] This could end up being stronger in the sense that it is a condition for lifting differential cohomology classes to differential K-theory, but we will leave that for future investigations.

Example 49 (D-brane charges) The charges of D-branes can, a priori, be taken to be given as a class in cohomology $\mathcal{Q}_H \in H^*(X;\mathbb{Q})$. Quantum effects requires some of these charges to be (up to shifts) to be in integral cohomology. However, in order to not discuss isomorphism classes of such physical objects but pinning down a particular physical object, one considers the charges to take values in differential cohomology, with Deligne cohomology being one such presentation: $\mathcal{Q}_{\widehat{H}} \in \widehat{H}^*(X;\mathbb{Z})$; see [19]. On the other hand, careful analysis reveals that the charges take values in K-theory rather than in cohomology $\mathcal{Q}_K \in K^i(X)$ for $i = 0, 1$ for type IIB/IIA; see [56; 29; 9]. Such a class exists if the cohomology charge satisfies $\mathrm{Sq}^3\mathcal{Q}_H = 0$. Again, at this stage, adding in the geometry requires the charges to take values in differential K-theory $\mathcal{Q}_{\widehat{K}} \in \widehat{K}^i(X)$. Our construction now allows for a characterization of when charges in Deligne cohomology lift to charges in differential K-theory, namely when they are annihilated by the third differential in the smooth AHSS, ie when $\widehat{\mathrm{Sq}}^3\mathcal{Q}_{\widehat{H}} = 0$.

4.3 Differential Morava K-theory

There are various interesting generalized cohomology theories that descend from complex cobordism, among which are Morava K-theory and Morava E-theory. Such theories can be defined using their coefficient rings, which in general are polynomials over finite or p–adic fields on generators whose dimension depends on the chromatic level and the prime p. As such, these kind of theories do not lend themselves directly to immediate geometric interpretation in contrast to the case of K-theory, which can be formulated via stable isomorphism classes of vector bundles.

However, recent work in [48] (generalizing some aspects of [5]) seems to give hope in that direction. Nevertheless, just because an entity is defined over a finite field does not automatically make it ineligible for differential refinement. In fact, we have recently demonstrated this [33] for the case of Steenrod cohomology operations, which are, a priori, $\mathbb{Z}/p$–valued operations. The main point there was that as long as these admit integral lifts, they do have a chance at a differential refinement. What we will seek here is something analogous: integral refinements of such generalized cohomology theories.

We will consider the integral Morava K-theory $\widetilde{K}(n)$ highlighted in [47; 60; 62]. Morava K-theory $K(n)$ is the mod p reduction of an integral (or p–adic) lift $\widetilde{K}(n)$ with coefficient ring $\widetilde{K}(n)_* = \mathbb{Z}_p[v_n, v_n^{-1}]$. This theory more closely resembles complex K-theory than is the case for the mod p versions (for $n = 1$, it is the p–completion of K-theory). The integral theory is much more suited to applications in physics [47; 60; 12; 62].

The Atiyah–Hirzebruch spectral sequence for Morava K-theory has been studied by Yagita in [72]; see also [47]. There is a spectral sequence converging to $K(n)^*(X)$ with E_2–term $E_2^{p,q} = H^p(X, K(n)^q)$. While this can be done for any prime, we will focus on the prime 2. In this case, the first possibly nontrivial differential is $d_{2^{n+1}-1}$; this is given by [72] as

$$d_{2^{n+1}-1}(xv_n^k) = Q_n(x)v_n^{k-1}.$$

Here Q_n is the n^{th} Milnor primitive at the prime 2, which we define inductively as $Q_0 = \mathrm{Sq}^1$, the Bockstein operation, and $Q_{j+1} = \mathrm{Sq}^{2^j} Q_j - Q_j \mathrm{Sq}^{2^j}$, where $\mathrm{Sq}^j\colon H^n(X;\mathbb{Z}_2) \to H^{n+j}(X;\mathbb{Z}_2)$ is the j^{th} Steenrod square. These operations are derivations

$$Q_j(xy) = Q_j(x)y + (-1)^{|x|} x Q_j(y).$$

The signs are of course irrelevant at $p=2$, but will become important in the integral version. Extensive discussion of the mod p Steenrod algebra in terms of these operations is given in [68].

The integral theory is also computable via an AHSS, which can be deduced from [47; 62]. There is an AHSS converging to $\widetilde{K}(n)^*(X)$ with $E_2^{p,q} = H^p(X, \widetilde{K}(n)^q)$. The first possibly nontrivial differential is $d_{2^{n+1}-1}$; this is given by

$$d_{2^{n+1}-1}(xv_n^k) = \widetilde{Q}_n(x)v_n^{k-1}.$$

Here $\widetilde{Q}_k\colon H^*(X;\mathbb{Z}) \to H^{*+2^{k+1}-1}(X;\mathbb{Z})$ is an integral cohomology operation lifting the Milnor primitive Q_k.

In order to consider differential refinement of Morava K-theory, we need geometric information encoded in differential forms, hence rational information. The rationalization of Morava K-theory $\widetilde{K}(n)$, like any reasonable spectrum, exists and can be thought of as localization at $\widetilde{K}(0) = H\mathbb{Q}$; see [8; 58]. We can, in the same way, localize at $\mathbb{R}$. More precisely, the localized theory is given by

$$\widetilde{K}_{\mathbb{R}}(n) = \widetilde{K}(n) \wedge M\mathbb{R},$$

where $M\mathbb{R}$ is an Eilenberg–Moore spectrum. We have an equivalence

$$\widetilde{K}_{\mathbb{R}}(n) \simeq H(\mathbb{Z}[v_n, v_n^{-1}] \otimes \mathbb{R})$$

and a Chern character map

$$\mathrm{ch}\colon \widetilde{K}(n) \to H(\mathbb{Z}[v_n, v_n^{-1}] \otimes \Omega^*).$$

Thus we can form the differential function spectrum $\mathrm{diff}(\widetilde{K}(n), \mathrm{ch})$, and we can form the associated AHSS. To see what form the spectral sequence takes, we need to discuss the *flat Morava K-theory* $\widetilde{K}_{U(1)}(n)$, defined by the fiber sequence

$$\widetilde{K}(n) \to \widetilde{K}(n) \wedge M\mathbb{R} \to \widetilde{K}_{U(1)}(n) := \widetilde{K}(n) \wedge MU(1).$$

$$
\begin{tikzcd}
0 & \prod_k \Omega_{\mathrm{cl}}^{k2(2^n-1)}(M) \oplus \mathbb{Z} \arrow[ddrr, "d_{2(2^n-1)}"] & & & \\
\vdots & & & & \\
-2^{n+1}+3 & & H^{2^{n+1}-3}(M;U(1)) & H^{2(2^n-1)}(M;U(1)) & \\
\vdots & & & & \\
-2^{n+2}+5 & & & & H^{4(2^n-1)}(M;\mathbb{R}/\mathbb{Z})
\end{tikzcd}
$$

Figure 5: $E_{2(2^n-1)}^{p,q}$ for Morava K-theory

This theory is periodic with period $2(2^n-1)$. Indeed, both $\widetilde{K}(n)$ and its rationalization are periodic, and we have a long exact sequence

$$\cdots \to \widetilde{K}(n)^m(M) \to (\widetilde{K}(n) \wedge M\mathbb{R})^m(M) \to \widetilde{K}^m_{U(1)}(n)(M) \to \widetilde{K}(n)^{m+1}(M) \to \cdots$$

relating the flat theory to both the rational and integral theories. This, in particular, gives the following identification.

Lemma 50 *The coefficients of flat Morava K-theory are given by*

$$\widetilde{K}_{U(1)}(n)^m(*) \simeq \begin{cases} U(1), & m = 2(2^n-1), \\ 0, & \textit{otherwise}. \end{cases}$$

Knowing the coefficients of the flat theory, we can write down the relevant nonzero terms on the $E_{2(2^n-1)}$–page of the corresponding spectral sequence in Figure 5, and the only nonzero differential is given by

$$d_{2(2^n-1)} \colon \prod_k \Omega_{\mathrm{cl}}^{k2(2^n-1)}(M) \oplus \mathbb{Z} \to H^{2(2^n-1)}(M;\mathbb{R}/\mathbb{Z}).$$

Just as in the case for differential K-theory (see Propositions 42 and 47), we have:

Proposition 51 *The group of permanent cycles in bidegree* $(0,0)$ *in the AHSS for* $\mathrm{diff}(\widetilde{K}(n), \mathrm{ch})$ *is a subgroup of certain closed forms with rational periods. More precisely, we have*

$$E_\infty^{0,0} \subset \prod_k \Omega_{\mathrm{cl},\mathbb{Q}}^{2k(2^n-1)}(M) \oplus \mathbb{Z}.$$

To identify the Čech cohomology groups with coefficients in $\widehat{K}(n)^0$, we make the identification (as we did for differential K-theory)

$$\widehat{K}(n)^0 \simeq \prod_k \Omega_{\mathrm{cl}}^{2k(2^n-1)} \oplus \underline{\mathbb{Z}}$$

on the site of Cartesian spaces. Again, using the sheaf condition over smooth manifolds, we have

$$H^p(M; \widehat{K}(n)^0) \simeq \prod_k \Omega_{\mathrm{cl}}^{2k(2^n-1)}(M) \oplus \mathbb{Z}.$$

We now consider the differential refinement of the (integrally lifted) Milnor primitive. As before, let $\Gamma_2\colon H^n(-;\mathbb{Z}/2) \to H^n(-;U(1))$ denote the map induced by the representation of $\mathbb{Z}/2$ as the square roots of unity, and let $\rho_2\colon \mathbb{Z} \to \mathbb{Z}/2$ denote the mod 2 reduction.

Lemma 52 *The integral Milnor primitive $\widetilde{Q}_n$ factors through the representation $\Gamma_2\colon \mathbb{Z}/2 \hookrightarrow U(1)$. That is, there exists an operation $\widehat{Q}_n$ such that*

$$Q_n\rho_2 = \rho_2\widetilde{Q}_n = \rho_2\beta\Gamma_2\widehat{Q}_n,$$

where β is the Bockstein for the exponential sequence.

Proof Recall first that $\rho_2\beta\Gamma_2 = \rho_2\widetilde{\beta} = \mathrm{Sq}^1$, where $\widetilde{\beta}$ is the Bockstein for the mod 2 reduction sequence. We can therefore rewrite the above equation as

$$Q_n\rho_2 = \rho_2\widetilde{Q}_n = \rho_2\beta\Gamma_2\widehat{Q}_n = \mathrm{Sq}^1\widehat{Q}_n,$$

and the existence of the class $\widehat{Q}_n$ holds if and only if $\mathrm{Sq}^1 Q_n\rho_2 = 0$. On the other hand, the existence of the integral lift $\widetilde{Q}_n$ immediately implies this condition. □

Again, let β and $\widetilde{\beta}$ denote the Bockstein homomorphism corresponding to the sequences $0 \to \mathbb{Z} \to \mathbb{R} \to \mathbb{R}/\mathbb{Z} \to 0$ and $0 \to \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}/2 \to 0$, respectively. Then the following can be proved in a similar way as we did for Proposition 43 in the case of differential K-theory.

Proposition 53 (odd differentials for Morava AHSS) *The $(2^{n+1}-1)$–differential in the AHSS for differential Morava K-theory is given by*

$$d_{2^{n+1}-1} = \begin{cases} \Gamma_2\widehat{Q}_n\rho_2\beta, & q < 0, \\ \widetilde{Q}_n, & q > 0, \\ 0, & q = 0. \end{cases}$$

Remark 54 (odd primes) The above discussion has been for the prime 2; that is, we are considering integral Morava K-theory as arising from lifting of the $p = 2$ Morava

K-theory. We can do the same for odd primes, leading to integral Morava K-theory lifted from an odd prime p. A similar discussion follows and we have an integral lift of the Milnor primitive at odd primes, as in Lemma 52. The differentials will be again given by these refinements of the Milnor primitive; ie Proposition 53 holds except that the primitives are defined using the Steenrod reduced power operations P^j. Precisely, Q_0 is the Bockstein homomorphism associated to reduction mod p sequence, and inductively $Q_{i+1} = P^{p^i} Q_i - Q_i P^{p^i}$. The operations P^j have been differentially refined in [33]. Hence the refinement of the Milnor primitives at odd primes will also follow. Then the $(p^{n+1}-1)$–differential in the AHSS for differential Morava K-theory is given by

$$d_{p^{n+1}-1} = \begin{cases} \Gamma_p \widehat{Q}_n \rho_p \beta, & q < 0, \\ \widetilde{Q}_n, & q > 0, \\ 0, & q = 0. \end{cases}$$

Example 55 (lifting fields to differential Morava K-theory) We build on Example 48 and aim to lift the cohomology classes beyond K-theory. In particular, for $x = \lambda = \frac{1}{2} p_1$ the first Spin characteristic class, we have $\hat{x} = \widehat{\lambda}$ the differential refinements of λ [61; 25] (which can be viewed as a lifted Wu class [41]), and we would have $\widehat{\mathrm{Sq}}^3 \widehat{\lambda} = 0$. This condition in differential cohomology can be viewed as a refinement of the condition $W_7 = \mathrm{Sq}^3 \lambda = 0$ leading to orientation with respect to integral Morava $K(2)$–theory (lifted from the prime $p = 2$) as shown in [47] and elaborated further in [12]. From the structure of the smooth AHSS in relation to the classical AHSS, one can extend various results to the differential case. For instance, one can generalize the statement in [47] on orientation to state that: *an oriented smooth* 10*–dimensional manifold is oriented with respect to differential (integrally lifted from* $p = 2$*) Morava* $K(2)$*–theory* $\widehat{K}(2)$ *if the class* $\widehat{W}_7 := \widehat{\mathrm{Sq}}^3 \widehat{\lambda}$ *is equal to* 0. The development of this, as well as the relation to refinements of characteristic classes, deserves a separate treatment and will be addressed elsewhere.

Remark 56 (i) Note that our construction allows for an AHSS for other spectra beyond the particular ones we discussed above. This holds for any spectrum which admits a rationalization, whose coefficients are known, and which can be lifted integrally in the sense that we discussed at the beginning of this section.

(ii) All the cohomology theories that we used in this paper can be twisted. Indeed, the construction in this paper can be generalized to construct an AHSS for twisted differential spectra [32], in the sense of [15], and using [34].

Acknowledgements The authors would like to thank Ulrich Bunke, Thomas Nikolaus and Craig Westerland for interesting discussions at the early stages of this project and Urs Schreiber for very useful comments on the first draft. We are grateful to the anonymous referee for a careful reading and for useful suggestions.

References

[1] **J F Adams**, *Stable homotopy and generalised homology*, Univ. of Chicago Press (1974) MR

[2] **D Arlettaz**, *The order of the differentials in the Atiyah–Hirzebruch spectral sequence*, K-Theory 6 (1992) 347–361 MR

[3] **M F Atiyah**, **F Hirzebruch**, *Vector bundles and homogeneous spaces*, from "Differential Geometry" (C B Allendoerfer, editor), Proc. Sympos. Pure Math. 3, Amer. Math. Soc., Providence, RI (1961) 7–38 MR

[4] **M F Atiyah**, **F Hirzebruch**, *Analytic cycles on complex manifolds*, Topology 1 (1962) 25–45 MR

[5] **N A Baas**, **B I Dundas**, **J Rognes**, *Two-vector bundles and forms of elliptic cohomology*, from "Topology, geometry and quantum field theory" (U Tillmann, editor), London Math. Soc. Lecture Note Ser. 308, Cambridge Univ. Press (2004) 18–45 MR

[6] **A Björner**, *Topological methods*, from "Handbook of combinatorics, II" (R L Graham, M Grötschel, L Lovász, editors), Elsevier Sci. B. V., Amsterdam (1995) 1819–1872 MR

[7] **R Bott**, **L W Tu**, *Differential forms in algebraic topology*, Graduate Texts in Mathematics 82, Springer (1982) MR

[8] **A K Bousfield**, *The localization of spectra with respect to homology*, Topology 18 (1979) 257–281 MR

[9] **J Brodzki**, **V Mathai**, **J Rosenberg**, **R J Szabo**, *D-branes, RR-fields and duality on noncommutative manifolds*, Comm. Math. Phys. 277 (2008) 643–706 MR

[10] **K S Brown**, *Abstract homotopy theory and generalized sheaf cohomology*, Trans. Amer. Math. Soc. 186 (1973) 419–458 MR

[11] **J-L Brylinski**, *Loop spaces, characteristic classes and geometric quantization*, Progress in Mathematics 107, Birkhäuser, Boston (1993) MR

[12] **L Buhné**, *Properties of integral Morava K-theory and the asserted application to the Diaconescu–Moore–Witten anomaly*, PhD thesis, Hamburg Univ. (2011)

[13] **U Bunke**, *Differential cohomology*, preprint (2012) arXiv

[14] **U Bunke**, **M Kreck**, **T Schick**, *A geometric description of differential cohomology*, Ann. Math. Blaise Pascal 17 (2010) 1–16 MR

[15] **U Bunke**, **T Nikolaus**, *Twisted differential cohomology*, preprint (2014) arXiv

[16] **U Bunke**, **T Nikolaus**, **M Völkl**, *Differential cohomology theories as sheaves of spectra*, J. Homotopy Relat. Struct. 11 (2016) 1–66 MR

[17] **U Bunke**, **T Schick**, *Smooth K-theory*, Astérisque 328, Soc. Math. France, Paris (2009) 45–135 MR

[18] **U Bunke**, **T Schick**, *Uniqueness of smooth extensions of generalized cohomology theories*, J. Topol. 3 (2010) 110–156 MR

[19] **A L Carey**, **S Johnson**, **M K Murray**, *Holonomy on D-branes*, J. Geom. Phys. 52 (2004) 186–216 MR

[20] **J Cheeger**, **J Simons**, *Differential characters and geometric invariants*, from "Geometry and topology" (J Alexander, J Harer, editors), Lecture Notes in Math. 1167, Springer (1985) 50–80 MR

[21] **D-E Diaconescu**, **G Moore**, **E Witten**, *E_8 gauge theory, and a derivation of K-theory from M-theory*, Adv. Theor. Math. Phys. 6 (2002) 1031–1134 MR

[22] **D Dugger**, *Combinatorial model categories have presentations*, Adv. Math. 164 (2001) 177–201 MR

[23] **J L Dupont**, **R Ljungmann**, *Integration of simplicial forms and Deligne cohomology*, Math. Scand. 97 (2005) 11–39 MR

[24] **H Esnault**, **E Viehweg**, *Lectures on vanishing theorems*, DMV Seminar 20, Birkhäuser, Basel (1992) MR

[25] **D Fiorenza**, **U Schreiber**, **J Stasheff**, *Čech cocycles for differential characteristic classes: an ∞–Lie theoretic construction*, Adv. Theor. Math. Phys. 16 (2012) 149–250 MR

[26] **A T Fomenko**, **D B Fuchs**, **V L Gutenmacher**, *Homotopic topology*, Akadémiai Kiadó, Budapest (1986) MR

[27] **D S Freed**, *Dirac charge quantization and generalized differential cohomology*, from "Surveys in differential geometry" (S-T Yau, editor), Surv. Differ. Geom. 7, International Press, Somerville, MA (2000) 129–194 MR

[28] **D S Freed**, **J Lott**, *An index theorem in differential K-theory*, Geom. Topol. 14 (2010) 903–966 MR

[29] **D S Freed**, **E Witten**, *Anomalies in string theory with D-branes*, Asian J. Math. 3 (1999) 819–851 MR

[30] **P Gajer**, *Geometry of Deligne cohomology*, Invent. Math. 127 (1997) 155–207 MR

[31] **A Gorokhovsky**, **J Lott**, *A Hilbert bundle description of differential K-theory*, preprint (2015) arXiv

[32] **D Grady**, **H Sati**, *AHSS for twisted differential spectra*, in preparation

[33] **D Grady**, **H Sati**, *Primary operations in differential cohomology*, preprint (2016) arXiv

[34] **D Grady**, **H Sati**, *Massey products in differential cohomology via stacks*, J. Homotopy Relat. Struct. (online publication May 2017)

[35] **P A Griffiths**, **J W Morgan**, *Rational homotopy theory and differential forms*, Progress in Mathematics 16, Birkhäuser, Boston (1981) MR

[36] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR

[37] **P Hekmati**, **M K Murray**, **V S Schlegel**, **R F Vozzo**, *A geometric model for odd differential K-theory*, Differential Geom. Appl. 40 (2015) 123–158 MR

[38] **P Hilton**, *General cohomology theory and K-theory*, London Math. Soc. Lecture Note Ser. 1, Cambridge Univ. Press (1971) MR

[39] **M-H Ho**, *The differential analytic index in Simons–Sullivan differential K-theory*, Ann. Global Anal. Geom. 42 (2012) 523–535 MR

[40] **M-H Ho**, *Remarks on flat and differential K-theory*, Ann. Math. Blaise Pascal 21 (2014) 91–101 MR

[41] **M J Hopkins**, **I M Singer**, *Quadratic functions in geometry, topology, and M-theory*, J. Differential Geom. 70 (2005) 329–452 MR

[42] **D Husemöller**, **M Joachim**, **B Jurčo**, **M Schottenloher**, *Basic bundle theory and K-cohomology invariants*, Lecture Notes in Physics 726, Springer (2008) MR

[43] **J F Jardine**, *Stable homotopy theory of simplicial presheaves*, Canad. J. Math. 39 (1987) 733–747 MR

[44] **J F Jardine**, *Local homotopy theory*, Springer (2015) MR

[45] **M Karoubi**, *Homologie cyclique et K-théorie*, Astérisque 149, Soc. Math. France, Paris (1987) MR

[46] **K R Klonoff**, *An index theorem in differential K-theory*, PhD thesis, The University of Texas at Austin (2008) MR Available at `http://search.proquest.com/docview/230710076`

[47] **I Kriz**, **H Sati**, *M-theory, type IIA superstrings, and elliptic cohomology*, Adv. Theor. Math. Phys. 8 (2004) 345–394 MR

[48] **J A Lind**, **H Sati**, **C Westerland**, *A higher categorical analogue of topological T-duality for sphere bundles*, preprint (2016) arXiv

[49] **J Lott**, *$\mathbf{R}/\mathbf{Z}$ index theory*, Comm. Anal. Geom. 2 (1994) 279–311 MR

[50] **J Lott**, *Secondary analytic indices*, from "Regulators in analysis, geometry and number theory" (A Reznikov, N Schappacher, editors), Progr. Math. 171, Birkhäuser, Boston (2000) 231–293 MR

[51] **J Lurie**, *Higher algebra*, prepublication book draft (2011) Available at `http://www.math.harvard.edu/~lurie/papers/HA.pdf`

[52] **C R F Maunder**, *The spectral sequence of an extraordinary cohomology theory*, Proc. Cambridge Philos. Soc. 59 (1963) 567–574 MR

[53] **J P May**, **J Sigurdsson**, *Parametrized homotopy theory*, Mathematical Surveys and Monographs 132, Amer. Math. Soc., Providence, RI (2006) MR

[54] **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Studies in Advanced Mathematics 58, Cambridge Univ. Press (2001) MR

[55] **H R Miller**, *On relations between Adams spectral sequences, with an application to the stable homotopy of a Moore space*, J. Pure Appl. Algebra 20 (1981) 287–312 MR

[56] **R Minasian**, **G Moore**, *K-theory and Ramond–Ramond charge*, J. High Energy Phys. (1997) art. id. 002, 7 pages MR

[57] **V V Prasolov**, *Elements of combinatorial and differential topology*, Graduate Studies in Mathematics 74, Amer. Math. Soc., Providence, RI (2006) MR

[58] **D C Ravenel**, *Localization with respect to certain periodic homology theories*, Amer. J. Math. 106 (1984) 351–414 MR

[59] **Y B Rudyak**, *On Thom spectra, orientability, and cobordism*, Springer (1998) MR

[60] **H Sati**, *Geometric and topological structures related to M-branes*, from "Superstrings, geometry, topology, and C^*–algebras" (R S Doran, G Friedman, J Rosenberg, editors), Proc. Sympos. Pure Math. 81, Amer. Math. Soc., Providence, RI (2010) 181–236 MR

[61] **H Sati**, **U Schreiber**, **J Stasheff**, *Twisted differential string and fivebrane structures*, Comm. Math. Phys. 315 (2012) 169–213 MR

[62] **H Sati**, **C Westerland**, *Twisted Morava K-theory and E-theory*, J. Topol. 8 (2015) 887–916 MR

[63] **U Schreiber**, *Differential cohomology in a cohesive infinity-topos*, preprint (2013) arXiv

[64] **S Schwede**, *Symmetric spectra*, unpublished manuscript (2012) Available at `http://www.math.uni-bonn.de/people/schwede/SymSpec-v3.pdf`

[65] **B Shipley**, *$H\mathbb{Z}$–algebra spectra are differential graded algebras*, Amer. J. Math. 129 (2007) 351–379 MR

[66] **J Simons**, **D Sullivan**, *Axiomatic characterization of ordinary differential cohomology*, J. Topol. 1 (2008) 45–56 MR

[67] **R M Switzer**, *Algebraic topology — homotopy and homology*, Grundl. Math. Wissen. 212, Springer (1975) MR

[68] **H Tamanoi**, *Q–subalgebras, Milnor basis, and cohomology of Eilenberg–Mac Lane spaces*, J. Pure Appl. Algebra 137 (1999) 153–198 MR

[69] **T Tradler**, **S O Wilson**, **M Zeinalian**, *An elementary differential extension of odd K-theory*, J. K-Theory 12 (2013) 331–361 MR

[70] **T Tradler**, **S O Wilson**, **M Zeinalian**, *Differential K-theory as equivalence classes of maps to Grassmannians and unitary groups*, New York J. Math. 22 (2016) 527–581 MR

[71] **M Upmeier**, *Algebraic structure and integration maps in cocycle models for differential cohomology*, Algebr. Geom. Topol. 15 (2015) 65–83 MR

[72] **N Yagita**, *On the Steenrod algebra of Morava K-theory*, J. London Math. Soc. 22 (1980) 423–438 MR

Department of Mathematics, New York University, Abu Dhabi
Abu Dhabi, United Arab Emirates

Department of Mathematics, University of Pittsburgh
Pittsburgh, PA, United States

djg9@nyu.edu, hsati@pitt.edu

Received: 3 June 2016 Revised: 11 October 2016

Algebraic & Geometric Topology 17 (2017) 2413–2428

Epimorphisms between 2–bridge knot groups and their crossing numbers

MASAAKI SUZUKI

Suppose that there exists an epimorphism from the knot group of a 2–bridge knot K onto that of another knot K'. We study the relationship between their crossing numbers $c(K)$ and $c(K')$. More specifically, it is shown that $c(K)$ is greater than or equal to $3c(K')$, and we estimate how many knot groups a 2–bridge knot group maps onto. Moreover, we formulate the generating function which determines the number of 2–bridge knot groups admitting epimorphisms onto the knot group of a given 2–bridge knot.

57M25; 57M27

1 Introduction

Let K be a knot and $G(K)$ the knot group, namely, the fundamental group of the exterior of K in S^3. We denote by $c(K)$ the crossing number of K. Recently, many authors have studied epimorphisms between knot groups. One of the main goals of their papers was Simon's conjecture: every knot group maps onto at most finitely many knot groups. For example, Boileau, Boyer, Reid and Wang [4] showed that Simon's conjecture is true for 2–bridge knots. Finally, Agol and Liu [2] proved that Simon's conjecture holds for all knots.

In Kitano and Suzuki [12] and Horie, Kitano, Matsumoto and Suzuki [10], the existence and nonexistence of a meridional epimorphism between knot groups of prime knots with up to 11 crossings are determined completely. We say that a homomorphism from $G(K)$ to $G(K')$ is *meridional* if a meridian of $G(K)$ is sent to a meridian of $G(K')$; see also Cha and Suzuki [7]. This result raises the following question: if there exists an epimorphism from $G(K)$ onto $G(K')$, then is $c(K)$ greater than or equal to $c(K')$? This question is also mentioned in Kitano and Suzuki [13]. If the answer is affirmative, then we obtain another proof for Simon's conjecture. This paper gives a partial affirmative answer for this question. That is to say, if there exists an epimorphism from the knot group of a 2–bridge knot K onto that of another knot K', then $c(K)$ is greater than or equal to $3c(K')$.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2413

In order to prove this result, we make use of the Ohtsuki–Riley–Sakuma construction [18]; these authors established a systematic construction of epimorphisms between 2–bridge knot groups. Additionally, Garrabrant, Hoste and Shanahan [9] gave necessary and sufficient conditions for any set of 2–bridge knots to have an upper bound with respect to the Ohtsuki–Riley–Sakuma construction. Conversely, it is shown that all epimorphisms between 2–bridge knot groups arise from the Ohtsuki–Riley–Sakuma construction, as a consequence of Agol's result announced in [1]. Aimi, Lee and Sakuma [3] give another proof for this result.

In this paper, we consider the crossing numbers of 2–bridge knots whose knot groups admit epimorphisms onto a 2–bridge knot group. By using this result, we estimate how many knot groups a 2–bridge knot group maps onto. Furthermore, we formulate the generating function which determines the number of 2–bridge knots K admitting epimorphisms from $G(K)$ onto the knot group of a given 2–bridge knot.

Throughout this paper, we do not distinguish a knot from its mirror image, since their knot groups are isomorphic and we discuss epimorphisms between knot groups. The numberings of the knots with up to 10 and 11 crossings follow Rolfsen's book [19] and the web page KnotInfo [6] by Cha and Livingston, respectively

2 2–bridge knot and continued fraction expansion

In this section, we recall some well-known results on 2–bridge knots. See [5; 17] in detail, for example.

A 2–bridge knot corresponds to a rational number $r = q/p \in \mathbb{Q}$; we denote the knot by $K(q/p)$. Schubert classified 2–bridge knots as follows.

Theorem 2.1 (Schubert) *Let $K(q/p)$ and $K(q'/p')$ be 2–bridge knots. These knots are equivalent if and only if the following conditions hold:*

(1) $p = p'$.

(2) *Either* $q \equiv \pm q' \pmod{p}$ *or* $qq' \equiv \pm 1 \pmod{p}$.

By using this theorem, it is sufficient to consider $r \in \mathbb{Q} \cap (0, \frac{1}{2})$. Note that $K(0)$ is the trivial link and that $K(\frac{1}{2})$ is the Hopf link. A rational number $q/p \in \mathbb{Q} \cap (0, \frac{1}{2})$ can be expressed as a continued fraction expansion

$$\frac{q}{p} = [a_1, a_2, \dots, a_{m-1}, a_m] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots \cfrac{1}{a_{m-1} + \cfrac{1}{a_m}}}}},$$

where $a_1 > 0$. Note that a rational number admits many continued fraction expansions. For example, we have $\frac{29}{81} = [3, -5, 4, 1, -2] = [2, 1, 3, 1, 5]$. It is easy to see that the following properties are satisfied. First, we can delete zeros in a continued fraction expansion by using the equation

$$[a_1, a_2, \dots, a_{i-2}, a_{i-1}, 0, a_{i+1}, a_{i+2}, \dots, a_m] \\ = [a_1, a_2, \dots, a_{i-2}, a_{i-1} + a_{i+1}, a_{i+2}, \dots, a_m].$$

If we consider a 2–bridge knot, we may assume that $a_1, a_m \neq \pm 1$, since

$$[a_1, a_2, \dots, a_{m-1}, \pm 1] = [a_1, a_2, \dots, a_{m-1} \pm 1]$$

and $K([a_1, a_2, \dots, a_m])$ is equivalent to $K([a_m, a_{m-1}, \dots, a_1])$ up to mirror image. Moreover, the euclidean algorithm allows us to take a continued fraction expansion such that all a_i in $[a_1, a_2, \dots, a_m]$ are positive.

If a rational number r is expressed as $[a_1, a_2, \dots, a_m]$ with $a_i > 0$ and $a_1, a_m \geq 2$, then the continued fraction expansion is called *standard*. By the above arguments, we can always take the standard continued fraction expansion of the rational number r for a 2–bridge knot $K(r)$. Furthermore, the standard continued fraction expansion gives us the unique continued fraction expansion of the rational number which corresponds to a 2–bridge knot in the following sense. Let $K(q/p)$ and $K(q'/p')$ be 2–bridge knots. Suppose that these rational numbers are written as the standard continued fraction expansions $q/p = [a_1, a_2, \dots, a_m]$ and $q'/p' = [a'_1, a'_2, \dots, a'_{m'}]$. It is known that $K(q/p)$ and $K(q'/p')$ are equivalent up to mirror image if and only if

$$(a_1, a_2, \dots, a_m) = (a'_1, a'_2, \dots, a'_{m'}) \text{ or } (a'_{m'}, a'_{m'-1}, \dots, a'_1).$$

Thistlethwaite [21], Kauffman [11] and Murasugi [15; 16] independently proved the first Tait conjecture. Hence, we can determine the crossing number of a 2–bridge knot by using the standard continued fraction expansion. Namely, the crossing number for the standard continued fraction $[a_1, a_2, \dots, a_m]$ is given by

$$c(K([a_1, a_2, \dots, a_m])) = \sum_{i=1}^{m} a_i.$$

3 Epimorphisms between 2–bridge knot groups

We have the following remarkable result about epimorphisms between 2–bridge knot groups: an epimorphism between 2–bridge knot groups is always meridional. Moreover, the rational numbers for these 2–bridge knots have the following relationship.

Theorem 3.1 (Ohtsuki, Riley and Sakuma [18], Agol [1], Aimi, Lee and Sakuma [3]) *Let $K(r), K(\widetilde{r})$ be 2–bridge knots, where $r=[a_1,a_2,\dots,a_m]$. If there exists an epimorphism $\varphi\colon G(K(\widetilde{r}))\to G(K(r))$, then φ is meridional and $\widetilde{r}$ can be written as*

$$(*)\qquad \widetilde{r}=[\varepsilon_1\boldsymbol{a},2c_1,\varepsilon_2\boldsymbol{a}^{-1},2c_2,\varepsilon_3\boldsymbol{a},2c_3,\varepsilon_4\boldsymbol{a}^{-1},2c_4,\dots,\varepsilon_{2n}\boldsymbol{a}^{-1},2c_{2n},\varepsilon_{2n+1}\boldsymbol{a}],$$

where $\boldsymbol{a}=(a_1,a_2,\dots,a_m)$, $\boldsymbol{a}^{-1}=(a_m,a_{m-1},\dots,a_1)$, $\varepsilon_i=\pm1$ $(\varepsilon_1=1)$, and $c_i\in\mathbb{Z}$.

Remark If a rational number $\widetilde{r}$ is expressed in the form $(*)$, then we say that $\widetilde{r}$ has an expansion of *type* $2n+1$ with respect to $\boldsymbol{a}=(a_1,a_2,\dots,a_m)$.

(1) In this paper, we do not need to consider an expression of type $2n$ with respect to $\boldsymbol{a}$, since $K([\varepsilon_1\boldsymbol{a},2c_1,\dots,2c_{2n-1},\varepsilon_{2n}\boldsymbol{a}^{-1}])$ is a 2–bridge link.

(2) If $c_i=0$ and $\varepsilon_i\cdot\varepsilon_{i+1}=-1$, then

$$\begin{aligned}\widetilde{r}&=[\dots,\varepsilon_{i-1}\boldsymbol{a}^{\pm1},2c_{i-1},\varepsilon_i\boldsymbol{a}^{\mp1},0,\varepsilon_{i+1}\boldsymbol{a}^{\pm1},2c_{i+1},\varepsilon_{i+2}\boldsymbol{a}^{\mp1},\dots]\\&=[\dots,\varepsilon_{i-1}\boldsymbol{a}^{\pm1},2c_{i-1},0,2c_{i+1},\varepsilon_{i+2}\boldsymbol{a}^{\mp1},\dots]\\&=[\dots,\varepsilon_{i-1}\boldsymbol{a}^{\pm1},2(c_{i-1}+c_{i+1}),\varepsilon_{i+2}\boldsymbol{a}^{\mp1},\dots].\end{aligned}$$

It follows that $\widetilde{r}$ has type $2n-1$. Hence we do not deal with the case $c_i=0$, $\varepsilon_i\cdot\varepsilon_{i+1}=-1$.

Example 3.2 For example, we consider a 2–bridge knot $K(\frac{5}{27})$. The rational number $\frac{5}{27}$ has continued fraction expansions

$$\tfrac{5}{27}=[5,2,2]=[3,0,3,-2,3].$$

The second expression implies that the crossing number of $K(\frac{5}{27})$ is 9. The last expression is of type 3 with respect to $\boldsymbol{a}=(3)$. Therefore the knot group $G(K(\frac{5}{27}))$ admits an epimorphism onto the trefoil knot group $G(3_1)=G(K(\frac{1}{3}))=G(K([3]))$. Similarly, we have

$$\tfrac{1}{9}=[9]=[3,0,3,0,3],\qquad \tfrac{19}{45}=[2,2,1,2,2]=[3,-2,3,-2,3].$$

It follows that there exist epimorphisms from $G(K(\frac{1}{9}))$ and $G(K(\frac{19}{45}))$ onto the trefoil knot group.

The previous papers [12] and [10] determined all the pairs of prime knots with up to 11 crossings which admit meridional epimorphisms between their knot groups. The results in those works coincide with the above examples. Note that $K(\frac{1}{9})=9_1$, $K(\frac{5}{27})=9_6$, and $K(\frac{19}{45})=9_{23}$.

In general, even if $[a_1, \ldots, a_m]$ is the standard continued fraction expansion, and $\tilde{r}$ is of type $2n+1$ with respect to $(a_1, \ldots, a_m)$, this expansion of $\tilde{r}$ may not be standard. However, we can get the standard continued fraction expansion and then determine the crossing number of $K(\tilde{r})$.

Theorem 3.3 *Let $[a_1, \ldots, a_m]$ be the standard continued fraction expansion. Suppose that a rational number $\tilde{r}$ has a continued fraction expansion of the form $(*)$ of type $2n+1$ with respect to $\boldsymbol{a} = (a_1, \ldots, a_m)$. Then the crossing number of $K(\tilde{r})$ is given by*

$$c(K(\tilde{r})) = (2n+1)|\boldsymbol{a}| + \sum_{i=1}^{2n} (2|c_i| - \psi(i) - \bar{\psi}(i)),$$

where $|\boldsymbol{a}| = \sum_{i=1}^{m} a_i$ and

$$\psi(i) = \begin{cases} 1 & \text{if } \varepsilon_i \cdot c_i < 0, \\ 0 & \text{if } \varepsilon_i \cdot c_i \geq 0, \end{cases} \qquad \bar{\psi}(i) = \begin{cases} 1 & \text{if } c_i \cdot \varepsilon_{i+1} < 0, \\ 0 & \text{if } c_i \cdot \varepsilon_{i+1} \geq 0. \end{cases}$$

Note that

$$\sum_{i=1}^{2n} (\psi(i) + \bar{\psi}(i))$$

is the number of sign changes. To prepare for Theorem 3.3, we prove the following lemma. Namely, negative integers in a continued fraction expansion can be changed into positive integers.

Lemma 3.4 *Let $a_1, \ldots, a_k, b_1, \ldots, b_l, c_1, \ldots, c_m$ be integers. We have four cases:*

(1) *If $l \geq 2$, then*

$$\begin{aligned}&[a_1, \ldots, a_k, -b_1, -b_2, \ldots, -b_{l-1}, -b_l, c_1, \ldots, c_m] \\ &\quad = [a_1, \ldots, , a_{k-1}, a_k - 1, 1, b_1 - 1, b_2, \ldots, b_{l-1}, b_l - 1, 1, c_1 - 1, c_2, \ldots, c_m].\end{aligned}$$

(2) *If $l = 1$ and $b_1 \geq 2$, then*

$$\begin{aligned}&[a_1, \ldots, a_k, -b_1, c_1, \ldots, c_m] \\ &\qquad = [a_1, \ldots, , a_{k-1}, a_k - 1, 1, b_1 - 2, 1, c_1 - 1, c_2, \ldots, c_m].\end{aligned}$$

(3) *If $l \geq 2$, then*

$$[a_1, \ldots, a_k, -b_1, \ldots, -b_l] = [a_1, \ldots, , a_{k-1}, a_k - 1, 1, b_1 - 1, b_2, \ldots, b_l].$$

(4) *If $l = 1$ and $b_1 \geq 2$, then*

$$[a_1, \ldots, a_k, -b_1] = [a_1, \ldots, , a_{k-1}, a_k - 1, 1, b_1 - 1].$$

Proof Recall the matrix representation of a continued fraction expansion (see for instance [14]). For a continued fraction $[x_1, x_2, \dots, x_m]$, we define p, q by

$$\begin{pmatrix} x_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} x_m & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p & * \\ q & * \end{pmatrix}.$$

It is known that we have an equality

$$[x_1, x_2, \dots, x_m] = \frac{q}{p}.$$

We will prove (1) by using the above matrix representation on both sides of the equation. Let A, C be the matrices defined by

$$A = \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix},$$

respectively, and define B by

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} b_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_{l-1} & 1 \\ 1 & 0 \end{pmatrix}.$$

The matrix representation of the left-hand side of (1) is

$$\begin{aligned}
&\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -b_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -b_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \\
&\qquad\qquad \cdot \begin{pmatrix} -b_{l-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -b_l & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix} \\
&= A \begin{pmatrix} -a_k b_1 + 1 & a_k \\ -b_1 & 1 \end{pmatrix} \begin{pmatrix} (-1)^l B_{11} & (-1)^{l+1} B_{12} \\ (-1)^{l+1} B_{21} & (-1)^l B_{22} \end{pmatrix} \begin{pmatrix} -b_l c_1 + 1 & -b_l \\ c_1 & 1 \end{pmatrix} C \\
&= A \begin{pmatrix} a_k b_1 - 1 & a_k \\ b_1 & 1 \end{pmatrix} \begin{pmatrix} (-1)^l B_{11} & (-1)^l B_{12} \\ (-1)^l B_{21} & (-1)^l B_{22} \end{pmatrix} \begin{pmatrix} b_l c_1 - 1 & b_l \\ c_1 & 1 \end{pmatrix} C \\
&= (-1)^l A \begin{pmatrix} a_k b_1 - 1 & a_k \\ b_1 & 1 \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} b_l c_1 - 1 & b_l \\ c_1 & 1 \end{pmatrix} C \\
&= (-1)^l \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \\
&\qquad\qquad \cdot \begin{pmatrix} b_1 - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_{l-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_l - 1 & 1 \\ 1 & 0 \end{pmatrix} \\
&\qquad\qquad \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix}.
\end{aligned}$$

The last expression is $(-1)^l$ times the matrix representation of the right-hand side of (1). Therefore the rational numbers on both sides of (1) coincide.

Next, we examine the matrix representation of the left-hand side of (2):

$$\begin{aligned}
&\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -b_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix} \\
&\qquad = A \begin{pmatrix} -a_k b_1 c_1 + a_k + c_1 & -a_k b_1 + 1 \\ -b_1 c_1 + 1 & -b_1 \end{pmatrix} C \\
&\qquad = (-1) A \begin{pmatrix} a_k b_1 c_1 - a_k - c_1 & a_k b_1 - 1 \\ b_1 c_1 - 1 & b_1 \end{pmatrix} C \\
&\qquad = (-1) \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_1 - 2 & 1 \\ 1 & 0 \end{pmatrix} \\
&\qquad\qquad \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 - 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} c_m & 1 \\ 1 & 0 \end{pmatrix}.
\end{aligned}$$

The last expression is also (-1) times the matrix representation of the right-hand side of (2). Hence these continued fraction expressions represent the same rational number.

A similar proof works for (3) and (4). □

Example 3.5 Suppose that the rational number $\frac{29}{81}$ is expressed as $[3, -5, 4, 1, -2]$. The above arguments show that

$$[3, -5, 4, 1, -2] = [2, 1, 3, 1, 3, 0, 1, 1] = [2, 1, 3, 1, 4, 1] = [2, 1, 3, 1, 5].$$

The last expression is the standard continued fraction expansion, and then the crossing number of $K(\frac{29}{81})$ is

$$c(K(\tfrac{29}{81})) = c(K([3, -5, 4, 1, -2])) = c(K([2, 1, 3, 1, 5])) = 12.$$

In Lemma 3.4, if all a_i, b_i, c_i are positive, then the integers on the right-hand sides of the equations are positive or zero. Hence, we can obtain the standard continued fraction expansion and determine the crossing number of a 2–bridge knot.

Corollary 3.6 *Let a_i, b_i, c_i be positive integers. If $l \neq 1$ or $b_1 \geq 2$, then we have*

(1) $c(K([a_1, \dots, a_k, -b_1, \dots, -b_l, c_1, \dots, c_m])) = \sum_{i=1}^{k} a_i + \sum_{i=1}^{l} b_i + \sum_{i=1}^{m} c_i - 2,$

(2) $c(K([a_1, \dots, a_k, -b_1, \dots, -b_l])) = \sum_{i=1}^{k} a_i + \sum_{i=1}^{l} b_i - 1.$

Corollary 3.6 suggests how to determine the crossing number without using the explicit standard continued fraction expansion. To be precise, it is sufficient to compute the sum of the absolute values in a continued fraction expansion and to count the number of sign changes. In the above example, the signs of components in $[3,-5,4,1,-2]$ are changed three times. Then the crossing number is

$$c(K([3,-5,4,1,-2])) = |3|+|-5|+|4|+|1|+|-2|-3 = 12.$$

These arguments prove Theorem 3.3.

Proof of Theorem 3.3 The sum of the absolute values of components in $\tilde{r}$ is

$$(2n+1)|\boldsymbol{a}| + \sum_{i=1}^{2n}(2|c_i|).$$

By Lemma 3.4, if the signs in a continued fraction expansion of $\tilde{r}$ are changed k times, then the crossing number of $K(\tilde{r})$ is decreased by k from the above value. The number of sign changes in $\tilde{r}$ is

$$\sum_{i=1}^{2n}(\psi(i)+\overline{\psi}(i))$$

by definition. Since $\tilde{r}$ is an expression of type $2n+1$ with respect to standard $\boldsymbol{a}$, we can apply Corollary 3.6. Therefore this completes the proof. □

We define $\overline{c_i}$ to be $2|c_i| - \psi(i) - \overline{\psi}(i)$. Then $\overline{c_i}$ is not negative.

Proposition 3.7 *Suppose that $\tilde{r}$ is as above. Then $\overline{c_i} \geq 0$ for $1 \leq i \leq 2n$.*

Proof If $c_i \neq 0$, then $2|c_i| \geq 2$. On the other hand, $\psi(i)$ and $\overline{\psi}(i)$ are 0 or 1, and then we get $\overline{c_i} \geq 0$. If $c_i = 0$, then $\psi(i) = 0$ and $\overline{\psi}(i) = 0$ by definition. Therefore $\overline{c_i} = 0$ in this case. □

4 Simon's conjecture

Simon's conjecture for 2–bridge knots is proved in [4], and for all knots in [2], as mentioned in Section 1. In this section, we investigate how many knot groups a 2–bridge knot group maps onto.

Let $\mathrm{EK}(n)$ be the maximal number of knots whose knot groups a 2–bridge knot group with n crossings admits epimorphisms onto. Theorem 3.3 and Proposition 3.7 imply the following, which is one of the main results in this paper. It gives us a rough estimate of $\mathrm{EK}(n)$.

Theorem 4.1 *Let $K(\widetilde{r})$ be a 2–bridge knot. If there exists an epimorphism from $G(K(\widetilde{r}))$ onto the knot group of another knot K, then*

$$c(K(\widetilde{r})) \geq 3c(K).$$

In particular, all the 2–bridge knots K with up to 8 crossings are minimal, that is to say, if $G(K)$ admits an epimorphism onto a knot group $G(K')$, then K' is equivalent to K or the trivial knot.

Proof By [4, Corollary 1.3] and [20, Proposition 2.4], if $G(K(\widetilde{r}))$ admits an epimorphism onto $G(K)$, then K is also a 2–bridge knot or the trivial knot. In the case that K is the trivial knot, the desired inequality obviously holds.

Next, we assume that K is a 2–bridge knot and that r is the corresponding rational number. Take the standard continued fraction expansion $[a_1, a_2, \dots, a_m]$ of r. Then $\widetilde{r}$ has an expansion of type $2n+1$ with respect to $\boldsymbol{a} = (a_1, a_2, \dots, a_m)$. By Theorem 3.3, we have

$$\begin{aligned} c(K(\widetilde{r})) &= (2n+1)|\boldsymbol{a}| + \sum_{i=1}^{2n} (2|c_i| - \psi(i) - \overline{\psi}(i)) \\ &= (2n+1)c(K) + \sum_{i=1}^{2n} \overline{c}_i \\ &\geq (2n+1)c(K) \qquad\qquad \text{(by Proposition 3.7)} \\ &\geq 3c(K). \end{aligned}$$

Furthermore, since a nontrivial knot has at least 3 crossings, all the 2–bridge knots with up to 8 crossings are minimal. □

Remark The previous paper [12] shows that there are seven knots with less than 9 crossings whose knot groups admit epimorphisms onto the trefoil knot group. To be precise, they are the 3–bridge knots 8_5, 8_{10}, 8_{15}, 8_{18}, 8_{19}, 8_{20}, 8_{21}. So the inequality of Theorem 4.1 does not hold for 3–bridge knots.

Ernst and Sumners [8] determined the number $\mathrm{TK}(n)$ of 2–bridge knots in terms of the crossing number $n \geq 3$ as follows:

$$\mathrm{TK}(n) = \begin{cases} \frac{1}{3}(2^{(n-3)} + 2^{(n-4)/2}) & \text{if } n \equiv 0 \pmod 4, \\ \frac{1}{3}(2^{(n-3)} + 2^{(n-3)/2}) & \text{if } n \equiv 1 \pmod 4, \\ \frac{1}{3}(2^{(n-3)} + 2^{(n-4)/2} - 1) & \text{if } n \equiv 2 \pmod 4, \\ \frac{1}{3}(2^{(n-3)} + 2^{(n-3)/2} + 1) & \text{if } n \equiv 3 \pmod 4. \end{cases}$$

Then we can estimate $\mathrm{EK}(n)$ by using Theorem 4.1:

$$\mathrm{EK}(n) \leq \sum_{k=3}^{\lfloor n/3 \rfloor} \mathrm{TK}(k).$$

These numbers are obtained as shown in the following table:

n	9–11	12–14	15–17	18–20	21–23	24–26	27–29	30–32	33–35	36–38	39–41
$\sum_{k=3}^{\lfloor n/3 \rfloor} \mathrm{TK}(k)$	1	2	4	7	14	26	50	95	186	362	714

In particular, we obtain that the knot groups of 2–bridge knots with 12, 13 or 14 crossings map onto at most two knot groups, which are the trefoil knot group $G(3_1)$ and the figure eight knot group $G(4_1)$. On the other hand, Garrabrant, Hoste and Shanahan studied an upper bound for a set of 2–bridge knots with respect to epimorphisms between their knot groups. We recall their arguments more precisely. Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_{2n})$ be a vector such that

(1) each a_i is in $\{-2, 0, 2\}$,

(2) $a_1 \neq 0$ and $a_{2n} \neq 0$,

(3) if $a_i = 0$, then $a_{i-1} = a_{i+1} = \pm 2$.

For such an $\boldsymbol{a}$, we call $[a_1, a_2, \ldots, a_{2n}]$ an *even standard continued fraction expansion*. If we consider $\boldsymbol{a} = (a_1, a_2, \ldots, a_{2n})$ up to the equivalence relations $\boldsymbol{a} = \pm\boldsymbol{b}$ and $\boldsymbol{a} = \pm\boldsymbol{b}^{-1}$, where $\boldsymbol{b}^{-1}$ is $\boldsymbol{b}$ read backwards, then a 2–bridge knot can be expressed uniquely as $K([a_1, a_2, \ldots, a_{2n}])$ by using an even standard continued fraction expansion:

Proposition 4.2 (Garrabrant, Hoste and Shanahan [9]) *Let $[a_1, a_2, \ldots, a_{2n}]$ and $[b_1, b_2, \ldots, b_{2n}]$ be even standard continued fraction expansions of the same length. If a 2–bridge knot group admits epimorphisms onto $G(K([a_1, a_2, \ldots, a_{2n}]))$ and $G(K([b_1, b_2, \ldots, b_{2n}]))$, then $(a_1, a_2, \ldots, a_{2n}) = (b_1, b_2, \ldots, b_{2n})$.*

For example, the trefoil is $3_1 = K([2, -2])$ and the figure eight knot is $4_1 = K([2, 2])$. Since the lengths of these even standard continued fraction expansions are the same, there does not exist a 2–bridge knot whose knot group admits epimorphisms onto $G(3_1)$ and $G(4_1)$ simultaneously, by Proposition 4.2. Similarly, a 2–bridge knot group maps onto the knot group of at most one of $\{5_1, 5_2, 6_1, 6_2, 6_3\}$, since

$$5_1 = K([2, -2, 2, -2]), \quad 5_2 = K([2, -2, 0, -2]),$$
$$6_1 = K([2, 0, 2, 2]), \quad 6_2 = K([2, 2, -2, 2]), \quad 6_3 = K([2, -2, -2, 2]).$$

In order to extend this argument, we consider the relationship between the length of an even standard continued fraction expansion and the crossing number.

Proposition 4.3 *Let $[a_1, a_2, \dots, a_{2n}]$ be an even standard continued fraction expansion. Then the crossing number $c(K([a_1, a_2, \dots, a_{2n}]))$ satisfies the inequalities*

$$2n + 1 \le c(K([a_1, a_2, \dots, a_{2n}])) \le 4n.$$

Proof First of all, we delete zeros in $(a_1, a_2, \dots, a_{2n})$ as before:

$$[a_1, a_2, \dots, a_{2n}] = [a'_1, a'_2, \dots, a'_{2n'}],$$

where $a'_i \in 2\mathbb{Z} \setminus \{0\}$. Let ℓ be the number of zeros in $(a_1, a_2, \dots, a_{2n})$. Then we have $2\ell = 2n - 2n'$ and

$$\sum_{i=1}^{2n'} (|a'_i| - 2) = 2\ell.$$

It follows that

$$\sum_{i=1}^{2n'} |a'_i| = 2\ell + 4n' = 2n + 2n'.$$

By the same argument as in the proof of Theorem 3.3, we obtain

$$c(K([a_1, a_2, \dots, a_{2n}])) = c(K([a'_1, a'_2, \dots, a'_{2n'}])) = \sum_{i=1}^{2n'} |a'_i| - k = \sum_{i=1}^{2n} |a_i| - k,$$

where k is the number of sign changes in $(a'_1, a'_2, \dots, a'_{2n'})$. Note that $0 \le k \le 2n' - 1$. (If all a'_i are positive or negative, then $k = 0$. If $a'_i \cdot a'_{i+1} < 0$ for all i $(0 \le i \le 2n' - 1)$, then $k = 2n' - 1$.) Since $|a_i| \le 2$, we have

$$\sum_{i=1}^{2n} |a_i| - k \le 4n.$$

Moreover, we obtain

$$\begin{aligned} \sum_{i=1}^{2n'} |a'_i| - k &= 2n + 2n' - k \\ &\ge 2n + 2n' - (2n' - 1) \\ &= 2n + 1. \end{aligned}$$

This completes the proof. □

By Proposition 4.2, if two distinct 2–bridge knots K, K' have even standard continued fraction expansions of the same length, then there does not exist a 2–bridge knot whose knot group maps onto $G(K)$ and $G(K')$. Combined with Proposition 4.2 and Proposition 4.3, we can estimate $\mathrm{EK}(n)$ more precisely.

Theorem 4.4 *The number* $\mathrm{EK}(n)$ *satisfies*

$$\mathrm{EK}(n) \leq \left\lfloor \frac{n-3}{6} \right\rfloor.$$

Proof Let K be a 2–bridge knot with n crossings. If $G(K)$ admits an epimorphism onto $G(K')$, then the crossing number of K' is at most $\lfloor n/3 \rfloor$, by Theorem 4.1. Let $[a_1, a_2, \dots, a_{2m}]$ be the even standard continued fraction expansion of K', namely $K' = K([a_1, a_2, \dots, a_{2m}])$. By Proposition 4.2, $\mathrm{EK}(n)$ is less than or equal to the number of the lengths of even standard continued fraction expansions. By Proposition 4.3, we have

$$2m \leq \lfloor n/3 \rfloor - 1.$$

Hence we obtain

$$\mathrm{EK}(n) \leq \left\lfloor \frac{\lfloor n/3 \rfloor - 1}{2} \right\rfloor = \left\lfloor \frac{n-3}{6} \right\rfloor. \qquad \square$$

For example, the knot group of a 2–bridge knot with 50 crossings maps onto at most seven distinct knot groups. Actually, we can get the precise number $\mathrm{EK}(n)$ for $n \leq 30$ by computer program:

$$\mathrm{EK}(n) = \begin{cases} 0 & \text{if } n = 3, 4, 5, 6, 7, 8, \\ 1 & \text{if } n = 9, 10, 11, 12, 13, 14, 18, 19, 20, 24, \\ 2 & \text{if } n = 15, 16, 17, 21, 22, 23, 25, 26, 27, 28, 29, 30. \end{cases}$$

In particular, $\mathrm{EK}(n)$ is less than 3 for all $n \leq 30$. On the other hand, it is easy to see that $G(K([45]))$ maps onto $G(K([3]))$, $G(K([5])$, $G(K([9]))$ and $G(K([15]))$. It follows that $\mathrm{EK}(45) \geq 4$.

Problem Does there exist a 2–bridge knot with less than 45 crossings whose knot group maps onto three (or four) distinct knot groups? In general, determine $\mathrm{EK}(n)$ explicitly for all $n \geq 31$.

5 The generating function

As shown in Example 3.2, there exist three distinct 2–bridge knots with 9 crossings whose knot groups admit epimorphisms onto the trefoil knot group. In this section, we generalize this result. Namely, for a given 2–bridge knot $K(r)$, we determine the number of 2–bridge knots $K(\widetilde{r})$ which admit epimorphisms $\varphi\colon G(K(\widetilde{r})) \to G(K(r))$, in terms of $c(K(\widetilde{r}))$.

Theorem 5.1 *For a given rational number* r, *we take the standard continued fraction expansion* $[a_1, a_2, \dots, a_m]$ *of* r *and define the generating function* f *as follows:*

(1) *If* $(a_1, a_2, \dots, a_m) \neq (a_m, \dots, a_2, a_1)$, *then*

$$f(r) = \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} 2^{2n} \binom{2n+k-1}{k} t^{(2n+1)c(K(r))+k}.$$

(2) *If* $(a_1, a_2, \dots, a_m) = (a_m, \dots, a_2, a_1)$, *then*

$$f(r) = \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} g(n,k) t^{(2n+1)c(K(r))+k},$$

where

$$g(n,k) = \begin{cases} 2^{2n-1}\binom{2n+k-1}{k} & \text{for } k \text{ odd,} \\ 2^{2n-1}\binom{2n+k-1}{k} + 2^{n-1}\binom{n+k/2-1}{k/2} & \text{for } k \text{ even.} \end{cases}$$

Here $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. *Then the number of 2–bridge knots* $K(\widetilde{r})$ *which admit epimorphisms* $\varphi\colon G(K(\widetilde{r})) \to G(K(r))$ *is the coefficient of* $t^{c(K(\widetilde{r}))}$ *in* $f(r)$.

Proof We will count the number of 2–bridge knots with $(2n+1)c(K(r))+k$ crossings which correspond to rational numbers of the form $(*)$. The crossing number $c(K(r))$ is $\sum_{i=1}^{m} a_i$. Compared with Theorem 3.3, we have

$$k = \sum_{i=1}^{2n} \overline{c_i} = \sum_{i=1}^{2n} 2|c_i| - \psi(i) - \overline{\psi}(i),$$

where $\overline{c_i} \geq 0$ by Proposition 3.7.

Suppose that $\overline{c_i} = j$ (≥ 0). Then $(\varepsilon_i \boldsymbol{a}^{\pm 1}, 2c_i, \varepsilon_{i+1}\boldsymbol{a}^{\mp 1})$, which is a part of $\widetilde{r}$, has the following possibilities:

(1) if j is even and $\varepsilon_i = 1$, then

$$(\varepsilon_i \boldsymbol{a}^{\pm 1}, 2c_i, \varepsilon_{i+1}\boldsymbol{a}^{\mp 1}) = (\boldsymbol{a}^{\pm 1}, j, \boldsymbol{a}^{\mp 1}) \text{ or } (\boldsymbol{a}^{\pm 1}, -(j+2), \boldsymbol{a}^{\mp 1});$$

(2) if j is even and $\varepsilon_i = -1$, then

$$(\varepsilon_i \boldsymbol{a}^{\pm 1}, 2c_i, \varepsilon_{i+1}\boldsymbol{a}^{\mp 1}) = (-\boldsymbol{a}^{\pm 1}, -j, -\boldsymbol{a}^{\mp 1}) \text{ or } (-\boldsymbol{a}^{\pm 1}, j+2, -\boldsymbol{a}^{\mp 1});$$

(3) if j is odd and $\varepsilon_i = 1$, then

$$(\varepsilon_i \boldsymbol{a}^{\pm 1}, 2c_i, \varepsilon_{i+1}\boldsymbol{a}^{\mp 1}) = (\boldsymbol{a}^{\pm 1}, j+1, -\boldsymbol{a}^{\mp 1}) \text{ or } (\boldsymbol{a}^{\pm 1}, -(j+1), -\boldsymbol{a}^{\mp 1});$$

(4) if j is odd and $\varepsilon_i = -1$, then

$$(\varepsilon_i \boldsymbol{a}^{\pm 1}, 2c_i, \varepsilon_{i+1}\boldsymbol{a}^{\mp 1}) = (-\boldsymbol{a}^{\pm 1}, j+1, \boldsymbol{a}^{\mp 1}) \text{ or } (-\boldsymbol{a}^{\pm 1}, -(j+1), \boldsymbol{a}^{\mp 1}).$$

Therefore $(\varepsilon_i \boldsymbol{a}^{\pm 1}, 2c_i, \varepsilon_{i+1}\boldsymbol{a}^{\mp 1})$ always has two possibilities. Besides, there are $\binom{2n+k-1}{k}$ cases for $(\overline{c}_1, \dots, \overline{c}_{2n})$, namely

$$(\overline{c}_1, \overline{c}_2, \dots, \overline{c}_{2n}) = (k, 0, \dots, 0), \quad (k-1, 1, \dots, 0), \quad \dots, \quad (0, 0, \dots, 0, k).$$

Hence there are $2^{2n}\binom{2n+k-1}{k}$ 2–bridge knots with $(2n+1)c(K(r)) + k$ crossings, and we get the generating function of (1).

In the case when $(a_1, \dots, a_m) = (a_m, \dots, a_1)$, we see

$$K([\varepsilon_1 \boldsymbol{a}, 2c_1, \varepsilon_2 \boldsymbol{a}^{-1}, \dots, 2c_{2n}, \varepsilon_{2n+1}\boldsymbol{a}]) = K([\varepsilon_{2n+1}\boldsymbol{a}, 2c_{2n}, \dots, \varepsilon_2 \boldsymbol{a}^{-1}, 2c_1, \varepsilon_1 \boldsymbol{a}]).$$

It implies that if $\widetilde{r}$ is not symmetric, that is, if $\widetilde{r}$ is not in the form

$$[\varepsilon_1 \boldsymbol{a}, 2c_1, \dots, 2c_n, \varepsilon_{n+1}\boldsymbol{a}^{\pm 1}, 2c_n, \dots, 2c_1, \varepsilon_1 \boldsymbol{a}],$$

we counted the same knot twice. Then the number of knots is

$$\frac{1}{2}\left(2^{2n}\binom{2n+k-1}{k} - 2^n\binom{n+k/2-1}{k/2}\right) + 2^n\binom{n+k/2-1}{k/2} = 2^{2n-1}\binom{2n+k-1}{k} + 2^{n-1}\binom{n+k/2-1}{k/2}.$$

Notice that if k is odd, then $\widetilde{r}$ must not be symmetric. As we saw in Section 2, if the standard continued fraction expansions are not the same, then the 2–bridge knots are not equivalent. We can get the standard fraction expansion of $\widetilde{r}$ by Lemma 3.4. It shows that these knots which are obtained by the Ohtsuki–Riley–Sakuma construction are not equivalent. This completes the proof. □

Example 5.2 First, we apply Theorem 5.1 to the trefoil knot. The generating function for the trefoil $K(\frac{1}{3}) = K([3])$ is

$$f(\tfrac{1}{3}) = 3t^9 + 4t^{10} + 7t^{11} + 8t^{12} + 11t^{13} + 12t^{14} + 25t^{15} + 48t^{16} + 103t^{17} + 180t^{18} + 309t^{19} + 472t^{20} + 743t^{21} + 1180t^{22} + 2045t^{23} + 3584t^{24} + 6391t^{25} + \cdots.$$

Then the number of 2–bridge knots with 9 crossings whose knot groups admit epimorphisms onto the trefoil knot group is the coefficient of t^9, which is 3. These 2–bridge knots are 9_1, 9_6, 9_{23}, as shown in Example 3.2.

Similarly, as shown in [12], there are four distinct 2–bridge knots with 10 crossings whose knot groups admit epimorphisms onto the trefoil knot group, namely 10_5, 10_9, 10_{32}, 10_{40}; as shown in [10], there are seven distinct such 2–bridge knots with 11 crossings, namely $11a_{117}$, $11a_{175}$, $11a_{176}$, $11a_{203}$, $11a_{236}$, $11a_{306}$, $11a_{355}$.

Another example shows the generating function for $5_2 = K(\frac{3}{7}) = K([2,3])$:

$$f(\tfrac{3}{7}) = 4t^{15} + 8t^{16} + 12t^{17} + 16t^{18} + 20t^{19} + 24t^{20} + 28t^{21} + 32t^{22} + 36t^{23} + 40t^{24} + 60t^{25} + 112t^{26} + 212t^{27} + 376t^{28} + 620t^{29} + 960t^{30} + \cdots .$$

Acknowledgments

The author wishes to express his thanks to Makoto Sakuma and Takayuki Morifuji for helpful communications. He also thanks the referee for some useful comments. This work is partially supported by KAKENHI grant No. 16K05159 and 15H03618 from the Japan Society for the Promotion of Science.

References

[1] **I Agol**, *The classification of non-free 2–parabolic generator Kleinian groups*, talk at Budapest Bolyai conference (2002)

[2] **I Agol**, **Y Liu**, *Presentation length and Simon's conjecture*, J. Amer. Math. Soc. 25 (2012) 151–187 MR

[3] **S Aimi**, **D Lee**, **M Sakuma**, *Parabolic generating pairs of 2–bridge link groups* In preparation

[4] **M Boileau**, **S Boyer**, **A W Reid**, **S Wang**, *Simon's conjecture for two-bridge knots*, Comm. Anal. Geom. 18 (2010) 121–143 MR

[5] **G Burde**, **H Zieschang**, **M Heusener**, *Knots*, 3rd edition, De Gruyter Studies in Mathematics 5, De Gruyter, Berlin (2014) MR

[6] **J Cha**, **C Livingston**, *KnotInfo: table of knot invariants*, electronic reference Available at `http://www.indiana.edu/~knotinfo`

[7] **J C Cha**, **M Suzuki**, *Non-meridional epimorphisms of knot groups*, Algebr. Geom. Topol. 16 (2016) 1135–1155 MR

[8] **C Ernst**, **D W Sumners**, *The growth of the number of prime knots*, Math. Proc. Cambridge Philos. Soc. 102 (1987) 303–315 MR

[9] **S M Garrabrant**, **J Hoste**, **P D Shanahan**, *Upper bounds in the Ohtsuki–Riley–Sakuma partial order on 2–bridge knots*, J. Knot Theory Ramifications 21 (2012) art. id. 1250084, 24 pages MR

[10] **K Horie**, **T Kitano**, **M Matsumoto**, **M Suzuki**, *A partial order on the set of prime knots with up to 11 crossings*, J. Knot Theory Ramifications 20 (2011) 275–303 MR

[11] **L H Kauffman**, *State models and the Jones polynomial*, Topology 26 (1987) 395–407 MR

[12] **T Kitano**, **M Suzuki**, *A partial order in the knot table*, Experiment. Math. 14 (2005) 385–390 MR

[13] **T Kitano**, **M Suzuki**, *Twisted Alexander polynomials and a partial order on the set of prime knots*, from "Groups, homotopy and configuration spaces" (N Iwase, T Kohno, R Levi, D Tamaki, J Wu, editors), Geom. Topol. Monogr. 13, Geom. Topol. Publ., Coventry (2008) 307–321 MR

[14] **L M Milne-Thomson**, *The calculus of finite differences*, Macmillan, London (1951) MR

[15] **K Murasugi**, *Jones polynomials and classical conjectures in knot theory*, Topology 26 (1987) 187–194 MR

[16] **K Murasugi**, *Jones polynomials and classical conjectures in knot theory, II*, Math. Proc. Cambridge Philos. Soc. 102 (1987) 317–318 MR

[17] **K Murasugi**, *Knot theory and its applications*, Birkhäuser, Boston (1996) MR

[18] **T Ohtsuki**, **R Riley**, **M Sakuma**, *Epimorphisms between 2–bridge link groups*, from "The Zieschang Gedenkschrift" (M Boileau, M Scharlemann, R Weidmann, editors), Geom. Topol. Monogr. 14, Geom. Topol. Publ., Coventry (2008) 417–450 MR

[19] **D Rolfsen**, *Knots and links*, Math. Lect. Ser. 7, Publish or Perish, Berkeley, CA (1976) MR

[20] **D S Silver**, **W Whitten**, *Knot group epimorphisms, II*, preprint (2008) arXiv

[21] **M B Thistlethwaite**, *A spanning tree expansion of the Jones polynomial*, Topology 26 (1987) 297–309 MR

Department of Frontier Media Science
School of Interdisciplinary Mathematical Sciences, Meiji University
4-21-1 Nakano, Tokyo 164-8525, Japan

macky@fms.meiji.ac.jp

Received: 15 June 2016 Revised: 1 October 2016

Algebraic & Geometric Topology 17 (2017) 2429–2480

Homotopy decompositions of gauge groups over real surfaces

MICHAEL WEST

We analyse the homotopy types of gauge groups of principal $U(n)$–bundles associated to pseudoreal vector bundles in the sense of Atiyah. We provide satisfactory homotopy decompositions of these gauge groups into factors in which the homotopy groups are well known. Therefore, we substantially build upon the low-dimensional homotopy groups as provided by Biswas, Huisman and Hurtubise.

55P15, 55Q52; 30F50

1 Introduction

Recently, the topology of gauge groups over real surfaces has received widespread interest due to their intimate ties with the moduli spaces of stable vector bundles; see Biswas, Huisman and Hurtubise [3] and Schaffhauser [8]. Indeed, there have been explicit calculations of some of the topological invariants of these gauge groups. For instance, real vector bundles over real surfaces were originally classified in [3] but more recently in Georgieva and Zinger [4]. Cohomology calculations of the classifying spaces appeared in Liu and Schaffhauser [6] and Baird [1; 2]. Furthermore, some of the low-dimensional homotopy groups were presented in [3]. The purpose of this paper is to extend the calculations of these homotopy groups by providing homotopy decompositions of the gauge groups into products of known factors.

In the coming section, we define our objects of interest and state their classification results. We go on to state the results of this paper, and then proofs are provided in Section 2. In Section 3, we present tables of homotopy groups and compare them to those provided in [3] in which we highlight a discrepancy.

Acknowledgements I would like to thank Professor Tom Baird and the referee of this paper for suggesting the proofs of Propositions 1.9 and 1.10. Additionally, I further thank Professor Baird for a suggested reformulation of Theorem 1.15 and for providing interesting conversations surrounding the topics in this paper. I would also like to thank Professor Stephen Theriault for suggesting this project, and for his continued support and guidance.

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2429

1.1 Definitions

The pair (X, σ), where X is a compact connected Riemann surface and σ is an antiholomorphic involution, will be called a *real surface*.

To a real surface (X, σ), we associate the triple $(g(X), r(X), a(X))$, where

- $g(X)$ is the genus of X;
- $r(X)$ is number of path components of the fixed set X^σ;
- $a(X) = 0$ if X/σ is orientable and $a(X) = 1$ otherwise.

We note that the path components of X^σ are each homeomorphic to S^1. The following classification of real surfaces was studied in Weichold [13].

Theorem 1.1 (Weichold) *Let (X, σ) and (X', σ') be real surfaces. Then there is a isomorphism $X \to X'$ (in the category of real surfaces) if and only if*

$$(g(X), r(X), a(X)) = (g(X'), r(X'), a(X')).$$

Furthermore, if a triple (g, r, a) satisfies one of the following conditions:

(1) *if $a = 0$, then $1 \leq r \leq g+1$ and $r \equiv (g+1) \bmod 2$;*

(2) *if $a = 1$, then $0 \leq r \leq g$;*

then there is a real surface (X, σ) such that $(g, r, a) = (g(X), r(X), a(X))$. □

Therefore, a real surface (X, σ) is completely determined by its triple (g, r, a), which we call the *type* of the real surface.

Let $\pi\colon P \to X$ be a principal $U(n)$–bundle over the underlying Riemann surface X of the real surface (X, σ). A *lift* of σ is a map $\widetilde{\sigma}\colon P \to P$ satisfying

(1) $\sigma\pi = \pi\widetilde{\sigma}$;

(2) $\widetilde{\sigma}(p \cdot g) = \widetilde{\sigma}(p) \cdot \overline{g}$ for all $p \in P, g \in U(n)$;

where $\overline{g}$ represents the entrywise complex conjugate of $g \in U(n)$. We remark that, due to property 2 of a lift, the fixed point set $P^{\widetilde{\sigma}}$ has the structure of a principal $O(n)$–bundle over the real points X^σ.

Let $\widetilde{\sigma}$ be a lift. Then we say that $(P, \widetilde{\sigma}) \to (X, \sigma)$ is a *real principal $U(n)$–bundle* (or *real bundle*) if $\widetilde{\sigma}$ further satisfies

(3) $\widetilde{\sigma}^2(p) = p$ for all $p \in P$;

or if n is even, we say that $(P,\widetilde{\sigma})\to(X,\sigma)$ is a *quaternionic principal $U(n)$–bundle* (or *quaternionic bundle*) if $\widetilde{\sigma}$ satisfies

(3′) $\widetilde{\sigma}^2(p)=p\cdot(-I_n)$ for all $p\in P$;

where I_n represents the $n\times n$ identity matrix. Such bundles were classified in [3].

Proposition 1.2 (Biswas, Huisman, Hurtubise) *Let (X,σ) be a type-(g,r,a) real surface, and denote its fixed components by X_i for $1\le i\le r$. Then real principal $U(n)$–bundles $(P,\widetilde{\sigma})\to(X,\sigma)$ are classified by the first Stiefel–Whitney classes of the restriction to bundles $P_i\to X_i$ over the fixed components*

$$\omega_1(P_i)\in H^1(X_i,\mathbb{Z}/2)\cong\mathbb{Z}/2,$$

and by the first Chern classes of the bundle over X,

$$c_1(P)\in H^2(X,\mathbb{Z})\cong\mathbb{Z},$$

subject to the relation

$$c_1(P)\equiv\sum w_1(P_i)\mod(2).$$

Furthermore, given any such characteristic classes there is a real principal $U(n)$–bundle that realises them. □

We write

$$(c,w_1,w_2,\dots,w_r):=(c_1(P),w_1(P_1),w_1(P_2),\dots,w_1(P_r)),$$

and we will refer to the tuple $(c,w_1,w_2,\dots,w_r)\in\mathbb{Z}\times\prod_r\mathbb{Z}_2$ as the *class* of the real principal $U(n)$–bundle $(P,\widetilde{\sigma})$.

Proposition 1.3 (Biswas, Huisman, Hurtubise) *Let (X,σ) be a real surface of type (g,r,a), and let n be even. Then quaternionic principal $U(n)$–bundles $(P,\widetilde{\sigma})\to(X,\sigma)$ are classified by their first Chern class which must be even. Furthermore, given any such Chern class, there is a quaternionic principal $U(n)$–bundle that realises it.* □

We recall that we only cater for quaternionic bundles of even rank. However, a similar result for the case when n is odd is also handled in [3].

Writing $c=c_1(P)$, we will therefore refer to $c\in2\mathbb{Z}$ as the *class* of the quaternionic principal $U(n)$–bundle $(P,\widetilde{\sigma})$.

Let $(P,\widetilde{\sigma}) \to (X,\sigma)$ be a real or quaternionic principal $U(n)$–bundle. An *automorphism* of $(P,\widetilde{\sigma})$ is a $U(n)$–equivariant map $\phi\colon P \to P$ such that the following diagrams commute:

$$\begin{array}{ccc} P & \xrightarrow{\phi} & P \\ \downarrow & & \downarrow \\ X & \xrightarrow{\mathrm{id}_X} & X \end{array} \qquad \text{and} \qquad \begin{array}{ccc} P & \xrightarrow{\phi} & P \\ \downarrow{\scriptstyle\widetilde{\sigma}} & & \downarrow{\scriptstyle\widetilde{\sigma}} \\ P & \xrightarrow{\phi} & P \end{array}$$

Let Map(P, P) denote the set of self maps of P endowed with the compact open topology.

Definition 1.4 The (*unpointed*) *gauge group* $\mathscr{G}(P,\widetilde{\sigma})$ is the subspace of Map(P, P) whose elements are automorphisms of $(P,\widetilde{\sigma})$.

It will be convenient to provide decompositions for certain subspaces of the gauge group.

Definition 1.5 Choose a basepoint $*_X$ of (X,σ) such that $\sigma(*_X) = *_X$ if $r > 0$. Then the (*single*)*-pointed gauge group* $\mathscr{G}^*(P,\widetilde{\sigma})$ consists of the elements of $\mathscr{G}(P,\widetilde{\sigma})$ that restrict to the identity above $*_X$.

Another pointed gauge group of interest was considered in [3]. Let (X,σ) be a real surface of type (g,r,a); then for each $1 \le i \le r$, choose a designated point $*_i$ contained in the fixed component X_i. Further, if $a = 1$, choose another designated point $*_{r+1}$ that is not fixed by the involution. By convention, we choose $*_1$ to be $*_X$ as in Definition 1.5.

Definition 1.6 The $(r+a)$*–pointed gauge group* $\mathscr{G}^{*(r+a)}(P,\widetilde{\sigma})$ consists of the elements of $\mathscr{G}(P,\widetilde{\sigma})$ that restrict to the identity above these $(r+a)$ designated points of (X,σ).

1.2 Main results for real bundles

In this section, we aim to present the main results pertaining to homotopy decompositions of gauge groups of real principal $U(n)$–bundles. To ease notation, we will sometimes use the following:

- $\mathscr{G}((g,r,a);(c,w_1,w_2,\dots,w_r))$ to represent the unpointed gauge group of a real bundle of class $(c,w_1,w_2,\dots,w_r)$ over a real surface of type (g,r,a);
- $\mathscr{G}^*((g,r,a);(c,w_1,w_2,\dots,w_r))$ to represent the single-pointed gauge group of the real bundle as above;
- $\mathscr{G}^{*(r+a)}((g,r,a);(c,w_1,w_2,\dots,w_r))$ to represent the $(r+a)$–pointed gauge group of the real bundle as above.

We first present the results relating to when gauge groups of different real bundles have the same homotopy type. For $(r+a)$–pointed gauge groups this is always the case.

Proposition 1.7 *Let $(P,\widetilde{\sigma})$ and (P',σ') be real principal $U(n)$–bundles over a real surface (X,σ) of arbitrary type (g,r,a), then there is a homotopy equivalence*

$$B\mathscr{G}^{*(r+a)}(P,\widetilde{\sigma}) \simeq B\mathscr{G}^{*(r+a)}(P',\sigma').$$

However, this is not necessarily the case for the single-pointed and unpointed gauge groups, although we do have the following results.

Proposition 1.8 *For any c, c', w_1, w_1', there is a homotopy equivalence*

$$B\mathscr{G}^*((g,r,a);(c,w_1,w_2,\dots,w_r)) \simeq B\mathscr{G}^*((g,r,a);(c',w_1',w_2,\dots,w_r)).$$

Proposition 1.9 *Let the following be classifying spaces of rank-n gauge groups. Then there are isomorphisms of gauge groups*

$$\mathscr{G}((g,r,a);(c,w_1,w_2,\dots,w_r)) \cong \mathscr{G}((g,r,a);(c+2n,w_1,w_2,\dots,w_r)).$$

Proposition 1.10 *Let n be odd. Then there are isomorphisms of rank-n gauge groups*

(1) $\mathscr{G}((g,r,a);(c,w_1,w_2,\dots,w_r)) \cong \mathscr{G}((g,r,a);(c,\sum_{i=1}^r w_i,0,\dots,0))$;

(2) $\mathscr{G}^*((g,r,a);(c,w_1,w_2,\dots,w_r)) \cong \mathscr{G}^*((g,r,a);(c,\sum_{i=1}^r w_i,0,\dots,0))$.

It would be better to provide stronger statements of Propositions 1.7 and 1.8, such as in the form of the isomorphisms of Propositions 1.9 and 1.10. Indeed, the proofs of the latter invoke a conceptually simple argument and it may be the case that Propositions 1.7 and 1.8 can be given stronger statements using a similar approach.

We now state homotopy decompositions for $(r+a)$–pointed gauge groups.

Theorem 1.11 *Let $(P,\widetilde{\sigma})$ be of arbitrary class. Then there are integral homotopy decompositions:*

type	decompositions for $\mathscr{G}^{*(r+a)}(P,\widetilde{\sigma})$
$(g,0,1)$ for g even	$\mathscr{G}^*((0,0,1);0) \times \prod_{g} \Omega U(n)$
$(g,0,1)$ for g odd	$\mathscr{G}^*((1,0,1);0) \times \prod_{g-1} \Omega U(n)$
$(g,r,0)$	$\Omega^2(U(n)/O(n)) \times \prod_{(g-r+1)+(r-1)} \Omega U(n) \times \prod_{r-1} \Omega O(n)$
$(g,r,1)$ $g-r$ even	$\mathscr{G}^*((1,1,1);(0,0)) \times \prod_{(g-r)+(r-1)+1} \Omega U(n) \times \prod_{r-1} \Omega O(n)$
$(g,r,1)$ $g-r$ odd	$\mathscr{G}^*((1,1,1);(0,0)) \times \prod_{(g-r-1)+(r-1)+2} \Omega U(n) \times \prod_{r-1} \Omega O(n)$

In the single-pointed case, we have to be a little more careful with regards to the class of the underlying real bundle. For the cases where $\mathcal{G}^{*(r+a)}(P,\widetilde{\sigma}) \neq \mathcal{G}^*(P,\widetilde{\sigma})$, that is when $r+a>1$, we have the following results.

Theorem 1.12 *Let n be odd or let $(P,\widetilde{\sigma})$ be of class $(c,w_1,0,\dots,0)$. Let $r+a>1$. Then there are integral homotopy decompositions:*

type	decompositions for $\mathcal{G}^*(P,\widetilde{\sigma})$
$(g,r,0)$	$\Omega^2(U(n)/O(n)) \times \prod_{g-r+1} \Omega U(n) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n))$
$(g,r,1)$ $g-r$ even	$\mathcal{G}^*((1,1,1);(0,0)) \times \prod_{g-r} \Omega U(n) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n))$
$(g,r,1)$ $g-r$ odd	$\mathcal{G}^*((1,1,1);(0,0)) \times \prod_{(g-r-1)+1} \Omega U(n) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n))$

The remaining cases seem to integrally indecomposable; however, we will obtain the following localised homotopy decompositions for odd-rank gauge groups.

Theorem 1.13 *Let $p \neq 2$ be prime and let n be odd. Then there are the following p–local homotopy equivalences*

(1) $\mathcal{G}^*((0,0,1);c) \simeq_p \Omega^2(U(n)/O(n)) \times \Omega(U(n)/O(n))$;

(2) $\mathcal{G}^*((1,0,1);c) \simeq_p \Omega^2(U(n)/O(n)) \times \Omega(U(n)/O(n)) \times \Omega U(n)$;

(3) $\mathcal{G}^*((1,1,1);(c,w_1)) \simeq_p \Omega^2(U(n)/O(n)) \times \Omega(U(n)/O(n)) \times \Omega O(n)$.

This result relies upon a self map of $U(n)/O(n)$ as studied in Harris [5], which is a p–local homotopy equivalence if and only if n is odd. Hence it seems to be difficult to provide such satisfactory decompositions in the even-rank case.

We move on to some integral homotopy decompositions for unpointed gauge groups. The reader is invited to compare the tables of Theorems 1.14 and 1.12.

Theorem 1.14 *Let $(P,\widetilde{\sigma})$ be of class $(c,w_1,w_2,\dots,w_r)$. Then there are integral homotopy decompositions:*

	type	decompositions for $\mathcal{G}(P,\widetilde{\sigma})$
(1)	$(g,r,0)$	$\mathcal{G}((r-1,r,0);(c,w_1,\dots,w_r)) \times \prod_{g-r+1} \Omega U(n)$
	$(g,r,1)$ $g-r$ even	$\mathcal{G}((r,r,1);(c,w_1,\dots,w_r)) \times \prod_{g-r} \Omega U(n)$
	$(g,r,1)$ $g-r$ odd	$\mathcal{G}((r+1,r,1);(c,w_1,\dots,w_r)) \times \prod_{g-r-1} \Omega U(n)$

(2) $$\mathscr{G}((2,1,1);(c,w_1)) \simeq \mathscr{G}((1,1,1);(c,w_1)) \times \Omega U(n).$$

Further, for $r \geq 1$ and when $(P,\widetilde{\sigma})$ is of class $(c,w_1,0,\dots,0)$ or n is odd, there are integral homotopy decompositions:

(3)

type	decompositions for $\mathscr{G}(P,\widetilde{\sigma})$
$(r-1,r,0)$	$\mathscr{G}((0,1,0);(c,\Sigma w_i)) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n))$
$(r,r,1)$	$\mathscr{G}((1,1,1);(c,\Sigma w_i)) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n))$
$(r+1,r,1)$	$\mathscr{G}((2,1,1);(c,\Sigma w_i)) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n))$

The remaining unfamiliar spaces in Theorem 1.14 seem to be integrally indecomposable; however, localising at particular primes permits further decompositions.

Theorem 1.15 *Let n be a positive integer and let p be a prime with $p \nmid n$.*

(1) *Let the following be gauge groups of rank n. Then there are p–local homotopy equivalences*
 (a) $\mathscr{G}((g,1,a);(c,0)) \simeq_p O(n) \times \mathscr{G}^*((g,1,a);(c,0))$;
 further, if $p \neq 2$ and n is odd, then there are p–local homotopy equivalences
 (b) $\mathscr{G}((0,0,1);c) \simeq_p \mathrm{SO}(n) \times \Omega^2(U(n)/\mathrm{SO}(n))$;
 (c) $\mathscr{G}((1,0,1);c) \simeq_p \mathrm{SO}(n) \times \Omega^2(U(n)/\mathrm{SO}(n)) \times \Omega U(n)$.

(2) *Let the following be gauge groups of rank p. Then there are p–local homotopy equivalences*
 (a) $\mathscr{G}((g,1,a);(c,0)) \simeq_p O(p) \times \mathscr{G}^*((g,1,a);(c,0))$;
 further, if $p \neq 2$, then there are p–local homotopy equivalences
 (b) $\mathscr{G}((0,0,1);c) \simeq_p \mathrm{SO}(p) \times \Omega^2(U(p)/\mathrm{SO}(p))$;
 (c) $\mathscr{G}((1,0,1);c) \simeq_p \mathrm{SO}(p) \times \Omega^2(U(p)/\mathrm{SO}(p)) \times \Omega U(p)$.

1.3 Main results for quaternionic bundles

To distinguish the notation of quaternionic gauge groups from the real case, we will use a subscript Q, for example $\mathscr{G}_{\mathrm{Q}}(P,\widetilde{\sigma})$. Further, to ease notation we will sometimes use the following:

- $\mathscr{G}_{\mathrm{Q}}((g,r,a);c)$ to represent the unpointed gauge group of a quaternionic bundle of class c over a real surface of type (g,r,a);
- $\mathscr{G}^*_{\mathrm{Q}}((g,r,a);c)$ to represent the single-pointed gauge group of the quaternionic bundle as above;

- $\mathscr{G}_{\mathrm{Q}}^{*(r+a)}((g,r,a);c)$ to represent the $(r+a)$–pointed gauge group of the quaternionic bundle as above.

We present results in the same order as we did in the real case. In the quaternionic case, the homotopy types of the pointed and $(r+a)$–pointed gauge groups are independent of the class of the bundle.

Proposition 1.16 *Let (X,σ) be a real surface of fixed type (g,r,a). Let $(P,\tilde{\sigma})$ and (P',σ') be quaternionic principal $U(2n)$–bundles over (X,σ). Then there are homotopy equivalences*

(1) $B\mathscr{G}_{\mathrm{Q}}^{*}(P,\tilde{\sigma}) \simeq B\mathscr{G}_{\mathrm{Q}}^{*}(P',\sigma')$;

(2) $B\mathscr{G}_{\mathrm{Q}}^{*(r+a)}(P,\tilde{\sigma}) \simeq B\mathscr{G}_{\mathrm{Q}}^{*(r+a)}(P',\sigma')$.

For the unpointed case, we have an analogue of Proposition 1.9.

Proposition 1.17 *Let (X,σ) be a real surface of fixed type (g,r,a) and let the following be gauge groups of quaternionic bundles of rank $2n$. Then for any even integer c, there is an isomorphism of topological groups*

$$\mathscr{G}_{\mathrm{Q}}((g,r,a);c) \cong \mathscr{G}_{\mathrm{Q}}((g,r,a);c+4n).$$

We now present homotopy decompositions for pointed gauge groups in the quaternionic case. The reader is invited to compare the following results to their real analogues.

Theorem 1.18 *Let $(P,\tilde{\sigma})$ be a quaternionic principal $U(2n)$–bundle of class c. Then there are integral homotopy decompositions:*

type	decompositions for $\mathscr{G}_{\mathrm{Q}}^{*(r+a)}(P,\tilde{\sigma})$
$(g,0,1)$ for g even	$\mathscr{G}_{\mathrm{Q}}^{*}((0,0,1);0) \times \prod_{g} \Omega U(2n)$
$(g,0,1)$ for g odd	$\mathscr{G}_{\mathrm{Q}}^{*}((1,0,1);0) \times \prod_{g-1} \Omega U(2n)$
$(g,r,0)$	$\Omega^2(U(2n)/\mathrm{Sp}(n)) \times \prod_{g} \Omega U(2n) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n)$
$(g,r,1)$ for $g-r$ even	$\mathscr{G}_{\mathrm{Q}}^{*}((1,1,1);0) \times \prod_{g} \Omega U(2n) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n)$
$(g,r,1)$ for $g-r$ odd	$\mathscr{G}_{\mathrm{Q}}^{*}((1,1,1);0) \times \prod_{g} \Omega U(2n) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n)$

For the cases where $\mathscr{G}_{\mathrm{Q}}^{*(r+a)}(P,\tilde{\sigma}) \neq \mathscr{G}_{\mathrm{Q}}^{*}(P,\tilde{\sigma})$, that is when $r+a>1$, we have:

Theorem 1.19 *For $(P,\widetilde{\sigma})$ of class c, there are integral homotopy decompositions:*

type	decompositions for $\mathscr{G}^*_{\mathrm{Q}}(P,\widetilde{\sigma})$
$(g,r,0)$	$\Omega^2(U(2n)/\mathrm{Sp}(n)) \times \prod_{g-r+1} \Omega U(2n) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n) \times \prod_{r-1} \Omega(U(2n)/\mathrm{Sp}(n))$
$(g,r,1)$ $g-r$ even	$\mathscr{G}^*_{\mathrm{Q}}((1,1,1);0) \times \prod_{g-r} \Omega U(2n) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n) \times \prod_{r-1} \Omega(U(2n)/\mathrm{Sp}(n))$
$(g,r,1)$ $g-r$ odd	$\mathscr{G}^*_{\mathrm{Q}}((1,1,1);0) \times \prod_{g-r} \Omega U(2n) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n) \times \prod_{r-1} \Omega(U(2n)/\mathrm{Sp}(n))$

Again, the remaining cases seem to be integrally indecomposable; however, we will obtain the following localised decompositions.

Theorem 1.20 *Let $p \neq 2$ be prime. Then there are p–local homotopy equivalences*

(1) $\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0) \simeq_p \Omega^2(U(2n)/\mathrm{Sp}(n)) \times \Omega(U(2n)/\mathrm{Sp}(n))$;

(2) $\mathscr{G}^*_{\mathrm{Q}}((1,0,1);0) \simeq_p \Omega^2(U(2n)/\mathrm{Sp}(n)) \times \Omega(U(2n)/\mathrm{Sp}(n)) \times \Omega U(2n)$;

(3) $\mathscr{G}^*_{\mathrm{Q}}((1,1,1);0) \simeq_p \Omega^2(U(2n)/\mathrm{Sp}(n)) \times \Omega(U(2n)/\mathrm{Sp}(n)) \times \Omega\,\mathrm{Sp}(n)$.

We now present homotopy decompositions for the unpointed case.

Theorem 1.21 *For $(P,\widetilde{\sigma})$ of class c, there are integral homotopy decompositions:*

type	decompositions for $\mathscr{G}_{\mathrm{Q}}(P,\widetilde{\sigma})$
$(g,0,1)$ g even	$\mathscr{G}_{\mathrm{Q}}((0,0,1);c) \times \prod_{g} \Omega U(n)$
$(g,0,1)$ g odd	$\mathscr{G}_{\mathrm{Q}}((1,0,1);c) \times \prod_{g-1} \Omega U(n)$
$(g,r,0)$	$\mathscr{G}_{\mathrm{Q}}((0,1,0);c) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n) \times \prod_{r-1} \Omega(U(2n)/\mathrm{Sp}(n)) \times \prod_{g-r+1} \Omega U(n)$
$(g,r,1)$	$\mathscr{G}_{\mathrm{Q}}((1,1,1);c) \times \prod_{r-1} \Omega\,\mathrm{Sp}(n) \times \prod_{r-1} \Omega(U(2n)/\mathrm{Sp}(n)) \times \prod_{g-r} \Omega U(n)$

The remaining unfamiliar spaces in Theorem 1.21 seem to be integrally fundamental; however, localising at a particular prime permits further decompositions.

Theorem 1.22 *Let n be a positive integer and let p be a prime such that $p \nmid 2n$. Let the following be gauge groups of a quaternionic bundle of rank $2n$. Then there are p–local homotopy equivalences*

(1) $\mathscr{G}_{\mathrm{Q}}((g,1,a);c) \simeq_p \mathrm{Sp}(n) \times B\mathscr{G}^*_{\mathrm{Q}}((g,1,a);c)$;

(2) $\mathscr{G}_{\mathrm{Q}}((0,0,1);c) \simeq_p \mathrm{Sp}(n) \times \Omega^2(U(2n)/\mathrm{Sp}(n))$;

(3) $\mathscr{G}_{\mathrm{Q}}((1,0,1);c) \simeq_p \mathrm{Sp}(n) \times \Omega^2(U(2n)/\mathrm{Sp}(n)) \times \Omega U(2n)$.

2 Proofs of statements

For the sake of clarity, we focus on the proofs of statements in the real case, and then we elaborate on some of the details in the quaternionic case in Section 2.5. We look to decompose the gauge groups by studying an equivariant mapping space as provided in [1].

Throughout our analysis, we think of real surfaces as $\mathbb{Z}_2$–spaces. For $\mathbb{Z}_2$–spaces Y and Z, let $\mathrm{Map}_{\mathbb{Z}_2}(Y,Z)$ denote the space of $\mathbb{Z}_2$–maps from Y to Z. We note that the fixed points of Y must be mapped to the fixed points of Z. If Y and Z are pointed, we denote a pointed version of this mapping space by $\mathrm{Map}^*_{\mathbb{Z}_2}(Y,Z)$. Further, recall the "basepoints" $*_i$ of (X,σ) from just before Definition 1.6. Let

$$A := \coprod_{i=1}^{r+1} *_i \amalg \sigma(*_{r+1}),$$

and let $\mathrm{Map}^{*(r+a)}_{\mathbb{Z}_2}(X,Z)$ denote the subspace of $\mathrm{Map}_{\mathbb{Z}_2}(X,Z)$ whose elements send A to $*_Z$.[1] Let $\bar{X}$ denote the cofibre of $A \hookrightarrow X$, and notice that there is a homeomorphism

$$\mathrm{Map}^{*(r+a)}_{\mathbb{Z}_2}(X,Z) \cong \mathrm{Map}^*_{\mathbb{Z}_2}(\bar{X},Z).$$

A universal real principal $U(n)$–bundle is given by

$$(EU(n),\tilde{\varsigma}) \to (BU(n),\varsigma),$$

where ς is induced by complex conjugation and hence $BU(n)^{\varsigma} = BO(n)$. Using this $\mathbb{Z}_2$–structure, [1] provides the following theorem.

Theorem 2.1 (Baird) *There are homotopy equivalences*

(1) $B\mathscr{G}(P,\tilde{\sigma}) \simeq \mathrm{Map}_{\mathbb{Z}_2}(X, BU(n); P)$;

(2) $B\mathscr{G}^*(P,\tilde{\sigma}) \simeq \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); P)$;

(3) $B\mathscr{G}^{*(r+a)}(P,\tilde{\sigma}) \simeq \mathrm{Map}^{*(r+a)}_{\mathbb{Z}_2}(X, BU(n); P) \cong \mathrm{Map}^*_{\mathbb{Z}_2}(\bar{X}, BU(n); P)$;

where on the right-hand side, we pick the path component of $\mathrm{Map}_{\mathbb{Z}_2}(X, BU(n))$ *that induces* $(P,\tilde{\sigma})$. □

The following lemma can be shown by adapting the proof in the nonequivariant case. We will frequently require this lemma throughout the paper.

[1] Of course, it may be necessary to assume that $*_Z$ is fixed by the $\mathbb{Z}_2$–action.

Lemma 2.2 *Let Y and Z be $\mathbb{Z}_2$–spaces with basepoints fixed by the action, and with Y locally compact Hausdorff. Then there are equivalences*

(1) $\Omega\,\mathrm{Map}^*_{\mathbb{Z}_2}(X,Y) \cong \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma X, Y)$;

(2) $\mathrm{Map}^*_{\mathbb{Z}_2}(X,\Omega Y) \cong \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma X, Y)$. □

Throughout this section, there are a number of $\mathbb{Z}_2$–spaces that will often appear; here we provide a dictionary:

- (X,id): any space X with the trivial involution;
- $(X\vee X,\mathrm{sw})$: the wedge $X\vee X$ equipped with the involution that swaps the factors;
- $(S^n,-\mathrm{id})$: the sphere S^n equipped with the antipodal involution;
- (S^n,he): the sphere S^n equipped with the involution that reflects along the equator.

2.1 Real surfaces as $\mathbb{Z}_2$–complexes

In order to provide homotopy decompositions for the gauge groups, it will prove useful to provide a $\mathbb{Z}_2$ CW-complex structure for real surfaces. The following is essentially a restatement of the structures provided in [3]. We let $\Sigma_{p,q}$ denote a Riemann surface of genus p with q open discs removed.

Type $(g,0,1)$ We first study the case where g is even. We can think of X as two copies of $\Sigma_{g/2,1}$ glued along their boundary components, each a copy of S^1. The involution restricted to S^1 is the antipodal map and extends to swap the two copies of $\Sigma_{g/2,1}$.

We give a CW-structure of X as follows: Let X^0 be two 0–cells, $*$ and $\sigma(*)$. There are $2g+2$ 1–cells

$$\alpha_1,\dots,\alpha_{g/2},\beta_1,\dots,\beta_{g/2},\gamma,$$
$$\sigma(\alpha_1),\dots,\sigma(\alpha_{g/2}),\sigma(\beta_1),\dots,\sigma(\beta_{g/2}),\sigma(\gamma).$$

The boundaries of α_i,β_i are glued to $*$, and the boundaries of $\sigma(\alpha_i),\sigma(\beta_i)$ are glued to $\sigma(*)$. One end of γ is glued to $*$ and the other to $\sigma(*)$, whilst the same is done for $\sigma(\gamma)$ with the opposite orientation. There are two 2–cells glued on, one with attaching map

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_{g/2}\beta_{g/2}\alpha_{g/2}^{-1}\beta_{g/2}^{-1}\gamma\sigma(\gamma),$$

and the other with the same attaching map but with α_i,β_i replaced with $\sigma(\alpha_i),\sigma(\beta_i)$ and $\gamma\sigma(\gamma)$ replaced with $\sigma(\gamma)\gamma$.

As the notation suggests, the involution swaps cells that differ by σ. In particular, this is a σ–equivariant CW-structure and hence descends to a CW-structure of X/σ.

Now assume that g is odd, and let $g' = (g-1)$. We see that X can be thought of as two copies of $\Sigma_{g'/2,2}$ glued along their boundaries; two copies of S^1 in X. The involution swaps these copies of S^1 but reverses orientations, and it extends to X to swap the two copies of $\Sigma_{g'/2,2}$.

There are two 0–cells, $*$ and $\sigma(*)$, and $2g$ 1–cells

$$\alpha_1,\dots,\alpha_{g'/2},\beta_1,\dots,\beta_{g'/2},\gamma,\delta,$$
$$\sigma(\alpha_1),\dots,\sigma(\alpha_{g'/2}),\sigma(\beta_1),\dots,\sigma(\beta_{g'/2}),\sigma(\gamma),\sigma(\delta),$$

where $\alpha_i,\beta_i,\sigma(\alpha_i),\sigma(\beta_i),\gamma,\sigma(\gamma)$ are glued as before, but the boundary of δ is glued to $*$ and $\sigma(\delta)$ to $\sigma(*)$. Now there are two 2–cells, one with boundary map

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_{g'/2}\beta_{g'/2}\alpha_{g'/2}^{-1}\beta_{g'/2}^{-1}\delta\gamma\sigma(\delta)\gamma^{-1}$$

and the other glued equivariantly. The cells δ and $\sigma(\delta)$ correspond to the copies of S^1 above, and here γ is a cell joining these copies of S^1.

Type $(g,r,0)$ Let the involution fix r circles and let $g' = \frac{1}{2}(g-r+1)$. Then X/σ is a $\Sigma_{g',r}$, and X can be thought of as two copies of $\Sigma_{g',r}$ glued along the r boundary components.

In this case, the basepoint is preserved under σ; however, X^0 is given r 0–cells, one for each fixed component. The 1–cells are then

$$\alpha_1,\dots,\alpha_{g'},\beta_1,\dots,\beta_{g'},\gamma_2,\dots,\gamma_r,\delta_1,\dots,\delta_r,$$
$$\sigma(\alpha_1),\dots,\sigma(\alpha_{g'}),\sigma(\beta_1),\dots,\sigma(\beta_{g'}),\sigma(\gamma_2),\dots,\sigma(\gamma_r),$$

where α_i,β_i are as before and γ_i joins the basepoint to the i^{th} fixed component which is represented by δ_i. One of the two 2–cells has attaching map

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_{g'}\beta_{g'}\alpha_{g'}^{-1}\beta_{g'}^{-1}\delta_1\gamma_2\delta_2\gamma_2^{-1}\cdots\gamma_r\delta_r\gamma_r^{-1},$$

and we again define the other one equivariantly.

Type $(g,r,1)$ for $r>0$ Let the involution fix r circles. We first consider the case where $g \equiv r \bmod 2$. Let $g' = \frac{1}{2}(g-r)$. Then X can be thought of as two copies of $\Sigma_{g',r+1}$ glued along the boundary components. The involution fixes the first r of these components whilst restricting to the antipodal map on the extra copy of S^1.

Now X^0 is given $r+2$ 0–cells $*_i$, one for each fixed component and two for the extra S^1. The 1–cells are then

$$\alpha_1,\dots,\alpha_{g'},\beta_1,\dots,\beta_{g'},\gamma_2,\dots,\gamma_{r+1},\delta_1,\dots,\delta_r,\delta,$$
$$\sigma(\alpha_1),\dots,\sigma(\alpha_{g'}),\sigma(\beta_1),\dots,\sigma(\beta_{g'}),\sigma(\gamma_2),\dots,\sigma(\gamma_{r+1}),\sigma(\delta),$$

where α_i,β_i are as before and γ_i joins the basepoint to the i^{th} boundary circle. Each fixed component is represented by δ_i, and δ joins $*_{r+1}$ to $*_{r+2}$; therefore, $\delta\sigma(\delta)$ represents the extra copy of S^1. One of the two 2–cells has attaching map

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_{g'}\beta_{g'}\alpha_{g'}^{-1}\beta_{g'}^{-1}\delta_1\gamma_2\delta_2\gamma_2^{-1}\cdots\gamma_r\delta_r\gamma_r^{-1}\gamma_{r+1}\delta\sigma(\delta)\gamma_{r+1}^{-1},$$

and we again define the other one equivariantly.

For the case where $g\equiv r+1 \bmod 2$, we let $g'=\frac{1}{2}(g-r-1)$. Now X can be thought of as two copies of $\Sigma_{g',r+2}$ glued along the boundary components. Again, the involution fixes r of these components, whilst swapping the final two copies of S^1, but reversing orientation.

Again X^0 is given $r+2$ 0–cells, one for each fixed component and one for each of the extra two copies of S^1. The 1–cells are then

$$\alpha_1,\dots,\alpha_{g'},\beta_1,\dots,\beta_{g'},\gamma_2,\dots,\gamma_{r+2},\delta_1,\dots,\delta_{r+1},$$
$$\sigma(\alpha_1),\dots,\sigma(\alpha_{g'}),\sigma(\beta_1),\dots,\sigma(\beta_{g'}),\sigma(\gamma_2),\dots,\sigma(\gamma_{r+2}),\sigma(\delta_{r+1}),$$

where α_i,β_i are as before and γ_i joins the basepoint to the i^{th} boundary circle. Each fixed component is represented by δ_i for $i\le r$, and δ_{r+1} and $\sigma(\delta_{r+1})$ represent the extra copies of S^1. One of the two 2–cells has attaching map

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_{g'}\beta_{g'}\alpha_{g'}^{-1}\beta_{g'}^{-1}\delta_1\gamma_2\delta_2\gamma_2^{-1}\cdots\gamma_{r+1}\delta_{r+1}\gamma_{r+1}^{-1}\gamma_{r+2}\sigma(\delta_{r+1})\gamma_{r+2}^{-1},$$

and we again define the other one equivariantly.

2.2 Equivalent components of mapping spaces

In this section, we aim to prove Propositions 1.7–1.10. The proofs are motivated from the analysis of nonequivariant mapping spaces found in [10].

Proof of Proposition 1.7 We study the actions of $\pi_2(BU(n))$ and $\pi_1(BO(n))$ on the components of $\mathrm{Map}^*_{\mathbb{Z}_2}(\overline{X},BU(n))$. In [10], an action of $\pi_2(BU(n))$ on the space $\mathrm{Map}(X,BU(n))$ was defined via

$$X\xrightarrow{\text{pinch}}X\vee S^2\xrightarrow{f\vee\alpha}BU(n)\vee BU(n)\xrightarrow{\text{fold}}BU(n) \tag{1}$$

with $\alpha\in\pi_2(BU(n))$ and $f\in\mathrm{Map}^*(X,BU(n))$.

We now consider the equivariant case for $r=0$. Let S^1 be the loop that is pinched under $\bar{X}\to\bar{X}\vee S^2$, similar to the first map in (1). Due to equivariance, we are also forced to pinch the loop $\sigma(S^1)$ producing an extra factor of S^2, and the action becomes

$$\bar{X}\xrightarrow{\text{pinch}}\bar{X}\vee S^2\vee\sigma(S^2)\xrightarrow{f\vee\alpha\vee\bar{\alpha}}BU(n)\vee BU(n)\vee BU(n)\xrightarrow{\text{fold}}BU(n),$$

where $\bar{\alpha}=\varsigma\alpha$. Since σ and ς are both orientation-reversing, the action of

$$\alpha\in\pi_2(BU(n))\cong\mathbb{Z}$$

alters the class $[f]$ by 2α. Hence for $2c\in[\bar{X},BU(n)]_{\mathbb{Z}_2}\cong 2\mathbb{Z}$, this action gives homotopy equivalences

$$\mathrm{Map}^*_{\mathbb{Z}_2}(X,BU(n);2c)\simeq\mathrm{Map}^*_{\mathbb{Z}_2}(X,BU(n);2c+2\alpha).$$

In particular, this gives the required homotopy equivalences for the case when $r=0$.

When $r>0$, the path components of $\mathrm{Map}^*_{\mathbb{Z}_2}(\bar{X},BU(n))$ are classified by the tuple

$$(c,w_1,w_2,\dots,w_r)\in\mathbb{Z}\times\prod_r\mathbb{Z}_2$$

subject to $c\equiv\sum_{i=1}^r w_i \mod 2$. We wish to construct an action of $\pi_1(BO(n))$ to alter each w_i. For $\beta\in\pi_1(BO(n))$, we note that the inclusion of the image of β into $BU(n)$ is nullhomotopic, so there is an extension $\beta'\colon D^2\to BU(n)$ of β. Now, consider (S^2,he) and denote the fixed equator by E, the upper hemisphere by U and the lower hemisphere by L. We can extend β to a map $\tilde{\beta}\colon(S^2,\mathrm{he})\to BU(n)$, where

$$\tilde{\beta}\,|_U=\beta'\quad\text{and}\quad\tilde{\beta}\,|_L=\varsigma\beta',$$

and therefore, $\tilde{\beta}|_E=\beta$. Due to the discussion preceding Proposition 4.1[2] in [3], the extension $\tilde{\beta}$ can be chosen so that the class $[\tilde{\beta}]\in\mathbb{Z}\times\mathbb{Z}_2$ is $(0,0)$ if β is trivial or $(\pm1,1)$ otherwise.

Let $(S^1,\mathrm{he})\hookrightarrow\bar{X}$ be an inclusion such that the fixed points of (S^1,he) are mapped to the i^{th} fixed component X_i of $\bar{X}$. As in (1), we apply the pinch map to this copy of (S^1,he) in $\bar{X}$ and hence produce a factor of (S^2,he). Now the action becomes

$$\bar{X}\xrightarrow{\text{pinch}}\bar{X}\vee(S^2,\mathrm{he})\xrightarrow{f\vee\tilde{\beta}}BU(n)\vee BU(n)\xrightarrow{\text{fold}}BU(n).$$

For $\tilde{\beta}$ of class $(\pm1,1)$, we conclude that this action gives a homotopy equivalence between the components $(c,w_1,w_2,\dots,w_r)$ and $(c\pm1,w_1,\dots,w_i+1,\dots,w_r)$. Combining the actions of $\pi_2(BU(n))$ and $\pi_1(BO(n))$ gives homotopy equivalences between all the components of $\mathrm{Map}^*_{\mathbb{Z}_2}(\bar{X},BU(n))$. □

[2]We note that Proposition 4.1 in [3] is stated as Proposition 1.2 in this paper.

Proof of Proposition 1.8 Recall from the preamble to Definition 1.6 that we chose $*_1$ as the basepoint of (X,σ). We define actions of $\pi_2(BU(n))$ and $\pi_1(BO(n))$ on $\mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n))$ in a similar fashion to the proof of Proposition 1.7, and this obtains the result. We cannot extend this result as in the $(r+a)$–pointed case due to the "unpointed" fixed circles. □

We cannot hope to use the actions of π_1 and π_2 on the unpointed mapping space due to the lack of basepoint. But, by tensoring the bundle $(P,\widetilde{\sigma})$ with a real $U(1)$–bundle, we can provide some equivalences between components.

Proof of Proposition 1.9 Let $\pi\colon (P,\widetilde{\sigma}) \to (X,\sigma)$ be a real principal $U(n)$–bundle of class $(c, w_1, w_2, \dots, w_r)$ over a real surface of type (g,r,a). The idea will be to tensor P with a real $U(1)$–bundle $\pi_Q\colon (Q,\tau) \to (X,\sigma)$ of class $(2,0,\dots,0)$.

Using the inclusion of the centre $U(1) \hookrightarrow U(n)$, there is a $U(1)$–action on $(P,\widetilde{\sigma})$. In the principal bundle setting, the tensor of $(P,\widetilde{\sigma})$ and (Q,τ) is the pullback

$$\begin{array}{ccc}
\big(\Delta^*(P\times_{U(1)} Q),\ \Delta^*(\widetilde{\sigma}\times\tau)\big) & \longrightarrow & (P\times_{U(1)} Q,\ \widetilde{\sigma}\times\tau) \\
\big\downarrow & & \big\downarrow{\scriptstyle\widetilde{\pi}} \\
(X,\sigma) & \xrightarrow{\quad\Delta\quad} & (X,\sigma)\times(X,\sigma)
\end{array}$$

where Δ is the diagonal map and $\widetilde{\pi} = \pi\times\pi_Q$. In a similar fashion to the discussion preceding Proposition 4.1 in [3], we calculate that $(c+2n, w_1, w_2, \dots, w_r)$ is the class of the pullback $(\Delta^*(P\times_{U(1)} Q), \Delta^*(\widetilde{\sigma}\times\tau))$.

We then define

$$\Theta\colon \mathscr{G}(P,\widetilde{\sigma}) \to \mathscr{G}\big(\Delta^*(P\times_{U(1)} Q),\ \Delta^*(\widetilde{\sigma}\times\tau)\big)$$

to be the map that sends $\phi\colon P\to P$ to $\Delta^*(\phi\times\mathrm{id})$. Then an inverse to Θ is defined in the same way as Θ, except that we replace the inclusion $U(1)\hookrightarrow U(n)$ with the conjugate inclusion defined via

$$a \mapsto \begin{pmatrix} \overline{a} & 0 & \cdots & 0 \\ 0 & \overline{a} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{a} \end{pmatrix}.$$ □

Proof of Proposition 1.10 Let $\pi\colon (P,\widetilde{\sigma}) \to (X,\sigma)$ be a real principal $U(n)$–bundle of class $(c, w_1, w_2, \dots, w_r)$ over a real surface of type (g,r,a). The statement is proven using the same method as Proposition 1.9, except that we tensor with a real $U(1)$–bundle $(\widetilde{Q},\widetilde{\tau})$ of class $\big(0, \sum_{i=2}^r w_i, w_2, \dots, w_r\big)$. If n is odd, the class of the pullback $(\Delta^*(P\times_{U(1)}\widetilde{Q}),\ \Delta^*(\widetilde{\sigma}\times\widetilde{\tau}))$ is then $\big(c, \sum_{i=1}^r w_i, 0, \dots, 0\big)$.

An isomorphism $\Theta\colon \mathscr{G}(P,\widetilde{\sigma}) \to \mathscr{G}(\Delta^*(P \times_{U(1)} \widetilde{Q}),\ \Delta^*(\widetilde{\sigma} \times \widetilde{\tau}))$ is then defined in the same way as for Proposition 1.9. □

2.3 Pointed gauge groups

In our analysis, it will be necessary to distinguish the following types of real surfaces:

(0) $r = 0$ ($\Longrightarrow$ $a = 1$);

(1) $r > 0$ and $a = 0$;

(2) $r > 0$ and $a = 1$.

Generally, we will analyse the gauge groups in order of ease; we first analyse the $(r+a)$–pointed gauge group, and then the single-pointed gauge group. Our results for the single-pointed gauge groups will then be used to analyse the unpointed case.

2.3.1 Integral decompositions For the underlying Riemann surface X of a real surface (X,σ), the attaching map $f\colon S^1 \to \bigvee_{2g} S^1$ of the top cell is a sum of Whitehead products, and hence the suspension Σf is nullhomotopic. In the real surface case, we see Whitehead products appearing in the attaching maps of Section 2.1. Therefore, we still see trivialities appearing in the suspension of these attaching maps, and these trivialities will provide a large class of homotopy decompositions.

We will use the notation as defined in Section 2.1, and furthermore, we require the following notation in this section. Let g' denote the number of 1–cells of X which are of the form α_i, β_i in X. Explicitly,

$$g' = \begin{cases} g - r + 1 & \text{when } a = 0, \\ g - r & \text{when } a = 1 \text{ and } g - r \text{ is even}, \\ g - r - 1 & \text{when } a = 1 \text{ and } g - r \text{ is odd}. \end{cases}$$

Proposition 2.3 *Let $X_{\alpha\beta} = \bigvee S^1$ be the 1–cells $\alpha_i, \sigma(\alpha_i), \beta_i, \sigma(\beta_i)$ in the decomposition of (X,σ). Then the map μ in the $\mathbb{Z}_2$–cofibration sequence*

$$X_{\alpha\beta} \hookrightarrow X \to X' \xrightarrow{\mu} \Sigma(X_{\alpha\beta})$$

is $\mathbb{Z}_2$–nullhomotopic.

Proof We recall that the attaching map of one of the 2–cells in a real surface of type $(g,r,0)$ is

$$\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_{g'}\beta_{g'}\alpha_{g'}^{-1}\beta_{g'}^{-1}\delta_1\gamma_2\delta_2\gamma_2^{-1}\cdots\gamma_r\delta_r\gamma_r^{-1}.$$

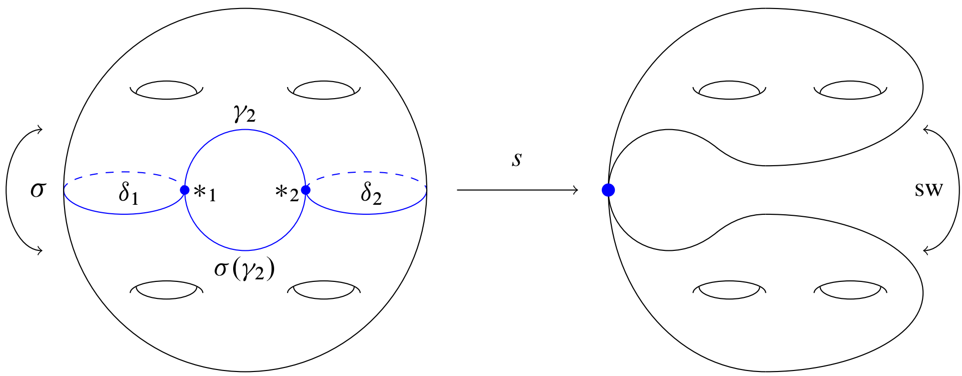


Figure 1: For a type-$(5,2,0)$ real surface, the map s collapses the 1–cells δ_1, δ_2, γ_2 and $\sigma(\gamma_2)$.

The attaching map involving the cells α_i and β_i is a sum of Whitehead products. The idea is to collapse the rest of the cells.

Now in the general case, let X be a type-(g,r,a) real surface, let $\Sigma_{g'/2}$ be a Riemann surface of genus $\frac{1}{2}g'$ and denote by

$$s\colon X \to (\Sigma_{g'/2} \vee \Sigma_{g'/2}, \mathrm{sw})$$

the map that collapses the 1–skeleton of X other than the cells $\alpha_i, \sigma(\alpha_i), \beta_i$ and $\sigma(\beta_i)$.

An example for the map s is illustrated in Figure 1. Note that four of the "holes" are undisturbed by s; these correspond to the 1–cells of the form $\alpha_i, \sigma(\alpha_i), \beta_i$ and $\sigma(\beta_i)$.

There is a commutative diagram

$$\begin{tikzcd}
X_{\alpha\beta} \arrow[r] \arrow[d, equal] & X \arrow[r] \arrow[d, "s"] & X' \arrow[r, "\mu"] \arrow[d, "s'"] & \Sigma(X_{\alpha\beta}) \arrow[d, equal] \\
X_{\alpha\beta} \arrow[r] & (\Sigma_{g'/2} \vee \Sigma_{g'/2}, \mathrm{sw}) \arrow[r] & (S^2 \vee S^2, \mathrm{sw}) \arrow[r, "\Sigma f \vee \overline{\Sigma f}"] & \Sigma(X_{\alpha\beta})
\end{tikzcd}$$

where the rows are $\mathbb{Z}_2$–cofibration sequences, s' is an induced map on cofibers and f is the attaching map of the Riemann surface $\Sigma_{g'/2}$. The $\mathbb{Z}_2$–triviality of μ therefore follows from the triviality of Σf. □

We deduce the following theorem which greatly contributes to Theorems 1.11 and 1.12.

Theorem 2.4 *With notation as above, there are homotopy equivalences*

(1) $\mathscr{G}^*(P,\tilde{\sigma}) \simeq \mathscr{G}^*((g-g',r,a);(c,w_1,\dots,w_r)) \times \prod_{g'} \Omega U(n)$;

(2) $\mathscr{G}^{(r+a)*}(P,\tilde{\sigma}) \simeq \mathscr{G}^{(r+a)*}((g-g',r,a);(c,w_1,\dots,w_r)) \times \prod_{g'} \Omega U(n)$.

Proof We use the notation of Proposition 2.3 and run through details for part (1). By Theorem 2.1, there is a homotopy fibration sequence

$$\mathscr{G}(P,\widetilde{\sigma}) \to \Omega\,\mathrm{Map}^*_{\mathbb{Z}_2}(X_{\alpha\beta}, BU(n)) \xrightarrow{\mu^*} \mathrm{Map}^*_{\mathbb{Z}_2}(X', BU(n); (c, w_1, \dots, w_r)),$$

and by Lemma 2.2, we can see that μ^* is induced from μ in Proposition 2.3. But Proposition 2.3 showed that μ is nullhomotopic, and the result follows. The proof for part (2) is similar. □

We note that for real surfaces of type $(g,0,1)$, Theorem 2.4 leaves only types $(0,0,1)$ and $(1,0,1)$ to consider. The gauge groups of these types seem to be integrally indecomposable and so we leave their analysis until later.

2.3.2 The case $r > 0$, $a = 0$ Although we restrict to the case $a = 0$, we will see that many of the methods in this section will also transfer to the case when $a = 1$.

Due to Theorem 2.4, we restrict to the case when (X,σ) is of type $(r-1, r, 0)$. For $(P,\widetilde{\sigma})$ of class $(0,0,\dots,0)$, we utilise Theorem 2.1 and Lemma 2.2, and obtain the equivalences

$$\mathscr{G}^{*r}(P,\widetilde{\sigma}) \simeq \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma(\bar{X}), BU(n));$$
$$\mathscr{G}^{*}(P,\widetilde{\sigma}) \simeq \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma(X), BU(n)).$$

The aim of this section is to prove Theorems 1.11 and 1.12 for types $(g,r,0)$, which is restated below.

Theorem 2.5 *Let $(P,\widetilde{\sigma})$ be a real bundle of class $(c, w_1, \dots, w_r)$ over a real surface (X,σ) of type $(r-1, r, 0)$. Then*

(1) *there is a homotopy equivalence*

$$\mathscr{G}^{*r}(P,\widetilde{\sigma}) \simeq \Omega^2(U(n)/O(n)) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega U(n);$$

(2) *if $w_i = 0$ for all $i > 1$ or if n is odd, then there is a homotopy equivalence*

$$\mathscr{G}^{*}(P,\widetilde{\sigma}) \simeq \Omega^2(U(n)/O(n)) \times \prod_{r-1} \Omega O(n) \times \prod_{r-1} \Omega(U(n)/O(n)).$$

Recall the $\mathbb{Z}_2$–structure of a type-$(g,r,0)$ surface in Section 2.1. In the following, X_γ will be the subcomplex of the 1–cells of X that are denoted by either γ_i or $\sigma(\gamma_i)$.

Proposition 2.6 *Let (X,σ) be as above. Then in the $\mathbb{Z}_2$–cofibration sequence*

$$X_\gamma \xrightarrow{\iota} X \to \widetilde{X} \xrightarrow{\mu'} \Sigma(X_\gamma),$$

there is a left $\mathbb{Z}_2$–homotopy inverse to ι. In particular, μ' is $\mathbb{Z}_2$–nullhomotopic.

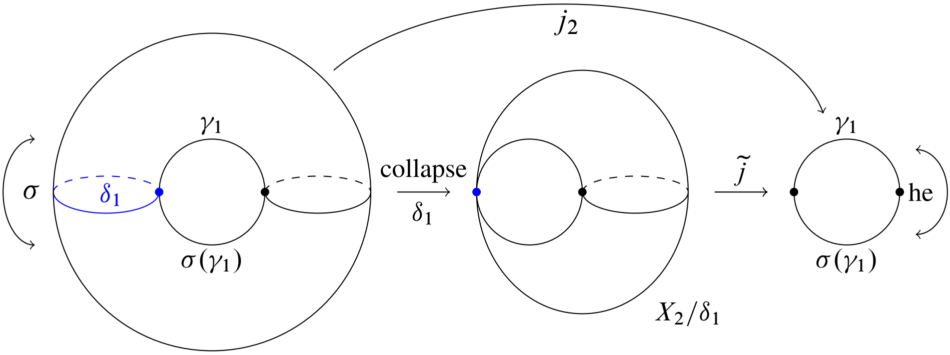


Figure 2: The map j_2 projects to the factor (S^1, he) and j_2 factors through X_2/δ_1.

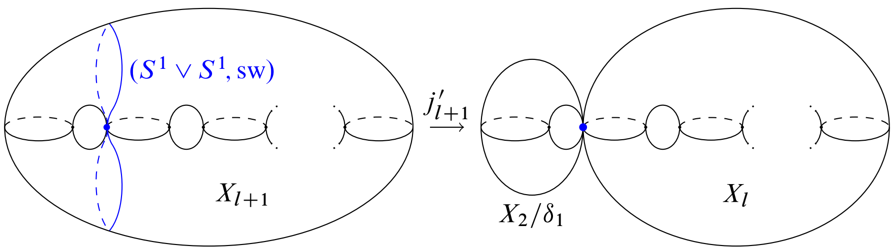


Figure 3: Collapse a copy of $(S^1 \vee S^1, \mathrm{sw})$ to obtain the wedge $X_2/\delta_1 \vee X_l$.

Proof We will use induction on r, the number of fixed circles of X. Let X_r denote a real surface of type $(r-1, r, 0)$, and let $(X_r)_\gamma$ be the subcomplex of X_r with 1–cells denoted by either γ_i or $\sigma(\gamma_i)$. We aim to define left homotopy inverses $j_r\colon X_r \to (X_r)_\gamma$ of ι for each r.

Note that the space $(X_r)_\gamma$ is the wedge $\bigvee_{r-1}(S^1, \mathrm{he})$, and hence the first nontrivial case is when $r = 2$. In this case, one can see that X_2 is the product

$$(S^1, \mathrm{id}) \times (S^1, \mathrm{he}).$$

We define j_2 to be the projection onto the second factor; Figure 2 illustrates this map.

For $r = l$, we assume that j_l exists. For $r = l + 1$, we first use a map j'_{l+1} that collapses a copy of $(S^1 \vee S^1, \mathrm{sw})$ in X_{l+1} such that the image is homeomorphic to $X_l \vee X_2/\delta_1$, where X_2/δ_1 is a copy of X_2 with the 1–cell δ_1 collapsed. The map j'_{l+1} is illustrated in Figure 3.

Figure 2 also shows that j_2 factors through the space X_2/δ_1. We therefore define j_{l+1} to be the composition

$$X_{l+1} \xrightarrow{j'_{l+1}} X_2/\delta_1 \vee X_l \xrightarrow{\tilde{j} \vee j_l} (X_{l+1})_\gamma,$$

where $\tilde{j}$ is defined in Figure 2. □

As an easy consequence of Proposition 2.6, we obtain the homotopy equivalences

$$\Sigma X \simeq \Sigma \tilde{X} \vee \Sigma X_\gamma \quad \text{and} \quad \Sigma \bar{X} \simeq \Sigma \tilde{X} \vee \Sigma \bar{X}_\gamma.$$

We shall see that the factors ΣX_γ and $\Sigma \bar{X}_\gamma$ give the factors $\prod_1^{r-1} \Omega U(n)$ and $\prod_1^{r-1} \Omega(U(n)/O(n))$, respectively, in Theorem 2.5, and that the factor $\Sigma \tilde{X}$ produces the factors $\Omega^2(U(n)/O(n)) \times \prod_1^{r-1} \Omega O(n)$. However, the map j_r automatically induces a map

$$\mathrm{Map}^*_{\mathbb{Z}_2}(X_\gamma, BU(n)) \to \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); (0, 0, \dots, 0)).$$

Hence we only obtain a splitting on the level of mapping spaces in this trivial case.

We now restrict to this trivial case for the rest of this section. For the other cases, Proposition 1.7 will then give results for Theorem 2.5(1) and Propositions 1.8 and 1.10 will give results for Theorem 2.5(2). We provide further decompositions at the level of the real surface to continue the proof of Theorem 2.5.

Proposition 2.7 *Let X_δ be the 1–cells in $\tilde{X}$ denoted by $\delta_2, \dots, \delta_r$. Then in the $\mathbb{Z}_2$–cofibration*

$$X_\delta \xrightarrow{\iota'} \tilde{X} \to (S^2, \mathrm{he}) \xrightarrow{\mu''} \Sigma(X_\delta),$$

the map μ'' is $\mathbb{Z}_2$–nullhomotopic.

Proof The space $\tilde{X}$ is the quotient of a type-$(r-1, r, 0)$ real surface with the 1–cells denoted by $\gamma_2, \dots, \gamma_r$ collapsed to a point. Recall that the attaching map of X is

$$\delta_1 \gamma_2 \delta_2 \gamma_2^{-1} \gamma_3 \delta_3 \gamma_3^{-1} \cdots \gamma_r \delta_r \gamma_r^{-1}, \tag{2}$$

and the induced attached map in $\tilde{X}$ becomes

$$\delta_1 \delta_2 \cdots \delta_r.$$

We conclude that $\tilde{X}$ is a sphere (S^2, he) with r of its fixed points identified.

Let U denote the upper "hemisphere" of $\tilde{X}$; it is homeomorphic to a disc with r of its boundary points identified, and notice that $\tilde{X} = U \cup \sigma(U)$. Now there is a deformation retract $H\colon U\times I\to U$ of U onto the wedge $\bigvee_{i=2}^{r}\delta_i$. Therefore, we define a left inverse to the map ι' via

$$x\mapsto\begin{cases}H(x,1) & \text{for } x\in U,\\ H(\sigma_{\widetilde{X}}(x),1) & \text{for } x\in\sigma_{\widetilde{X}}(U),\end{cases}$$

and the result follows. □

We deduce that

$$\Sigma\widetilde{X}\simeq\Sigma X_\delta\vee\Sigma(S^2,\mathrm{he}).$$

The factor $\Sigma X_\delta=\bigvee_1^{r-1}(S^1,\mathrm{id})$ provides the factor $\prod_1^{r-1}\Omega O(n)$ for both cases in Theorem 2.5. We now show that the spaces $\Sigma(S^2,\mathrm{he})$ and ΣX_γ provide the other factors.

Lemma 2.8 *There are homotopy equivalences*

(1) $\mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma X_\gamma,BU(n))\simeq\prod_{r-1}\Omega(U(n)/O(n))$;

(2) $\mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma\bar{X}_\gamma,BU(n))\simeq\prod_{r-1}\Omega U(n)$.

Proof The space $\Sigma(X_\gamma)$ is the same as the wedge $\bigvee_{r-1}\Sigma(S^1,\mathrm{he})$. Looking at the r–pointed case, the 0–skeleton of $\Sigma(X_\gamma)$ is collapsed, and the space $\Sigma\overline{(X_\gamma)}$ becomes the wedge $\Sigma\bigvee_{r-1}(S^1\vee S^1,\mathrm{sw})$. This shows part (2) of the lemma.

For part (1), we introduce a pullback similar to the pullbacks used in [1]. The space $\mathrm{Map}^*_{\mathbb{Z}_2}((S^1,\mathrm{he}),BU(n))$ fits into the following pullback diagram:

$$\begin{tikzcd}
\mathrm{Map}^*_{\mathbb{Z}_2}((S^1,\mathrm{he}),BU(n)) \arrow[r,"\tilde{u}"] \arrow[d,"\tilde{r}"] & \mathrm{Map}^*(D^1,BU(n)) \arrow[d,"r"] \\
\mathrm{Map}^*_{\mathbb{Z}_2}((S^0,\mathrm{id}),BU(n)) \arrow[r,"u"] & \mathrm{Map}^*(S^0,BU(n))
\end{tikzcd}$$

Here $\tilde{r}$ restricts to the fixed points of (S^1,he), and $\tilde{u}$ restricts to the upper hemisphere of (S^1,he) and then forgets about equivariance. Since

$$\mathrm{Map}^*_{\mathbb{Z}_2}((S^0,\mathrm{id}),BU(n))\simeq BO(n),$$

the map u is just the inclusion $BO(n)\hookrightarrow BU(n)$, and hence the homotopy fibre of u is $U(n)/O(n)$. Since r is a fibration, the square is also a homotopy pullback. We note that the space $\mathrm{Map}^*(D^1,BU(n))$ is contractible, and so the result follows. □

Lemma 2.9 *There is a homotopy equivalence*

$$\mathrm{Map}^*_{\mathbb{Z}_2}((S^2,\mathrm{he}), BU(n); (0,0)) \simeq \Omega(U(n)/O(n))_0,$$

where $\Omega(U(n)/O(n))_0$ *denotes the connected component of* $\Omega(U(n)/O(n))$ *containing the basepoint.*

Proof There is a similar pullback as in Lemma 2.8:

$$\begin{array}{ccc} \mathrm{Map}^*_{\mathbb{Z}_2}((S^2,\mathrm{he}), BU(n)) & \xrightarrow{\tilde{u}} & \mathrm{Map}^*(D^2, BU(n)) \\ \downarrow{\scriptstyle \tilde{r}} & & \downarrow{\scriptstyle r} \\ \mathrm{Map}^*_{\mathbb{Z}_2}((S^1,\mathrm{id}), BU(n)) & \xrightarrow{u} & \mathrm{Map}^*(S^1, BU(n)) \end{array}$$

This time the map u is homotopic to the inclusion $O(n) \hookrightarrow U(n)$, and so the homotopy fibre of u is $\Omega(U(n)/O(n))$. The space $\mathrm{Map}^*(D^2, BU(n))$ is contractible, and so there is an equivalence

$$\mathrm{Map}^*_{\mathbb{Z}_2}((S^2,\mathrm{he}), BU(n)) \simeq \Omega(U(n)/O(n)),$$

and the result follows. □

Proof of Theorem 2.5 For (1), it is enough to deal with the trivial component of $\mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n))$ by Proposition 1.7. Using a similar method to the proof of Theorem 2.4, we have that Proposition 2.6 and Lemma 2.8 contribute the factor $\prod_{r-1}\Omega U(n)$, Proposition 2.7 contributes the factor $\prod_{r-1}\Omega O(n)$ and Lemma 2.9 contributes the factor $\Omega^2(U(n)/O(n))$.

For (2), the proof is similar, but one has to be careful with the nontrivial components. □

2.3.3 The case $r>0$, $a=1$ We use the techniques and notation of the previous section. In particular, let $(P,\tilde{\sigma})$ be a bundle of class $(0,0,\dots,0)$ over a real surface (X,σ) of type $(g,r,1)$. We first note that by Proposition 2.3, we can restrict to the cases

$$g = r \quad \text{or} \quad g = r+1. \tag{3}$$

With these cases in mind, the main aim will be to prove the following theorem which is a restatement of Theorems 1.11 and 1.12 for real surfaces of type $(g,r,1)$.

Theorem 2.10 *For notation as above and* g *as in* (3)*, there are homotopy equivalences*

(1) $\mathscr{G}^*(P,\tilde{\sigma}) \simeq \mathscr{G}^*((g-r+1,1,1);(0,0)) \times \prod_{r-1}\Omega O(n) \times \prod_{r-1}\Omega(U(n)/O(n))$;

(2) $\mathscr{G}^{*r+1}(P,\tilde{\sigma}) \simeq \mathscr{G}^{*2}((g-r+1,1,1);(0,0)) \times \prod_{r-1}\Omega O(n) \times \prod_{r-1}\Omega U(n)$.

We note that after we have proven the above theorem, the only cases we have left to analyse will be gauge groups over real surfaces of type $(2,1,1)$ and type $(1,1,1)$.

For the proof of the theorem, we will essentially follow the methods of the previous section. Let X_γ denote the subcomplex of X consisting of the 1–cells denoted by either γ_i or $\sigma(\gamma_i)$ for $2 \le i \le r$.

Proposition 2.11 *Let (X,σ) be as above. Then in the $\mathbb{Z}_2$–cofibration sequence*

$$X_\gamma \xrightarrow{\kappa} X \to \widetilde{X} \xrightarrow{\nu} \Sigma(X_\gamma),$$

the map ν is $\mathbb{Z}_2$–nullhomotopic.

Proof We define a left inverse to κ. First, in X, collapse the cells

$$\gamma_{r+1}, \sigma(\gamma_{r+1}), \delta_{r+1}, \sigma(\delta_{r+1})$$

and the cells $\gamma_{r+2}, \sigma(\gamma_{r+2})$ if they exist. We are left with a space $\mathbb{Z}_2$–homeomorphic to a real surface of type $(r-1,r,0)$; we now use the map j_r as defined in the proof of Proposition 2.6. □

The proof of the next proposition is identical to that of Proposition 2.7 except we exchange (S^2, he) for a real surface X' of type either $(2,1,1)$ or $(1,1,1)$.

Proposition 2.12 *Let X_δ be the 1–cells in $\widetilde{X}$ denoted by $\delta_2,\dots,\delta_r$. Then in the $\mathbb{Z}_2$–cofibration*

$$X_\delta \xrightarrow{\kappa'} \widetilde{X} \to X' \xrightarrow{\nu'} \Sigma(X_\delta),$$

the map ν' is $\mathbb{Z}_2$–nullhomotopic. □

Proof of Theorem 2.10 This follows from Lemma 2.8 together with Propositions 2.12 and 2.11. □

From Theorem 2.10, we reduce our study to the gauge groups

$$\mathcal{G}^*((1,1,1);(0,0)) \quad \text{and} \quad \mathcal{G}^*((2,1,1);(0,0));$$
$$\mathcal{G}^{*2}((1,1,1);(0,0)) \quad \text{and} \quad \mathcal{G}^{*2}((2,1,1);(0,0)).$$

The following theorem provides the remaining integral homotopy decompositions that we can obtain for these gauge groups. The theorem contributes to results in the last two rows of Theorem 1.11 and the last row in Theorem 1.12.

Theorem 2.13 *There are integral homotopy equivalences*

(1) $\mathscr{G}^{*2}((1,1,1);(0,0)) \simeq \mathscr{G}^{*}((1,1,1);(0,0)) \times U(n)$;

(2) $\mathscr{G}^{*2}((2,1,1);(0,0)) \simeq \mathscr{G}^{*}((1,1,1);(0,0)) \times U(n) \times U(n)$;

(3) $\mathscr{G}^{*}((2,1,1);(0,0)) \simeq \mathscr{G}^{*}((1,1,1);(0,0)) \times U(n)$.

We analyse the structure of a type-$(2,1,1)$ real surface X'.

Proposition 2.14 *Let X' be a type-$(2,1,1)$ real surface, and let X'_γ be the 1–cells $\gamma_2, \gamma_3, \sigma(\gamma_2), \sigma(\gamma_3)$ of X'. Then in the $\mathbb{Z}_2$–cofibration*

$$X'_\gamma \xrightarrow{\kappa''} X' \to X'/X'_\gamma \xrightarrow{\nu''} \Sigma(X'_\gamma),$$

the map ν'' is $\mathbb{Z}_2$–nullhomotopic.

Proof We define a left inverse to κ''. In X', collapse the cell δ_1, and then collapse a copy of $(S^1 \vee S^1, \mathrm{sw})$ so that $X'/{\sim}$ is the wedge $((\Sigma_1/{\sim}) \vee (\Sigma_1/{\sim}), \mathrm{sw})$, where $(\Sigma_1/{\sim})$ is a torus with δ_1 collapsed. We now project to $(S^1 \vee S^1, \mathrm{sw})$ as we did in the proof of Proposition 2.6; in fact, the left inverse is similar to the map j_3 from this proposition. □

In the following, we show that the space X'/X'_γ is $\mathbb{Z}_2$–homotopy equivalent to a $(1,1,1)$ real surface (X,σ). We first recall the $\mathbb{Z}_2$–decomposition of (X,σ). The 0–skeleton X^0 is given three 0–cells $*_i$ for $1 \le i \le 3$. The 1–cells are then

$$\delta_1, \delta, \sigma(\delta), \gamma_2, \sigma(\gamma_2),$$

where the fixed circle is represented by δ_1, and δ joins $*_2$ to $*_3$; therefore, $\delta\sigma(\delta)$ represents the copy of $(S^1, -\mathrm{id})$. The 1–cell γ_2 joins $*_1$ to $*_2$, and $\sigma(\gamma_2)$ joins $*_1$ to $*_3$. One of the two 2–cells has attaching map

$$\delta_1\gamma_2\delta\sigma(\delta)\gamma_2^{-1},$$

and we define the other one equivariantly.

On the other hand, the space X'/X'_γ has an induced $\mathbb{Z}_2$–complex structure as follows. There is one 0–cell $*$, to which we attach the 1–cells

$$\delta'_1, \delta' \quad \text{and} \quad \sigma(\delta').$$

There are two 2–cells, one of which is attached to the above 1–skeleton via

$$\delta_1\delta'\sigma(\delta'),$$

and the other is glued equivariantly. However, the subcomplex given by $\gamma_2 \cup \sigma(\gamma_2)$ of (X, σ) is $\mathbb{Z}_2$–contractible, and therefore, (X, σ) is homotopy equivalent to the $\mathbb{Z}_2$–complex structure of X'/X'_γ.

Proof of Theorem 2.13(2) and (3) By Proposition 2.14, we obtain the homotopy equivalences

$$\Sigma X' \simeq \Sigma X'_\gamma \vee \Sigma X'/X'_\gamma;$$
$$\Sigma \bar{X}' \simeq \Sigma \bar{X}'_\gamma \vee \Sigma X'/X'_\gamma.$$

In the first case, the factor $\Sigma X'_\gamma$ is the same as the suspension of $(S^1 \vee S^1, \mathrm{sw})$. We see that collapsing the 0–skeleton of $\Sigma X'_\gamma$ provides the suspension of $\bigvee_2 (S^1 \vee S^1, \mathrm{sw})$, and hence this corresponds to the factor $\Sigma \bar{X}'_\gamma$ in the second equivalence. The result follows. □

Proof of Theorem 2.13(1) We use the $\mathbb{Z}_2$–structure provided after Proposition 2.14. In this 2–pointed case, we identify the three 0–cells $*_1, *_2, *_3$ to produce $\bar{X}$. Let

$$X_\gamma = \gamma_2 \cup \sigma(\gamma_2),$$

and let $\bar{X}_\gamma$ be the image in the quotient $\bar{X}$. There is a left inverse to the inclusion

$$\bar{X}_\gamma \hookrightarrow \bar{X}$$

using a similar map to j_2 in the proof of Proposition 2.6. Therefore, there is a homotopy equivalence

$$\Sigma \bar{X} \simeq \Sigma \bar{X}_\gamma \vee \Sigma(\bar{X}/\bar{X}_\gamma),$$

but by the comments after Proposition 2.14, the factor $\Sigma(\bar{X}/\bar{X}_\gamma)$ is $\mathbb{Z}_2$–homotopy equivalent to the suspension of a real surface of type $(1, 1, 1)$. This finishes the proof. □

2.3.4 Nonintegral decompositions By the previous sections, we have reduced our study of the pointed gauge groups to those over real surfaces of the types

$$(0, 0, 1), \quad (1, 0, 1) \quad \text{and} \quad (1, 1, 1).$$

These spaces seem fundamental in some way, and for the single-pointed case we do not obtain any further integral decompositions.

However, one may expect these spaces to become easier to examine when we choose to invert 2 since the involution has order 2 and the 2–torsion in $O(n)$ vanishes. This turns out to be the case, and we will find that localising at a prime $p \neq 2$ will prove particularly fruitful.

In the coming sections, we aim to prove Theorem 1.13, dealing with each part in turn. The proof of each part is quite laborious, but we only provide full details for part (1). We outline the main parts of the proof of Theorem 1.13(1):

- The existence of the pullback (4) gives the existence of the map (6).
- We use an argument of [5] to prove that (6) is a p–local homotopy equivalence for primes $p \neq 2$.
- We calculate the homotopy fibre of qr in (6).

The proofs of Theorem 1.13(2) and (3) will then invoke similar methods.

The case (0, 0, 1) Let $(S^2, -\mathrm{id})$ be a real surface of type $(0,0,1)$. By Proposition 1.8, all of the pointed gauge groups over $(S^2, -\mathrm{id})$ are homotopy equivalent, so we assume that $(P, \widetilde{\sigma})$ is of class 0. In this section, we aim to prove the following theorem which is a restatement of Theorem 1.13(1).

Theorem 2.15 *For a prime $p \neq 2$ and odd n, there is a p–local homotopy equivalence*

$$\mathscr{G}^*(P, \widetilde{\sigma}) \simeq_p \Omega(U(n)/O(n)) \times \Omega^2(U(n)/O(n)).$$

Let $u\colon B\mathscr{G}^*(P, \widetilde{\sigma}) \to \mathrm{Map}^{*2}(D^2, BU(n))$ be the map that restricts to the upper hemisphere of $(S^2, -\mathrm{id})$ and forgets about equivariance. Let

$$r\colon B\mathscr{G}^*(P, \widetilde{\sigma}) \to \mathrm{Map}^*_{\mathbb{Z}_2}((S^1 \vee S^1, \mathrm{sw}), BU(n))$$

be the map restricting to the 1–skeleton of $(S^2, -\mathrm{id})$. These maps fit into the pullback

$$\begin{tikzcd} B\mathscr{G}^*(P, \widetilde{\sigma}) \arrow[r, "u"] \arrow[d, "r"] & \mathrm{Map}^{*2}(D^2, BU(n)) \arrow[d, "r'"] \\ \mathrm{Map}^*_{\mathbb{Z}_2}((S^1 \vee S^1, \mathrm{sw}), BU(n)) \arrow[r, "u'"] & \mathrm{Map}^*(S^1 \vee S^1, BU(n)) \end{tikzcd} \tag{4}$$

where r' restricts to the 1–skeleton and u' forgets about equivariance.

Let $\widehat{\varsigma}\colon U(n) \to U(n)$ denote complex conjugation and note that u' is homotopic to the map $\overline{\Delta}\colon U(n) \to U(n) \times U(n)$, where $\overline{\Delta}(\alpha) = (\alpha, \widehat{\varsigma}\alpha)$. Also note that the map r' is homotopic to the map $\Delta^{-1}\colon U(n) \to U(n) \times U(n)$, where $\Delta^{-1}(\alpha) = (\alpha, \alpha^{-1})$. Let Q be the strict pullback of $\overline{\Delta}$ and Δ^{-1} as in the following diagram:

$$\begin{tikzcd} Q \arrow[r, "\pi_2"] \arrow[d, "\pi_1"'] & U(n) \arrow[d, "\Delta^{-1}"] \\ U(n) \arrow[r, "\overline{\Delta}"] & U(n) \times U(n) \end{tikzcd}$$

We will see that Q retracts off $B\mathscr{G}^*(P,\widetilde{\sigma})$ after inverting the prime 2.

The map r' in diagram (4) is a fibration, and hence this diagram is a homotopy pullback. Therefore, there is an induced homotopy commuting diagram

(5)
$$\begin{tikzcd}
Q \arrow[dr, dotted, "\widetilde{\pi}"] \arrow[drr, bend left, "\pi_2"] \arrow[ddr, bend right, "\pi_1"'] & & \\
& B\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "u"] \arrow[d, "r"] & U(n) \arrow[d, "\Delta^{-1}"] \\
& U(n) \arrow[r, "\bar{\Delta}"] & U(n)\times U(n)
\end{tikzcd}$$

where we have replaced the pullback square (4) with a homotopy equivalent square.

Lemma 2.16 *The pullback Q is homeomorphic to $U(n)/O(n)$.*

Proof The pullback Q is the space

$$\{A\in U(n)\mid A^{-1}=\widehat{\varsigma}(A)\}.$$

Let $f\colon U(n)\to Q$ be defined by $f(A)=A\widehat{\varsigma}(A)^{-1}$. For matrices $A\in U(n)$ and $W\in U(n)^{\widehat{\varsigma}}=O(n)$, we have

$$(AW)\widehat{\varsigma}(AW)^{-1}=AW\widehat{\varsigma}(W^{-1})\widehat{\varsigma}(A^{-1})=A\widehat{\varsigma}(A)^{-1}$$

since $\widehat{\varsigma}$ is a homomorphism. Hence f induces a map $f'\colon U(n)/O(n)\to Q$.

We show that f' is a bijection. For injectivity, let $A,B\in U(n)$, and suppose that $A\widehat{\varsigma}(A)^{-1}=B\widehat{\varsigma}(B)^{-1}$. Then

$$I_n=B^{-1}A\widehat{\varsigma}(A)^{-1}\widehat{\varsigma}(B)=(B^{-1}A)\widehat{\varsigma}(B^{-1}A)^{-1}$$

for $I_n\in U(n)$ the identity matrix. Hence $B^{-1}A\in U(n)^{\widehat{\varsigma}}$, and so $AU(n)^{\widehat{\varsigma}}\equiv BU(n)^{\widehat{\varsigma}}$.

For surjectivity, let $A\in Q$. Then A is symmetric, and due to the Autonne–Takagi factorisation (see [14]), there is a unitary matrix P such that $A=PDP^t$, where D is a diagonal matrix with real entries. Let $\sqrt{D}$ be a diagonal matrix (hence an element of Q) in $U(n)$ such that $\sqrt{D}^2=D$. We have

$$A=P\sqrt{D}\sqrt{D}P^t=P\sqrt{D}\widehat{\varsigma}(P\sqrt{D})^{-1},$$

and therefore, $f'((P\sqrt{D})O(n))=A$.

The map f' is therefore a continuous bijection, and since $U(n)/O(n)$ is compact and Q is Hausdorff, it is a homeomorphism. □

The above diagram and Lemma 2.16 give the following composition

$$\varphi\colon U(n)/O(n) \xrightarrow{f'} Q \xrightarrow{\tilde{\pi}} B\mathscr{G}^*(P,\tilde{\sigma}) \xrightarrow{r} U(n) \xrightarrow{q} U(n)/O(n) \tag{6}$$

for q the quotient map. From the properties of π_1, we see that φ is homotopic to a map that sends an element $AO(n)$ to $AA^tO(n)$. For odd n, [5] showed that the related map

$$\mathrm{SU}(n)/\mathrm{SO}(n) \to \mathrm{SU}(n)/\mathrm{SO}(n), \quad A\,\mathrm{SO}(n) \mapsto AA^t\,\mathrm{SO}(n), \tag{7}$$

is a homotopy equivalence when localised at a prime $p \neq 2$. Our aim is to show that the same is true for φ.

Lemma 2.17 *For a prime $p \neq 2$, there is an p–local homotopy equivalence*

$$U(n)/O(n) \simeq_p U(n)/\mathrm{SO}(n).$$

Proof Consider the following pullback diagram where the downward arrows represent taking universal covers:

$$\begin{tikzcd}
U(n)/\mathrm{SO}(n) \arrow[r] \arrow[d] & B\,\mathrm{SO}(n) \arrow[r] \arrow[d] & BU(n) \arrow[d, equal] \\
U(n)/O(n) \arrow[r] \arrow[d] & BO(n) \arrow[r] \arrow[d] & BU(n) \\
K(\mathbb{Z}_2,1) \arrow[r, equal] & K(\mathbb{Z}_2,1) &
\end{tikzcd}$$

The result immediately follows. □

We now show that $U(n)/\mathrm{SO}(n)$ further decomposes into the product

$$\mathrm{SU}(n)/\mathrm{SO}(n) \times S^1.$$

The map $B\,\mathrm{SO}(n) \to BU(n)$ factors through $B\,\mathrm{SU}(n)$. Hence we obtain the following commutative diagram which defines the maps i and j:

$$\begin{tikzcd}
 & U(n) \arrow[r, equal] \arrow[d] & U(n) \arrow[d, "f"] \\
\mathrm{SU}(n)/\mathrm{SO}(n) \arrow[r, "i"] \arrow[d, equal] & U(n)/\mathrm{SO}(n) \arrow[r, "j"] \arrow[d] & S^1 \arrow[d] \\
\mathrm{SU}(n)/\mathrm{SO}(n) \arrow[r] & B\,\mathrm{SO}(n) \arrow[r] \arrow[d] & B\,\mathrm{SU}(n) \arrow[d] \\
 & BU(n) \arrow[r, equal] & BU(n)
\end{tikzcd} \tag{8}$$

It is not too much more work to show the following lemma.

Lemma 2.18 *There is a homotopy equivalence*

$$\eta\colon \mathrm{SU}(n)/\mathrm{SO}(n)\times S^1 \xrightarrow{\simeq} U(n)/\mathrm{SO}(n).$$

Proof There is a right inverse l to the map f and there is an action of $U(n)$ on $U(n)/\mathrm{SO}(n)$; hence the composition

$$\eta\colon S^1\times \mathrm{SU}(n)/\mathrm{SO}(n) \xrightarrow{l\times i} U(n)\times U(n)/\mathrm{SO}(n) \xrightarrow{\text{“action”}} U(n)/\mathrm{SO}(n)$$

is the required homotopy equivalence. □

Let φ be the composition in (6), and then define

$$s\colon U(n)/\mathrm{SO}(n)\to U(n)/\mathrm{SO}(n)$$

to be the composition

$$U(n)/\mathrm{SO}(n) \xrightarrow{\simeq} U(n)/O(n) \xrightarrow{\varphi} U(n)/O(n) \xrightarrow{\simeq} U(n)/\mathrm{SO}(n).$$

Our aim is to show that s restricts to the factors $\mathrm{SU}(n)/\mathrm{SO}(n)$ and S^1 in a nice enough way.

Lemma 2.19 *There exist maps*

$$s''\colon \mathrm{SU}(n)/\mathrm{SO}(n)\to \mathrm{SU}(n)/\mathrm{SO}(n) \quad \text{and} \quad s'\colon S^1\to S^1$$

such that the following is a homotopy commuting square:

$$\begin{tikzcd}
\mathrm{SU}(n)/\mathrm{SO}(n)\times S^1 \arrow[r, "s''\times s'"] \arrow[d, "\eta"] & \mathrm{SU}(n)/\mathrm{SO}(n)\times S^1 \arrow[d, "\eta"] \\
U(n)/\mathrm{SO}(n) \arrow[r, "s"] & U(n)/\mathrm{SO}(n)
\end{tikzcd}$$

Furthermore, these maps can be chosen such that s'' is homotopic to the map

$$A\,\mathrm{SO}(n)\mapsto AA^t\,\mathrm{SO}(n),$$

and s' is homotopic to the map $x\mapsto x^2$.

Proof Let $\tilde{s}\colon \mathrm{SU}(n)/\mathrm{SO}(n)\times S^1\to \mathrm{SU}(n)/\mathrm{SO}(n)\times S^1$ be the composition

$$\mathrm{SU}(n)/\mathrm{SO}(n)\times S^1 \xrightarrow{\eta} U(n)/\mathrm{SO}(n) \xrightarrow{s} U(n)/\mathrm{SO}(n) \xrightarrow{\eta^{-1}} \mathrm{SU}(n)/\mathrm{SO}(n)\times S^1$$

for a homotopy inverse η^{-1} of η. Let $\iota\colon \mathrm{SU}(n)/\mathrm{SO}(n) \to \mathrm{SU}(n)/\mathrm{SO}(n) \times S^1$ and $\kappa\colon S^1 \to \mathrm{SU}(n)/\mathrm{SO}(n) \times S^1$ be the inclusions. We note that ι is homotopic to

$$\mathrm{SU}(n)/\mathrm{SO}(n) \xrightarrow{i} U(n)/\mathrm{SO}(n) \xrightarrow{\eta^{-1}} \mathrm{SU}(n)/\mathrm{SO}(n) \times S^1,$$

where i is as in diagram (8). By the way the homotopy equivalences are defined in Lemmas 2.17 and 2.18, we see that the composition si is homotopic to

$$B\,\mathrm{SO}(n) \mapsto BB^t\,\mathrm{SO}(n) \quad \text{for } B \in \mathrm{SU}(n),$$

and hence the image of this map lands in the image of i. We deduce that $\tilde{s}\iota$ has image in $\mathrm{SU}(n)/\mathrm{SO}(n)$, and we define

$$s'' = \tilde{s}\iota.$$

Similarly, $\tilde{s}\kappa$ has image in S^1 and we define $s' = \tilde{s}\kappa$. We see that s'' is homotopic to a map defined via $A\,\mathrm{SO}(n) \mapsto AA^t\,\mathrm{SO}(n)$, and that s' is homotopic to the map $x \mapsto x^2$. □

We immediately obtain the following homotopy commuting diagram where the rows are homotopy fibrations:

$$\begin{tikzcd}
\mathrm{SU}(n)/\mathrm{SO}(n) \arrow[r, "i"] \arrow[d, "s''"] & U(n)/\mathrm{SO}(n) \arrow[r] \arrow[d, "s"] & S^1 \arrow[d, "s'"] \\
\mathrm{SU}(n)/\mathrm{SO}(n) \arrow[r, "i"] & U(n)/\mathrm{SO}(n) \arrow[r] & S^1
\end{tikzcd} \tag{9}$$

By Lemma 2.19, the map s'' is homotopic to the map in (7), and hence it is a p–local equivalence when n is odd and $p \neq 2$ is a prime. We note that s' is also a p–local equivalence. Finally, the spaces in (9) are connected; hence s is also a p–local equivalence. We are now able to deduce the following.

Proposition 2.20 *With the notation as in (6), we let F be the homotopy fibre of $qr\colon B\mathcal{G}^*(P,\tilde{\sigma}) \to U(n)/O(n)$. Then for n odd and for any prime $p \neq 2$, there is a p–local homotopy equivalence*

$$\mathcal{G}^*(P,\tilde{\sigma}) \simeq_p \Omega(U(n)/O(n)) \times \Omega F.$$

Proof Recall the maps f' and $\tilde{\pi}$ from (6). Then the above discussion has shown that $\tilde{\pi} f'$ provides a p–local homotopy section to the homotopy fibration

$$F \to B\mathcal{G}^*(P,\tilde{\sigma}) \xrightarrow{qr} U(n)/O(n),$$

and the result follows. □

Therefore, to prove Theorem 2.15 it only remains to identify the fibre F.

Proposition 2.21 *For any prime $p \neq 2$, there is a p–local homotopy equivalence*

$$F \simeq_p \Omega(U(n)/O(n)).$$

Proof The map qr from (6) is defined as a composition; hence there is a homotopy commutative diagram

$$\begin{tikzcd}
F \arrow[r] \arrow[d] & B\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "qr"] \arrow[d, "r"] & U(n)/O(n) \arrow[d, equal] \\
O(n) \arrow[r] & U(n) \arrow[r, "q"] & U(n)/O(n)
\end{tikzcd}$$

where the left square is a homotopy pullback square. The map r is a fibration since it is induced by $i\colon (S^1,-\mathrm{id}) \hookrightarrow (S^2,-\mathrm{id})$, the inclusion of the meridian copy of $(S^1,-\mathrm{id})$ into $(S^2,-\mathrm{id})$. Therefore, the space F is homotopy equivalent to the strict pullback of $O(n) \to U(n) \xleftarrow{r} B\mathscr{G}^*(P,\widetilde{\sigma})$, which is the relative mapping space

$$\mathrm{Map}^*_{\mathbb{Z}_2}\big(((S^2,-\mathrm{id}),(S^1,-\mathrm{id})),\ (BU(n), BO(n)); 0\big).$$

We will associate another pullback square with this description of F. There is a map $T\colon F \to \mathrm{Map}^*_{\mathbb{Z}_2}\big((S^2,-\mathrm{id}),(BU(n),\mathrm{id}); 0\big)$ given by

$$T(f)(x) = \begin{cases} f(x) & \text{for } x \text{ in the upper hemisphere including the equator,} \\ f(-\mathrm{id}(x)) & \text{for } x \text{ in the lower hemisphere excluding the equator.} \end{cases}$$

Let $i\colon (S^1,-\mathrm{id}) \hookrightarrow (S^2,-\mathrm{id})$ be defined as above. Then i induces the following homotopy pullback diagram:

$$\begin{tikzcd}
F \arrow[r, "T"] \arrow[d, "i^*"'] \arrow[dr, phantom, "\lrcorner", very near start] & \mathrm{Map}^*_{\mathbb{Z}_2}((S^2,-\mathrm{id}),(BU(n),\mathrm{id}); 0) \arrow[d, "i^*"] \\
O(n) \arrow[r, hook] & U(n)
\end{tikzcd}$$

There is a homeomorphism

$$\mathrm{Map}^*_{\mathbb{Z}_2}\big((S^2,-\mathrm{id}),(BU(n),\mathrm{id}); 0\big) \cong \mathrm{Map}^*(\mathbb{R}P^2, BU(n); 0),$$

but for a prime $p \neq 2$, the space $\mathbb{R}P^2$ is p–locally contractible. Therefore, p–locally, we have identified the space F as the fibre of the inclusion $O(n) \to U(n)$, and the result follows. $\square$

Proof of Theorem 2.15 Use Propositions 2.20 and 2.21. $\square$

The case $(1,0,1)$ Let (T,τ) be a real surface of type $(1,0,1)$, and since all pointed gauge groups over (T,τ) are homotopy equivalent, we restrict to the case where $(P,\widetilde{\sigma})$

is a bundle of class 0 over (T,τ). We will use similar techniques to the even genus case to obtain the following theorem, which is a restatement of Theorem 1.13(2).

Theorem 2.22 *For a prime $p \neq 2$ and n odd, there is a p–local homotopy equivalence*

$$\mathscr{G}^*(P,\widetilde{\sigma}) \simeq_p \Omega(U(n)/O(n)) \times \Omega^2(U(n)/O(n)) \times \Omega U(n).$$

Proof Let $u\colon B\mathscr{G}^*(P,\widetilde{\sigma}) \to \mathrm{Map}^*(C, BU(n))$ be the map that forgets about equivariance and restricts to the upper half of (T,τ), which is homeomorphic to a cylinder C. Let i be the inclusion of the boundary circles of C. Then i induces a pullback

$$\begin{tikzcd} B\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "u"] \arrow[d, "r"] & \mathrm{Map}^*(C, BU(n)) \arrow[d, "r'"] \\ \mathrm{Map}^*_{\mathbb{Z}_2}((S^1\vee S^1, \mathrm{sw}), BU(n)) \arrow[r, "u'"] & \mathrm{Map}^*(S^1 \sqcup S^1, BU(n)) \end{tikzcd} \tag{10}$$

where $r' = i^*$ and r is the restriction to the 1–skeleton of (X,σ).

In a similar fashion to the way we obtained diagram (5), we replace (10) with a homotopy equivalent square and obtain the diagram:

$$\begin{tikzcd} Q \arrow[dr, dotted] \arrow[drr, bend left] \arrow[ddr, bend right] & & \\ & B\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "u"] \arrow[d, "r"] & U(n) \arrow[d, "\Delta^{-1}"] \\ & U(n) \arrow[r, "\bar{\Delta}"] & U(n)\times LBU(n) \end{tikzcd}$$

Here $LBU(n)$ is the free loop space of $U(n)$, and Q is the strict pullback of the diagram

$$U(n) \xrightarrow{\bar{\Delta}} U(n)\times LBU(n) \xleftarrow{\Delta^{-1}} U(n).$$

Hence Q is again the symmetric matrices in $U(n)$. We deduce that $U(n)/O(n)$ also p–locally retracts off $B\mathscr{G}^*(P,\widetilde{\sigma})$.

It is clear that, as in the even case, there is a similar description for the fibre F of the map $B\mathscr{G}^*(P,\widetilde{\sigma}) \to U(n)/O(n)$. The space F fits into the following pullback diagram:

$$\begin{tikzcd} F \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & \mathrm{Map}^*_{\mathbb{Z}_2}\big((T,\tau),(BU(n),\mathrm{id});0\big) \arrow[d, "\bar{r}"] \\ O(n) \arrow[r, hook] & U(n) \end{tikzcd}$$

We note that if we let K be a Klein bottle, then there is an homeomorphism

$$\mathrm{Map}^*_{\mathbb{Z}_2}\big((T,\tau),(BU(n),\mathrm{id});0\big)\cong \mathrm{Map}^*(K,BU(n);0).$$

The map $\overline{r}$ is induced by the inclusion $S^1\hookrightarrow K$ which on fundamental groups induces the quotient

$$\mathbb{Z}\to\mathbb{Z}\times\mathbb{Z}_2,\quad a\mapsto(0,[a]_2),$$

onto the right factor. We see that for a prime $p\neq 2$, the map $\overline{r}$ is p–locally nullhomotopic, and we obtain

$$\Omega F\simeq_p \Omega^2(U(n)/O(n))\times\Omega\,\mathrm{Map}^*(K,BU(n);0).$$

Now for $p\neq 2$ prime, we have a p–local homotopy equivalence $K\simeq_p S^1$ because K is a $K(\mathbb{Z}\times\mathbb{Z}_2,1)$. Therefore, the space $\Omega\,\mathrm{Map}^*(K,BU(n);0)$ is homotopy equivalent to $\Omega U(n)$ when localised away from 2, and Theorem 2.22 follows. □

The case $(1,1,1)$ Let (X,σ) be a real surface of type $(1,1,1)$. For convenience, we choose $(P,\widetilde{\sigma})$ to be a bundle of class $(0,0)$ over (X,σ). We use a very similar method to the previous sections to prove the following theorem. This theorem is a more general statement than Theorem 1.13(3), whose statement claims to only be valid for odd n.

Theorem 2.23 *For any prime $p\neq 2$, there is a p–local homotopy equivalence*

$$\mathscr{G}^*(P,\widetilde{\sigma})\simeq_p\mathscr{G}^*((S^2,-\mathrm{id});0)\times\Omega O(n).$$

Proof We first recall the $\mathbb{Z}_2$–decomposition of (X,σ). The 0–skeleton X^0 is given three 0–cells $*_i$ for $1\le i\le 3$. The 1–cells are then

$$\delta_1,\ \delta,\ \sigma(\delta),\ \gamma_2,\ \sigma(\gamma_2),$$

where the fixed circle is represented by δ_1, and δ joins $*_2$ to $*_3$; therefore, $\delta\sigma(\delta)$ represents the copy of $(S^1,-\mathrm{id})$. The 1–cell γ_2 joins $*_1$ to $*_2$, and $\sigma(\gamma_2)$ joins $*_1$ to $*_3$. One of the two 2–cells has attaching map

$$\delta_1\gamma_2\delta\sigma(\delta)\gamma_2^{-1},$$

and we define the other one equivariantly.

Since the subspace $\gamma_2\cup\sigma(\gamma_2)$ is $\mathbb{Z}_2$–contractible, we amend the above decomposition to have only three 1–cells $\delta_1,\delta,\sigma(\delta)$ and amend the attaching map to

$$\delta_1\delta\sigma(\delta).$$

We obtain a pullback similar to that of the previous section:

$$\begin{tikzcd}
B\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "u"] \arrow[d, "r"] & \mathrm{Map}^{*3}(D^2, BU(n)) \arrow[d, "r'"] \\
\mathrm{Map}^*_{\mathbb{Z}_2}\big((S^1,\mathrm{id})\vee(S^1\vee S^1,\mathrm{sw}),\ BU(n);w_1\big) \arrow[r, "u'"] & \mathrm{Map}^*(S^1\vee S^1\vee S^1,\ BU(n))
\end{tikzcd}$$

where r is the restriction to the 1–skeleton of (X,σ), and u restricts to one of the 2–cells and forgets about equivariance.

In a similar fashion to the way we obtained diagram (5), we obtain the diagram

$$\begin{tikzcd}
O(n) \arrow[dr, dotted, "f_3"] \arrow[drr, bend left, "f_2"] \arrow[ddr, bend right, "f_1"'] & & \\
& B\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "u"] \arrow[d, "r"] & U(n)\times U(n) \arrow[d, "r'"] \\
& \mathrm{SO}(n)\times U(n) \arrow[r, "u'"] & U(n)\times U(n)\times U(n)
\end{tikzcd} \tag{11}$$

where f_1, f_2 and f_3 are to be defined momentarily.

The map $r'\colon U(n)\times U(n)\to U(n)\times U(n)\times U(n)$ is the map

$$r'(A,B)=(B^{-1}A^{-1},A,B),$$

and the map $u'\colon \mathrm{SO}(n)\times U(n)\to U(n)\times U(n)\times U(n)$ is the map

$$u'(C,D)=(C,D,\overline{D}).$$

We can hence define maps $f_1\colon O(n)\to\mathrm{SO}(n)\times U(n)$ and $f_2\colon O(n)\to U(n)\times U(n)$ by

$$f_1(X)=(X^{-2},X) \quad \text{and} \quad f_2(Y)=(Y,Y)$$

such that $u'f_1=r'f_2$. Since (11) is a homotopy pullback, there exists a map

$$f_3\colon O(n)\to B\mathscr{G}^*(P,\widetilde{\sigma})$$

such that the composition

$$\chi\colon O(n)\xrightarrow{f_3} B\mathscr{G}^*(P,\widetilde{\sigma})\xrightarrow{r} O(n)\times U(n)\xrightarrow{p_1} O(n)$$

sends an element X to X^{-2}. Then observe that χ has image lying in $\mathrm{SO}(n)$, and therefore, when χ is restricted to $\mathrm{SO}(n)$, it is the inverse of the H–space squaring map. We conclude that the restriction of χ to $\mathrm{SO}(n)$ is a p–local homotopy equivalence for $p\neq 2$, and therefore, $\mathrm{SO}(n)$ retracts off $B\mathscr{G}(P,\widetilde{\sigma})$.

The map $p_1 r$ is just the restriction to the fixed points of the involution. Hence the fibre of this map is the space $B\mathscr{G}^*((0,0,1);0)$, which we have already studied. We finish by noting that $\Omega\,\mathrm{SO}(n)$ and $\Omega O(n)$ are homeomorphic. $\square$

2.4 Unpointed gauge groups

In the last section, we showed that certain trivialities of the attaching map of the top cells of X led to homotopy decompositions in the pointed case. We will see that these decompositions somewhat extend to the unpointed case.

2.4.1 Integral decompositions Let (X,σ) be a real surface of type (g,r,a). In the following proposition, g' will denote the number of α_i and β_i cells in the description of (X,σ) in Section 2.1. Explicitly,

$$g' = \begin{cases} g-r+1 & \text{when } a=0, \\ g-r & \text{when } a=1 \text{ and } g-r \text{ is even}, \\ g-r-1 & \text{when } a=1 \text{ and } g-r \text{ is odd}. \end{cases}$$

We now present Proposition 2.24 which is a restatement of Theorem 1.14(1).

Proposition 2.24 *There are homotopy equivalences*

$$\mathscr{G}((g,r,a);(c,w_1,\dots,w_r)) \simeq \mathscr{G}((g-g',r,a);(c,w_1,\dots,w_r)) \times \prod_{g'} \Omega U(n).$$

Proof In essence, we follow the proof of [11, Proposition 2.1]. For convenience, we write

$$(c,\overline{w}) := (c,w_1,\dots,w_r).$$

Let $X_{\alpha\beta} = \bigvee_{g'}(S^1 \vee S^1, \mathrm{sw})$ be subcomplex of X represented by $\alpha_i, \sigma(\alpha_i), \beta_i, \sigma(\beta_i)$. Recall the $\mathbb{Z}_2$–cofibration sequence of Proposition 2.3:

$$X_{\alpha\beta} \hookrightarrow X \xrightarrow{q} X' \xrightarrow{\mu} \Sigma(X_{\alpha\beta}).$$

Then the map q induces the diagram

$$\begin{array}{ccccccc}
\Omega B & \xrightarrow{\partial_{(c,\overline{w})}} & \mathrm{Map}^*_{\mathbb{Z}_2}(X', BU(n);(c,\overline{w})) & \longrightarrow & \mathrm{Map}_{\mathbb{Z}_2}(X', BU(n);(c,\overline{w})) & \xrightarrow{\mathrm{ev}} & B \\
\| & & \downarrow{\scriptstyle q^*} & & \downarrow{\scriptstyle q^*} & & \| \\
\Omega B & \xrightarrow{\varphi_{(c,\overline{w})}} & \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n);(c,\overline{w})) & \longrightarrow & \mathrm{Map}_{\mathbb{Z}_2}(X, BU(n);(c,\overline{w})) & \xrightarrow{\mathrm{ev}} & B
\end{array}$$

where

$$B = \begin{cases} BU(n) & \text{if } r=0, \\ BO(n) & \text{otherwise}. \end{cases}$$

The equation $\varphi_{(c,\overline{w})} = q^*\partial_{(c,\overline{w})}$ results in the diagram

$$\begin{tikzcd}
 & \mathrm{Map}^*(\Sigma(X), BU(n); (c,\overline{w})) \arrow[r, equal] \arrow[d] & \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma X, BU(n); (c,\overline{w})) \arrow[d, "(\Sigma i)^*"] \\
\mathscr{G}(g-g') \arrow[r, "h'"] \arrow[d, equal] & \mathscr{G}((g,r,a); (c,\overline{w})) \arrow[r, "h"] \arrow[d] & \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma(X_{\alpha\beta}), BU(n)) \arrow[d, "\mu^*"] \\
\mathscr{G}(g-g') \arrow[r] & \Omega B \arrow[r, "\partial_{(c,\overline{w})}"] \arrow[d, "\varphi_{(c,\overline{w})}"] & \mathrm{Map}^*_{\mathbb{Z}_2}(X', BU(n); (c,\overline{w})) \arrow[d, "q^*"] \\
 & \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); (c,\overline{w})) \arrow[r, equal] & \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); (c,\overline{w}))
\end{tikzcd}$$

which defines the maps h and h', and in which $\mathscr{G}(g-g') := \mathscr{G}((g-g', r, a); (c,\overline{w}))$. By Proposition 2.3, the map μ^* is trivial. Hence there is a section to the map $(\Sigma i)^*$, so there is also a section to h, and the result follows. $\square$

The quotient map q in Proposition 2.24 induced an isomorphism on π_0 between

$$\mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); (c,\overline{w})) \quad \text{and} \quad \mathrm{Map}^*_{\mathbb{Z}_2}(X', BU(n); (c,\overline{w})).$$

However, for a fixed cell δ_i of (X,σ), the quotient map $\tilde{q}\colon X \to X/\delta_i$ automatically induces the map

$$\mathrm{Map}_{\mathbb{Z}_2}(X/\delta_i, BU(n)) \xrightarrow{q^*} \mathrm{Map}_{\mathbb{Z}_2}(X, BU(n); 0),$$

hence the requirement for $w_i = 0$ in Theorem 1.14(3). Whilst there is an equivalence

$$\mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); (c,0)) \simeq \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); (c,1)),$$

there is not necessarily an equivalence in the unpointed case in general. Hence there is not enough information to guarantee the commutativity of the diagram needed to induce a homotopy decomposition.

Omitting such nontrivialities allows further splittings; let X_1 be a subset of the 1–cells of X such that

(1) if there is a fixed cell $\delta_i \subset X_1$, then $w_i = 0$;

(2) for appropriate components, the induced map

$$g^*\colon \mathrm{Map}^*_{\mathbb{Z}_2}(\Sigma X_1, BU(n); (\overline{w})) \to \mathrm{Map}^*_{\mathbb{Z}_2}(X/X_1, BU(n); (c,\overline{w}))$$

is $\mathbb{Z}_2$–nullhomotopic.

Under these assumptions, it is clear that the methods in the previous proposition would yield further homotopy decompositions.

Proof of Theorem 1.14(2) and (3) The above conditions apply to the 1–cells considered in Propositions 2.6, 2.11 and 2.14 for bundles of arbitrary type.

Additionally, the conditions are satisfied by the 1–cells considered in Propositions 2.7 and 2.12 for bundles of type $(c, w_1, 0, \dots, 0)$. When n is odd, we can take advantage of Proposition 1.10 to obtain the table in Theorem 1.14(3). We have now finished the proof of Theorem 1.14. □

2.4.2 Analysing the boundary map Let $(P, \widetilde{\sigma}) \to (X, \sigma)$ be a real bundle of class $(c, w_1, \dots, w_r)$ over a real surface (X, σ) of type (g, r, a). Let

$$B = \begin{cases} BO(n) & \text{if } r > 0, \\ BU(n) & \text{otherwise,} \end{cases}$$

and consider the homotopy fibration sequence induced from the map that evaluates at the basepoint of X:

$$\mathscr{G}(P, \widetilde{\sigma}) \to \Omega B \xrightarrow{\partial_P} \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); P) \to \mathrm{Map}_{\mathbb{Z}_2}(X, BU(n); P) \to B. \tag{12}$$

Since $\mathscr{G}(P, \widetilde{\sigma})$ appears as the homotopy fibre of the boundary map ∂_P, we aim to gather information about $\mathscr{G}(P, \widetilde{\sigma})$ by studying ∂_P. Our method will involve comparing ∂_P to a map arising from a similar homotopy fibration sequence found in [12]. This approach is particularly fruitful when X^σ is nonempty, that is, when $r > 0$. We reserve analysis of the $r = 0$ cases not handled by Section 2.4.1 to later sections, however, we will require discussion from this section and Section 2.3.4.

Note that

$$\pi_0\big(\mathrm{Map}(S^2, BU(n))\big) \cong \mathbb{Z},$$

and for $d \in \mathbb{Z}$, we obtain a fibration sequence

$$U(n) \xrightarrow{\partial_d} \mathrm{Map}^*(S^2, BU(n); d) \to \mathrm{Map}(S^2, BU(n); d) \to BU(n). \tag{13}$$

The trivialities of the map ∂_d were extensively studied in [12]. We state the relevant results from this paper.

Theorem 2.25 (Theriault) *Let p be a prime, and let*

$$\partial_d \colon U(n) \to \mathrm{Map}^*(S^2, BU(n); d)$$

be as in (13). Then

(1) *if $p \nmid n$, then ∂_d is p–locally trivial;*

(2) *if $n = p$ with $p \mid d$, then ∂_d is p–locally trivial.*

The case $n = p \nmid d$ was also studied in [12]; the map ∂_d is not p–locally trivial, but the homotopy fibre was identified. The following two propositions adapt some of the trivialities of ∂_d to our setting.

Proposition 2.26 *Fix $d \in \mathbb{Z}$, and let ∂_d be the boundary map in (13). Let $(P, \tilde{\sigma})$ be a real principal $U(n)$–bundle of class $(2d, 0, \dots, 0)$ over a real surface of type (g, r, a). Let*

$$\partial_P \colon \Omega B \to B\mathscr{G}^*((g, r, a); (2d, 0, \dots, 0))$$

be the boundary map of the evaluation fibration in (12). For a prime q, if ∂_d is (q–locally) trivial, then

(1) *if $r > 0$, then ∂_P is (q–locally) trivial;*

(2) *if $r = 0$, then the composition*

$$O(n) \hookrightarrow U(n) \xrightarrow{\partial_P} B\mathscr{G}^*((g, r, a); (2d, 0, \dots, 0))$$

is (q–locally) trivial.

Proof The key will be to compare both maps to another evaluation boundary map involving the $\mathbb{Z}_2$–space $Y = (S^2 \vee S^2, \mathrm{sw})$. Note that components of $\mathrm{Map}^*_{\mathbb{Z}_2}(Y, BU(n))$ are classified by even integers.

Let $S^2 \xrightarrow{i_1} S^2 \vee S^2 = Y$ be the inclusion onto the left factor, and note that this is not a $\mathbb{Z}_2$–map. The following diagram commutes:

$$\begin{tikzcd} O(n) \arrow[r, "\overline{\partial}_{2d}"] \arrow[d, hook] & \mathrm{Map}^*_{\mathbb{Z}_2}(Y, BU(n); 2d) \arrow[r] \arrow[d, "i_1^*"] & \mathrm{Map}_{\mathbb{Z}_2}(Y, BU(n); 2d) \arrow[r] \arrow[d] & BO(n) \arrow[d, hook] \\ U(n) \arrow[r, "\partial_d"] & \mathrm{Map}^*(S^2, BU(n); d) \arrow[r] & \mathrm{Map}(S^2, BU(n); d) \arrow[r] & BU(n) \end{tikzcd} \tag{14}$$

Now there is an inverse to i_1^* which sends a map f in $\mathrm{Map}^*(S^2, BU(n); d)$ to the composition

$$S^2 \vee S^2 \xrightarrow{f \vee f} BU(n) \vee BU(n) \xrightarrow{\mathrm{id} \vee \sigma_{BU(n)}} BU(n) \vee BU(n) \xrightarrow{\text{fold}} BU(n),$$

which is $\mathbb{Z}_2$–equivariant because the involution on $S^2 \vee S^2$ swaps the factors. Note that the map induced on the unpointed mapping spaces does not have an inverse because the basepoint of Y must land in $BO(n)$. We conclude that if ∂_d is q–locally trivial, then so is $\overline{\partial}_{2d}$.

Let $q\colon X \to Y$ be the map that collapses the 1–skeleton of the real surface (X, σ). We obtain the following commutative diagram:

$$\begin{tikzcd}
O(n) \arrow[r, "\overline{\partial}_{2d}"] \arrow[d, "f"] & \mathrm{Map}^*_{\mathbb{Z}_2}(Y, BU(n); 2d) \arrow[r] \arrow[d, "q^*"] & \mathrm{Map}_{\mathbb{Z}_2}(Y, BU(n); 2d) \arrow[r] \arrow[d] & BO(n) \arrow[d] \\
\Omega B \arrow[r, "\partial_P"] & \mathrm{Map}^*_{\mathbb{Z}_2}(X, BU(n); P) \arrow[r] & \mathrm{Map}_{\mathbb{Z}_2}(X, BU(n); P) \arrow[r] & B
\end{tikzcd}$$

The map f is an equivalence if $r > 0$ and is the inclusion $O(n) \hookrightarrow U(n)$ otherwise. Since $\overline{\partial}_{2d}$ is (q–locally) trivial, the result follows. □

Proposition 2.27 *Let p be a prime such that $p \nmid d$, and let $(P, \widetilde{\sigma})$ be a real principal $U(p)$–bundle of class $(2d, 0, \dots, 0)$ over a real surface of type (g, r, a). Let*

$$\partial_P\colon \Omega B \to B\mathscr{G}^*((g, r, a); (2d, 0, \dots, 0))$$

be the boundary map of the evaluation fibration. Then

(1) *if $r > 0$, then ∂_P is p–locally trivial;*

(2) *if $r = 0$, then the composition*

$$O(p) \hookrightarrow U(p) \xrightarrow{\partial_P} B\mathscr{G}^*((g, r, a); (2d, 0, \dots, 0))$$

is p–locally trivial.

Proof We assume that $p \nmid d$ is a prime and that all spaces and maps are localised at p. Let $Y = (S^2 \vee S^2, \mathrm{sw})$ be as above. Then there is a homotopy commuting diagram

$$\begin{tikzcd}
O(p) \arrow[rrr, "\overline{\partial}_{2d}"] \arrow[d, hook, "i"] & & & \mathrm{Map}^*_{\mathbb{Z}_2}(Y, BU(n); 2d) \arrow[d, "\simeq"] \\
U(p) \arrow[rrr, "\partial_d"] \arrow[d, equal] & & & \Omega U(p)_0 \\
U(p) \arrow[rrr, "\partial_1"] \arrow[d, "e"] & & & \Omega U(p)_0 \arrow[u, "d"] \arrow[d, "(\Omega e)_0"] \\
\prod_{i=0}^{p-1} S^{2i+1} \arrow[r, "\mathrm{proj}"] & S^{2p-1} \arrow[r, "\alpha"] & \Omega S^3 \arrow[r, "\mathrm{incl}"] & \prod_{j=1}^{p-1} \Omega S^{2j+1}
\end{tikzcd} \tag{15}$$

where the top square is from diagram (14) and the bottom two squares are found in [12], specifically in Proposition 4.1 and the proof of Theorem 1.1(b) and (c).

The d^{th} power map $d\colon \Omega U(p)_0 \to \Omega U(p)_0$ is a homotopy equivalence because $p \nmid d$. Furthermore, the maps e and $(\Omega e)_0$ are homotopy equivalences provided in [9]. Now for $p \neq 2$ prime, there is a p–local homotopy equivalence

$$\mathrm{SO}(p) \simeq_p \prod_{i=1}^{(p-1)/2} S^{4i-1},$$

and furthermore, the inclusion $O(p) \hookrightarrow U(p)$ is in fact the inclusion of these factors into $\prod_{i=0}^{p-1} S^{2i+1}$. We conclude that the composition

$$\chi\colon O(p) \hookrightarrow U(p) \to \prod_{i=0}^{p-1} S^{2i+1} \xrightarrow{\text{proj}} S^{2p-1} \tag{16}$$

is nullhomotopic, and therefore, so is $\overline{\partial}_{2d}$.

For $p = 2$, the space $O(2)$ is homeomorphic to $S^1 \amalg S^1$. Since χ in (16) has target space S^3, we conclude that χ and hence $\overline{\partial}_{2d}$ are nullhomotopic in this case, too. The result then follows in a similar way to the last paragraph in the proof of Proposition 2.26. □

Proof of Theorem 1.15(1a) and (2a) Theorem 2.25 and Proposition 2.26 immediately obtain (1a). Similarly, Theorem 2.25 and Proposition 2.26 obtain (2a) when $p \mid d$, and Proposition 2.27 then gives the remaining case when $p \nmid d$. □

2.4.3 The case $(0,0,1)$ We restrict to analysing gauge groups above real surfaces of type $(0,0,1)$. Fix an even integer c. Then we wish to analyse the boundary map ∂_c of the evaluation fibration.

For a $\mathbb{Z}_2$–space A, let $\overline{\Delta}\colon A \to A \times A$ be the composition

$$A \xrightarrow{\Delta} A \times A \xrightarrow{\mathrm{id} \times \sigma_A} A \times A. \tag{17}$$

Let u be the composition

$$B\mathscr{G}^*((0,0,1);c) \xrightarrow{\simeq} \mathrm{Map}^*_{\mathbb{Z}_2}(S^2, BU(n); c) \xrightarrow{\tilde{u}} \mathrm{Map}^{*2}(D^2, BU(n)) \xrightarrow{\simeq} U(n),$$

where $\tilde{u}$ restricts to the upper hemisphere of $(S^2, -\mathrm{id})$ and forgets about equivariance except at $*$ and $\sigma(*)$. The last equivalence follows since D^2 with two points identified is homotopy equivalent to S^1. The maps u and $\overline{\Delta}$ are the same as in (5), and they fit into the commutative diagram

$$\begin{tikzcd}
U(n) \arrow[r, "\partial_c"] \arrow[d, "\overline{\Delta}"] & B\mathscr{G}^*((0,0,1);c) \arrow[r] \arrow[d, "u"] & B\mathscr{G}((0,0,1);c) \arrow[r] \arrow[d] & BU(n) \arrow[d, "\overline{\Delta}"] \\
U(n) \times U(n) \arrow[r, "\xi"] & U(n) \arrow[r] & \mathrm{Map}(D^2, BU(n)) \arrow[r, "\mathrm{ev}_2"] & BU(n) \times BU(n)
\end{tikzcd}$$

where ev_2 evaluates at two antipodal points on the boundary of D^2, and ζ is defined via this diagram.

Since D^2 is contractible, the map ev_2 is homotopic to the diagonal map

$$\Delta\colon BU(n) \to BU(n) \times BU(n).$$

Therefore, the map ζ is homotopic to the map defined by $(A, B) \mapsto AB^{-1}$. Let $f\colon U(n) \to U(n)$ be defined as $f(A) = AA^t$ and we conclude that $u\partial_c \simeq f$.

After localising the map f at a prime $p \neq 2$, we have the composition

$$\mathrm{SO}(n) \times U(n)/\,\mathrm{SO}(n) \xrightarrow{f} \mathrm{SO}(n) \times U(n)/\,\mathrm{SO}(n) \xrightarrow{p_2} U(n)/\,\mathrm{SO}(n), \tag{18}$$

where p_2 is the projection map. Recall the map φ from (6) and compare with f. For $p \neq 2$, we showed that φ is a p–local homotopy equivalence, and we conclude that restricting the composition (18) to the factor $U(n)/\,\mathrm{SO}(n)$ also obtains a p–local homotopy equivalence. We have shown the following proposition.

Proposition 2.28 *Let n be odd. Then localised at a prime $p \neq 2$, the following composition is a homotopy equivalence:*

$$U(n)/\,\mathrm{SO}(n) \hookrightarrow U(n) \xrightarrow{\partial_c} B\mathscr{G}^*((0,0,1);c) \xrightarrow{u} U(n) \to U(n)/\,\mathrm{SO}(n). \qquad \square$$

With this proposition, we have enough ammunition to prove Theorem 1.15(1b) and (2b).

Proof of Theorem 1.15(1b) and (2b) We first prove part (1b). Localise at a prime $p \neq 2$ such that $p \nmid n$, and reconsider the fibration sequence

$$\mathscr{G}((0,0,1);c) \to \mathrm{SO}(n) \times U(n)/\,\mathrm{SO}(n) \xrightarrow{\partial_c} B\mathscr{G}^*((0,0,1);c).$$

By Proposition 2.28, the factor $U(n)/\,\mathrm{SO}(n)$ retracts off $B\mathscr{G}^*((0,0,1);c)$, and by Proposition 2.26(2) the factor $\mathrm{SO}(n)$ retracts off $\mathscr{G}((0,0,1);c)$ under a lift

$$l\colon \mathrm{SO}(n) \to \mathscr{G}((0,0,1);c)$$

of the inclusion $\mathrm{SO}(n) \hookrightarrow U(n)$. Then the composition

$$\begin{aligned}\mathrm{SO}(n) \times \Omega^2(U(n)/\,\mathrm{SO}(n)) &\xrightarrow{l\times\mathrm{id}} \mathscr{G}((0,0,1);c) \times \Omega^2(U(n)/\,\mathrm{SO}(n)) \\ &\xrightarrow{\text{“action”}} \mathscr{G}((0,0,1);c)\end{aligned}$$

is a homotopy equivalence, and the result follows. The proof of part (2b) is similar. $\square$

2.4.4 The case $(1,0,1)$ We now analyse unpointed gauge groups above a real surface (T,τ) of type $(1,0,1)$. We use a similar method to the $(0,0,1)$ case and adopt some of its notation.

As in the proof of Theorem 2.22, let $u'\colon B\mathscr{G}^*((1,0,1);c)\to \mathrm{Map}^*(C,BU(n))$ be the map that forgets about equivariance and restricts to the upper half of (T,τ), which is homeomorphic to a cylinder C. Let $\overline{\Delta}$ be as in (17). Then we obtain the diagram

$$\begin{tikzcd}
U(n) \arrow[r, "\partial_c"] \arrow[d, "\overline{\Delta}"] & B\mathscr{G}^*((1,0,1);c) \arrow[r] \arrow[d, "u'"] & B\mathscr{G}((1,0,1);c) \arrow[r] \arrow[d] & BU(n) \arrow[d, "\overline{\Delta}"] \\
U(n)\times U(n) \arrow[r, "\zeta'"] & \mathrm{Map}^{*2}(C,BU(n)) \arrow[r] & \mathrm{Map}(C,BU(n)) \arrow[r, "\mathrm{ev}_2"] & BU(n)\times BU(n)
\end{tikzcd}$$

where ev_2 is another double evaluation map; viewing C as a subcomplex of (T,τ), the map ev_2 evaluates at the basepoint $*_1$ and its image under the involution $\tau(*_1)$. Again, the map ζ' is defined via this diagram.

As in the previous case, we aim to study the homotopy type of the map $\zeta'\overline{\Delta}$. However, it is not immediately clear on the homotopy type of the "boundary" map ζ'. We note that $C\simeq S^1$ under a deformation retract fixing $*_1$ and taking $\tau(*_1)$ to $*_1$. Therefore, if we let $LBU(n)$ be the free loop space of $BU(n)$, we deduce that there is a homotopy commutative diagram

$$\begin{tikzcd}
\mathrm{Map}(C,BU(n)) \arrow[r, "\mathrm{ev}_2"] \arrow[d, "\simeq"] & BU(n)\times BU(n) \\
LBU(n) \arrow[r, "\mathrm{ev}"] & BU(n) \arrow[u, "\Delta"']
\end{tikzcd}$$

where ev evaluates at the basepoint $*_1$ and Δ is the diagonal map. Given that $\Delta\,\mathrm{ev}$ is a composition, we obtain the homotopy commutative diagram

$$\begin{tikzcd}
 & U(n)\times U(n) \arrow[r, equal] \arrow[d, "\zeta'"] & U(n)\times U(n) \arrow[d, "\tilde{\xi}"] \\
U(n) \arrow[r, "h'"] \arrow[d, equal] & \mathrm{Map}^{*2}(C,BU(n)) \arrow[r, "h"] \arrow[d] & U(n) \arrow[d, "*"] \\
U(n) \arrow[r] & LBU(n) \arrow[r, "\mathrm{ev}"] \arrow[d, "\Delta\,\mathrm{ev}"] & BU(n) \arrow[d, "\Delta"] \\
 & BU(n)\times BU(n) \arrow[r, equal] & BU(n)\times BU(n)
\end{tikzcd}$$

where the map $*$ is the inclusion of the homotopy fibre, which is nullhomotopic. The middle square is a homotopy pullback, and hence the maps h and h' are defined using this diagram.

By the triviality of the middle right vertical, there is a right homotopy inverse i to h and a left inverse q to h'. Therefore, the space $\mathrm{Map}^{*2}(C, BU(n))$ is homotopy equivalent to the product $U(n) \times U(n)$.[3] Therefore, the homotopy type of

$$\zeta' \colon U(n) \times U(n) \to \mathrm{Map}^{*2}(C, BU(n))$$

can be determined by studying $q\zeta'$ and $h\zeta'$. It is clear that $q\zeta' \sim *$ and $h\zeta' \sim \widetilde{\zeta}$. However, $\widetilde{\zeta}$ is the same as the map $\zeta \colon U(n) \times U(n) \to U(n)$ in case $(0,0,1)$, and therefore, it is homotopic to the map $(A, B) \mapsto AB^{-1}$.

We conclude that ζ' is homotopic to a map

$$U(n) \times U(n) \to U(n) \times U(n), \quad (A, B) \mapsto (I_n, AB^{-1}).$$

Proof of Theorem 1.15(1c) and (2c) We first prove part (1c). Let $p \neq 2$ be a prime with $p \nmid n$. Then localised at p, in the same way as Proposition 2.28, we see that the factor $U(n)/\mathrm{SO}(n)$ in

$$U(n) \simeq_p U(n)/\mathrm{SO}(n) \times \mathrm{SO}(n)$$

retracts off $B\mathscr{G}^*((1,0,1); c)$ via

$$U(n)/\mathrm{SO}(n) \hookrightarrow U(n) \xrightarrow{\partial_c} B\mathscr{G}^*((1,0,1); c) \xrightarrow{u'} U(n) \to U(n)/\mathrm{SO}(n).$$

Additionally, by Proposition 2.26(2), the factor $\mathrm{SO}(n)$ retracts off the gauge group $\mathscr{G}((1,0,1); c)$. We then find the required homotopy equivalence as in the proof of Theorem 1.15(1b). The proof of (2c) is similar. □

2.5 The quaternionic case

From here on, we restrict to the quaternionic case. Again, our method of attack will be to study some mapping spaces related to these gauge groups. In fact, these mapping spaces are the same as in the real case, except $BU(2n)$ is endowed with an involution so that

$$(EU(2n), \widetilde{\varsigma}_{\mathrm{Q}}) \to (BU(2n), \varsigma_{\mathrm{Q}})$$

is a universal quaternionic bundle. Recall that in the real case, the involution ς was induced by complex conjugation $\widehat{\varsigma} \colon U(n) \to U(n)$. In this case, the involution ς_{Q} is

[3] Of course, this can be seen directly by studying the homotopy type of C.

induced from the homomorphism

$$\widehat{\varsigma}_{\mathrm{Q}}\colon U(n) \to U(n), \quad A \mapsto J^{-1}AJ,$$

where

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 0 \end{pmatrix}.$$

Most of the results in the real case come from geometric properties of (X,σ); hence we will see that these results transfer to the quaternionic setting without too much hassle. Furthermore, since $(BG)^{\varsigma_{\mathrm{Q}}} = B\,\mathrm{Sp}(n)$, we will see that a number of results will be easier to prove due to the high connectivity of $B\,\mathrm{Sp}(n)$.

For the $\mathbb{Z}_2$–space $(BU(2n), \varsigma_{\mathrm{Q}})$ as above, we write

$$\mathrm{Map}_{\mathrm{Q}}(X, BU(2n)) := \mathrm{Map}_{\mathbb{Z}_2}(X, BU(2n))$$

to distinguish from the real case, and use similar notation for the pointed cases. Now let $\bar{X}$ be as in the preamble to Theorem 2.1, and we state the quaternionic analogue of Theorem 2.1.

Theorem 2.29 *Let $(P,\tilde{\sigma})$ be a quaternionic principal $U(2n)$–bundle of class c over a real surface (X,σ) of type (g,r,a). Then there are homotopy equivalences*

(1) $B\,\mathscr{G}_{\mathrm{Q}}(P,\tilde{\sigma}) \simeq \mathrm{Map}_{\mathrm{Q}}(X, BU(2n); P)$;

(2) $B\,\mathscr{G}^*_{\mathrm{Q}}(P,\tilde{\sigma}) \simeq \mathrm{Map}^*_{\mathrm{Q}}(X, BU(2n); P)$;

(3) $B\,\mathscr{G}^{*(r+a)}_{\mathrm{Q}}(P,\tilde{\sigma}) \simeq \mathrm{Map}^{*(r+a)}_{\mathrm{Q}}(X, BU(2n); c) \simeq \mathrm{Map}^*_{\mathrm{Q}}(\bar{X}, BU(2n); P)$;

where on the right-hand side, we pick the path component of $\mathrm{Map}_{\mathrm{Q}}(X, BU(n))$ that induces $(P,\tilde{\sigma})$.

We now sketch the proofs for the results in Section 1.3.

Proof of 1.16 We use the action of $\pi_2(BU(2n))$ on $[(X,\sigma),(BU(2n),\varsigma_{\mathrm{Q}})]_{\mathbb{Z}_2}$ as presented in the proof of Proposition 1.7. □

As in the real case, the lack of a $\pi_2(BU(2n))$ action means that we cannot provide an analogue for $B\,\mathscr{G}_{\mathrm{Q}}(P,\tilde{\sigma})$.

Proof of 1.17 The idea is to tensor the quaternionic bundle $(P, \widetilde{\sigma})$ with a real $U(1)$–bundle $\pi_Q\colon (Q,\tau) \to (X,\sigma)$ of class $(2,0,\dots,0)$. The required isomorphism of gauge groups is then defined as in the proof of Proposition 1.9. □

We sketch the proofs for the results related to homotopy decompositions of the gauge groups.

Proof of 1.18 and 1.19 The proof is similar to those in Sections 2.3.1–2.3.3, except that in this case, $BU(2n)^{\varsigma_{\mathrm{Q}}} = B\,\mathrm{Sp}(n)$. We recall that decompositions involving fixed circles in the real case needed to be handled delicately, but this does not occur in the quaternionic case due to the high connectivity of $B\,\mathrm{Sp}(n)$. □

Our aim is to now prove Theorem 1.20 using a similar method to that of Theorem 2.15. Localised at a prime $p \neq 2$ and for n odd, we obtained a p–local decomposition in the real case due to the fact that the p–local homotopy equivalence

$$U(n)/O(n) \to U(n)/O(n), \quad AO(n) \mapsto AA^t O(n),$$

factored through the space $B\,\mathscr{G}^*((0,0,1);0)$. We shall see that a similar map involving $U(2n)/\,\mathrm{Sp}(n)$ also factors through the quaternionic analogue of this gauge group. Let

$$u\colon B\,\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0) \to \mathrm{Map}^{*2}(D^2, BU(2n))$$

be the map that restricts to the upper hemisphere of $(S^2, -\mathrm{id})$ and forgets about equivariance, considering the image as landing in $\mathrm{Map}^{*2}(D^2, BU(2n))$. Let

$$r\colon B\,\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0) \to \mathrm{Map}^*_{\mathbb{Z}_2}((S^1 \vee S^1, \mathrm{sw}), BU(2n))$$

be the map restricting to the 1–skeleton of $(S^2, -\mathrm{id})$. We obtain a homotopy commuting diagram similar to diagram (5):

$$\begin{tikzcd}
Q \arrow[dr, dotted] \arrow[drr, bend left] \arrow[ddr, bend right] & & \\
& B\,\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0) \arrow[r, "u"] \arrow[d, "r"] & U(2n) \arrow[d, "\Delta^{-1}"] \\
& U(2n) \arrow[r, "\Delta^{\mathrm{Q}}"] & U(2n) \times U(2n)
\end{tikzcd}$$

where Δ^{Q} is the map $A \mapsto (A, \hat{\varsigma}_{\mathrm{Q}} A)$. Here Q is the strict pullback of the diagram

$$U(2n) \xrightarrow{\Delta^{\mathrm{Q}}} U(2n) \times U(2n) \xleftarrow{\Delta^{-1}} U(2n),$$

and $B\,\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0)$ is the homotopy pullback of the same diagram. Once again, we aim to show that Q retracts off $B\,\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0)$.

Lemma 2.30 *The pullback Q is homeomorphic to $U(2n)/\operatorname{Sp}(n)$.*

Proof This is essentially the same proof as Lemma 2.16, but we must elaborate on the details for surjectivity of the map

$$f'\colon U(2n)/\operatorname{Sp}(n)\to Q,\quad A\operatorname{Sp}(n)\mapsto A\hat{\varsigma}_{\mathrm{Q}}(A)^{-1}.$$

It can be shown that a matrix A is in Q if and only if AJ is skew-symmetric, and hence due to the Youla lemma [14], there is a unitary matrix P such that $AJ=PJP^t$. Therefore,

$$A=PJP^tJ^{-1}=P(J^{-1}\bar{P}J)^{-1}=f'(P\operatorname{Sp}(n)),$$

and the result follows. □

Similar to the map in (6), we obtain the composition

(19) $$\varphi\colon U(2n)/\operatorname{Sp}(n)\xrightarrow{f'}Q\to B\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0)\xrightarrow{r}U(2n)\xrightarrow{q}U(2n)/\operatorname{Sp}(n),$$

where q is the quotient map. The map φ sends an element $A\operatorname{Sp}(n)$ to the element $A\hat{\varsigma}_{\mathrm{Q}}(A)^{-1}\operatorname{Sp}(n)$. It was shown in [5] that the related map

(20) $$s'\colon \mathrm{SU}(2n)/\operatorname{Sp}(n)\to\mathrm{SU}(2n)/\operatorname{Sp}(n),\quad A\operatorname{Sp}(n)\mapsto A\hat{\varsigma}_{\mathrm{Q}}(A)^{-1}\operatorname{Sp}(n),$$

is a homotopy equivalence when localised at a prime $p\neq 2$.

Clearly, there are analogue statements to Lemmas 2.18 and 2.19 and Proposition 2.21.

Lemma 2.31 *There is a homotopy equivalence*

$$\eta\colon U(2n)/\operatorname{Sp}(n)\times S^1\xrightarrow{\simeq}U(2n)/\operatorname{Sp}(n).$$ □

Lemma 2.32 *There exist maps*

$$s''\colon \mathrm{SU}(n)/\mathrm{SO}(n)\to\mathrm{SU}(n)/\mathrm{SO}(n)\quad\textit{and}\quad s'\colon S^1\to S^1$$

such that the following is a homotopy commuting square:

$$\begin{tikzcd}
\mathrm{SU}(2n)/\operatorname{Sp}(n)\times S^1 \arrow[r, "s''\times s'"] \arrow[d, "\eta"] & \mathrm{SU}(2n)/\operatorname{Sp}(n)\times S^1 \arrow[d, "\eta"] \\
U(2n)/\operatorname{Sp}(n) \arrow[r, "s"] & U(2n)/\operatorname{Sp}(n)
\end{tikzcd}$$

Furthermore, s'' and s' are p–local equivalences. □

Proposition 2.33 *Let F be the homotopy fibre of the composition*

$$B\mathscr{G}^*_{\mathrm{Q}}((0,0,1);0)\xrightarrow{r}U(2n)\xrightarrow{q}U(2n)/\operatorname{Sp}(n).$$

Then for any prime $p \neq 2$, there is a p–local homotopy equivalence

$$F \simeq_p \Omega(U(2n)/\operatorname{Sp}(n)). \qquad \square$$

Proof of Theorem 1.20(1) For a prime $p \neq 2$, we have shown that there is a p–local section to the principal homotopy fibration

$$\Omega^2(U(2n)/\operatorname{Sp}(n)) \to \mathscr{G}^*_{\mathrm{Q}}((0,0,1);0) \xrightarrow{\Omega(qr)} \Omega(U(2n)/\operatorname{Sp}(n)),$$

and the result follows. $\square$

Proof of Theorem 1.20(2) and (3) These follow using the same proofs as Theorems 2.22 and 2.23. $\square$

In the unpointed case, the theorems involving integral decompositions follow immediately from the real case.

Proof of Theorem 1.21 The results presented in Section 2.3.1 do not depend on the fixed point set of the involution on $BU(n)$, and hence Theorem 1.21 follows immediately. $\square$

We proceed to prove the quaternionic analogues of Section 2.4.2. Let

$$B = \begin{cases} B\operatorname{Sp}(n) & \text{if } r > 0, \\ BU(2n) & \text{otherwise,} \end{cases}$$

and recall the evaluation fibration

$$\Omega B \xrightarrow{\partial_P} \operatorname{Map}^*_{\mathrm{Q}}(X, BU(n); P) \to \operatorname{Map}_{\mathrm{Q}}(X, BU(n); P) \to B. \tag{21}$$

The following proposition can be proven using the same method as Proposition 2.26.

Proposition 2.34 *Fix $d \in \mathbb{Z}$ and let ∂_d be the boundary map in (13). Denote by*

$$\partial_P \colon \Omega B \to B\mathscr{G}^*_{\mathrm{Q}}((g,r,a);2d)$$

the boundary map of the evaluation fibration as in (21). If ∂_d is (q–locally) trivial, then

(1) *if $r > 0$, then ∂_P is (q–locally) trivial;*

(2) *if $r = 0$, then the following composition is (q–locally) trivial:*

$$\operatorname{Sp}(n) \hookrightarrow U(2n) \xrightarrow{\partial_P} B\mathscr{G}^*_{\mathrm{Q}}((g,r,a);2d). \qquad \square$$

Proof of Theorem 1.22(1) Let p be a prime such that $p \nmid 2n$. Then by Theorem 2.25, the map ∂_{2n} is p–locally trivial. The result then follows from Proposition 2.34. $\square$

real	$\pi_0(\mathscr{G}^{*(r+a)}(P,\tilde{\sigma}))$	$\pi_0(\mathscr{G}(P,\tilde{\sigma}))$	$\pi_1(\mathscr{G}^{*(r+a)}(P,\tilde{\sigma}))$	$\pi_1(\mathscr{G}(P,\tilde{\sigma}))$
$n>2$	$\mathbb{Z}^{g+a}\times(\mathbb{Z}_2)^r$	$\mathbb{Z}^g\times(\mathbb{Z}_2)^{r+1}$	$\underline{\mathbb{Z}}$	$\underline{\mathbb{Z}}\times(\mathbb{Z}_2)^r$
$n=2$	$\mathbb{Z}^{g+a+r}$	$\mathbb{Z}^{g+r}\times\mathbb{Z}_2$	$\mathbb{Z}$	$\mathbb{Z}^{r+1}$
$n=1$	$\mathbb{Z}^{g+a}$	$\mathbb{Z}^g\times\mathbb{Z}_2$	0	0
quat. rank $2n$	$\mathbb{Z}^{g+a}$	$\mathbb{Z}^g\times(\mathbb{Z}_2)^a$	$\underline{\mathbb{Z}}$	$\underline{\mathbb{Z}}$

Table 1: Results of [3]: the low-dimensional homotopy groups of rank n gauge groups above a real surface of type (g,r,a). The underlined entries disagree with the author's results.

Proof of Theorem 1.22(2) and (3) This is similar to the proofs of Theorem 1.15(1b) and (1c). We do require that $p\neq 2$, but this is automatic with the assumption that $p\nmid 2n$. $\square$

3 Tables of homotopy groups

We present homotopy groups of the $(r+a)$–pointed and unpointed gauge groups. We only present these for the trivial components, that is,

- $(c,w_1,\dots,w_r)=(0,0,\dots,0)$ for real bundles;
- $c=0$ for quaternionic bundles;

with the understanding that results can be obtained for different components using the results in Section 1. Specifically, in the $(r+a)$–pointed case, we can obtain results for all of the components using Propositions 1.7 and 1.16, and in the unpointed case, we can obtain results for some of the different components using Propositions 1.9, 1.10 and 1.17.

We first recall the status of the calculation of the homotopy groups before this paper; that is, we present the low-dimensional homotopy groups from [3] in Table 1.

From the results in Sections 1.2 and 1.3, we can see that our homotopy decompositions usually contain factors involving $U(n)$, $O(n)$ and $\mathrm{Sp}(n)$. Due to Bott periodicity, it is easy to calculate some of the higher homotopy groups for high-rank gauge groups. We present such results in Tables 2 and 3 where η is defined via

$$\eta=\eta(g,r,a)=\begin{cases}1 & \text{if } r>0 \text{ and } a=1,\\ 0 & \text{otherwise.}\end{cases}$$

Some of the results in Table 2 are a consequence of localised homotopy equivalences and hence may provide incomplete information. To highlight these localised results we use the following notation:

	$\mathcal{G}^{*(r+a)}(P,\widetilde{\sigma})$	$\mathcal{G}(P,\widetilde{\sigma})$
π_{8j}	$\mathbb{Z}^{g-1}\times\mathbb{Z}_2^{r-1}\times(\mathbb{Z}^{1+a})_p\times(\mathbb{Z}_2^{1+\eta})_p$	$\mathbb{Z}^{g-1}\times\mathbb{Z}_2^{r-1}\times(\mathbb{Z})_p\times(\mathbb{Z}_2^{1+\eta})_p$
π_{8j+1}	$(\mathbb{Z}_2^{1+a})_p$	$\mathbb{Z}_2^{r-1}\times(\mathbb{Z}_2^{2+\eta})_p$
π_{8j+2}	$\mathbb{Z}^{g+r-2}\times(\mathbb{Z}^{1+\eta})_p\times(\mathbb{Z}_2^{a})_p$	$\mathbb{Z}^{g-1}\times\mathbb{Z}_2^{r-1}\times(\mathbb{Z})_p\times(\mathbb{Z}_2^{\eta})_p$
π_{8j+3}	$(\mathbb{Z})_p$	$(\mathbb{Z}^2)_p$
π_{8j+4}	$\mathbb{Z}^{g-1}\times(\mathbb{Z}^{1+a})_p$	$\mathbb{Z}^{g-1}\times(\mathbb{Z})_p$
π_{8j+5}	0	0
π_{8j+6}	$\mathbb{Z}^{g+r-2}\times(\mathbb{Z}^{1+\eta})_p$	$\mathbb{Z}^{g-1}\times(\mathbb{Z}^{1-\eta})_p$
π_{8j+7}	$\mathbb{Z}_2^{r-1}\times(\mathbb{Z})_p\times(\mathbb{Z}_2^{\eta})_p$	$\mathbb{Z}_2^{r-1}\times(\mathbb{Z}^2)_p\times(\mathbb{Z}_2^{\eta})_p$

Table 2: Homotopy groups for high-rank gauge groups of real bundles, that is, the homotopy groups π_i when the rank $n > i+2$. The results in the first two rows correspond to the top row in Table 1.

- groups surrounded by $(-)_p$ are understood to have come from p–local homotopy equivalences where p and the rank n of the gauge groups satisfy the requirements of Theorems 1.13 and 1.15.

Similarly, some of the results in Table 3 are a consequence of localised homotopy equivalences and hence may provide incomplete information. To highlight these localised results we use the following notation:

- groups surrounded by $(-)_p$ are understood to have come from p–local homotopy equivalences where p is prime and the rank $2n$ of the gauge groups satisfy the requirements of Theorems 1.20 and 1.22.

Due to the properties of Bott periodicity, Table 3 is a translation of Table 2. We note that additional calculations can be made for the lower-rank cases. We point the reader to [7, Section 3.2] where explicit homotopy groups of some of the relevant factors can be found.

We note that the author's results disagree with the $\mathbb{Z}$–summands underlined in Table 1. In the pointed case, this $\mathbb{Z}$–summand arises in [3] by studying a fibration arising from restricting the gauge group to the 1–skeleton of the real surface.

For example, the corresponding fibration for a type-$(g,r,0)$ real surface is

$$\Omega^2U(n)\to\mathcal{G}^*(P,\widetilde{\sigma})\to\prod_1^g\Omega U(n)\times\prod_1^r\Omega O(n),$$

	$\mathscr{G}_{\mathrm{Q}}^{*(r+a)}(P,\widetilde{\sigma})$	$\mathscr{G}_{\mathrm{Q}}(P,\widetilde{\sigma})$
π_{8j}	$\mathbb{Z}^{g-1}\times(\mathbb{Z}^{1+a})_p$	$\mathbb{Z}^{g-1}\times(\mathbb{Z})_p$
π_{8j+1}	0	0
π_{8j+2}	$\mathbb{Z}^{g+r-2}\times(\mathbb{Z}^{1+\eta})_p$	$\mathbb{Z}^{g-1}\times(\mathbb{Z}^{1-\eta})_p$
π_{8j+3}	$\mathbb{Z}_2^{r-1}\times(\mathbb{Z})_p\times(\mathbb{Z}_2^{\eta})_p$	$\mathbb{Z}_2^{r-1}\times(\mathbb{Z}^2)_p\times(\mathbb{Z}_2^{\eta})_p$
π_{8j+4}	$\mathbb{Z}^{g-1}\times\mathbb{Z}_2^{r-1}\times(\mathbb{Z}^{1+a})_p\times(\mathbb{Z}_2^{1+\eta})_p$	$\mathbb{Z}^{g-1}\times\mathbb{Z}_2^{r-1}\times(\mathbb{Z})_p\times(\mathbb{Z}_2^{1+\eta})_p$
π_{8j+5}	$(\mathbb{Z}_2^{1+a})_p$	$\mathbb{Z}_2^{r-1}\times(\mathbb{Z}_2^{2+\eta})_p$
π_{8j+6}	$\mathbb{Z}^{g+r-2}\times(\mathbb{Z}^{1+\eta})_p\times(\mathbb{Z}_2^{a})_p$	$\mathbb{Z}^{g-1}\times\mathbb{Z}_2^{r-1}\times(\mathbb{Z})_p\times(\mathbb{Z}_2^{\eta})_p$
π_{8j+7}	$(\mathbb{Z})_p$	$(\mathbb{Z}^2)_p$

Table 3: Homotopy groups for high-rank gauge groups of quaternionic bundles, that is, the homotopy groups π_i when the rank $2n > \frac{1}{4}(i+1)$. The results in the first two rows correspond to the bottom row in Table 1.

and we obtain the exact sequence

$$0 \to \pi_2(\mathscr{G}^*(P,\widetilde{\sigma})) \xrightarrow{\nu} \mathbb{Z}^{g+r} \to \mathbb{Z} \xrightarrow{\mu} \pi_1(\mathscr{G}^*(P,\widetilde{\sigma})) \to 0.$$

The claim in [3] is that the map μ can be thought in terms of the classification of bundles over $S^2 \wedge X$. Further, since μ is induced by a map that collapses the 1–skeleton of X, the map μ is essentially providing an identification of the second Chern class, and hence is an isomorphism.

The author agrees that this argument holds in the nonequivariant case. Indeed, if we consider X as a Riemann surface, we obtain that $S^2 \wedge X$ is a wedge of spheres, and then μ is induced by a map that collapses all but the top copy of S^4.

However, we now demonstrate that $\pi_1(\mathscr{G}^*(P,\widetilde{\sigma}))$ cannot contain a $\mathbb{Z}$–summand, at least for the type-$(0,1,0)$ case. We assume that $\pi_1(\mathscr{G}^*(P,\widetilde{\sigma}))$ contains a $\mathbb{Z}$–summand, and that subsequently the map μ is an isomorphism. Therefore, ν is an isomorphism, and we recall that it is induced by the map r' which restricts to the 1–skeleton of X. The map r' fits into the commutative diagram

$$\begin{tikzcd}
\mathscr{G}^*(P,\widetilde{\sigma}) \arrow[r, "u'"] \arrow[d, "r'"] & \Omega\,\mathrm{Map}^*(D^2, BU(n)) \arrow[d, "r"] \\
\Omega\,\mathrm{Map}^*_{\mathbb{Z}_2}((S^1,\mathrm{id}), BU(n);0) \arrow[r, "u"] & \Omega\,\mathrm{Map}^*(S^1, BU(n))
\end{tikzcd}$$

where u' is the map that forgets about equivariance and restricts to the upper hemisphere of X.

Now u is homotopic to the inclusion $\Omega O(n) \hookrightarrow \Omega U(n)$, and hence, by assumption, the induced map

$$u_* \nu = (ur')_* \colon \mathbb{Z} \cong \pi_2(\mathscr{G}^*(P, \tilde{\sigma})) \to \pi_2(\Omega U(n)) \cong \mathbb{Z}$$

is multiplication by 2. But ru' is nullhomotopic because it factors through the contractible space $\Omega \operatorname{Map}^*(D^2, BU(n))$, and we obtain a contradiction. We conclude that μ cannot be an isomorphism.

It remains to show that the other underlined entries in Table 1 cannot contain $\mathbb{Z}$–summands. However, these entries were obtained from the calculation in the pointed case, and therefore, we argue that these cannot contain $\mathbb{Z}$–summands either.

References

[1] **T Baird**, *Moduli spaces of vector bundles over a real curve: $\mathbb{Z}/2$–Betti numbers*, Canad. J. Math. 66 (2014) 961–992 MR

[2] **T J Baird**, *Cohomology of the moduli space of rank two, odd degree vector bundles over a real curve*, Symmetry Integrability Geom. Methods Appl. 12 (2016) art. id. 072, 18 pages MR

[3] **I Biswas**, **J Huisman**, **J Hurtubise**, *The moduli space of stable vector bundles over a real algebraic curve*, Math. Ann. 347 (2010) 201–233 MR

[4] **P Georgieva**, **A Zinger**, *On the topology of real bundle pairs over nodal symmetric surfaces*, Topology Appl. 214 (2016) 109–126 MR

[5] **B Harris**, *On the homotopy groups of the classical groups*, Ann. of Math. 74 (1961) 407–413 MR

[6] **C-C M Liu**, **F Schaffhauser**, *The Yang–Mills equations over Klein surfaces*, J. Topol. 6 (2013) 569–643 MR

[7] **M Mimura**, *Homotopy theory of Lie groups*, from "Handbook of algebraic topology" (I M James, editor), North-Holland, Amsterdam (1995) 951–991 MR

[8] **F Schaffhauser**, *Moduli spaces of vector bundles over a Klein surface*, Geom. Dedicata 151 (2011) 187–206 MR

[9] **J-P Serre**, *Groupes d'homotopie et classes de groupes abéliens*, Ann. of Math. 58 (1953) 258–294 MR

[10] **W A Sutherland**, *Function spaces related to gauge groups*, Proc. Roy. Soc. Edinburgh Sect. A 121 (1992) 185–190 MR

[11] **S D Theriault**, *Odd primary homotopy decompositions of gauge groups*, Algebr. Geom. Topol. 10 (2010) 535–564 MR

[12] **S D Theriault**, *Homotopy decompositions of gauge groups over Riemann surfaces and applications to moduli spaces*, Internat. J. Math. 22 (2011) 1711–1719 MR

[13] **G Weichold**, *Über symmetrische Riemann'sche Flächen und die Periodicitätsmoduln der zugehörigen Abel'schen Normalintegrale erster Gattung*, Dissertation, Universität Leipzig (1883)

[14] **D C Youla**, *A normal form for a matrix under the unitary congruence group*, Canad. J. Math. 13 (1961) 694–704 MR

University of Southampton
Building 54, Salisbury Rd, Southampton, SO17 1BJ, United Kingdom

michaelwtwest@gmail.com

Received: 20 June 2016 Revised: 2 November 2016

Algebraic & Geometric Topology 17 (2017) 2481–2498

Coarse medians and Property A

JÁN ŠPAKULA
NICK WRIGHT

We prove that uniformly locally finite quasigeodesic coarse median spaces of finite rank and at most exponential growth have Property A. This offers an alternative proof of the fact that mapping class groups have Property A.

20F65; 30L05

1 Introduction

Coarse median spaces and groups were invented by Bowditch [4; 5; 6] as (we are guessing here) a device offering a unified approach to hyperbolic groups and mapping class groups.

Indeed, hyperbolic groups are precisely coarse median groups of rank 1 [4, Theorem 2.1], and mapping class groups are instances of coarse median groups of finite rank [4, Theorem 2.5].

Furthermore, groups that are relatively hyperbolic with respect to a collection of coarse median groups are again coarse median [5]. This provides more examples of coarse median groups, for instance geometrically finite Kleinian groups and Sela's limit groups.

The coarse median approach to these classes of groups is quite powerful: in this series of papers, Bowditch uses it to give unified proofs of some properties, for instance the rapid decay property and quadratic isoperimetric inequality, and to compute the dimension of asymptotic cones.

Intuitively, a coarse median space is a metric space endowed with a ternary structure (a map assigning a point to every triple of points), which is metrically a controlled amount away from being an actual median structure. (Finite) sets with an actual median structure are just (vertex sets of) CAT(0) cube complexes. Hence one may loosely regard coarse median structures as coarse versions of (metrized) CAT(0) cube complexes. This analogy works exactly in the "rank one" situation, where the CAT(0) cube complexes are trees, and hyperbolic groups are "coarsely tree-like". For the actual definitions, see Section 2.

Published: 3 August 2017   DOI: 10.2140/agt.2017.17.2481

The main result of this piece is that quasigeodesic coarse median spaces of finite rank, which are uniformly locally finite and have at most exponential growth, have Yu's Property A. For proving Property A we use a criterion which is an adaptation of Brown and Ozawa's proof [9] that hyperbolic groups act amenably on the boundary. As a side-effect, we obtain a quick proof of Property A for finite-dimensional CAT(0) cube complexes, a fact originally established by Brodzki, Campbell, Guentner, Niblo and Wright [8] by a different, more combinatorial method. Our proof for coarse median spaces is a coarsification of this short argument.

As a consequence, we obtain an alternative proof of the result that mapping class groups have Property A (ie are exact), originally proved by Hamenstädt [12] and Kida [15].

Finally, we would like to mention a related notion of *hierarchically hyperbolic spaces* (and groups), developed recently by Behrstock, Hagen and Sisto [1; 3]. While this property is stronger (see [1, Section 7]), and somewhat more involved than coarse medians, it is also substantially more powerful: it implies even finite asymptotic dimension; see Behrstock, Hagen and Sisto [2]. Having finite asymptotic dimension is a strictly stronger property than Property A. We close off with a question: do coarse median groups of finite rank have finite asymptotic dimension?

The structure of the paper is as follows: In Section 2 we recall the relevant definitions and facts. Section 3 explains Brown and Ozawa's criterion for Property A. In Section 4 we outline the quick proof of Property A for CAT(0) cube complexes. In Section 5 we establish some facts about (metric) median algebras; and finally Section 6 contains the proof of the main result.

Acknowledgements Špakula thanks Goulnara Arzhantseva for her encouragement, continuing support, and the initial impetus for this work.

2 Preliminaries

2.1 CAT(0) cube complexes

We recall the notions related to CAT(0) cube complexes. For details, please consult [7; 16].

A *cube complex* is a polyhedral complex in which the cells are Euclidean cubes of side length one, the attaching maps are isometries identifying the faces of a given cube with cubes of lower dimension and the intersection of two cubes is a common face of each. One-dimensional cubes are called *edges*; and the complex is *finite-dimensional* if there is a bound on the dimension of its cubes.

Recall that we can endow a cube complex with a naturally defined *geodesic metric*. Furthermore, we can endow the set of vertices of a cube complex with an *edge-path* metric; in the finite-dimensional case, this metric is coarsely equivalent to (the restriction of) the geodesic metric [8, Proposition 1.7].

A cube complex is a CAT(0) *cube complex* if the underlying topological space is simply connected and the complex satisfies Gromov's *link condition* [10]. In the finite-dimensional case, this is equivalent to asking that the geodesic metric should satisfy the CAT(0) inequality [7].

A *hyperplane* H (or a *wall*) is a geometric hyperplane which cuts each cube that it intersects exactly in half. Such an H divides the vertex set into two path-connected subspaces, which are referred to as *half-spaces*. Two hyperplanes *cross* if each of the four possible intersections of the associated half-spaces is nonempty. We say that H *separates* two vertices if every edge-path connecting them crosses H. For two sets of vertices A and B, we shall write $A\,|_H\,B$ if H separates every vertex in A from every vertex in B, ie A and B are in different half-spaces determined by H. The *interval* $[x,y]$ between two vertices x and y is the intersection of all half-spaces containing both vertices.

Every n–dimensional cube in a CAT(0) cube complex defines n pairwise intersecting hyperplanes (which it *crosses*) and, conversely, a collection of n pairwise intersecting hyperplanes defines a unique n–cube (which crosses exactly these hyperplanes).

Note that the set of vertices of a CAT(0) cube complex is a median algebra in the sense defined below — the median of three points x, y, z is the unique vertex in the intersection $[x,y]\cap[y,z]\cap[z,x]$; see [19]. Equivalently the median of x, y, z is the unique point lying on a geodesic between x and y, on a geodesic between y and z and a geodesic between z and x. Furthermore, the notions of an interval, wall, etc, are the same whether defined as here, or using the median structure (below).

In a CAT(0) cube complex, each collection of pairwise intersecting hyperplanes determines a unique cube and, conversely, each cube (of dimension k) provides k pairwise intersecting hyperplanes. A *cube path* from a vertex x to a vertex y in a CAT(0) cube complex X is a sequence of cubes $C_0,\dots,C_n$ such that x is a vertex of C_0, y is a vertex of C_n, and every two consecutive cubes intersect in exactly one vertex. A *normal cube path* from x to y is a cube path from x to y such that every hyperplane separating x and y is crossed exactly once, with the maximal number of hyperplanes crossed at each step [16]. Note that if X is finite-dimensional, then $\frac{1}{d}\rho(x,y)\le n\le\rho(x,y)$, where d is the dimension of X and ρ denotes the edge-path distance. We also refer to the sequence of the common vertices between the consecutive cubes on the normal cube path as the normal cube path.

2.2 Metric median algebras

We summarise the notions that we need for this paper. For a more thorough account on median structures, we refer to [4; 6]. The median algebras can be thought of as an abstraction of CAT(0) cube complexes — every finite median algebra *is* actually the vertex set of a finite CAT(0) cube complex. While one direction of this link works in general, median algebras can be "larger" (for instance $\mathbb{R}$–trees are also median algebras).

A *median algebra* is a set, Φ, equipped with a ternary operation, $\mu\colon \Phi^3 \to \Phi$, such that for all $a, b, c, d, e \in \Phi$ we have

(M1) $\mu(a,b,c) = \mu(b,c,a) = \mu(b,a,c)$,

(M2) $\mu(a,a,b) = a$, and

(M3) $\mu(a,b,\mu(c,d,e)) = \mu(\mu(a,b,c), \mu(a,b,d), e)$.

While this is the formal definition, we prefer to think about finite median algebras as the vertex sets of finite CAT(0) cube complexes (with the natural median structure).

Given $a, b \in \Phi$, the *interval* $[a,b]$ is defined to be $[a,b] = \{c \in \Phi \mid \mu(a,b,c) = c\}$. A subset $H \subset \Phi$ is *convex* if $[a,b] \subset H$ for all $a, b \in H$.

For $A \subset \Phi$, define the *convex hull*, denoted by $\mathrm{hull}(A)$, to be the smallest convex subset of Φ containing A. Note that $\mathrm{hull}(\{a,b\}) = [a,b]$ for $a, b \in \Phi$. Furthermore, for $A \subset \Phi$, define the *join*, $J(A) = \bigcup_{a,b\in A}[a,b]$. Continuing inductively, we put $J^0(A) = J(A)$ and $J^i(A) = J(J^{i-1}(A))$. In general, there always exists some $p \in \mathbb{N}$ such that $J^p(A) = \mathrm{hull}(A)$, and moreover we know that p can be taken to be no larger than the rank of Φ; see [4, Lemma 5.5].

A *wall*, W, is a partition $\{H^-(W), H^+(W)\}$ of Φ into two nonempty convex subsets. We say that two walls W and W' *cross* if each of the sets $H^-(W) \cap H^-(W')$, $H^-(W) \cap H^+(W')$, $H^+(W) \cap H^-(W')$ and $H^+(W) \cap H^+(W')$ is nonempty.

We say that Φ has *rank at most d* if there is no collection of $d+1$ pairwise crossing walls of Φ.

By a *topological median algebra* we mean a topological space Φ endowed with a structure of a median algebra $\mu\colon \Phi^3 \to \Phi$ such that μ is continuous in the induced topology. When the topology on Φ comes from a metric ρ, we say that Φ is a *metric median algebra*.

Let Φ be a metric median algebra as above. We also recall one of the conditions used in [6] to obtain the embedding result:

(L2) There exists $K \geq 1$ such that for all $a, b, c, d \in \Phi$, $\rho(\mu(a,b,c), \mu(a,b,d)) \leq K\rho(c,d)$.

Let us mention that the main embedding result of [6] states that if a metric median algebra Φ satisfies (L2), is Lipschitz path-connected and is ν–colourable,[1] then it bilipschitzly embeds into a product of ν $\mathbb{R}$–trees.

2.3 Coarse median spaces

In this subsection, we recall the definitions and facts related to coarse medians. For more details, we refer to [4; 6].

Let (X,ρ) be a metric space, and let $\mu\colon X^3 \to X$ be a ternary operation. We say that μ is a *coarse median* and that (X,ρ,μ) is a *coarse median space* if the following conditions hold:

(C1) There are constants $K \geq 1$ and $H(0) \geq 0$ such that for all $a,b,c,a',b',c' \in X$ we have

$$\rho(\mu(a,b,c),\mu(a',b',c')) \leq K(\rho(a,a')+\rho(b,b')+\rho(c,c'))+H(0).$$

(C2) There is a function $H\colon \mathbb{N} \to [0,\infty)$ with the following property: Suppose that $A \subseteq X$ with $1 \leq |A| \leq p < \infty$. Then there is a finite median algebra (Π,μ_Π) and maps $\pi\colon A \to \Pi$ and $\sigma\colon \Pi \to X$ such that for all $x,y,z \in \Pi$ we have

$$\rho(\sigma\mu_\Pi(x,y,z),\mu(\sigma x,\sigma y,\sigma z)) \leq H(p)$$

and

$$\rho(a,\sigma\pi a) \leq H(p)$$

for all $a \in A$.

We refer to K and H as the *parameters* of (X,ρ,μ).

Without loss of generality, we may assume that μ satisfies the axioms (M1) and (M2), by [4, page 22].

We say that X has *rank at most* d if we can always choose Π to have rank at most d.

Let us recall the asymptotic cones from [4, Section 9; 6, Section 8]: Let (X,ρ,μ) be a coarse median space, let (r_n) be a sequence of positive reals such that $r_n \to \infty$, let $(x_n) \subset X$ be a sequence of points in X, and finally fix a nonprincipal ultrafilter on $\mathbb{N}$. With this data, we can construct an ultralimit $(X_\infty,\rho_\infty,\mu_\infty)$ of pointed coarse median spaces $((X,\rho/r_n,\mu),x_n)$. This ultralimit is referred to as an *asymptotic cone* of X (with the given data), and it is a complete[2] metric median algebra satisfying (L2) (with the constant K being the same as in the definition of coarse median). Moreover, if X has rank at most d, then X_∞ also has rank at most d.

[1] This is a more restrictive version of rank, which is equivalent to rank for intervals.

[2] The completeness here refers to the metric.

2.4 Property A

Property A is a coarse geometric property of metric spaces (or more generally coarse spaces), first defined by G Yu [21] as a criterion that (for discrete countable groups) implies the coarse Baum–Connes conjecture, and hence the Novikov conjecture. The catchphrase here is "nonequivariant amenability" or "coarse amenability". Since its inception, many equivalent formulations were discovered, including analytic (exactness of the reduced group C^*–algebra, nuclearity of the uniform Roe algebra [11; 18]) and dynamical (admitting an amenable action on a compact topological space [13]).

We shall recall one of the possible definitions (the one used in Proposition 3.1) for completeness; we refer to [20] for the whole spectrum.

Let (X,ρ) be a uniformly locally finite discrete metric space. We say that X has *Property A* if for all $R,\varepsilon>0$ there exists a map $\xi\colon X\to\ell^1(X)$ from X into the Banach space $\ell^1(X)$ such that

- $\|\xi(x)\|_1=1$ for all $x\in X$;
- for all $x,y\in X$ with $\rho(x,y)\le R$, we have $\|\xi(x)-\xi(y)\|_1\le\varepsilon$;
- there exists $S>0$ such that $\xi(x)$ is supported in the closed ball $B(x;S)$ around x with radius S for each $x\in X$.

2.5 Geodesicity

We shall say that a metric space (X,ρ) is *quasigeodesic* if there exist constants G_1 and G_2 such that there exists a (G_1,G_2)*–quasigeodesic* between any pair of points in X. Note that when X is a quasigeodesic coarse median space, all its asymptotic cones are Lipschitz path-connected. This is required for applying the embedding result of Bowditch [6] (and is a blanket assumption in [4; 6]).

3 A criterion

We extract a criterion from a proof of Brown and Ozawa [9, Theorem 5.3.15] for proving Property A. Its proof is just an excerpt from [9], which is in turn inspired by [14].

Proposition 3.1 *Let X be a uniformly finite, discrete metric space. Suppose that we have an assignment of a set $S(x,k,l)\subset X$ to every $l\in\mathbb{N}$, $k\in\{1,\dots,3l\}$ and $x\in X$ such that:*

(i) *For every $l\in\mathbb{N}$ there exists $S_l>0$ such that $S(x,k,l)\subset B(x,S_l)$ for all $x\in X$ and $k\in\{1,\dots,3l\}$.*

(ii) *For every $x, y \in X$, $l \geq \rho(x,y)$ and $k \in \{l+1, \dots, 2l\}$, we have inclusions $S(x, k-\rho(x,y), l) \subset S(x,k,l) \cap S(y,k,l)$ and $S(x,k,l) \cup S(y,k,l) \subset S(x, k+\rho(x,y), l)$.*

(iii) *There exists a function p such that $|S(x,k,l)| \leq p(l)$ for every $x \in X$, $l \in \mathbb{N}$ and $k \in \{1, \dots, 3l\}$ with $\lim_{n\to\infty} p(n)^{1/n} = 1$.*

Then X has Property A.

To have some mental picture, let us recall that Brown and Ozawa apply this criterion to hyperbolic groups Γ, defining the sets as follows: fix a point $u \in \partial\Gamma$. Given x, k and l, the set $S(x,k,l)$ consists of points that are exactly $3l$ steps along a geodesic between a point within the k–ball around x and u. With this definition, the conditions (i) and (ii) follow from the triangle inequality, and (iii) uses the stability of geodesics in hyperbolic spaces (in this case p can be taken to be a linear function).

Proof Consider the Banach space $\ell^1(X)$ and for $A \subset X$ denote by $\chi_A \in \ell^1(X)$ the normalised characteristic function of A. Given $n \in \mathbb{N}$ and $x \in X$, define

$$\xi_n(x) = \frac{1}{n} \sum_{k=n+1}^{2n} \chi_{S(x,k,n)}.$$

Note that $\|\xi_n(x)\| = 1$ and $\operatorname{supp}(\xi_n(x)) \subset B(x, S_n)$ for all $x \in X$ by (i).

To establish Property A, we use the formulation from [20, Theorem 1.2.4(2)], recalled also in Section 2.4. We need to show that, for a fixed m, we have

$$\lim_{n\to\infty} \sup_{\rho(x,y)=m} \|\xi_n(x) - \xi_n(y)\| = 0.$$

First, observe that for any $A, B \subset X$, we have

$$\|\chi_A - \chi_B\| = 2\left(1 - \frac{|A \cap B|}{\max\{|A|, |B|\}}\right) \leq 2\left(1 - \frac{|A \cap B|}{|A \cup B|}\right).$$

Take $x, y \in X$ with $\rho(x,y) = m$ and assume $n \geq 2m$. Then for any $k \in \{n+1, \dots, 2n\}$, applying (ii),

$$\|\chi_{S(x,k,n)} - \chi_{S(y,k,n)}\| \leq 2\left(1 - \frac{|S(x,k-m,n)|}{|S(x,k+m,n)|}\right).$$

Consequently,

$$\|\xi_n(x) - \xi_n(y)\| \leq \frac{1}{n} \sum_{k=n+1}^{2n} \|\chi_{S(x,k,n)} - \chi_{S(y,k,n)}\|$$

$$
\begin{aligned}
&\le 2\Bigl(1-\frac{1}{n}\sum_{k=n+1}^{2n}\frac{|S(x,k-m,n)|}{|S(x,k+m,n)|}\Bigr)\\
&\le 2\Bigl(1-\Bigl(\prod_{k=n+1}^{2n}\frac{|S(x,k-m,n)|}{|S(x,k+m,n)|}\Bigr)^{1/n}\Bigr)\\
&= 2\Bigl(1-\Bigl(\frac{\prod_{j=n+1-m}^{n+m}|S(x,j,n)|}{\prod_{j=2n+1-m}^{2n+m}|S(x,j,n)|}\Bigr)^{1/n}\Bigr)\\
&\le 2(1-p(n)^{-2m/n}).
\end{aligned}
$$

We have used the inequality between the arithmetic and geometric mean in the middle step, magic cancellation of many terms in the penultimate step, and the last step uses (iii) plus a simple estimate of the sizes of sets by 1 from below. By (iii), the last expression converges to 0 as n converges to ∞. We are done. □

Remark 3.2 In condition (iii), we ask for a bound in terms of l. However, it is apparent from the proof that a bound in terms of k with an analogous property also suffices.

Remark 3.3 It is clear from the proof of the proposition that we need to define the sets $S(x,k,l)$ only for an infinite sequence of indices l (and the corresponding $k\in\{1,\dots,3l\}$), not necessarily for all $l\in\mathbb{N}$.

4 CAT(0) cube complexes

Proposition 3.1 allows us to quickly prove that finite-dimensional $\mathrm{CAT}(0)$ cube complexes have Property A. This was first proved in [8] using a more combinatorial approach.

Proposition 4.1 *Let X be a finite-dimensional* $\mathrm{CAT}(0)$ *cube complex. Then X has Property A.*

Proof Fix a base vertex $x_0\in X$. Given a vertex $x\in X$, $l\in\mathbb{N}$ and $k\in\{1,\dots,3l\}$, consider the normal cube path from y to x_0, where $\rho(y,x)\le k$. We define the set $S(x,k,l)$ to contain the $3l^{\text{th}}$ vertex on such a normal cube path (or x_0 if we run out of space). Note that conditions (i) and (ii) from Proposition 3.1 are automatically satisfied, courtesy of the triangle inequality. To be more precise, if $z\in S(x,k,l)$, then $\rho(x,z)\le 6ld$, where $d=\dim(X)$.

To prove condition (iii) of Proposition 3.1, we shall argue that if $z\in S(x,k,l)$, then $z\in[x,x_0]$. Or, equivalently:

Claim *Every half-space containing both x and x_0 contains also z.*

Each hyperplane that we need to consider (ie such that one of the associated half-spaces contains both x and x_0) either separates x and x_0 from y or it does not. In the latter case, the same half-space also clearly contains z, so it remains to deal with the former case.

Denote by $C_0, C_1, \dots, C_m$ the normal cube path from y to x_0, and denote by $y = v_0, v_1, \dots, v_m = x_0$ the vertices on this cube path. We shall argue that any hyperplane separating y from x and x_0 is "used" within the first $\rho(x, y)$ steps on the cube path. Suppose that the cube C_i does not cross any hyperplane H with $\{y\} \,|_H\, \{x, x_0\}$. Hence every hyperplane K crossing C_i satisfies $\{y, x\} \,|_K\, \{x_0, v_{i+1}\}$. If there was a hyperplane L with $\{y\} \,|_L\, \{x, x_0\}$ which was not "used" before C_i on the cube path, then necessarily $\{y, v_{i+1}\} \,|_L\, \{x, x_0\}$, hence L crosses all the hyperplanes K crossing C_i. This contradicts the maximality of this step on the normal cube path. Thus there is no such L, and so all the hyperplanes H with $\{y\} \,|_H\, \{x, x_0\}$ must be crossed within the first $\rho(x, y)$ steps (as there at most $\rho(x, y)$ such hyperplanes).

Since z is the $3l^{\text{th}}$ vertex on the cube path and $\rho(x, y) \le k \le 3l$, all the hyperplanes H with $\{y\} \,|_H\, \{x, x_0\}$ must have been crossed before z. Thus any such H actually also satisfies $\{y\} \,|_H\, \{x, x_0, z\}$. We have proved our claim.

Coming back to showing condition (iii) of Proposition 3.1, observe that the interval $[x, x_0]$ embeds isometrically into the cube complex $\mathbb{R}^d$ [8, Theorem 1.14], hence it has polynomial growth. This means that as $S(x, k, l) \subset [x, x_0] \cap B(x, 6ld)$, its cardinality is bounded by a polynomial in l (of degree d). This finishes the proof. □

For the record, we note that dropping the finite-dimensionality assumption renders the statement false, namely infinite-dimensional CAT(0) cube complexes do not have Property A; this follows from [17], as they contain isometric copies of $(\mathbb{Z}/2\mathbb{Z})^n$ for arbitrarily large n.

5 Median algebras

Definition 5.1 Let Φ be a median algebra. Let $n \ge 2$ and $x_1, \dots, x_n, b \in \Phi$. Define

$$\mu(x_1; b) := x_1$$

and inductively, for $1 \le k < n-1$,

$$\mu(x_1, \dots, x_{k+1}; b) := \mu(\mu(x_1, \dots, x_k; b), x_{k+1}, b).$$

Note that this definition "agrees" with the original median map μ, since $\mu(x_1, x_2; b) = \mu(x_1, x_2, b)$.

Intuitively, $\mu(x_1,\dots,x_n;b)$ should be thought of as a projection of b onto the set $\operatorname{hull}\{x_1,\dots,x_n\}$, just as $\mu(x_1,x_2,b)$ is the projection of b onto $[x_1,x_2]$. However, we do not prove this in this note (but see Lemma 5.3).

Lemma 5.2 *The μ symbol from Definition 5.1 is symmetric in $x_1,\dots,x_n$.*

Proof Recalling that interchanging the points in $\mu(\cdot,\cdot,\cdot)$ is one of the axioms of a median algebra, it is clearly sufficient to prove that, for $n=3$, we can switch x_2 with x_3. However, applying axioms of median algebras, we have that

$$\begin{aligned}\mu(x_1,x_2,x_3;b) &= \mu(\mu(x_1,x_2,b),x_3,b)\\ &= \mu(\mu(x_3,b,x_1),\mu(x_3,b,b),x_2)\\ &= \mu(\mu(x_1,x_3,b),x_2,b)\\ &= \mu(x_1,x_3,x_2;b).\end{aligned}$$

We are done. □

In fact, it is easy to see that $[x_1,b]\cap[x_2,b]=[\mu(x_1,x_2,b),b]$ and then by induction that $\bigcap_{k=1}^n[x_k,b]=[\mu(x_1,\dots,x_n;b),b]$.

Lemma 5.3 *Let Φ be a median algebra. Let $x_1,\dots,x_n,b\in\Phi$.*

(i) *A wall separates $x_1,\dots,x_n$ from b if and only if it separates $\mu(x_1,\dots,x_n;b)$ from b.*

(ii) *If $\mu(x_1,\dots,x_{n-1};b)\neq\mu(x_1,\dots,x_n;b)$ then there exists a wall separating $x_1,\dots,x_{n-1}$ from x_n and b.*

(iii) *If, in addition, Φ has rank at most d, then there exists $\{y_1,\dots,y_k\}\subseteq\{x_1,\dots,x_n\}$ with $k\le d$ such that $\mu(y_1,\dots,y_k;b)=\mu(x_1,\dots,x_n;b)$.*

(iv) *If $a\in\Phi$ and $x_1,\dots,x_n\in[a,b]$, then $\{x_1,\dots,x_n\}\subset[a,\mu(x_1,\dots,x_n;b)]$.*

Proof Since $\mu(x,y,b)\in[x,y]=J(\{x,y\})$, we can easily prove by induction that $\mu(x_1,\dots,x_n;b)\in J^{n-1}(\{x_1,\dots,x_n\})\subset\operatorname{hull}\{x_1,\dots,x_n\}$. The "only if" statement in (i) follows. For the converse, assume for contradiction that there exists a wall W that separates b from $\mu(x_1,\dots,x_n;b)$ but does not separate b from (say, using Lemma 5.2) x_n. As half-spaces are convex, the whole interval $[b,x_n]$ is in the same half-space as b. But as $\mu(\cdot,x_n,b)\in[b,x_n]$, this contradicts the assumption that b is separated from $\mu(x_1,\dots,x_n;b)=\mu(\mu(x_1,\dots,x_{n-1};b),x_n,b)$.

For (ii), write $c=\mu(x_1,\dots,x_{n-1};b)$ and note that $\mu(x_1,\dots,x_n;b)=\mu(c,x_n,b)$. Hence $c\neq\mu(x_1,\dots,x_n;b)$ implies $c\notin[x_n,b]$. As $\{c\}$ and $[x_n,b]$ are convex, this implies, by [4, Lemma 6.1], that there is a wall separating c from x_n and b. By (i), this wall separates $x_1,\dots,x_{n-1}$ from x_n and b.

For (iii), we proceed by contradiction. Assume that there are at least $d+1$ points in $\{x_1,\dots,x_n\}$ which cannot be removed from the expression $\mu(x_1,\dots,x_n;b)$ without changing the result. The previous part of the lemma, together with Lemma 5.2, implies that there exist at least $d+1$ different walls which all intersect (for instance, if one wall separates $x_1,\dots,x_{n-1}$ from x_n and b, and another one separates $x_1,\dots,x_{n-2},x_n$ from x_{n-1} and b, they clearly intersect). This contradicts the rank assumption (see [4, Proposition 6.2]).

Part (iv) follows by induction from the following statement: if $x,y\in[a,b]$, then $x,y\in[a,\mu(x,y,b)]$ (so in particular $[a,x]\subset[a,\mu(x,y,b)]$). It is of course sufficient to prove that $x\in[a,\mu(x,y,b)]$, which is done using median axioms as follows:

$$\mu(a,x,\mu(x,y,b))=\mu(\mu(a,x,x),\mu(a,x,b),y)=\mu(x,x,y)=x.$$

We are done. □

Lemma 5.4 *Let Φ be a median algebra and let $a,b,x,y\in\Phi$ satisfy $x,y\in[a,b]$. Then $y\in[a,x]$ implies $x\in[y,b]$.*

Proof We compute

$$\mu(y,b,x)=\mu(\mu(a,x,y),b,x)=\mu(\mu(b,x,a),\mu(b,x,x),y)=\mu(x,x,y)=x,$$

using the median axioms and the lemma's assumptions. □

Proposition 5.5 *Let Φ be a topological median algebra of rank at most d. Given an interval $[a,b]\subset\Phi$ and a compact set $C\subset[a,b]$, there exists $h_1,\dots,h_d\in C$, such that $C\subset[a,\mu(h_1,\dots,h_d;b)]$.*

If Φ is moreover a metric median algebra satisfying the condition (L2), then we have $\rho(a,\mu(h_1,\dots,h_d;b))\le 3^dK^d\max_{1\le i\le d}\rho(a,h_i)\le 3^dK^d\sup_{h\in C}\rho(a,h)$.

Proof Consider the compact space C^d. Given a tuple $\xi\in C^d$, write $\mu(\xi;b)$ for the short. Given $\xi\in C^d$, define

$$A_\xi=\{\eta\in C^d\mid \mu(\eta;b)\in[\mu(\xi;b),b]\}=\{\eta\in C^d\mid [a,\mu(\xi;b)]\subset[a,\mu(\eta;b)]\}.$$

Note that the two conditions are equivalent: if $\mu(\eta;b)\in[\mu(\xi;b),b]$, then by Lemma 5.4 $\mu(\xi;b)\in[a,\mu(\eta;b)]$, hence $[a,\mu(\xi;b)]\subset[a,\mu(\eta;b)]$. Conversely, the last inclusion implies $\mu(\xi;b)\in[a,\mu(\eta;b)]$, so again by Lemma 5.4 we have $\mu(\eta;b)\in[\mu(\xi;b),b]$.

Observe that each set A_ξ is closed: in fact it is exactly the inverse image of the closed[3] set $[\mu(\xi;b),b]\subset\Phi$ under the continuous map $\mu(\cdot\,;b)\colon C^d\to\Phi$.

Finally, the collection $\{A_\xi \mid \xi\in C^d\}$ of subsets of C^d has the finite intersection property. Given $\xi,\eta\in C^d$, by Lemma 5.3(iii) there is $\omega\in C^d$ such that $\mu(\omega;b)=\mu(\xi\cup\eta;b)$. However, from the definition of μ and Lemma 5.2, we know that $\mu(\xi\cup\eta;b)$ belongs to both $[\mu(\xi;b),b]$ and $[\mu(\eta;b),b]$. In other words, $\omega\in A_\xi\cap A_\eta$.

Now, as C^d is compact, there exists $\zeta\in\bigcap_{\xi\in C^d}A_\xi$. Thus $[a,\mu(\xi;b)]\subset[a,\mu(\zeta;b)]$ for all $\xi\in C^d$. In particular, by Lemma 5.3(iv), $\xi\subset[a,\mu(\xi;b)]\subset[a,\mu(\zeta;b)]$ for all $\xi\in C^d$, so $[a,\mu(\zeta;b)]$ contains all the points of C. Now just enumerate ζ as $h_1,\dots,h_d$.

For the second part of the proposition, we do inductive estimates using (L2). Write $T=\max_{1\le i\le d}\rho(a,h_i)$. Then, as the first step, $\rho(a,\mu(h_1;b))=\rho(a,h_1)\le T$. We show inductively that $\rho(a,\mu(h_1,\dots,h_i;b))\le 3^{i-1}K^{i-1}T$. Assuming this inequality for i, writing $\mu(h_1,\dots,h_i;b)=g_i$ we estimate

$$\begin{aligned}\rho(a,\mu(h_1,\dots,h_{i+1};b)) &= \rho(a,\mu(g_i,h_{i+1};b))\\ &\le \rho(a,g_i)+\rho(\mu(g_i,g_i,b),\mu(g_i,h_{i+1},b))\\ &\le 3^{i-1}K^{i-1}T+K\rho(g_i,h_{i+1})\\ &\le 3^{i-1}K^{i-1}T+K(3^{i-1}K^{i-1}T+T)\\ &= T(3^{i-1}K^i+3^{i-1}K^{i-1}+K)\le 3^iK^iT.\end{aligned}$$

We are done. □

6 Coarse medians

We shall adapt the idea of "moving deep into the interval" from the CAT(0) cube complex setting to the more general coarse median spaces.

To explain the idea, consider two points a and b and the context-appropriate notion of the interval $[a,b]$. In the CAT(0) cube complex case, we have moved deep into this interval by stepping sufficiently far along the cube path from a to b. In the coarse median case we shall, roughly speaking, be looking for "the other end" of the convex hull of $B(a,l)\cap[a,b]$ (see the second bullet in Corollary 6.3). Along the lines of [6], this is done by going to the asymptotic cone (where the results of Section 5 can be applied).

We begin by fixing a fair amount of notation.

[3] Intervals are closed in topological median algebras; this just uses continuity of μ.

For the rest of this section, when we say that X is a coarse median space, we mean that X is a coarse median space, with metric denoted by ρ, the median function denoted by μ, and with parameters K and H. These will be fixed throughout.

When convenient, we shall be using the notation $x \sim_s y$ for $\rho(x,y) \le s$.

Given $\tau \ge 0$ and $a, b \in X$, we shall denote by $[a,b]_\tau$ the coarse interval between a and b, ie $[a,b]_\tau := \{x \in X \mid \mu(a,b,x) \sim_\tau x\}$.

We will denote by $\lambda \ge 0$ a constant (depending only on K and H) such that for all $x, y, z \in X$ we have $\mu(x,y,z) \in [x,y]_\lambda$; its existence is proved in [6, Lemma 9.2].

Recall that since the median axiom (M3) holds exactly in median algebras, it does hold in coarse median spaces up to a constant $\gamma \ge 0$ depending only on the parameters K and H of the coarse median structure (actually $\gamma = 3K(3K+2)H(5)+(3K+2)H(0)$). By this we mean $\mu(x,y,\mu(z,v,w)) \sim_\gamma \mu(\mu(x,y,z),\mu(x,y,v),w)$ for all $x,y,z,v,w \in X$. We shall be using γ and this fact throughout this section.

Fixing some more notation, given $r, t \ge 0$, let

$$\begin{aligned} L_1(r) &= (K+1)r + K\lambda + \gamma + 2H(0), \\ L_2(r) &= (K+2)r + H(0), \\ L_3(r,t) &= 3^d K^d rt + r. \end{aligned}$$

The point is that L_1 and L_2 are linear functions of r, and L_3 is linear in r with t fixed, and bounded by a linear function of rt (for $t \ge 1$).

Lemma 6.1 *Let X be a coarse median space, $r \ge 0$, and let $a, b \in X$ and $x \in [a,b]_\lambda$. Then $[a,x]_r \subset [a,b]_{L_1(r)}$.*

Proof Let $z \in [a,x]_r$. Thus $\mu(a,x,z) \sim_r z$ and by assumption also $\mu(a,b,x) \sim_\lambda x$. Hence,

$$\begin{aligned} \mu(a,b,z) &\sim_{Kr+H(0)} \mu(a,b,\mu(a,x,z)) \\ &\sim_\gamma \mu(\mu(a,b,a),\mu(a,b,x),z) \\ &\sim_{K\lambda+H(0)} \mu(a,x,z) \sim_r z. \end{aligned}$$

Thus $\rho(\mu(a,b,z),z) \le (K+1)r + K\lambda + \gamma + 2H(0) = L_1(r)$, which means by definition that $z \in [a,b]_{L_1(r)}$. □

In what follows, r can be thought of as "a scale" and t as "a distance". In other words, the statements can read as "given a distance (t) at which we want the space to behave, there exists a scale (r_t) such that on all larger scales ($r \ge r_t$) it behaves as a median space, with an error proportional to r".

Proposition 6.2 *Let X be a quasigeodesic coarse median space of rank at most d. For every $\kappa > 0$ and $t > 0$, there exists $r_t > 0$ such that for all $r \geq r_t$, $a, b \in X$ and $A \subset B(a, rt) \cap [a,b]_\kappa$ with* $\operatorname{sep}(A) \geq r$, *there exists $h \in [a,b]_{L_1(r)}$ such that*

- *$\rho(a,h) \leq L_3(r,t)$, and*
- *$A \subset [a,h]_r$.*

Corollary 6.3 *Let X be a quasigeodesic coarse median space of rank at most d. For every $\kappa > 0$ and $t > 0$, there exists $r_t > 0$ such that for all $r \geq r_t$ and $a, b \in X$ there exists $h \in [a,b]_{L_1(r)}$ such that*

- *$\rho(a,h) \leq L_3(r,t)$, and*
- *$B(a, rt) \cap [a,b]_\kappa \subset [a,h]_{L_2(r)}$.*

Proof This readily follows from Proposition 6.2, by noting that we may choose A to be a maximal r–separated subset of $B(a,rt) \cap [a,b]_\kappa$. Then any point $x \in B(a,rt) \cap [a,b]_\kappa$ is at most r–far from a point $a_x \in A$, hence the condition $A \subset [a,h]_r$ implies that

$$\mu(a,h,x) \sim_{Kr+H(0)} \mu(a,h,a_x) \sim_r a_x \sim_r x.$$

Since we wrote $Kr + H(0) + r + r = L_2(r)$, the above reads $x \in [a,h]_{L_2(r)}$. $\square$

Proof of Proposition 6.2 We proceed by contradiction: Suppose that for some κ and t the statement is not true, ie there exists a sequence $0 < r_1 < r_2 < \cdots \in \mathbb{R}$ and for each $n \in \mathbb{N}$ there exist $a_n, b_n \in X$ and $A_n \subset [a_n, b_n]_\kappa \cap B(a_n, r_n t)$ with $\operatorname{sep}(A_n) \geq r_n$ such that for all $h \in [a_n, b_n]_{L_1(r_n)}$ with $\rho(a_n, h) \leq L_3(r_n, t)$ there exists $x \in A_n$ with $\rho(x, \mu(a_n, h, x)) > r_n$.

It follows from [6, Lemma 9.7] that we can assume that the cardinalities $|A_n|$ are uniformly bounded by a constant p depending on K, H, d, κ and t.

The next step is to argue that we can arrange that the distances from a_n to b_n are linear in r_n.

Claim *There exist constants $\delta_1, \delta_2, \kappa_1 \geq 0$ (depending only on K, H, κ, t and p) and points $b_n' \in [a_n, b_n]_\lambda$ such that $\rho(a_n, b_n') \leq \delta_1 r_n + \delta_2$ and $A_n \subset [a_n, b_n']_{\kappa_1}$.*

The claim follows from [6, Lemma 9.6], which says that in our situation there are constants ζ, ξ and κ' (depending only on K, H, κ, t and p) and points $c_n, d_n \in X$ such that $A_n \subset [c_n, d_n]_{\kappa'}$ and $\operatorname{diam}(A_n \cup \{c_n, d_n\}) \leq \zeta \operatorname{diam}(A_n) + \xi \leq 2\zeta r_n t + \xi$. Since $A_n \subset B(a_n, r_n t)$, by the proof of that lemma we can assume that $c_n = a_n$ for every n. Finally, we define $b_n' = \mu(a_n, b_n, d_n) \in [a_n, b_n]_\lambda$ and check that

$$b_n' = \mu(a_n, b_n, d_n) \sim_{K(2\zeta r_n t + \xi) + H(0)} \mu(d_n, b_n, d_n) = d_n \sim_{2\zeta r_n t + \xi} a_n,$$

and, for every $x \in A_n$ (so that $\mu(a_n, b_n, x) \sim_\kappa x$ and $\mu(a_n, d_n, x) \sim_{\kappa'} x$),

$$\begin{aligned}\mu(a_n, b'_n, x) &= \mu(a_n, \mu(a_n, b_n, d_n), x)\\ &\sim_\gamma \mu(\mu(a_n, x, b_n), \mu(a_n, x, d_n), a_n)\\ &\sim_{K(\kappa+\kappa')+H(0)} \mu(x, x, a_n) = x.\end{aligned}$$

So, altogether, we put $\kappa_1 = K(\kappa+\kappa') + H(0) + \gamma$, $\delta_1 = 2(K+1)\zeta t$ and $\delta_2 = (K+1)\xi + H(0)$ and the claim is proved.

We have now set up the situation so that we can conclude the proof by going to the asymptotic cone.

Let $(X_\infty, \rho_\infty, \mu_\infty)$ be an asymptotic cone of X, with the sequence of scales (r_n), basepoints (a_n) and any nonprincipal ultrafilter on $\mathbb{N}$.

The sequences (a_n) and (b'_n) determine points $a, b \in X_\infty$ (with $\rho_\infty(a,b) \le \delta_1$), and the intervals $[a_n, b'_n]_{\kappa_1}$ converge to the interval $[a,b]$ in X_∞. Also the sets A_n converge to a (finite, 1–separated) set $A \subset [a,b] \cap B(a,t)$.

By Proposition 5.5, there exists $h \in [a,b]$ such that $A \subset B(a,t) \cap [a,b] \subset [a,h]$ and $\rho_\infty(a,h) \le 3^d K^d t$. This implies that we have a sequence of points $h_n \in X$, eventually in $[a_n, b'_n]_{r_n}$,[4] such that $\lim \rho(a_n, h_n)/r_n \le 3^d K^d t$, thus eventually $\rho(a_n, h_n) \le 3^d K^d r_n t + r_n = L_3(r_n, t)$.

Since $b'_n \in [a_n, b_n]_\lambda$ and $h_n \in [a_n, b'_n]_{r_n}$, by Lemma 6.1 we have that $h_n \in [a_n, b_n]_{L_1(r_n)}$. Hence, by our original assumption, there (eventually) exist points $x_n \in A_n$ with $\rho(x_n, \mu(a_n, h_n, x_n)) > r_n$. The sequence of x_n yields a point $x \in A$ such that $\rho_\infty(x, \mu_\infty(a, h, x)) \ge 1$. This point witnesses that $A \not\subset [a,h]$, which is a contradiction. □

Lemma 6.4 *Let X be a coarse median space. There exist constants $\alpha, \beta \ge 0$ (depending only on the parameters of the coarse median structure) such that the following holds: Let $a, b, h, m \in X$ and $r \ge 0$ satisfy $m \in [a,h]_{L_2(r)}$ and $h \in [a,b]_{L_1(r)}$. Then $p = \mu(m, b, h)$ satisfies $\rho(h, p) \le \alpha r + \beta$.*

Proof Note that the assumptions say that $m \sim_{L_2(r)} \mu(a, h, m)$ and $h \sim_{L_1(r)} \mu(a, b, h)$. We estimate

$$\begin{aligned}p = \mu(m, b, h) &\sim_{KL_2(r)+H(0)} \mu(\mu(a, h, m), b, h)\\ &\sim_\gamma \mu(\mu(b, h, h), \mu(b, h, a), m)\\ &\sim_{KL_1(r)+H(0)} \mu(h, h, m) = h.\end{aligned}$$

Altogether, $\rho(h, p) \le K(L_1(r) + L_2(r)) + 2H(0) + \gamma$, which is a linear function of r. □

[4] Since $\mu(a, b, h) = h$, we have $\rho(h_n, \mu(a_n, b'_n, h_n))/r_n \to 0$, hence the claim.

Theorem 6.5 *Let X be a uniformly locally finite, quasigeodesic, of at most exponential growth, coarse median space of finite rank. Then X has Property A.*

Proof The proof follows the idea of our proof for CAT(0) cube complexes, which relies on Proposition 3.1. We shall verify its assumptions. Let $\alpha, \beta > 0$ be the constants from Lemma 6.4 and fix a basepoint $x_0 \in X$.

We now apply Corollary 6.3 for $\kappa = \lambda$ and all $t \in \mathbb{N}$ to obtain a sequence $r_t \in \mathbb{N}$, so that the conclusion of the corollary holds. Furthermore, we can choose the r_t inductively to arrange that the sequence $t \mapsto l_t = (tr_t - H(0))/(3K)$ for $t \in \mathbb{N}$ is increasing.

For a moment, fix $x \in X$, $t \in \mathbb{N}$ and $k \in \{1, \dots, 3l_t\}$. For every $y \in B(x,k)$, Corollary 6.3 applied to $a = y$, $b = x_0$ and $r = r_t$ produces for us a point $h_y \in [y, x_0]_{L_1(r_t)}$. We collect these points into the set

$$S(x,k,l_t) = \{h_y \in X \mid y \in B(x,k)\}.$$

Loosely speaking, the set $S(x,k,l_t)$ contains one point associated to each $y \in B(x,k)$, which should be thought of as being "tr_t–deep" inside the interval $[y, x_0]_\lambda$.

Defined like this, condition (ii) of Proposition 3.1 is automatic, (i) follows from the first bullet of Corollary 6.3, and finally (iii) requires some checking:

Take $y \in B(x,k)$, with notation as above. Write $m_y = \mu(x, y, x_0)$. Then $m_y \in [y, x_0]_\lambda$ and

$$\rho(y, m_y) \le K\rho(x,y) + H(0) \le K \cdot 3l_t + H(0) = tr_t.$$

Thus the second bullet of Corollary 6.3 implies $m_y \in [y, h_y]_{L_2(r_t)}$. Since we also know that $h_y \in [y, x_0]_{L_1(r_t)}$, Lemma 6.4 implies that the point $p_y = \mu(m_y, x_0, h_y) \in [m_y, x_0]_\lambda$ satisfies $\rho(h_y, p_y) \le \alpha r_t + \beta$. As $m_y = \mu(x, y, x_0) \in [x, x_0]_\lambda$, Lemma 6.1 implies $p_y \in [x, x_0]_{L_1(\lambda)}$.

To summarise, for each $h_y \in S(x,k,l_t)$ we can associate a point $p_y \in [x, x_0]_{L_1(\lambda)}$ satisfying $\rho(h_y, p_y) \le \alpha r_t + \beta$, and consequently also

$$\rho(x, p_y) \le \rho(x,y) + \rho(y, h_y) + \rho(h_y, p_y) \le 3l_t + 3^d K^d tr_t + r_t + \alpha r_t + \beta,$$

which clearly depends linearly on l_t. Hence, by [6, Proposition 9.8], the number of possible points p_y is bounded by $P(l_t)$ for some polynomial P (depending only on H, K, d and uniform local finiteness of X). Since we assume at most exponential growth of X, it follows that the cardinality of $S(x,k,l_t)$ is at most $P(l_t)c'c^{r_t}$ for some constants $c, c' \ge 1$. Finally, as $l_t \to \infty$ means by definition also $t \to \infty$, it is easy to see that also $r_t/l_t \to 0$, thus $(P(l_t)c'c^{r_t})^{1/l_t} \to 1$. This finishes the proof of condition (iii) of Proposition 3.1 and we are done. □

References

[1] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, II: Combination theorems and the distance formula*, preprint (2015) arXiv

[2] **J Behrstock**, **M F Hagen**, **A Sisto**, *Asymptotic dimension and small-cancellation for hierarchically hyperbolic spaces and groups*, Proc. Lond. Math. Soc. 114 (2017)

[3] **J Behrstock**, **M Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, Geom. Topol. 21 (2017) 1731–1804 MR

[4] **B H Bowditch**, *Coarse median spaces and groups*, Pacific J. Math. 261 (2013) 53–93 MR

[5] **B H Bowditch**, *Invariance of coarse median spaces under relative hyperbolicity*, Math. Proc. Cambridge Philos. Soc. 154 (2013) 85–95 MR

[6] **B H Bowditch**, *Embedding median algebras in products of trees*, Geom. Dedicata 170 (2014) 157–176 MR

[7] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR

[8] **J Brodzki**, **S J Campbell**, **E Guentner**, **G A Niblo**, **N J Wright**, *Property A and* CAT(0) *cube complexes*, J. Funct. Anal. 256 (2009) 1408–1431 MR

[9] **N P Brown**, **N Ozawa**, *C^*–algebras and finite-dimensional approximations*, Graduate Studies in Mathematics 88, Amer. Math. Soc., Providence, RI (2008) MR

[10] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR

[11] **E Guentner**, **J Kaminker**, *Exactness and the Novikov conjecture*, Topology 41 (2002) 411–418 MR

[12] **U Hamenstädt**, *Geometry of the mapping class groups, I: Boundary amenability*, Invent. Math. 175 (2009) 545–609 MR

[13] **N Higson**, **J Roe**, *Amenable group actions and the Novikov conjecture*, J. Reine Angew. Math. 519 (2000) 143–153 MR

[14] **V A Kaimanovich**, *Boundary amenability of hyperbolic spaces*, from "Discrete geometric analysis" (M Kotani, T Shirai, T Sunada, editors), Contemp. Math. 347, Amer. Math. Soc., Providence, RI (2004) 83–111 MR

[15] **Y Kida**, *Classification of the mapping class groups up to measure equivalence*, Proc. Japan Acad. Ser. A Math. Sci. 82 (2006) 4–7 MR

[16] **G A Niblo**, **L D Reeves**, *The geometry of cube complexes and the complexity of their fundamental groups*, Topology 37 (1998) 621–633 MR

[17] **P W Nowak**, *Coarsely embeddable metric spaces without Property A*, J. Funct. Anal. 252 (2007) 126–136 MR

[18] **N Ozawa**, *Amenable actions and exactness for discrete groups*, C. R. Acad. Sci. Paris Sér. I Math. 330 (2000) 691–695 MR

[19] **M A Roller**, *Poc sets, median algebras and group actions: an extended study of Dunwoody's construction and Sageev's theorem*, preprint (1998) Available at http://www.personal.soton.ac.uk/gan/Roller.pdf

[20] **R Willett**, *Some notes on property A*, from "Limits of graphs in group theory and computer science" (G Arzhantseva, A Valette, editors), EPFL Press, Lausanne (2009) 191–281 MR

[21] **G Yu**, *The coarse Baum–Connes conjecture for spaces which admit a uniform embedding into Hilbert space*, Invent. Math. 139 (2000) 201–240 MR

Department of Mathematical Sciences, University of Southampton
Southampton, SO17 1BJ, United Kingdom

jan.spakula@soton.ac.uk, n.j.wright@soton.ac.uk

Received: 15 July 2016

Algebraic & Geometric Topology 17 (2017) 2499–2510

Geometric embedding properties of Bestvina–Brady subgroups

HUNG CONG TRAN

We compute the relative divergence of right-angled Artin groups with respect to their Bestvina–Brady subgroups and the subgroup distortion of Bestvina–Brady subgroups. We also show that for each integer $n \geq 3$, there is a free subgroup of rank n of some right-angled Artin group whose inclusion is not a quasi-isometric embedding. The corollary answers the question of Carr about the minimum rank n such that some right-angled Artin group has a free subgroup of rank n whose inclusion is not a quasi-isometric embedding. It is well known that a right-angled Artin group A_Γ is the fundamental group of a graph manifold whenever the defining graph Γ is a tree with at least three vertices. We show that the Bestvina–Brady subgroup H_Γ in this case is a horizontal surface subgroup.

20F65, 20F67; 20F36

1 Introduction

For each Γ a finite simplicial graph, the associated *right-angled Artin group* A_Γ has generating set S the vertices of Γ, and relations $st = ts$ whenever s and t are adjacent vertices. If Γ is nonempty, there is a homomorphism from A_Γ onto the integers that takes every generator to 1. The *Bestvina–Brady subgroup* H_Γ is defined to be the kernel of this homomorphism.

Bestvina–Brady subgroups were introduced by Bestvina and Brady [2] to study the finiteness properties of subgroups of right-angled Artin groups. One result in [2] is that the Bestvina–Brady subgroup H_Γ is finitely generated if and only if the graph Γ is connected. This fact is a motivation to study the geometric connection between a right-angled Artin group and its Bestvina–Brady subgroup. More precisely, we examine the relative divergence of right-angled Artin groups with respect to their Bestvina–Brady subgroups and the subgroup distortion of Bestvina–Brady subgroups.

Theorem 1.1 *Let Γ be a connected, finite, simplicial graph with at least two vertices. Let A_Γ be the associated right-angled Artin group and H_Γ the Bestvina–Brady subgroup. Then the relative divergence $\mathrm{Div}(A_\Gamma, H_\Gamma)$ and the subgroup distortion $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ are both linear if Γ is a join graph. Otherwise, they are both quadratic.*

Published: 3 August 2017　　DOI: 10.2140/agt.2017.17.2499

In the above theorem, we can see that the relative divergence $\mathrm{Div}(A_\Gamma, H_\Gamma)$ and the subgroup distortion $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ are equivalent. In general, we show that the relative divergence is always dominated by the subgroup distortion for any pair of finitely generated groups (G, H), where H is a normal subgroup of G such that the quotient group G/H is an infinite cyclic group see Proposition 4.3.

Carr [3] proved that nonabelian two-generator subgroups of right-angled Artin groups are quasi-isometrically embedded free groups. In his paper, he also showed an example of a distorted free subgroup of a right-angled Artin group. However, the minimum rank n such that some right-angled Artin group has a free subgroup of rank n whose inclusion is not a quasi-isometric embedding was still unknown; see [3]. The following corollary of Theorem 1.1 answers this question.

Corollary 1.2 *For each integer $n \geq 3$, there is a right-angled Artin group containing a free subgroup of rank n whose inclusion is not a quasi-isometric embedding.*

We remark that a special case of Theorem 1.1 can also be derived as a consequence of previous work by Hruska and Nguyen [6] on distortion of surfaces in graph manifolds. They showed that every virtually embedded horizontal surface in a 3–dimensional graph manifold has quadratic distortion. This led us to prove the following theorem, which implies that many Bestvina–Brady subgroups are also horizontal surface subgroups.

Theorem 1.3 *If Γ is a finite tree with at least three vertices, then the associated right-angled Artin group A_Γ is a fundamental group of a graph manifold, and the Bestvina–Brady subgroup H_Γ is a horizontal surface subgroup.*

It is well known that a right-angled Artin group A_Γ is the fundamental group of a graph manifold whenever the defining graph Γ is a tree with at least three vertices. However, the fact that the Bestvina–Brady subgroup H_Γ is a horizontal subgroup does not seem to be recorded in the literature. With the use of Theorem 1.3, we see that Theorem 1.1 can be viewed as a generalization of a special case of the quadratic distortion theorem of Hruska and Nguyen. Moreover, Theorem 1.3 combined with the Hruska–Nguyen theorem gives an alternative proof of Corollary 1.2.

Acknowledgements I would like to thank Professor Christopher Hruska and Hoang Nguyen for very helpful comments and suggestions, especially their help on the proof of Theorem 1.3. Hoang Nguyen helped me with the background of graph manifolds and horizontal surfaces. He also showed me his joint work with Hruska on subgroup distortion of horizontal surface subgroups, and their work is a motivation for this article. I also thank the referee for advice that improved the exposition of the paper.

2 Right-angled Artin groups and Bestvina–Brady subgroups

Definition 2.1 Given a finite simplicial graph Γ, the associated *right-angled Artin group* A_Γ has generating set S the vertices of Γ, and relations $st = ts$ whenever s and t are adjacent vertices.

Let S_1 be a subset of S. The subgroup of A_Γ generated by S_1 is a right-angled Artin group A_{Γ_1}, where Γ_1 is the induced subgraph of Γ with vertex set S_1 (ie Γ_1 is the union of all edges of Γ with both endpoints in S_1). The subgroup A_{Γ_1} is called a *special subgroup* of A_Γ.

Definition 2.2 Let Γ be a finite simplicial graph with the set S of vertices. Let T be a torus of dimension $|S|$ with edges labeled by the elements of S. Let X_Γ denote the subcomplex of T consisting of all faces whose edge labels span a complete subgraph in Γ (or equivalently, mutually commute in A_Γ). X_Γ is called the *Salvetti complex*.

Remark 2.3 The fundamental group of X_Γ is A_Γ. The universal cover $\widetilde{X}_\Gamma$ of X_Γ is a CAT(0) cube complex with a free, cocompact action of A_Γ. Obviously, the 1–skeleton of $\widetilde{X}_\Gamma$ is the Cayley graph of A_Γ with respect to the generating set S.

Definition 2.4 Let Γ be a finite simplicial graph. Let $\Phi\colon A_\Gamma \to \mathbb{Z}$ be an epimorphism which sends all the generators of A_Γ to 1 in $\mathbb{Z}$. The kernel H_Γ of Φ is called the *Bestvina–Brady subgroup*.

Remark 2.5 There is a natural continuous map $f\colon X_\Gamma \to S^1$ which induces the homomorphism $\Phi\colon A_\Gamma \to \mathbb{Z}$. Moreover, it is not hard to see that the lifting map $\tilde{f}\colon \widetilde{X}_\Gamma \to \mathbb{R}$ is an extension of Φ.

Theorem 2.6 (Bestvina and Brady [2] and Dicks and Leary [4]) *Let Γ be a finite simplicial graph. The Bestvina–Brady subgroup H_Γ is finitely generated if and only if Γ is connected. Moreover, the set T of all elements of the form st^{-1} whenever s and t are adjacent vertices form a finite generating set for H_Γ. Furthermore, if Γ is a tree with n edges, then the Bestvina–Brady subgroup H_Γ is a free group of rank n.*

Definition 2.7 Let Γ_1 and Γ_2 be two graphs; the *join* of Γ_1 and Γ_2 is a graph obtained by connecting every vertex of Γ_1 to every vertex of Γ_2 by an edge.

Let J be a complete subgraph of Γ which decomposes as a nontrivial join. We call A_J a *join subgroup* of A_Γ.

Let Γ be a finite simplicial graph with vertex set S, and let g an element of A_Γ. A *reduced word* for g is a minimal-length word in the free group $F(S)$ representing g.

Given an arbitrary word representing g, one can obtain a reduced word by a process of "shuffling" (ie interchanging commuting elements) and canceling inverse pairs. Any two reduced words for g differ only by shuffling. For an element $g \in A_\Gamma$, a cyclic reduction of g is a minimal-length element of the conjugacy class of g. If w is a reduced word representing g, then we can find a cyclic reduction $\overline{g}$ by shuffling commuting generators in w to get a maximal-length word u such that $w = u\overline{w}u^{-1}$. In particular, g itself is *cyclically reduced* if and only if every shuffle of w is cyclically reduced as a word in the free group $F(S)$.

3 Relative divergence, geodesic divergence and subgroup distortion

Before we define the concepts of relative divergence, geodesic divergence and subgroup distortion, we need to build the tools to measure them, namely the notions of domination and equivalence.

Definition 3.1 Let $\mathcal{M}$ be the collection of all functions from $[0,\infty)$ to $[0,\infty]$. Let f and g be arbitrary elements of $\mathcal{M}$. *The function f is dominated by the function g*, denoted by $f \preceq g$, if there are positive constants A, B, C and D such that $f(x) \leq Ag(Bx) + Cx$ for all $x > D$. Two functions f and g are *equivalent*, denoted by $f \sim g$, if $f \preceq g$ and $g \preceq f$.

Remark 3.2 A function f in $\mathcal{M}$ is *linear, quadratic or exponential* if f is respectively equivalent to a degree-one polynomial, a degree-two polynomial or a function of the form a^{bx+c}, where $a > 1$ and $b > 0$.

Definition 3.3 Let $\{\delta^n_\rho\}$ and $\{\delta'^n_\rho\}$ be two families of functions of $\mathcal{M}$, indexed over $\rho \in (0,1]$ and positive integers $n \geq 2$. *The family $\{\delta^n_\rho\}$ is dominated by the family $\{\delta'^n_\rho\}$*, denoted by $\{\delta^n_\rho\} \preceq \{\delta'^n_\rho\}$, if there exists a constant $L \in (0,1]$ and a positive integer M such that $\delta^n_{L\rho} \preceq \delta'^{Mn}_\rho$ for all ρ and n. Two families $\{\delta^n_\rho\}$ and $\{\delta'^n_\rho\}$ are *equivalent*, denoted by $\{\delta^n_\rho\} \sim \{\delta'^n_\rho\}$, if $\{\delta^n_\rho\} \preceq \{\delta'^n_\rho\}$ and $\{\delta'^n_\rho\} \preceq \{\delta^n_\rho\}$.

Remark 3.4 A family $\{\delta^n_\rho\}$ is dominated by (or dominates) a function f in $\mathcal{M}$ if $\{\delta^n_\rho\}$ is dominated by (or dominates) the family $\{\delta'^n_\rho\}$ where $\delta'^n_\rho = f$ for all ρ and n. The equivalence between a family $\{\delta^n_\rho\}$ and a function f in $\mathcal{M}$ can be defined similarly. Thus, a family $\{\delta^n_\rho\}$ is linear, quadratic, exponential, etc if $\{\delta^n_\rho\}$ is equivalent to a function f with said property.

Definition 3.5 Let X be a geodesic space and A a subspace of X. Let r be any positive number.

(1) $N_r(A) = \{x \in X \mid d_X(x, A) < r\}$.
(2) $\partial N_r(A) = \{x \in X \mid d_X(x, A) = r\}$.
(3) $C_r(A) = X - N_r(A)$.
(4) Let $d_{r,A}$ be the induced length metric on the complement of the r–neighborhood of A in X. If the subspace A is clear from context, we use the notation d_r instead of $d_{r,A}$.

Definition 3.6 Let (X, A) be a pair of metric spaces. For each $\rho \in (0, 1]$ and positive integer $n \geq 2$, we define a function $\delta^n_\rho\colon [0, \infty) \to [0, \infty]$ as follows:

For each r, let $\delta^n_\rho(r) = \sup d_{\rho r}(x_1, x_2)$ where the supremum is taken over all $x_1, x_2 \in \partial N_r(A)$ such that $d_r(x_1, x_2) < \infty$ and $d(x_1, x_2) \leq nr$. The family of functions $\{\delta^n_\rho\}$ is *the relative divergence* of X with respect to A, denoted by $\mathrm{Div}(X, A)$.

We now define the concept of relative divergence of a finitely generated group with respect to a subgroup.

Definition 3.7 Let G be a finitely generated group with subgroup H. We define *the relative divergence* of G with respect to H, denoted by $\mathrm{Div}(G, H)$, to be the relative divergence of the Cayley graph $\Gamma(G, S)$ with respect to H for some finite generating set S.

Remark 3.8 The concept of relative divergence was introduced by the author [9] with the name upper relative divergence. The relative divergence of geodesic spaces is a pair quasi-isometry invariant concept. This implies that the relative divergence of a finitely generated group does not depend on the choice of finite generating sets.

Definition 3.9 The *divergence* of a bi-infinite geodesic α, denoted by Div_α, is a function $g\colon (0, \infty) \to (0, \infty)$ such that for each positive number r, the value $g(r)$ is the infimum on the lengths of all paths, outside the open ball about $\alpha(0)$ with radius r, connecting $\alpha(-r)$ and $\alpha(r)$.

The following lemma is deduced from the proof of Corollary 4.8 in [1].

Lemma 3.10 *Let Γ be a connected, finite, simplicial graph with at least two vertices. Assume that Γ is not a join. Let g be a cyclically reduced element in A_Γ that does not lie in any join subgroup. Then the divergence of the bi-infinite geodesic $\cdots ggggg \cdots$ is at least quadratic.*

Definition 3.11 Let G be a group with a finite generating set S and H a subgroup of G with a finite generating set T. The *subgroup distortion* of H in G is the function

$$\mathrm{Dist}^H_G\colon (0, \infty) \to (0, \infty), \quad \mathrm{Dist}^H_G(r) = \max\{|h|_T : h \in H,\ |h|_S \leq r\}.$$

Remark 3.12 It is well known that the concept of distortion does not depend on the choice of finite generating sets.

4 Connection between subgroup distortion and relative divergence

Lemma 4.1 *Let H be a finitely generated group with finite generating set T and ϕ in* $\mathrm{Aut}(H)$. *Let $G = \langle H, t / tht^{-1} = \phi(h)\rangle$ and $S = T \cup \{t\}$.*

(1) *Each element in G can be written uniquely in the form ht^n, where h is a group element in H.*

(2) *The set S is a finite generating set of G, and*

$$d_S(ht^m, h't^n) \geq |m-n| \quad \text{and} \quad d_S(ht^m, Ht^n) = |m-n|.$$

Proof The statement (1) is well known, and we only need to prove statement (2). Let ψ be the map from G to $\mathbb{Z}$ by sending element t to 1 and each generator in T to 0. It is not hard to see that ψ is a group homomorphism. We first show that the absolute value of $\psi(g)$ is at most the length of g with respect to S for each group element g in G. In fact, let $w_1 t^{n_1} w_2 t^{n_2} \cdots w_k t^{n_k}$ be the shortest word in S that represents g, where each w_i is a word in T. Therefore,

$$\psi(g) = n_1 + n_2 + \cdots + n_k$$

and

$$|g|_S = (\ell(w_1) + \ell(w_2) + \cdots + \ell(w_k)) + (|n_1| + |n_2| + \cdots + |n_k|).$$

This implies that the absolute value of $\psi(g)$ is at most the length of g with respect to S. The distance between two elements ht^m and $h't^n$ is the length of the group element $g = (ht^m)^{-1}h't^n$. Obviously, $\psi(g) = n - m$. Therefore, the distance between two elements ht^m and $h't^n$ is at least $|m-n|$. This fact directly implies that the distance between ht^m and any element in Ht^n is at least $|m-n|$. Also, ht^n is an element in Ht^n, and the distance between ht^m and ht^n is at most $|m-n|$. Therefore, the distance between ht^m and Ht^n is exactly $|m-n|$. □

Lemma 4.2 *Let H be a finitely generated group with finite generating set T and ϕ in $Aut(H)$. Let $G = \langle H, t / tht^{-1} = \phi(h)\rangle$ and $S = T \cup \{t\}$. Let n be an arbitrary positive integer, and let x and y be two points in $\partial N_n(H)$. Then there is a path outside $N_n(H)$ connecting x and y if and only if the pair (x, y) is of either the form $(h_1 t^n, h_2 t^n)$ or $(h_1 t^{-n}, h_2 t^{-n})$, where h_1 and h_2 are elements in H.*

Proof By Lemma 4.1, the pair (x, y) must be of the form $(h_1 t^{m_1}, h_2 t^{m_2})$, where $|m_1| = |m_2| = n$. We first assume that $m_1 m_2 < 0$. Let γ be an arbitrary path connecting x and y. By Lemma 4.1, we observe that if two vertices ht^m and $h't^{m'}$ of γ are consecutive, then $|m-m'| \leq 1$. Therefore, there exists a vertex of γ that belongs to H. Thus, there is no path outside $N_n(H)$ connecting x and y.

If $m_1 = m_2$, then x and y both lie in the same coset $t_{m_1}H$. Therefore, there is a path α with all vertices in $t_{m_1}H$ connecting x and y. By Lemma 4.1 again, α must lie outside $N_n(H)$. Therefore, the pair (x,y) is of either the form (h_1t^n, h_2t^n) or (h_1t^{-n}, h_2t^{-n}). □

Proposition 4.3 *Let H be a finitely generated group and $G = \langle H, t / tht^{-1} = \phi(h)\rangle$, where ϕ in $Aut(H)$. Then* $\mathrm{Div}(G,H) \preceq \mathrm{Dist}_G^H$.

Proof Let T be a finite generating set of H, and let $S = T \cup \{t\}$. Then S is a finite generating set of G. Suppose that $\mathrm{Div}(G,H) = \{\delta_\rho^n\}$. We will show that $\delta_\rho^n(r) \leq \mathrm{Dist}_G^H(nr)$ for every positive integer r.

Indeed, let x and y be arbitrary points in $\partial N_r(H)$ such that $d_{r,H}(x,y) < \infty$ and $d_S(x,y) \leq nr$. By Lemma 4.2, x and y both lie in the same coset t^mH, where $|m| = r$. Therefore, there is a path γ with all vertices in t^mH connecting x and y, and the length of γ is at most $\mathrm{Dist}_G^H(nr)$. By Lemma 4.1 again, the path γ must lie outside $N_r(H)$. Therefore, $d_{\rho r,H}(x,y) \leq \mathrm{Dist}_G^H(nr)$. Thus, $\delta_\rho^n(r) \leq \mathrm{Dist}_G^H(nr)$. This implies that $\mathrm{Div}(G,H) \preceq \mathrm{Dist}_G^H$. □

5 Relative divergence of right-angled Artin groups with respect to Bestvina–Brady subgroups and subgroup distortion of Bestvina–Brady subgroups

From now on, we let Γ be a finite, connected, simplicial graph with at least two vertices. Let A_Γ be the associated right-angled Artin group and H_Γ its Bestvina–Brady subgroup. Let X_Γ be the associated Salvetti complex and $\widetilde{X}_\Gamma$ its universal covering. We consider the 1–skeleton of $\widetilde{X}_\Gamma$ as a Cayley graph of A_Γ, and the vertex set S of Γ as a finite generating set of A_Γ. By Theorem 2.6, we can choose the set T of all elements of the form st^{-1} whenever s and t are adjacent vertices as a finite generating set for H_Γ. Let Φ and $\tilde{f}$ be the group homomorphism and continuous map as in Remark 2.5.

Lemma 5.1 *Let M be the diameter of Γ. Let a and b be arbitrary vertices in S. For each integer m, the length of a^mb^{-m} with respect to T is at most $M|m|$.*

Proof Since the diameter of Γ is M, we can choose a positive integer $n \leq M$ and $n+1$ generators $s_0, s_1, \ldots, s_n$ in S such that the following conditions hold:

(1) $s_0 = a$ and $s_n = b$.

(2) s_i and s_{i+1} commute for $i \in \{0, 1, 2, \dots, n-1\}$.

Obviously,

$$\begin{aligned} a^m b^{-m} = s_0^m s_n^{-m} &= (s_0^m s_1^{-m})(s_1^m s_2^{-m})(s_2^m s_3^{-m}) \cdots (s_{n-2}^m s_{n-1}^{-m})(s_{n-1}^m s_n^{-m}) \\ &= (s_0 s_1^{-1})^m (s_1 s_2^{-1})^m (s_2 s_3^{-1})^m \cdots (s_{n-2} s_{n-1}^{-1})^m (s_{n-1} s_n^{-1})^m. \end{aligned}$$

Also, $s_{i-1}s_i^{-1}$ belongs to T. Therefore, the length of $a^m b^{-m}$ with respect to T is at most $n|m|$. This implies the length of $a^m b^{-m}$ with respect to T is at most $M|m|$. □

Proposition 5.2 *The subgroup distortion* $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ *is dominated by a quadratic function. Moreover,* $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ *is linear when* Γ *is a join.*

Proof We first show that $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ is dominated by a quadratic function. Let n be an arbitrary positive integer and h be an arbitrary element in H_Γ such that $|h|_S \leq n$. We can write $h = s_1^{m_1} s_2^{m_2} s_3^{m_3} \cdots s_k^{m_k}$ such that:

(1) Each s_i lies in S, $|m_i| \geq 1$ and $|m_1| + |m_2| + |m_3| + \cdots + |m_k| \leq n$.

(2) $m_1 + m_2 + m_3 + \cdots + m_k = 0$.

Obviously, we can rewrite h as follows:

$$h = (s_1^{m_1} s_2^{-m_1})(s_2^{(m_1+m_2)} s_3^{-(m_1+m_2)}) \cdots (s_{k-1}^{(m_1+m_2+\cdots+m_{k-1})} s_k^{-(m_1+m_2+\cdots+m_{k-1})}).$$

Let M be the diameter of Γ. By Lemma 5.1, we have

$$\begin{aligned} |h|_T &\leq M|m_1| + M|m_1+m_2| + \cdots + M|m_1+m_2+\cdots+m_{k-1}| \\ &\leq M|m_1| + M(|m_1|+|m_2|) + \cdots + M(|m_1|+|m_2|+\cdots+|m_{k-1}|) \\ &\leq M(k-1)n \leq Mn^2. \end{aligned}$$

Therefore, the distortion function $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ is bounded above by Mn^2.

We now assume that Γ is a join of Γ_1 and Γ_2. We need to prove that the distortion $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ is linear. Let n be an arbitrary positive integer and h be an arbitrary element in H_Γ such that $|h|_S \leq n$. Since A_Γ is the direct product of A_{Γ_1} and A_{Γ_2}, we can write $h = (a_1^{m_1} a_2^{m_2} \cdots a_k^{m_k})(b_1^{n_1} b_2^{n_2} \cdots b_\ell^{n_\ell})$ such that:

(1) Each a_i is a vertex of Γ_1 and each b_j is a vertex of Γ_2.

(2) $(|m_1| + |m_2| + \cdots + |m_k|) + (|n_1| + |n_2| + \cdots + |n_\ell|) \leq n$.

(3) $(m_1 + m_2 + \cdots + m_k) + (n_1 + n_2 + \cdots + n_\ell) = 0$.

Let $m = m_1 + m_2 + \cdots + m_k$. Then $n_1 + n_2 + \cdots + n_\ell = -m$ and $|m| \leq n$. Let a be a vertex in Γ_1 and b a vertex in Γ_2. Since a commutes with each b_j, b commutes with each a_i, and a and b commute, we can rewrite h as follows:

$$\begin{aligned} h &= (a_1^{m_1} a_2^{m_2} \cdots a_k^{m_k} b^{-m})(b^m a^{-m})(a^m b_1^{n_1} b_2^{n_2} \cdots b_\ell^{n_\ell}) \\ &= (a_1 b^{-1})^{m_1} (a_2 b^{-1})^{m_2} \cdots (a_k b^{-1})^{m_k} (ba^{-1})^m (ab_1^{-1})^{-n_1} (ab_2^{-1})^{-n_2} \cdots (ab_\ell^{-1})^{-n_\ell}. \end{aligned}$$

Also, ab_j^{-1}, $a_i b^{-1}$ and ba^{-1} all belong to T. Therefore,

$$|h|_T \le (|m_1| + |m_2| + \cdots + |m_k|) + (|n_1| + |n_2| + \cdots + |n_\ell|) + |m| \le 2n.$$

Therefore, the distortion function $\mathrm{Dist}_{A_\Gamma}^{H_\Gamma}$ is bounded above by $2n$. □

Proposition 5.3 *If Γ is not a join graph, then the relative divergence* $\mathrm{Div}(A_\Gamma, H_\Gamma)$ *is at least quadratic.*

Proof Let J be a maximal join in Γ, and let v be a vertex not in J. Let g in A_J be the product of all vertices in J. Let $n = \Phi(g)$ and let $h = gv^{-n}$. Then h is an element in H_Γ. Since J is a maximal join in Γ and v is a vertex not in J, we see that h does not lie in any join subgroup. Also, h is a cyclically reduced element. Therefore, the divergence of the bi-infinite geodesic $\alpha = \cdots hhhhh \cdots$ is at least quadratic by Lemma 3.10.

Let t be an arbitrary generator in S and $k = |h|_S$. We can assume that $\alpha(0) = e$, $\alpha(km) = h^m$ and $\alpha(-km) = h^{-m}$. In order to prove that the relative divergence $\mathrm{Div}(A_\Gamma, H_\Gamma)$ is at least quadratic, it is sufficient to prove each function δ_ρ^n dominates the divergence function of α for each $n \ge 2k+2$.

Indeed, let r be an arbitrary positive integer. Let $x = h^{-r} t^r$ and $y = h^r t^r$. By a similar argument as in Lemmas 4.1 and 4.2, the two points x and y both lie in $\partial N_r(H_\Gamma)$, and $d_{r,H_\Gamma}(x,y) < \infty$. Moreover,

$$d_S(x,y) \le d_S(x, h^{-r}) + d_S(h^{-r}, h^r) + d_S(h^r, y) \le r + 2kr + r \le (2k+2)r \le nr.$$

Let γ be an arbitrary path outside $N_{\rho r}(H)$ connecting x and y. Obviously, the path γ must lie outside the open ball $B(\alpha(0), \rho r)$. It is obvious that we can connect x and h^{-r} by a path γ_1 of length r which lies outside $B(\alpha(0), \rho r)$. Similarly, we can connect y and h^r by a path γ_2 of length r which lies outside $B(\alpha(0), \rho r)$. Let γ_3 be the subsegment of α connecting $\alpha(-\rho r)$ and h^{-r}. Let γ_4 be the subsegment of α connecting $\alpha(\rho r)$ and h^r. It is not hard to see the lengths of γ_3 and γ_4 are both $(k-\rho)r$.

Let $\overline{\gamma} = \gamma_3 \cup \gamma_1 \cup \gamma \cup \gamma_2 \cup \gamma_4$. Then $\overline{\gamma}$ is a path that lies outside $B(\alpha(0), \rho r)$ connecting $\alpha(-\rho r)$ and $\alpha(\rho r)$. Therefore, the length of $\overline{\gamma}$ is at least $\mathrm{Div}_\alpha(\rho r)$. Also,

$$\ell(\overline{\gamma}) = \ell(\gamma_3) + \ell(\gamma_1) + \ell(\gamma) + \ell(\gamma_2) + \ell(\gamma_5) = \ell(\gamma) + 2(k - \rho + 1)r.$$

Thus,

$$\ell(\gamma) \ge \mathrm{Div}_\alpha(\rho r) - 2(k - \rho + 1)r.$$

This implies that

$$d_{\rho r, H_\Gamma}(x,y) \geq \mathrm{Div}_\alpha(\rho r) - 2(k-\rho+1)r.$$

Therefore,

$$\delta^n_\rho(r) \geq \mathrm{Div}_\alpha(\rho r) - 2(k-\rho+1)r.$$

Thus, the relative divergence $\mathrm{Div}(A_\Gamma, H_\Gamma)$ is at least quadratic. □

The following theorem is deduced from Propositions 4.3, 5.2 and 5.3.

Theorem 5.4 *Let Γ be a connected, finite, simplicial graph with at least two vertices. Let A_Γ be the associated right-angled Artin group and H_Γ the Bestvina–Brady subgroup. Then the relative divergence $\mathrm{Div}(A_\Gamma, H_\Gamma)$ and the subgroup distortion $\mathrm{Dist}^{H_\Gamma}_{A_\Gamma}$ are both linear if Γ is a join graph. Otherwise, they are both quadratic.*

Corollary 5.5 *For each integer $n \geq 3$, there is a right-angled Artin group containing a free subgroup of rank n whose inclusion is not a quasi-isometric embedding.*

Proof For each positive integer $n \geq 3$, let Γ be a tree with n edges such that Γ is not a join graph. By the above theorem, the distortion of H_Γ in the right-angled Artin group A_Γ is quadratic. Also, H_Γ is the free group of rank n by Theorem 2.6. □

6 Connection to horizontal surface subgroups

Definition 6.1 A *graph manifold* is a compact, irreducible, connected orientable 3–manifold M that can be decomposed along $\mathcal{T}$ into finitely many Seifert manifolds, where $\mathcal{T}$ is the canonical toric decomposition of Johannson [8] and of Jaco and Shalen [7]. We call the collection $\mathcal{T}$ its JSJ-decomposition in M, and each element in $\mathcal{T}$ its JSJ-torus.

Definition 6.2 If M is a Seifert manifold, a properly immersed surface $g\colon S \looparrowright M$ is *horizontal* if $g(S)$ is transverse to the Seifert fibers everywhere. In the case where M is a graph manifold, a properly immersed surface $g\colon S \looparrowright M$ *horizontal* if $g(S) \cap P_v$ is horizontal for every Seifert component P_v.

Theorem 6.3 *If Γ is a finite tree with at least three vertices, then the associated right-angled Artin group A_Γ is a fundamental group of a graph manifold, and the Bestvina–Brady subgroup H_Γ is a horizontal surface subgroup.*

Proof First, we construct the graph manifold M whose fundamental group is A_Γ. Let v be a vertex of Γ of degree $k \geq 2$. Let $u_1, u_2, \ldots, u_k$ be all elements in $\ell k(v)$. Let Σ_v be a punctured disk with k holes whose boundaries are labeled by elements

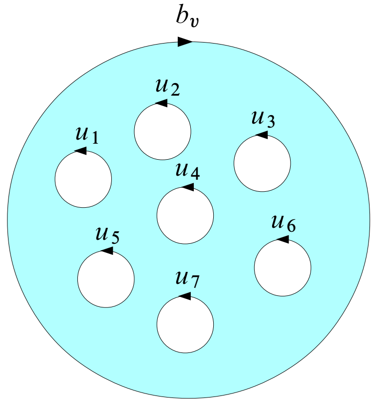


Figure 1: A punctured disk Σ_v when the degree of v in Γ is 7

in $\ell k(v)$. We also label the outside boundary component of Σ_v by b_v; see Figure 1. Obviously, $\pi_1(\Sigma_v)$ is the free group generated by $u_1, u_2, \dots, u_k$.

Let $P_v = \Sigma_v \times S^1_v$, where we label the circle factor in P_v by v. Obviously, each P_v is a Seifert manifold. Moreover, for each u_i in $\ell k(v)$, the Seifert manifold P_v contains the torus $S^1_{u_i} \times S^1_v$ as a component of its boundary.

We construct the graph manifold by gluing pairs of Seifert manifolds (P_{v_1}, P_{v_2}) along their tori $S^1_{v_1} \times S^1_{v_2}$ whenever v_1 and v_2 are adjacent vertices in Γ. We observe that the pair of such regions are glued together by switching fiber and base directions. It is not hard to see that the fundamental group of M is the right-angled Artin group A_Γ.

We now construct the horizontal surface S in M with the Bestvina–Brady subgroup H_Γ as its fundamental group. We first construct the horizontal surface S_v on each Seifert piece $P_v = \Sigma_v \times S^1_v$, where v is a vertex of Γ of degree $k \geq 2$.

We remind the reader that Σ_v is a punctured disk with k holes whose boundaries are labeled by the elements $u_1, u_2, \dots, u_k$ in $\ell k(v)$. We also label the outside boundary component of Σ_v by b_v; see Figure 1. We label the circle factor in P_v by v.

Let S_v be a copy of the punctured disk Σ_v. However, we relabel all inside circles by $c_1, c_2, \dots, c_k$ and the outside circle by c_v. We will construct a map $(g, h)\colon S_v \to \Sigma_v \times S^1_v$ as follows:

(1) The map g is the identity map that maps each c_i to u_i and c_v to b_v.

(2) The map h has degree -1 on boundary component c_i and degree k on c_v.

We now construct the map h with the above properties. We observe that the fundamental group of S_v is generated by $c_1, c_2, \dots, c_k$ and c_v with a unique relator $c_1 c_2 c_3 \cdots c_k c_v = e$. Here we abused notation for the presentation of $\pi_1(S_v)$. By that presentation of $\pi_1(S_v)$, we can see that there is a group homomorphism ϕ from $\pi_1(S_v)$ to $\mathbb{Z}$ that maps each c_i to -1 and c_v to k. By [5, Proposition 1B.9], the group homomorphism ϕ is induced by a map h from S_v to S^1_v. Therefore, we constructed the desired map h.

Finally, we identify the surface S_v with its image via the map (g,h). By construction, $\pi_1(S_v)$ is the subgroup of $\pi_1(P_v)$ generated by elements $u_1v^{-1}, u_2v^{-1}, \dots, u_kv^{-1}$. We observe that if we glue pair of Seifert manifolds (P_{v_1}, P_{v_2}) along their tori $S^1_{v_1} \times S^1_{v_2}$, the pair of horizontal surfaces (S_{v_1}, S_{v_2}) will be matched up along their boundaries in $S^1_{v_1} \times S^1_{v_2}$. Therefore, we constructed a horizontal surface S in M. By the Van Kampen theorem, the fundamental group of S is generated by all elements of the form st^{-1} whenever s and t are adjacent vertices in Γ. In other words, $\pi_1(S)$ is the Bestvina–Brady subgroup by Theorem 2.6. □

References

[1] **J Behrstock**, **R Charney**, *Divergence and quasimorphisms of right-angled Artin groups*, Math. Ann. 352 (2012) 339–356 MR

[2] **M Bestvina**, **N Brady**, *Morse theory and finiteness properties of groups*, Invent. Math. 129 (1997) 445–470 MR

[3] **M Carr**, *Two-generator subgroups of right-angled Artin groups are quasi-isometrically embedded*, preprint (2014) arXiv

[4] **W Dicks**, **I J Leary**, *Presentations for subgroups of Artin groups*, Proc. Amer. Math. Soc. 127 (1999) 343–348 MR

[5] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR

[6] **G C Hruska**, **H T Nguyen**, *Distortion of surfaces in graph manifolds*, preprint (2017) arXiv

[7] **W H Jaco**, **P B Shalen**, *Seifert fibered spaces in 3–manifolds*, Mem. Amer. Math. Soc. 220 (1979) MR

[8] **K Johannson**, *Homotopy equivalences of 3–manifolds with boundaries*, Lecture Notes in Mathematics 761, Springer (1979) MR

[9] **H C Tran**, *Relative divergence of finitely generated groups*, Algebr. Geom. Topol. 15 (2015) 1717–1769 MR

Department of Mathematics, The University of Georgia
Athens, GA, United States

hung.tran@uga.edu

Received: 23 August 2016 Revised: 20 October 2016

Algebraic & Geometric Topology 17 (2017) 2511–2522

Non-L–space integral homology 3–spheres with no nice orderings

XINGHUA GAO

We give infinitely many examples of non-L–space irreducible integer homology 3–spheres whose fundamental groups do not have nontrivial $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ representations.

57M50; 57M25, 57M27

1 Introduction

Before stating the main result, I will review some definitions. A rational homology 3–sphere Y is called an *L–space* if $\mathrm{rk}\,\widehat{\mathrm{HF}}(Y) = |H_1(Y;\mathbb{Z})|$, ie its Heegaard Floer homology is minimal. An L–space does not admit any coorientable taut foliation, by Bowden [1], Kazez and Roberts [13] and Ozsváth and Szabó [15]. A nontrivial group G is called left-orderable if there exists a strict total ordering of G invariant under left multiplication. Boyer, Gordon and Watson conjectured in [2] that an irreducible rational homology 3–sphere is a non-L–space if and only if its fundamental group is left-orderable. A stronger conjecture states that for an irreducible $\mathbb{Q}$–homology 3–sphere, being a non-L–space, having left-orderable fundamental group and admitting a coorientable taut foliation are the same (see eg Culler and Dunfield [5]).

To show the fundamental group $\pi_1(Y)$ of a 3–manifold Y is orderable, it is most common to consider $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ representations of $\pi_1(Y)$. In fact, in many cases, $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ representations are sufficient to define an order on $\pi_1(Y)$; see [5]. However, Theorem 1 in this paper shows that, even in the case of non-L–space integral homology spheres, orders coming from $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ are not enough to prove the conjecture of Boyer, Gordon and Watson.

It is conjectured that any integer homology 3–sphere different from the 3–sphere admits an irreducible representation in $\mathrm{SU}_2(\mathbb{C})$ (see eg Kirby's problem list [14, Problem 3.105]). Zentner [20] showed that if one enlarges the target group to $\mathrm{SL}_2(\mathbb{C})$, then every such integral homology 3–sphere has an irreducible representation. By contrast, I will give examples where there are no irreducible $\mathrm{PSL}_2(\mathbb{R})$ representations. Let $\mathcal{M}$ be the manifold $m137$ — see Callahan, Hildebrand and Weeks [3] — and $\mathcal{M}(1,n)$ be the integral homology sphere obtained by $(1,n)$ Dehn fillings on $\mathcal{M}$. The main result of this paper states:

Published: 3 August 2017　　DOI: 10.2140/agt.2017.17.2511

Theorem 1 *For all $n \ll 0$, the manifold $\mathcal{M}(1,n)$ is a hyperbolic integral homology 3–sphere where*

(a) $\pi_1(\mathcal{M}(1,n))$ *does not have a nontrivial* $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ *representation;*

(b) $\mathcal{M}(1,n)$ *is not an L–space.*

This means that we can not produce an order on $\pi_1(\mathcal{M}(1,n))$ simply by pulling back the action of $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ on $\mathbb{R}$.

Section 2 is devoted to proving Theorem 1(a). Let $X_0(\mathcal{M})$ be the component of the $\mathrm{SL}_2(\mathbb{C})$ character variety of $\mathcal{M}$ containing the character of an irreducible representation (see Culler and Shalen [7] for the definition). Here is an outline of the approach. Let $X_{0,\mathbb{R}}(\mathcal{M})$ be the real points of $X_0(M)$. Define $[\rho] \in X_{0,\mathbb{R}}(\mathcal{M})$ and denote by s the trace of $\rho(\lambda)$, where λ is the homological longitude of $\mathcal{M}$. The proof is divided into two parts. In the first part, I show that points on the $|s| < 2$ components of $X_{0,\mathbb{R}}(\mathcal{M})$ all correspond to $\mathrm{SU}_2(\mathbb{C})$ representations, while points on the $|s| > 2$ components correspond to $\mathrm{SL}_2(\mathbb{R})$ representations. In the second part, I show that $\mathrm{SL}_2(\mathbb{R})$ representations of $\pi_1(\mathcal{M})$ give rise to no $\mathrm{SL}_2(\mathbb{R})$ representations of $\pi_1(\mathcal{M}(1,n))$ when $n \ll 0$. This part of the proof is basically analyzing real solutions to the A–polynomial of $\mathcal{M}$ under the relation $\mu\lambda^n = 1$ given by $(1,n)$ Dehn filling, where μ is a choice of meridian of $\partial\mathcal{M}$.

In Section 3, by applying techniques in the paper by Rasmussen and Rasmussen [17] and Gillespie [11], I show that none of the $(1,n)$ Dehn fillings on $m137$ is an L–space, completing the proof of Theorem 1.

Acknowledgements The author was partially supported by NSF grants DMS-1510204, and Campus Research Board grant RB15127. I would like to pay special thanks to my advisor, Nathan Dunfield for suggesting me this problem and offering me extraordinary help. I would also like to thank the referee for detailed and helpful comments and suggestions.

2 $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ representations

I will prove Theorem 1(a) in this section.

SnapPy [6] gives us the following presentation of the fundamental group of $\mathcal{M} = m137$:

$$\pi_1(\mathcal{M}) = \langle \alpha, \beta \mid \alpha^3\beta^2\alpha^{-1}\beta^{-3}\alpha^{-1}\beta^2 \rangle.$$

The peripheral system of $\mathcal{M}$ can be represented as

$$\{\mu, \lambda\} = \{\alpha^{-1}\beta^2\alpha^4\beta^2, \alpha^{-1}\beta^{-1}\} = \{\beta^2\lambda^{-1}\beta^{-3}\lambda^{-1}\beta^2, \lambda\},$$

where λ is the homological longitude and μ is a choice of meridian. Then we can rewrite the fundamental group as

$$\pi_1(\mathcal{M}) = \langle \lambda, \beta \mid \beta^{-1}\lambda^{-1}\beta^{-1}\lambda^{-1}\beta^2\lambda = \lambda\beta^{-2}\lambda^{-1}\beta^2 \rangle, \tag{2-1}$$

and the meridian becomes $\mu = \beta^2\lambda^{-1}\beta^{-3}\lambda^{-1}\beta^2$ under this presentation.

Remark The triangulation of $m137$ we used (included in [9]) to get these presentations is different from SnapPy's default triangulation. We got it by performing random Pachner moves on the default triangulation in SnapPy. In particular, our notations for longitude and meridian in the peripheral system are meridian and longitude, respectively, in SnapPy's default notations.

We will first look at irreducible $\mathrm{SL}_2(\mathbb{C})$ representations of the fundamental group of $\mathcal{M}$ before we look at those of Dehn fillings of $\mathcal{M}$. Denote by $X(\mathcal{M})$ the $\mathrm{SL}_2(\mathbb{C})$ character variety of $\mathcal{M}$, that is, the geometric invariant theory quotient

$$\mathrm{Hom}(\pi_1(\mathcal{M}), \mathrm{SL}_2(\mathbb{C}))/\!\!/\mathrm{SL}_2(\mathbb{C}).$$

It is an affine variety [7]. Suppose $\rho\colon \pi_1(\mathcal{M}) \longrightarrow \mathrm{SL}_2(\mathbb{C})$ is a representation of the fundamental group of $\mathcal{M}$. Recall that a representation ρ of G in $\mathrm{SL}_2(\mathbb{C})$ is irreducible if the only subspaces of $\mathbb{C}^2$ invariant under $\rho(G)$ are $\{0\}$ and $\mathbb{C}^2$ [7]. This is equivalent to saying that ρ can't be conjugated to a representation by upper triangular matrices. Otherwise ρ is called reducible. We will call a character irreducible (reducible) if the corresponding representation is irreducible (reducible).

First, I determine which components of $X(\mathcal{M})$ contain characters of irreducible representations. Computation with SnapPy [6] shows that the Alexander polynomial $\Delta_{\mathcal{M}}$ of $m137$ is 1, which has no root. So there are no reducible nonabelian representations [4, Section 6.1]. Therefore all the reducible representations are abelian. Since $H_1(\mathcal{M}) = \mathbb{Z}$, there is only one such component and it is parametrized by the image of β and is isomorphic to $\mathrm{Hom}(\mathbb{Z}, \mathrm{SL}_2(\mathbb{C}))/\!\!/\mathrm{SL}_2(\mathbb{C}) \simeq \mathbb{C}$. Moreover, it is disjoint from any component of $X(\mathcal{M})$ containing the character of an irreducible representation [4, Section 6.2]. For more details, we refer the readers to Tillmann's note [19], where he studied $m137$ as an example.

If an abelian representation of $\pi_1(\mathcal{M})$ induces an abelian representation of $\pi_1(\mathcal{M}(1,n))$ then it factors through the abelianization $\mathrm{ab}\big(\pi_1(\mathcal{M}(1,n))\big) = 1$. So they correspond to trivial $\mathrm{SL}_2(\mathbb{C})$ representations and we don't need to worry about them.

Now we consider components of $X(\mathcal{M})$ that contain the character of an irreducible representation. We have:

Lemma 2 *There is a single component $X_0(\mathcal{M})$ of $X(\mathcal{M})$ containing an irreducible character. The functions $s = \operatorname{tr}\rho(\lambda) = \operatorname{tr}\rho(\alpha^{-1}\beta^{-1}) = \operatorname{tr}\rho(\alpha\beta)$ and $t = \operatorname{tr}\rho(\beta)$ give complete coordinates on $X_0(\mathcal{M})$, which is the curve in $\mathbb{C}^2$ cut out by*

$$(-2-3s+s^3)t^4 + (4+4s-s^2-s^3)t^2 - 1 = 0.$$

Moreover, $w := \operatorname{tr}\rho(\lambda\beta) = \operatorname{tr}\rho((\lambda\beta)^{-1}) = t - 1/(t(s+1))$.

Proof Let $X_0(\mathcal{M})$ be $X(\mathcal{M}) - \{\text{reducible characters}\}$. From the discussion above, we know that all the reducible characters form a single component of $X(\mathcal{M})$ and this component is disjoint from any other component of $X(\mathcal{M})$. So $X_0(\mathcal{M})$ is Zariski closed. We will show later that $X_0(\mathcal{M})$ is actually an irreducible algebraic variety, as claimed in the lemma.

Suppose $[\rho] \in X_0(\mathcal{M})$. So ρ is an irreducible representation. By conjugating ρ if necessary, we can assume that ρ has the form

$$\rho(\lambda) = \begin{pmatrix} z & 1 \\ 0 & 1/z \end{pmatrix}, \quad \rho(\beta) = \begin{pmatrix} x & 0 \\ y & 1/x \end{pmatrix}.$$

From the relator of $\pi_1(\mathcal{M})$ in (2-1) we have $\rho(\beta)^{-1}\rho(\lambda)^{-1}\beta^{-1}\rho(\lambda)^{-1}\rho(\beta)^2\rho(\lambda) = \rho(\lambda)\rho(\beta)^{-2}\rho(\lambda)^{-1}\rho(\beta)^2$. Comparing the entries of the matrices on both sides, we get four equations. These four equations together with $s = z + 1/z$, $t = x + 1/x$ and $w = zx + z^{-1}x^{-1} + y$ form a system $\mathcal{S}$ which defines $X_0(\mathcal{M})$. By computing a Gröbner basis of this system, SageMath [18] gives the following generators of the radical ideal $I = I(X_0(\mathcal{M}))$:

$$stw - t^2 - w^2 - s + 2, \tag{2-2}$$

$$t^3 - w^3 + st - sw - 2t + w, \tag{2-3}$$

$$st^2 - tw - w^2 - s + 1, \tag{2-4}$$

$$sw^3 - s^2t + s^2w - t^2w - tw^2 + st - sw + t. \tag{2-5}$$

Subtracting (2-4) from (2-2), we get

$$w = t - \frac{1}{t(s+1)}. \tag{2-6}$$

Eliminating w, we get a defining equation for $X_0(\mathcal{M})$:

$$\begin{aligned} 0 &= (-2-3s+s^3)t^4 + (4+4s-s^2-s^3)t^2 - 1 \\ &= (s-2)(s+1)^2t^4 - (s-2)(s+2)(s+1)t^2 - 1. \end{aligned} \tag{2-7}$$

Thus, we can think of $X_0(\mathcal{M})$ as living in $\mathbb{C}^2$.

To prove the lemma, we must show that $X_0(M)$ is irreducible or, equivalently, the polynomial $P(s,t) := (s-2)(s+1)^2t^4 - (s-2)(s+2)(s+1)t^2 - 1$ in (2-7) does not factor in $\mathbb{C}[s,t]$. Assume $P(s,t)$ factors. Suppose it factors as

$$(at^2+bt+c)\Big(dt^2+et-\frac{1}{c}\Big)=adt^4+(ae+bd)t^3+\Big(cd-\frac{a}{c}+be\Big)t^2+\Big(ce-\frac{b}{c}\Big)t-1,$$

where $a,b,d,e \in \mathbb{C}[s]$ and $c \in \mathbb{C}-\{0\}$. Setting the coefficients of t and t^3 to be 0, we get $b = c^2e$ and $ae = -c^2de$. If $e \neq 0$, then $a = -c^2d$. But this is impossible as $ad = (s-2)(s+1)^2$ is a polynomial in s of odd degree. So $e = 0$ and it follows that $b = 0$. Comparing the coefficients of t^2 and t^4, we get

$$ad = (s-2)(s+1)^2, \tag{2-8}$$

$$cd - \frac{a}{c} = -(s-2)(s+2)(s+1). \tag{2-9}$$

So $\mathrm{degree}(a) + \mathrm{degree}(d) = 3$ and $\max\{\mathrm{degree}(a), \mathrm{degree}(d)\} \geq 3$, which implies exactly one of a and d has degree 3 and the other has degree 0. Without loss of generality, we can assume that $\mathrm{degree}(a) = 3$ and $\mathrm{degree}(d) = 0$. Multiplying both sides of (2-9) by c, we get $a = c^2d + c(s-2)(s+2)(s+1)$. So the coefficient of s^3 in a is c. Comparing with the coefficient of s^3 in (2-8), we see that $d = 1/c$. Eliminating a and d gives us an equality $1 + (s-2)(s+2)(s+1) = (s-2)(s+1)^2$, which does not hold.

Otherwise, suppose $P(s,t)$ factors as

$$(at+c)\Big(bt^3+dt^2+et-\frac{1}{c}\Big)=abt^4+(ad+cb)t^3+(cd+ae)t^2+\Big(ce-\frac{a}{c}\Big)t-1,$$

where $a,b,d,e \in \mathbb{C}[s]$ and $c \in \mathbb{C}-\{0\}$. Setting the coefficients of t and t^3 to be 0, we get $a = c^2e$ and $b = ced$. Comparing the coefficients of t^2 and t^4, we get

$$c^3de^2 = (s-2)(s+1)^2, \tag{2-10}$$

$$cd + c^2e^2 = -(s-2)(s+2)(s+1). \tag{2-11}$$

So $\mathrm{degree}(d) + 2\,\mathrm{degree}(e) = 3$ and $\max\{\mathrm{degree}(d), 2\,\mathrm{degree}(e)\} \geq 3$, which implies $\mathrm{degree}(d) = 3$ and $\mathrm{degree}(e) = 0$. Comparing the coefficients of s^3 in (2-10) and (2-11), we know that $c^2e^2 = -1$. Plugging into (2-10), we get $cd = (s-2)(s+1)^2$, which when plugging into (2-11) implies $c^2e^2 = -(s+1)(s-2)$, a contradiction. So $P(s,t)$ is irreducible over $\mathbb{C}$. Therefore, $X_0(\mathcal{M})$ has only one component. $\square$

To find irreducible $\mathrm{SL}_2(\mathbb{R})$ representations of $\pi_1(\mathcal{M})$, we need to check all real points on $X_0(\mathcal{M})$, which correspond to real solutions of (2-7). Notice that (2-7) has no solutions when $s = -1$ or 2, so (2-7) is a quadratic equation in t^2. In order for t to be real, t^2 has to be real and nonnegative. Then first we need the discriminant to be nonnegative. That is,

$$\Delta_1 = (s+1)^2(s-2)(s^3+2s^2-4s-4) \geq 0.$$

So $s \in U := (-\infty, p_1] \cup [p_2, p_3] \cup (2, \infty)$, where $p_1 \approx -2.9032$, $p_2 \approx -0.8061$ and $p_3 \approx 1.7093$ are the three roots of the cubic polynomial s^3+2s^2-4s-4.

The following lemma will help us determine when a $\mathrm{SL}_2(\mathbb{C})$ representation of $\pi_1(\mathcal{M})$ can be conjugated into $\mathrm{SL}_2(\mathbb{R})$ by simply checking where it lies on the character variety.

Lemma 3 *The set of real points $X_{0,\mathbb{R}}(\mathcal{M}) = X_0(\mathcal{M}) \cap \mathbb{R}^2$ of $X_0(\mathcal{M})$ has 6 connected components:*

- *Points on the two components with $|s| < 2$ correspond to $\mathrm{SU}_2(\mathbb{C})$ representations.*
- *Points on the four components with $|s| > 2$ correspond to $\mathrm{SL}_2(\mathbb{R})$ representations.*

Remark The above lemma shows that, in our case, the absolute value of one character being smaller than 2 implies that the representation is $\mathrm{SU}_2(\mathbb{C})$. But, in general, this is not true.

To prove this lemma, we need to determine when $[\rho] \in X_{0,\mathbb{R}}(\mathcal{M})$ corresponds to $\rho \in \mathrm{SU}_2(\mathbb{C})$ and when it corresponds to $\rho \in \mathrm{SL}_2(\mathbb{R})$. It can't be in both because otherwise it would be reducible [5, Lemma 2.10] and we know $X_0(\mathcal{M})$ contains only irreducible characters. The tool we use is a reformulation of Proposition 3.1 in [12], which states that given three angles $\theta_i \in [0, \pi]$, $i = 1, 2, 3$, there exist three $\mathrm{SU}_2(\mathbb{C})$ matrices C_i satisfying $C_1C_2C_3 = I$ with eigenvalues $\exp(\pm i\theta_i)$, respectively, if and only if these angles satisfy

$$|\theta_1 - \theta_2| \leq \theta_3 \leq \min\{\theta_1 + \theta_2,\ 2\pi - (\theta_1 + \theta_2)\}. \tag{2-12}$$

We want to rewrite the above inequality in terms of traces of C_1, C_2 and C_3. We have the following lemma:

Lemma 4 *Suppose $t_1, t_2, t_3 \in (-2, 2)$ are the traces of $C_1, C_2, C_3 \in \mathrm{SL}_2(\mathbb{C})$ which satisfy $C_1C_2C_3 = I$. Then C_1, C_2 and C_3 are simultaneously conjugate in $\mathrm{SU}_2(\mathbb{C})$ if and only if*

$$(2t_3 - t_1t_2)^2 \leq (4 - t_1^2)(4 - t_2^2).$$

Proof Suppose $t_1 = 2\cos\theta_1$, $t_2 = 2\cos\theta_2$ and $t_3 = 2\cos\theta_3$ with $\theta_1, \theta_2, \theta_3 \in [0, \pi]$.

If $0 \leq \theta_1 + \theta_2 \leq \pi$, then the inequality (2-12) becomes $|\theta_1 - \theta_2| \leq \theta_3 \leq \theta_1 + \theta_2$. Taking cosines, we get $\cos(\theta_1 + \theta_2) \leq \cos\theta_3 \leq \cos(\theta_1 - \theta_2)$.

If $\pi \le \theta_1+\theta_2 \le 2\pi$, then the inequality becomes $|\theta_1-\theta_2| \le \theta_3 \le 2\pi-(\theta_1+\theta_2)$. Taking cosines, we also get $\cos(\theta_1+\theta_2) \le \cos\theta_3 \le \cos(\theta_1-\theta_2)$.

Using the relations $t_1 = 2\cos\theta_1$, $t_2 = 2\cos\theta_2$ and $t_3 = 2\cos\theta_3$, we get in both cases that

$$\frac{t_1t_2}{4} - \sqrt{\left(1-\frac{t_1^2}{4}\right)\left(1-\frac{t_2^2}{4}\right)} \le \frac{t_3}{2} \le \frac{t_1t_2}{4} + \sqrt{\left(1-\frac{t_1^2}{4}\right)\left(1-\frac{t_2^2}{4}\right)}.$$

Then

$$-\sqrt{\left(1-\frac{t_1^2}{4}\right)\left(1-\frac{t_2^2}{4}\right)} \le \frac{t_3}{2} - \frac{t_1t_2}{4} \le \sqrt{\left(1-\frac{t_1^2}{4}\right)\left(1-\frac{t_2^2}{4}\right)}.$$

So we have

$$\left|\frac{t_3}{2} - \frac{t_1t_2}{4}\right| \le \sqrt{\left(1-\frac{t_1^2}{4}\right)\left(1-\frac{t_2^2}{4}\right)}.$$

Squaring both sides and simplifying, we get

$$(2t_3 - t_1t_2)^2 \le (4-t_1^2)(4-t^2),$$

as desired. □

With the criterion of Lemma 4 in hand, we now can prove Lemma 3.

Proof of Lemma 3 The six components correspond to $s \in (-\infty, p_1] \cup [p_2, p_3] \cup (2, \infty)$ and $t \in (-\infty, 0) \cup (0, \infty)$.

Set $C_1 = \rho(\lambda)$, $C_2 = \rho(\beta)$ and $C_3 = \rho(\beta^{-1}\lambda^{-1}) = \rho((\lambda\beta)^{-1})$. Then $t_1 = s$, $t_2 = t$ and $t_3 = w$. Applying Lemma 4 we have

$$(2w - st)^2 \le (4-s^2)(4-t^2). \tag{2-13}$$

Plugging (2-6) into (2-13) and simplifying,

$$(s-2)^2t^2 + \frac{4(s-2)}{s+1} + \frac{4}{t^2(s+1)^2} \le (4-s^2)(4-t^2).$$

Multiplying both sides by $t^2(s+1)^2$, we get

$$(s+1)^2(s-2)^2t^4 + 4(s-2)(s+1)t^2 + 4 \le (4-s^2)(s+1)^2(4-t^2)t^2,$$

which simplifies to

$$-(s+1)^2(s-2)t^4 + (s^2+3s+3)(s-2)(s+1)t^2 + 1 \le 0.$$

Plugging in (2-7), we get

$$(s+1)^3(s-2)t^2 \le 0,$$

which always holds when $s \in (p_2, p_3) \approx (-0.8061, 1.7093) \subset (-2, 2)$.

So, points on $X_{0,\mathbb{R}}(\mathcal{M})$ correspond to $SU_2(\mathbb{C})$ representations if and only if $|s|<2$ and correspond to $SL_2(\mathbb{R})$ representations if and only if $|s|>2$. $\square$

Proof of Theorem 1(a) Lemma 3 tells us a $SL_2(\mathbb{C})$ representation ρ of $m137$ is real if and only if eigenvalues of $\rho(\lambda)$ are real. Moreover, the condition $\mu\lambda^n=1$ forces the eigenvalues of $\rho(\mu)$ to also be real in this case. So we could restrict our attention to $|s|>2$ and look at the A–polynomial instead (see eg [4] for the definition of the A–polynomial). Recall that z is an eigenvalue of $\rho(\lambda)$. Denote by m the eigenvalue of $\rho(\mu)$ which shares its eigenvector with z. The A–polynomial of $m137$ is computed by SageMath [18] as

$$\begin{aligned}(z^4+2z^5+3z^6+z^7-z^8-3z^9-2z^{10}-z^{11})+m^2(-1-3z-2z^2-z^3\\ +2z^4+4z^5+z^6+4z^7+z^8+4z^9+2z^{10}-z^{11}-2z^{12}-3z^{13}-z^{14})\\ +m^4(-z^3-2z^4-3z^5-z^6+z^7+3z^8+2z^9+z^{10}).\end{aligned}$$

Write $A=-1-2z-3z^2-z^3+z^4+3z^5+2z^6+z^7=(z-1)(z^2+z+1)^3$ and $B=1+3z+2z^2+z^3-2z^4-4z^5-z^6-4z^7-z^8-4z^9-2z^{10}+z^{11}+2z^{12}+3z^{13}+z^{14}$. So the A–polynomial can be simplified as $-z^4A-Bm^2+z^3Am^4$. We are interested in the real solutions of

$$-z^4A-Bm^2+z^3Am^4=0. \tag{2-14}$$

Now consider the $(1,n)$ Dehn filling on $m137$. Then we are adding an extra relation $\mu\lambda^n=1$, which is $\rho(\mu)\rho(\lambda)^n=I$ under the representation ρ, ie

$$\rho(\mu)=\rho(\lambda)^{-n}=\begin{pmatrix} z^{-n} & * \\ 0 & z^n\end{pmatrix}.$$

Restricting to $\partial\mathcal{M}$ gives us the relation $m=z^{-n}$.

When n is negative, we shall write $n'=-n$. So we have $m=z^{n'}$. Plugging into (2-14) and dividing both sides by z^4, we get

$$-A-Bz^{2n'-4}+Az^{4n'-1}=0. \tag{2-15}$$

We will show the following lemma is true, completing the proof of Theorem 1(a).

Lemma 5 *Equation (2-15) has no real solutions when n' is large enough.*

Proof Define $F(z)=A(z^{4n'-1}-1)-Bz^{2n'-4}$. We'll show $F(z)>0$.

First notice that $A=0$ only when $z=1$. And $A>0$ when $z>1$ while $A<0$ when $z<1$. The polynomial B has 6 real roots, which are all simple: -2.3396, -1.4121, -0.7082, -0.4274, 0.8684, 1.1516 (rounded to the fourth digit).

As we saw earlier, the domain for s is

$$U := (-\infty, p_1) \cup [p_2, p_3] \cup (2, \infty) \approx (-\infty, -2.9032] \cup [-0.8061, 1.7093] \cup (2, \infty).$$

So the $|s| > 2$ condition restricts s to $(-\infty, p_1] \cup (2, \infty)$. Then

$$z \in V := (-\infty, -2.5038] \cup [-0.3994, 0) \cup (0, 1) \cup (1, \infty).$$

Notice that $z^7 A(1/z) = -A(z)$ and $z^{14} B(1/z) = B(z)$. Interchanging z with $1/z$ in $F(z)$ gives us $F(1/z) = A(1/z)(z^{-(4n'-1)} - 1) - B(1/z) z^{-(2n'-4)} = F(z)/z^{4n'+6}$. So we can assume $|z| < 1$.

Case 1 $(0.8684 \le z < 1)$ In this case, we have $A(z) < 0$, $B(z) \le 0$ and $z^{4n'-1} - 1 < 0$. So $F(z) > 0$.

Case 2 $(-0.3994 \le z < 0.8684$ and $z \ne 0)$ In this case, we have $A(z) < C_5 < 0$ and $C_6 > B(z) > 0$ for some constants C_5 and C_6. When n' is large enough, we have $|C_5| \times |(z^{4n'-1} - 1)| > C_6 z^{2n'-4}$. So $A(z^{4n'-1} - 1) = |A| \times |(z^{4n'-1} - 1)| > B z^{2n'-4}$ and it follows that $F(z) > 0$.

Therefore, when $n' = -n$ is large enough, we always have $F(z) > 0$ on the domain V. So (2-15) has no real solution when $n' \gg 0$. □

It follows from the above lemma that (2-14) has no real solution when $n \ll 0$ and thus the equality $\rho(\mu)\rho(\lambda)^n = I$ does not hold for $n \ll 0$.

From all the discussion above, we can now conclude that $\mathcal{M}(1, n)$ has no nontrivial $\mathrm{SL}_2(\mathbb{R})$ representation and thus no nontrivial $\mathrm{PSL}_2(\mathbb{R})$ representation for $n \ll 0$. Since the first Betti number of $\mathcal{M}(1, n)$ is 0, the lift of a trivial $\mathrm{PSL}_2(\mathbb{R})$ representation of $\pi_1(\mathcal{M}(1, n))$ into $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ will be trivial. So all representations of $\pi_1(\mathcal{M}(1, n))$ into $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ are trivial for $n \ll 0$, proving Theorem 1(a). □

In contrast, when n is positive there are examples of nontrivial $\mathrm{SL}_2(\mathbb{R})$ representations.

Plugging $m = z^{-n}$ into (2-14) and multiplying both sides by z^{4n-3}, we get

$$-A + B z^{2n-3} + A z^{4n+1} = 0.$$

Similarly define $G(z) = A(z^{4n+1} - 1) + B z^{2n-3}$. Since $G(1) = -4$ and $G(0.8684) > 0$, $G(z)$ must have at least one root in $[0.8684, 1)$. So $\pi_1(\mathcal{M}(1, n))$ has at least one nontrivial $\mathrm{SL}_2(\mathbb{R})$ representation for any $n > 0$. They lift to $\widetilde{\mathrm{PSL}}_2(\mathbb{R})$ representations, since the Euler number of any representation of an integral homology sphere vanishes [10, Section 6].

3 No L–space fillings

In this section, I will prove Theorem 1(b) using results from Gillespie [11], which are based on Rasmussen and Rasmussen [17]. In fact, I will show that none of the nonlongitudinal fillings of $m137$ is an L–space. The homology groups in this section are all homology with integral coefficients.

Suppose Y is a compact connected 3–manifold with a single torus as boundary. I will follow Gillespie's notation [17]. Define the set of slopes on ∂Y as

$$\mathcal{S}l(Y) = \{a \in H_1(\partial Y) \mid a \text{ is primitive}\}/\pm 1.$$

Define the set of L–space filling slopes of Y

$$\mathcal{L}(Y) = \{a \in \mathcal{S}l(Y) \mid Y(a) \text{ is an L–space}\}.$$

Moreover, Y is said to have genus 0 if $H_2(Y, \partial Y)$ is generated by a surface of genus 0.

We will use [11, Theorem 1.2], which is stated as:

Theorem 6 *The following are equivalent:*

(1) $\mathcal{L}(Y) = \mathcal{S}l(Y) - \{l\}$.

(2) Y *has genus* 0 *and has an L–space filling.*

Proof of Theorem 1(b) Let $l \in \mathcal{S}l(\mathcal{M})$ be the homological longitude. In our case l can be taken to be $[\lambda]$. I will show that none of the $(1, n)$ fillings to $\mathcal{M}$ is an L–space.

I will find one non-L–space filling first. SnapPy [6] shows that $(1, -1)$ filling on the knot 8_{20} complement with homological framing is homeomorphic to $m011(2, 3)$, which is also homeomorphic to $\mathcal{M}(1, -3)$. Ozsváth and Szabó showed that if some $(1, p)$ Dehn filling of a knot complement in S^3 with homological framing is an L–space, then the Alexander polynomial of the knot has coefficients ± 1 [16, Corollary 1.3]. We can compute with SnapPy [6] that the Alexander polynomial of 8_{20} is $x^4 - 2x^3 + 3x^2 - 2x + 1$. So $\mathcal{M}(1, -3)$ is not an L–space. Therefore,

$$-3l + [\mu] \notin \mathcal{L}(\mathcal{M}) \neq \mathcal{S}l(\mathcal{M}) - \{l\} \ni -3l + [\mu],$$

By Theorem 6, either $\mathcal{M}$ has no L–space fillings or $\mathcal{M}$ has positive genus.

The manifold $\mathcal{M}$ can be viewed as the complement of a knot K in $S^2 \times S^1$ [8]. This knot K intersects each S^2 three times. So $[K] \neq 0$ in $H_1(S^2 \times S^1; \mathbb{Z})$. It follows that $H_2(\mathcal{M}, \partial\mathcal{M})$ is generated by genus 0 surface $(S^2 \times \{P\}) \cap \mathcal{M}$ for generic point P on K. So $\mathcal{M}$ has genus 0, which forces $\mathcal{M}$ to have no L–space filling. Therefore, none of the integral homology spheres $\mathcal{M}(1, n)$ is an L–space. □

References

[1] **J Bowden**, *Approximating C^0–foliations by contact structures*, Geom. Funct. Anal. 26 (2016) 1255–1296 MR

[2] **S Boyer**, **C M Gordon**, **L Watson**, *On L-spaces and left-orderable fundamental groups*, Math. Ann. 356 (2013) 1213–1245 MR

[3] **P J Callahan**, **M V Hildebrand**, **J R Weeks**, *A census of cusped hyperbolic 3–manifolds*, Math. Comp. 68 (1999) 321–332 MR

[4] **D Cooper**, **M Culler**, **H Gillet**, **D D Long**, **P B Shalen**, *Plane curves associated to character varieties of 3–manifolds*, Invent. Math. 118 (1994) 47–84 MR

[5] **M Culler**, **N M Dunfield**, *Orderability and Dehn filling*, preprint (2016) arXiv

[6] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy, a computer program for studying the geometry and topology of 3–manifolds* Available at `http://snappy.computop.org`

[7] **M Culler**, **P B Shalen**, *Varieties of group representations and splittings of 3–manifolds*, Ann. of Math. 117 (1983) 109–146 MR

[8] **N M Dunfield**, *Examples of non-trivial roots of unity at ideal points of hyperbolic 3–manifolds*, Topology 38 (1999) 457–465 MR

[9] **X Gao**, *Non L–space integral homology 3–spheres with no nice orderings*, preprint (2016) arXiv

[10] **E Ghys**, *Groups acting on the circle*, Enseign. Math. 47 (2001) 329–407 MR

[11] **T J Gillespie**, *L–space fillings and generalized solid tori*, preprint (2016) arXiv

[12] **L C Jeffrey**, **J Weitsman**, *Bohr–Sommerfeld orbits in the moduli space of flat connections and the Verlinde dimension formula*, Comm. Math. Phys. 150 (1992) 593–630 MR

[13] **W H Kazez**, **R Roberts**, *Approximating $C^{1,0}$–foliations*, from "Interactions between low-dimensional topology and mapping class groups" (R I Baykur, J Etnyre, U Hamenstädt, editors), Geom. Topol. Monogr. 19, Geom. Topol. Publ. (2015) 21–72 MR

[14] **R Kirby**, *Problems in low-dimensional topology*, preprint (1995) Available at `https://math.berkeley.edu/~kirby/`

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

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

[17] **J Rasmussen**, **S D Rasmussen**, *Floer simple manifolds and L–space intervals*, preprint (2015) arXiv

[18] **W A Stein**, **et al**, *Sage mathematics software (Version 6.10)*, The Sage Development Team (2015) Available at http://www.sagemath.org

[19] **S Tillmann**, *Varieties associated to* 3*–manifolds: finding hyperbolic structures of and surfaces in* 3*–manifolds*, lecture notes (2003) Available at http://www.maths.usyd.edu.au/u/tillmann/research.html

[20] **R Zentner**, *Integer homology* 3*–spheres admit irreducible representations in* $\mathrm{SL}(2,\mathbb{C})$, preprint (2016) arXiv

Department of Mathematics, University of Illinois, Urbana-Champaign
Urbana, IL 61801, United States

xgao29@illinois.edu

Received: 5 October 2016 Revised: 7 January 2017 msp

Algebraic & Geometric Topology 17 (2017) 2523–2542

Noncommutative formality implies commutative and Lie formality

BASHAR SALEH

Over a field of characteristic zero we prove two formality conditions. We prove that a dg Lie algebra is formal if and only if its universal enveloping algebra is formal. We also prove that a commutative dg algebra is formal as a dg associative algebra if and only if it is formal as a commutative dg algebra. We present some consequences of these theorems in rational homotopy theory.

55P62

1 Introduction

Formality is an important concept in rational homotopy theory (see Deligne, Griffiths, Morgan and Sullivan [5]), deformation quantization (see Kontsevich [12]), deformation theory (see Goldman and Millson [8]) and other branches of mathematics where differential graded homological algebra is used. The notion of formality exists in many categories, eg the category of (commutative) dg associative algebras and the category of dg Lie algebras. An object A in such a category is called formal if there exists a zigzag of quasi-isomorphisms connecting A with its cohomology $H(A)$,

$$A \xleftarrow{\sim} B_1 \xrightarrow{\sim} \cdots \xleftarrow{\sim} B_n \xrightarrow{\sim} H(A).$$

A functor between categories in which the notion of formality exists may or may not preserve formal objects. For example, over a field of characteristic zero, it is known that the universal enveloping algebra functor $U\colon \mathbf{DGL}_{\Bbbk} \to \mathbf{DGA}_{\Bbbk}$ preserves formal objects; see Félix, Halperin and Thomas [6, Theorem 21.7]. That means that the formality of a dg Lie algebra (dgl) L implies the formality of UL (as a dg associative algebra (dga)). But what about the reversed relation? Does the formality of UL imply the formality of L? In this paper we show that this holds for dg Lie algebras over a field of characteristic zero.

Theorem 1.1 *A dg Lie algebra L over a field of characteristic zero is formal if and only if its universal enveloping algebra UL is formal as a dga.*

Among the results in the spirit of Theorem 1.1, there is a theorem by Aubry and Lemaire [1] saying that two dgl morphisms $f, g\colon L \to L'$ are homotopic if and only if

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2523

$U(f), U(g)\colon UL \to UL'$ are homotopic. The author does not think that the result by Aubry and Lemaire implies Theorem 1.1 or vice versa.

Milnor and Moore [17] showed that, over a field of characteristic zero, the universal enveloping algebra defines an equivalence of categories between the category of dg Lie algebras and the category of connected cocommutative dg Hopf algebras. This equivalence together with Theorem 1.1 and the fact that a dgl morphism $f\colon L \to L'$ is a quasi-isomorphism if and only if $U(f)\colon UL \to UL'$ is a quasi-isomorphism (see Félix, Halperin and Thomas [6, Theorem 21.7(ii)]) gives that a connected cocommutative dg Hopf algebra is formal as a connected cocommutative dg Hopf algebra if and only if it is formal as a dga.

We demonstrate a topological consequence of Theorem 1.1. The rational homotopy type of a simply connected space X is algebraically modeled by Quillen's dg Lie algebra $\lambda(X)$ over the rationals [18]. The space X is called coformal if $\lambda(X)$ is a formal dgl. It is known that there exists a zigzag of quasi-isomorphisms connecting $U\lambda(X)$ to the algebra $C_*(\Omega X, \mathbb{Q})$ of singular chains on the Moore loop space of X; see Félix, Halperin and Thomas [6, Chapter 26]. From Theorem 1.1 the following corollary is immediate:

Corollary 1.2 *Let X be a simply connected space. Then X is coformal if and only if $C_*(\Omega X; \mathbb{Q})$ is formal as a dga.*

Our second formality result is concerning the forgetful functor from the category of commutative dgas (cdgas) to the category of dgas. This functor preserves formality; a cdga which is formal as a cdga is obviously formal as a dga. Again, we ask whether this relation is reversible or not. We will prove that over a field of characteristic zero the answer is positive.

Theorem 1.3 *Let A be a cdga over a field of characteristic zero. Then A is formal as dga if and only if it is formal as a cdga.*

Recall that a space X is called rationally formal if the Sullivan–de Rham algebra $A_{\mathrm{PL}}(X; \mathbb{Q})$ is formal as a cdga; see Félix, Halperin and Thomas [6, Chapter 12]. In that case the rational homotopy type of X is a formal consequence of its cohomology $H^*(X; \mathbb{Q})$, meaning that $H^*(X; \mathbb{Q})$ determines the rational homotopy type of X. Moreover, it is known that there exists a zigzag of quasi-isomorphisms connecting $A_{\mathrm{PL}}(X; \mathbb{Q})$ with the singular cochain algebra $C^*(X; \mathbb{Q})$ of X [6, Theorem 10.9]. An immediate topological consequence is the following corollary:

Corollary 1.4 *A space X is rationally formal if and only if the singular cochain algebra $C^*(X; \mathbb{Q})$ of X is formal as a dga.*

Overview

The reader is assumed to be familiar with the theory of operads and with the notions of A_∞–, C_∞–, and L_∞–algebras. We refer the reader to Keller [11], Loday and Vallette [14] and Markl, Shnider and Stasheff [16] for introductions to these subjects.

In Section 2 we review Baranovsky's universal enveloping construction [2] on the category of L_∞–algebras. The construction is a generalization of the universal enveloping algebra functor and is an important ingredient in the proof of Theorem 1.1. In Section 3 we present an obstruction theory for formality in different categories. The obstructions will be cohomology classes of certain cohomology groups. The obstruction theory together with Baranovsky's universal enveloping will give us tools to compare the concept of formality in $\mathbf{DGA}_{\Bbbk}$ and $\mathbf{DGL}_{\Bbbk}$ ($\operatorname{char} \Bbbk = 0$). This will be treated in Section 4 and will finally yield a proof of Theorem 1.1. In Section 5 we prove Theorem 1.3.

The reader interested only in Theorem 1.1 may skip Section 5, whilst the reader only interested in Theorem 1.3 may skip Sections 2 and 4.

Conventions

- S_k denotes the symmetric group on k letters.
- The Koszul sign of a permutation $\sigma \in S_k$ acting on $v_1 \cdots v_k \in V^{\otimes k}$ (where V is a graded vector space) is given by the following rule: The Koszul sign of an adjacent transposition that permutes x and y is given by $(-1)^{|x||y|}$. This is then extended multiplicatively to all of S_k (recall that the set of adjacent transpositions generates S_k).
- The suspension sV of a graded vector space V is the graded vector space given by $sV^i = V^{i+1}$. The suspension of a cochain complex (C, d) is the cochain complex $(sC, -sds^{-1})$.
- A standing assumption will be that $\Bbbk$ is a field of characteristic zero. We will only consider (co)algebras and (co)operads over fields of characteristic zero.

Acknowledgments I would like to thank my advisor Alexander Berglund for his invaluable guidance during the preparation of this paper and also for proposing the topics treated here. I would also like to thank Stephanie Ziegenhagen for her careful reading of this paper and her comments and suggestions. Thanks to Kaj Börjeson and Felix Wierstra for introducing me to many concepts that I had very little knowledge about. Finally, I would like to thank Victor Protsak for helping me with the proof of Proposition 4.2 (at MathOverflow).

2 Baranovsky's universal enveloping for L_∞–algebras

The proof of Theorem 1.1 will partly rely on a construction by Baranovsky [2] that generalizes the universal enveloping algebra construction to L_∞–algebras.

Applying Baranovsky's universal enveloping (denoted by U^{Bar}) to an L_∞–algebra $(L, \{l_i\})$ gives an A_∞–algebra $U^{\mathrm{Bar}}(L, \{l_i\}) = (\Lambda L, \{m_i\})$, where ΛL is the underlying graded vector space of the symmetric algebra on L. Applying U^{Bar} to an L_∞–morphism $\phi\colon L \to L'$ gives an A_∞–morphism $U^{\mathrm{Bar}}(\phi)\colon U^{\mathrm{Bar}}(L) \to U^{\mathrm{Bar}}(L')$.

U^{Bar} is not a functor since it fails to preserve compositions (ie in general $U^{\mathrm{Bar}}(\psi\circ\phi) \neq U^{\mathrm{Bar}}(\psi)\circ U^{\mathrm{Bar}}(\phi)$). However, the restriction of U^{Bar} to $\mathbf{DGL}_{\Bbbk} \subset \infty$–$L_\infty$–**alg** (here ∞–L_∞–**alg** denotes the category of L_∞–algebras with ∞–morphisms) coincides with the usual universal enveloping algebra functor, denoted by U.

We record some properties of U^{Bar}.

Theorem 2.1 *Let $(L, \{l_i\})$ be an L_∞–algebra with universal enveloping*

$$U^{\mathrm{Bar}}(L, \{l_i\}) = (\Lambda L, \{m_i\}).$$

The following properties hold:

(a) *$m_1\colon \Lambda L \to \Lambda L$ is the symmetrization of l_1 (ie $m_1 = \Lambda(l_1)$).*

(b) *If $\phi\colon (L, \{l_i\}) \to (L', \{l'_i\})$ is an L_∞–quasi-isomorphism, then*

$$U^{\mathrm{Bar}}(\phi)\colon (\Lambda L, \{m_i\}) \to (\Lambda L', \{m'_i\})$$

is an A_∞–quasi-isomorphism.

(c) *The map $m_j\colon (\Lambda L)^{\otimes j} \to \Lambda L$ depends only on L, $l_1, l_2, \dots, l_j$. In particular, if $(L, \{k_i\})$ is another L_∞–algebra structure on the same vector space L with $l_j = k_j$ for $j = 1, 2, \dots, d$, then $U^{\mathrm{Bar}}(L, \{k_i\}) = (\Lambda L, \{n_i\})$ with $n_j = m_j$ for $j = 1, 2, \dots, d$.*

(d) *Let $v_1, \dots, v_j \in L \subset U^{\mathrm{Bar}}L$. Then*

$$l_j(v_1 \cdots v_j) = \sum_{\sigma\in S_j} \gamma(\sigma; v_1, \dots, v_j) m_j(v_{\sigma^{-1}(1)} \cdots v_{\sigma^{-1}(j)}),$$

where $\gamma(\sigma; v_1, \dots, v_n)$ is the product of the sign of the permutation σ and the Koszul sign obtained by applying σ on $v_1 \cdots v_j$.

(e) *The restriction $U^{\mathrm{Bar}}|_{\mathbf{DGL}_{\Bbbk}}$ of U^{Bar} to $\mathbf{DGL}_{\Bbbk}$ coincides with the ordinary universal enveloping algebra functor.*

Properties (a)–(c) are not explicitly stated in [2], so we will briefly recall Baranovsky's construction in order to prove these properties.

A summary of the construction

Given a complex (V,d), let $T_a^*(V)$ (resp. $T_c^*(V)$) and $\Lambda_a^*(V)$ (resp. $\Lambda_c^*(V)$) denote the tensor and symmetric algebras (resp. coalgebras) on V with (co)differential corresponding to the unique (co)derivation extension of d.

Let $(L,\{l_i\})$ be an L_∞–algebra. We start by considering the complex (L,l_1) and construct from it two coalgebras, $(T_c^*(s\overline{\Lambda_a^*(L)}),d^\circ)$ and $\bigl(T_c^*\bigl(s\overline{\Omega(\Lambda_c^*(sL))}\bigr),\delta^\circ\bigr)$ (where Ω denotes the cobar construction and $\overline{(\cdot)}$ denotes the augmentation ideal).

Baranovsky shows that there exists a coalgebra contraction from $T_c^*\bigl(s\overline{\Omega(\Lambda_c^*(sL))}\bigr)$ to $T_c^*(s\overline{\Lambda_a^*(L)})$

$$H \circlearrowright T_c^*\bigl(s\overline{\Omega(\Lambda_c^*(sL))}\bigr) \underset{G}{\overset{F}{\rightleftarrows}} T_c^*(s\overline{\Lambda_a^*(L)}) \quad . \tag{2-1}$$

By comparing $T_c^*\bigl(s\overline{\Omega(\Lambda_c^*(sL))}\bigr)$ with the cobar–bar construction on the Chevalley–Eilenberg construction on the L_∞–algebra $(L,\{l_i\})$, denoted by $B\Omega C(L)$, we see that they only differ by their differentials. The differential δ of $B\Omega C(L)$ is given by

$$\delta = \delta^\circ + t_\mu + t_L,$$

where t_μ is the part that encodes the multiplication on $\Omega C(L)$ and $t_L = t_2 + t_3 + \cdots$ encodes the L_∞–structure on L with t_i encoding l_i. Applying the basic perturbation lemma to the perturbation $t_\mu + t_L$ of the contraction above results in a new differential $d = (d^\circ)_{t_\mu + t_L}$ on $T_c^*(s\overline{\Lambda_a^*(L)})$, which corresponds to an A_∞–algebra structure on ΛL, which will be Baranovsky's universal enveloping $U^{\mathrm{Bar}}(L,\{l_i\})$.

Geometric grading

Baranovsky introduces a geometric grading on $B\Omega C(L)$ by first declaring that an element of $s^{-1}\Lambda_c^k(sL)$ is of geometric degree $k-1$ and then extends the grading to $B\Omega C(L)$ by the following rule: the geometric degree of $\alpha\otimes\beta$ is the sum of the geometric degrees of α and β. The maps in the contraction (2-1) and the perturbations t_μ and $t_L = t_2 + t_3 + \cdots$ satisfy some conditions regarding the geometric grading:

- The image of G belongs to the geometric degree 0 part.
- H increases the geometric degree by 1.
- t_μ preserves the geometric degree.
- t_i decreases the geometric degree by $i-1$ and vanishes on elements of geometric degree $< i-1$.

Proof of Theorem 2.1 By the basic perturbation lemma (stated in [2, Lemma 2]) we have that the differential $d = (d^\circ)_{t_\mu + t_L}$ is given by

$$d = d^\circ + F\left(\sum_{i\geq 0}((t_\mu + t_L)H)^i\right)(t_\mu + t_L)G.$$

Since the image of G belongs to the geometric degree 0 part and since $t_L = t_2 + t_3 + \cdots$ vanishes on elements of geometric degree 0, we may rewrite the differential as

$$d = d^\circ + F\left(\sum_{i\geq 0}(t_\mu H + t_2 H + t_3 H + \cdots)^i\right)t_\mu G. \tag{2-2}$$

The terms in the differential above that correspond to $m_n\colon U^{\mathrm{Bar}}(L)^{\otimes n} \to U^{\mathrm{Bar}}(L)$ are those terms that contain t_μ exactly $n-1$ times (see the proof of [2, Theorem 3] for the details).

(a) Since d° is the only term in (2-2) that does not contain t_μ as a factor, we have that d° is the part of the differential d that corresponds to $m_1\colon U^{\mathrm{Bar}}(L) \to U^{\mathrm{Bar}}(L)$. One can easily see that d° corresponds to $\Lambda(l_1)\colon \Lambda L \to \Lambda L$.

(b) By [2, Theorem 3.i] we have that the first component $U^{\mathrm{Bar}}(\phi)_1$ of $U^{\mathrm{Bar}}(\phi)$ is given by $\Lambda(\phi_1)$, where ϕ_1 is the first component of ϕ. In order to show that $U^{\mathrm{Bar}}(\phi)$ is an A_∞–quasi-isomorphism, we need to show that

$$U^{\mathrm{Bar}}(\phi)_1 = \Lambda(\phi_1)\colon (\Lambda L, m_1) \to (\Lambda L', m_1') \tag{2-3}$$

is a quasi-isomorphism of complexes. Since ϕ is an L_∞–quasi-isomorphism, it follows that $\phi_1\colon (L, l_1) \to (L', l_1')$ is a quasi-isomorphism of complexes. By (a), m_1 and m_1' are given by symmetrizations of l_1 and l_1', respectively, which means that the map in (2-3) is obtained by applying the symmetrization functor $\Lambda(-)$ on $\phi_1\colon (L, l_1) \to (L', l_1')$. Over a field $\Bbbk$ of characteristic zero we have that the symmetrization functor $\Lambda(-)$ preserves quasi-isomorphisms (since $L \otimes_{\Bbbk} -$ is exact and taking S_n–coinvariants is also exact), and (b) follows.

(c) Firstly, H depends only on L and l_1, by [2, Theorem 1]. Moreover, we have that $t_\mu H$ increases the geometric degree by 1 while $t_i H$ decreases the geometric degree by $i-2$. Furthermore, $t_i H$ vanish on elements of degree $< i-2$. That means if there exists a nonzero term containing $t_i H$, then $t_\mu H$ has to occur at least $i-2$ times before $t_i H$ (ie to the right of $t_i H$).

We have that m_n corresponds to those nonzero terms that contain t_μ exactly $n-1$ times, which is equivalent to those terms that contain $t_\mu H$ exactly $n-2$ times. These terms cannot contain any $t_i H$ where $i > n$ (since they are nonzero). From this and the fact that t_i is completely encoded by l_i, claim (c) follows.

(d)–(e) See [2, Theorem 3.vii] and [2, Theorem 3.v], respectively. □

3 Minimal $\mathscr{P}_\infty$–algebras and obstructions to formality

Given an algebraic operad $\mathscr{P}$, we have that the cohomology of a dg $\mathscr{P}$–algebra has an induced dg $\mathscr{P}$–algebra structure with a trivial differential [14, Proposition 6.3.5]. Thus, the notion of formality makes sense in the category of dg $\mathscr{P}$–algebras.

If $\mathscr{P}$ is a Koszul operad, we denote by $\mathscr{P}_\infty$ the operad obtained by applying the cobar construction on the Koszul dual cooperad of $\mathscr{P}$ [14, Chapter 10]. The category of $\mathscr{P}_\infty$–algebras with $\mathscr{P}_\infty$–morphisms (denoted by ∞–$\mathscr{P}_\infty$–**alg**) contains the category of $\mathscr{P}$–algebras as a subcategory and has some properties that the category of $\mathscr{P}$–algebras lacks, eg that quasi-isomorphisms are invertible up to homotopy.

Theorem 3.1 [14, Theorem 11.4.9] *Let $\mathscr{P}$ be a Koszul operad over a field of characteristic zero and let A be a dg $\mathscr{P}$–algebra. Then A is formal as a $\mathscr{P}$–algebra if and only if there exists a $\mathscr{P}_\infty$–algebra quasi-isomorphism $A \to H(A)$.*

In this paper we will be interested in algebras over the operads $\mathscr{A}ss$, $\mathscr{C}om$ and $\mathscr{L}ie$, which are all Koszul. From now on, $\mathscr{P}$ is either $\mathscr{A}ss$, $\mathscr{C}om$ or $\mathscr{L}ie$, which means that a dg $\mathscr{P}$–algebras is either a dga, cdga or dgl, and that a $\mathscr{P}_\infty$–algebra is either an A_∞–, C_∞– or L_∞–algebra.

We denote the Koszul dual operad of $\mathscr{P}$ by $\mathscr{P}^!$ (recall that $\mathscr{A}ss^! = \mathscr{A}ss$, $\mathscr{C}om^! = \mathscr{L}ie$ and $\mathscr{L}ie^! = \mathscr{C}om$). We have that a $\mathscr{P}_\infty$–algebra structure on a vector space A is a collection $(A, \{b_n\})$, where $b_n\colon \mathscr{P}^!(n) \otimes_{S_n} A^{\otimes n} \to A$ for $n \geq 1$ are linear maps of degree $n-2$ that satisfy certain compatibility conditions (see [14]). A dg $\mathscr{P}$–algebra (A, b_1, b_2) may be regarded as a $\mathscr{P}_\infty$–algebra by identifying (A, b_1, b_2) with $(A, b_1, b_2, 0, 0, \dots)$. A morphism of $\mathscr{P}_\infty$–algebras $\phi\colon (A, \{b_n\}) \to (A', \{b'_n\})$ is a collection $\phi = (\phi_n)$, where the ϕ_n are maps $\mathscr{P}^!(n) \otimes A^{\otimes n} \to A'$ of degree $n-1$ that satisfy certain conditions.

Given an operad $\mathscr{P}$ there is a notion of the operadic cochain complex $C^*_{\mathscr{P}}(A)$ of a $\mathscr{P}$–algebra A, where $C^n_{\mathscr{P}}(A) = \operatorname{Hom}(\mathscr{P}^!(n) \otimes_{S_n} A^{\otimes n}, A)$ (see [14, Chapter 12] for details). We have that $C^*_{\mathscr{A}ss}(A)$ is the Hochschild cochain complex of A, $C^*_{\mathscr{C}om}(C)$ is the Harrison cochain complex of C, and $C^*_{\mathscr{L}ie}(L)$ is the Chevalley–Eilenberg cochain complex of L. Since we will consider $\mathscr{P}$–algebras with nontrivial homological grading, the operadic cohomology will be endowed with a nontrivial homological grading, and $C^{n,p}_{\mathscr{P}}(A)$ will denote the part of $\operatorname{Hom}(\mathscr{P}^!(n) \otimes_{S_n} A^{\otimes n}, A)$ that is of homological degree $p \in \mathbb{Z}$.

The main goal of this section is to present an obstruction theory for formality in $\mathbf{DGA}_{\Bbbk}$, $\mathbf{CDGA}_{\Bbbk}$ and $\mathbf{DGL}_{\Bbbk}$ over any field $\Bbbk$ of characteristic zero. This obstruction theory is presumably well-known to experts, but we will recall it and formulate it in a way that is suitable for the context of this paper. In order to do that we need to recall some results by Kadeishvili [10] on minimal A_∞–algebras and the Hochschild cochain complex, and minimal C_∞–algebras and the Harrison cochain complex. The ideas of Kadeishvili apply also to minimal L_∞–algebras and the Chevalley–Eilenberg cochain complex (we leave the details to the reader).

Minimal $\mathscr{P}_\infty$–algebras

We will now present some results by Kadeishvili [10].

Definition 3.2 Let $\mathscr{P} = \mathscr{A}ss$, $\mathscr{C}om$ or $\mathscr{L}ie$. A $\mathscr{P}_\infty$–algebra $(H, \{b_i\})$ is called *minimal* if $b_1 = 0$.

Given a minimal $\mathscr{P}_\infty$–algebra $(H, 0, b_2, b_3, \dots)$, we have that $\mathcal{H} = (H, 0, b_2)$ is a $\mathscr{P}$–algebra, and therefore it makes sense to consider the operadic cochain complex $C^*_{\mathscr{P}}(\mathcal{H})$ of $\mathcal{H}$.

Proposition 3.3 [10] *Suppose $\mathscr{P} = \mathscr{A}ss$, $\mathscr{C}om$ or $\mathscr{L}ie$. Then the following holds:*

(a) *Let $(H, \{b_i\})$ and $(H, \{b'_i\})$ be two minimal $\mathscr{P}_\infty$–algebras with $b_2 = b'_2$ and let $\phi = (\mathrm{id}, 0, \dots, 0, \phi_k, \phi_{k+1}, \dots)\colon (H, \{b_i\}) \to (H, \{b'_i\})$ be a $\mathscr{P}_\infty$–algebra isomorphism. The formal sums*

$$\bar{b} = b_3 + b_4 + \cdots, \quad \bar{b}' = b'_3 + b'_4 + \cdots, \quad \bar{\phi} = \phi_k + \phi_{k+1} + \cdots$$

in $C^*_{\mathscr{P}}(\mathcal{H})$, where $\mathcal{H} = (H, 0, b_2)$, satisfy the equality

$$\bar{b} - \bar{b}' = \partial_{\mathscr{P}}(\bar{\phi}) + (\text{elements in } C^{\geq k+2}(\mathcal{H})),$$

where $\partial_{\mathscr{P}}$ is the differential of $C^*_{\mathscr{P}}(\mathcal{H})$.

(b) *Let $(H, \{b_i\})$ be a minimal $\mathscr{P}_\infty$–algebra, and let $\{\phi_n \in C^{n,n-2}_{\mathscr{P}}(\mathcal{H})\}_{n\geq 2}$ be any collection of maps. Then there exists a minimal $\mathscr{P}_\infty$–algebra $(H, \{b'_i\})$ with $b'_2 = b_2$ such that $\phi = (\mathrm{id}, \phi_2, \phi_3, \dots)$ is a $\mathscr{P}_\infty$–algebra isomorphism $(H, \{b_i\}) \to (H, \{b'_i\})$.*

Obstruction to formality

We will, in the spirit of Halperin and Stasheff [9], present an obstruction theory for $\mathscr{P}$–algebra formality that is presumably well-known to experts. However, the author could not find in the literature an exposition that was optimized for the context of this paper. Obstructions to formality in $\mathbf{CDGA}_{\Bbbk}$ are treated in [9] and obstructions to formality in $\mathbf{DGL}_{\Bbbk}$ are treated in [15].

We start by recalling an easy consequence of the homotopy transfer theorem for $\mathscr{P}_\infty$–algebras, where $\mathscr{P}$ is a Koszul operad.

Proposition 3.4 *Let $\mathscr{P} = \mathscr{A}ss$, $\mathscr{C}om$ or $\mathscr{L}ie$ and let $(A, \bar{b}_1, \bar{b}_2)$ be a dg $\mathscr{P}$–algebra. Then there exists a $\mathscr{P}_\infty$–algebra structure $(H(A), 0, b_2, b_3, \dots)$ on the underlying vector space of the cohomology $H(A)$ such that*

- (i) *$b_2\colon H(A)^{\otimes 2} \to H(A)$ is the induced $\mathscr{P}$–algebra multiplication on the cohomology $H(A)$, and*
- (ii) *$(A, \bar{b}_1, \bar{b}_2)$ is $\mathscr{P}_\infty$–quasi-isomorphic to $(H(A), 0, b_2, b_3, \dots)$.*

Proof Since $\mathscr{A}ss$, $\mathscr{C}om$ and $\mathscr{L}ie$ are all Koszul, the theorem follows easily from the homotopy transfer theorem for $\mathscr{P}_\infty$–algebras (see [14, Section 10.3] or [4]). □

Remark 3.5 $(A, \bar{b}_1, \bar{b}_2)$ is formal if and only if there exists a $\mathscr{P}_\infty$–algebra quasi-isomorphism $(H(A), 0, b_2, b_3, \dots) \to (H(A), 0, b_2)$ (recall that quasi-isomorphisms are invertible up to homotopy in the category of $\mathscr{P}_\infty$–algebras). Thus, an obstruction theory for quasi-isomorphisms $(H, 0, b_2, b_3, \dots) \to (H, 0, b_2)$ is an obstruction theory for formality.

Now we are ready to formulate the main theorem of this section.

Theorem 3.6 *Assume $\mathscr{P} = \mathscr{A}ss$, $\mathscr{C}om$ or $\mathscr{L}ie$ and that $\mathcal{H} = (H, 0, b_2)$ is a dg $\mathscr{P}$–algebra with trivial differential. Given a $\mathscr{P}_\infty$–algebra of the form $(H, 0, b_2, b_3, \dots)$, there is an associated sequence of cohomology classes $[b_3], [b'_4], [b''_5], \dots,$ where $[b_k^{(k-3)}] \in H_{\mathscr{P}}^{k,k-1}(\mathcal{H})$. This sequence is either an infinite sequence of vanishing cohomology classes, or finite and terminating in a nonzero cohomology class $[b_k^{(k-3)}]$. There exists a $\mathscr{P}_\infty$–algebra quasi-isomorphism $(H, 0, b_2, b_3, \dots) \to (H, 0, b_2)$ if and only if $[b_3], [b'_4], [b''_5], \dots$ is an infinite sequence of vanishing cohomology classes.*

This theorem will follow easily from the following proposition:

Proposition 3.7 *Assume $\mathscr{P} = \mathscr{A}ss$, $\mathscr{C}om$ or $\mathscr{L}ie$. Let $\mathcal{H} = (H, 0, m_2)$ be a given minimal dg $\mathscr{P}$–algebra.*

- (a) *Let $H_\alpha = (H, 0, m_2, 0, \dots, 0, m_k, m_{k+1}, \dots)$ with $k \geq 3$ be a $\mathscr{P}_\infty$–algebra that is quasi-isomorphic to $\mathcal{H} = (H, 0, m_2)$. Then m_k is a boundary in $C^*_{\mathscr{P}}(\mathcal{H})$, ie $[m_k] = 0$ in $H^*_{\mathscr{P}}(\mathcal{H})$.*
- (b) *Given a $\mathscr{P}_\infty$–algebra $H_\alpha = (H, 0, m_2, 0, \dots, 0, m_k, m_{k+1}, \dots)$, if $[m_k] = 0$ in $H^*_{\mathscr{P}}(\mathcal{H})$, ie $m_k = \partial_{\mathscr{P}}(\phi_{k-1})$ for some $\phi_{k-1} \in C^{k-1}_{\mathscr{P}}(\mathcal{H})$, then H_α is quasi-isomorphic to some $\mathscr{P}_\infty$–algebra H_β of the form*

$$H_\beta = (H, 0, m_2, 0, \dots, 0, m'_{k+1}, m'_{k+2}, \dots)$$

Remark 3.8 If all obstructions from Theorem 3.6 vanish, we will get a sequence of quasi-isomorphisms

$$(H, 0, m_2, m_3, \dots) \to (H, 0, m_2, 0, m'_4, m'_5, \dots) \to (H, 0, m_2, 0, 0, m''_5, m''_6, \dots) \to \cdots.$$

One can easily see that the colimit of this diagram is $(H, 0, m_2, 0, \dots)$. Since quasi-isomorphisms between minimal $\mathscr{P}_\infty$–algebras are isomorphisms, it follows that $(H, 0, m_2, m_3, \dots) \to (H, 0, m_2, 0, \dots)$ is an isomorphism, hence a quasi-isomorphism.

Proof (a) By Lemma A.5, there exists a morphism

$$\phi = (\mathrm{id}, 0, \dots, 0, \phi_{k-1}, \phi_k, \dots)\colon H_\alpha \to \mathcal{H}.$$

It follows from Proposition 3.3(a) that

$$m_k + m_{k+1} + \cdots = (\partial_{\mathscr{P}}(\phi_{k-1}) + \partial_{\mathscr{P}}(\phi_k) + \cdots) + (\text{elements in } C^{\geq k+1}_{\mathscr{P}}(\mathcal{H})).$$

Collecting the elements of $C^k_{\mathscr{P}}(\mathcal{H})$ from both sides of the equality gives that $m_k = \partial_{\mathscr{P}}(\phi_{k-1})$.

(b) By Proposition 3.3(b) there exists a $\mathscr{P}_\infty$–algebra $H_\beta = (H, 0, m_2, m'_3, m'_4, \dots)$ such that

$$(\mathrm{id}, 0, \dots, 0, \phi_{k-1}, 0, \dots)\colon H_\alpha \to H_\beta$$

is a $\mathscr{P}_\infty$–algebra isomorphism. By Proposition 3.3(a) we have that

$$(m_k + m_{k+1} + \cdots) - (m'_3 + m'_4 + \cdots) = \partial_{\mathscr{P}}(\phi_{k-1}) + (\text{elements in } C^{\geq k+1}_{\mathscr{P}}(\mathcal{H})).$$

We see from the equality that $m'_3, \dots, m'_{k-1}$ vanish. We also see that $m_k - m'_k = \partial_{\mathscr{P}}(\phi_{k-1})$, giving that $m'_k = 0$. This completes the proof. □

4 Proof of Theorem 1.1

We used the language of operadic cohomology in the obstruction theory for formality in the previous section. We will compare different cohomology theories corresponding to different operads in order to compare the concept of formality in different categories. Recall that $H^*_{\mathscr{A}ss}$ and $H^*_{\mathscr{L}ie}$ correspond to the Hochschild and the Chevalley–Eilenberg cohomologies, respectively. The Hochschild cochain complex of an associative algebra A with coefficients in A will be denoted by $C^*_{\mathrm{Hoch}}(A)$ and its cohomology will be denoted by $HH^*(A)$. The Chevalley–Eilenberg cochain complex of a Lie algebra L with coefficients in L will be denoted by $C^*_{\mathrm{CE}}(L)$ and its cohomology will be denoted by $H^*_{\mathrm{CE}}(L)$. We will also work with the Chevalley–Eilenberg cochain complex of a Lie algebra with coefficients in a left L–module M different from L, which will be denoted by $C^*_{\mathrm{CE}}(L, M)$; its cohomology will be denoted by $H^*_{\mathrm{CE}}(L, M)$.

Hochschild and Chevalley–Eilenberg cohomology

Recall that the universal enveloping algebra UL of a dg Lie algebra L is explicitly given by

$$UL = T^*_a(L)/\big(ab - (-1)^{|a||b|}ba - [a, b] \mid a, b \in L\big).$$

A Lie algebra L is of course a left module over itself via $g.h = [g,h]$.

Let UL^{ad} denote the left L–module structure on UL given by

$$g.m = g \otimes m - (-1)^{|g||m|} m \otimes g$$

for $g \in L$ and $m \in UL$ (where m is of some homogenous degree $|m|$). This makes the inclusion $L \hookrightarrow UL^{\mathrm{ad}}$ a map of left L–modules.

Lemma 4.1 [13, Lemma 3.3.3] *There exists a cochain map*

$$\mathrm{Alt}\colon C^*_{\mathrm{Hoch}}(UL) \to C^*_{\mathrm{CE}}(L, UL^{\mathrm{ad}})$$

from the Hochschild cochain complex of UL to the Chevalley–Eilenberg cochain complex of L with coefficients in UL^{ad}. If $f \in C^n_{\mathrm{Hoch}}(UL) = \mathrm{Hom}_{\Bbbk}(UL^{\otimes n}, UL)$, then $\mathrm{Alt}(f) \in C^n_{\mathrm{CE}}(L, UL^{\mathrm{ad}}) = \mathrm{Hom}_{\Bbbk}(L^{\wedge n}, UL^{\mathrm{ad}})$ is given by

$$\mathrm{Alt}(f)(l_1 \wedge \cdots \wedge l_n) = \sum_{\sigma \in S_n} \gamma(\sigma, l_1, \ldots, l_n) f(l_{\sigma^{-1}(1)} \otimes \cdots \otimes l_{\sigma^{-1}(n)}),$$

where $\gamma(\sigma; l_1, \ldots, l_n)$ is the product of the sign of σ and the Koszul sign obtained by applying σ on $l_1 \cdots l_j$.

By the map above we have a tool for comparison of cohomology classes in $HH^*(UL)$ and $H^*_{\mathrm{CE}}(L, UL^{\mathrm{ad}})$. However, the obstruction theory for formality in $\mathbf{DGL}_{\Bbbk}$ was expressed in terms of cohomology classes in $H^*_{\mathrm{CE}}(L)$ (ie $H^*_{\mathrm{CE}}(L,L)$). In the next proposition we show that the inclusion $L \hookrightarrow UL^{\mathrm{ad}}$ of left L–modules induces an injection $H^*_{\mathrm{CE}}(L,L) \to H^*_{\mathrm{CE}}(L, UL^{\mathrm{ad}})$ in cohomology.

Proposition 4.2 *The inclusion of L–modules $L \hookrightarrow UL^{\mathrm{ad}}$ induces an injection*

$$H^*_{\mathrm{CE}}(L,L) \to H^*_{\mathrm{CE}}(L, UL^{\mathrm{ad}})$$

in cohomology.

Proof We start by recalling the Poisson algebra structure on $\Lambda_a L$ (see [13, Section 3.3.4]). The Poisson bracket $\{-,-\}$ on $\Lambda_a L$ is determined by the following two properties: (i) $\{g,h\} = [g,h]$ for $g, h \in L$, and (ii) $\{-,-\}$ is a derivation in each variable. Now we may give ΛL a left L–module structure, given by $g.\alpha = \{g, \alpha\}$. With this L–module structure, the Poincaré–Birkhoff–Witt isomorphism $\eta\colon \Lambda L \to UL^{\mathrm{ad}}$ is an L–module morphism [13, Lemma 3.3.5]. In particular, ΛL and UL^{ad} are isomorphic as L–modules. Since L is a direct summand of the L–module ΛL, it follows by the L–module isomorphism above that L is also a direct summand of UL^{ad}. Hence, there is a projection $UL^{\mathrm{ad}} \twoheadrightarrow L$ of L–modules, and therefore id_L may be decomposed as $L \hookrightarrow UL^{\mathrm{ad}} \twoheadrightarrow L$. This in turn gives a decomposition

$$\mathrm{id}_{H^*_{\mathrm{CE}}(L,L)}\colon H^*_{\mathrm{CE}}(L,L) \to H^*_{\mathrm{CE}}(L,UL^{\mathrm{ad}}) \to H^*_{\mathrm{CE}}(L,L).$$

Thus, $H^*_{\mathrm{CE}}(L,L) \to H^*_{\mathrm{CE}}(L,UL^{\mathrm{ad}})$ must be injective. □

The proof

In this section it will be necessary to be able to distinguish between a dg Lie algebra $(L,\bar{l}_1,\bar{l}_2)$ and the underlying vector space L. Therefore we will denote the Lie algebra structure by $\mathcal{L}$ and the underlying vector space by L. We will denote the Lie algebra structure on the cohomology of $\mathcal{L}$ by $H(\mathcal{L})$ while its underlying vector space will be denoted by $H(L)$. We make the same distinction between $U\mathcal{L}$ and UL.

Lemma 4.3 [6, Theorem 21.7] *Suppose* $\mathrm{char}\,\Bbbk = 0$ *and* $\mathcal{L} \in \mathbf{DGL}_\Bbbk$. *Then there exists a natural isomorphism* $UH(\mathcal{L}) \cong H(U\mathcal{L})$ *of algebras.*

It follows directly from the lemma that $U\colon \mathbf{DGL}_\Bbbk \to \mathbf{DGA}_\Bbbk$ preserves formality. Thus, what is left to show in order to prove Theorem 1.1 is that if $U\mathcal{L}$ is formal in $\mathbf{DGA}_\Bbbk$, then $\mathcal{L}$ is formal in $\mathbf{DGL}_\Bbbk$. In the language of A_∞– and L_∞–algebras, we need to prove the following:

Theorem 4.4 *Let* $\Bbbk$ *be a field of characteristic zero and let* $\mathcal{L} \in \mathbf{DGL}_\Bbbk$. *If there exists an* A_∞*–quasi-isomorphism* $U\mathcal{L} \to H(U\mathcal{L})$ *then there exists an* L_∞*–quasi-isomorphism* $\mathcal{L} \to H(\mathcal{L})$.

Proof Let $\mathcal{L}$ be on the form $\mathcal{L} = (L,\bar{l}_1,\bar{l}_2)$ and let $U\mathcal{L} = (UL,\bar{m}_1,\bar{m}_2)$ be its universal enveloping algebra.

By the homotopy transfer theorem for L_∞–algebras (Proposition 3.4) there exists an L_∞–algebra $(H^*(L),0,l_2,l_3,\dots)$ and an L_∞–quasi-isomorphism

$$\phi\colon \mathcal{L} \to (H(L),0,l_2,l_3,\dots).$$

Applying U^{Bar} to ϕ gives an A_∞–quasi-isomorphism

$$U^{\mathrm{Bar}}(\phi)\colon U^{\mathrm{Bar}}(\mathcal{L}) \to U^{\mathrm{Bar}}(H(L),0,l_2,l_3,\dots) \tag{4-1}$$

(recall from Theorem 2.1(b) that U^{Bar} preserves quasi-isomorphisms). By Theorem 2.1(e), $U^{\mathrm{Bar}}(\mathcal{L})$ is the ordinary universal enveloping algebra $U\mathcal{L}$ of $\mathcal{L}$. Let us analyze the A_∞–structure of $U^{\mathrm{Bar}}(H^*(L),0,l_2,l_3,\dots)$.

Claim $U^{\mathrm{Bar}}(H(L),0,l_2,l_3,\dots)$ *is an* A_∞*–algebra* $(H(UL),0,m_2,m_3,\dots)$ *whose* 2*–truncation,* $(H(UL),0,m_2)$*, is isomorphic to the cohomology algebra* $H(U\mathcal{L})$ *of the universal enveloping algebra* $U\mathcal{L}$.

Proof Since $U^{\mathrm{Bar}}|_{\mathbf{DGL}_{\Bbbk}} = U$, the following holds:

$$\begin{aligned} U^{\mathrm{Bar}}(H(L),0,l_2) &= UH(\mathcal{L}) \\ &= H(U\mathcal{L}) && \text{(by Lemma 4.3)} \\ &= H(UL,\bar{m}_1,\bar{m}_2) \\ &= (H(UL),0,m_2). \end{aligned}$$

Now it follows by Theorem 2.1(c) that

$$U^{\mathrm{Bar}}(H(L),0,l_2,l_3,\dots) = (H(UL),0,m_2,m_3,\dots). \qquad \square$$

Thus the quasi-isomorphism in (4-1) is a map of the form

$$U^{\mathrm{Bar}}(\phi)\colon (UL,\bar{m}_1,\bar{m}_2) \to (H^*(UL),0,m_2,m_3,\dots).$$

Since $U\mathcal{L}$ is formal, it follows that $(H(UL),0,m_2,m_3,\dots)$ is A_∞–quasi-isomorphic to $H(U\mathcal{L}) = (H(UL),0,m_2)$. It follows by Proposition 3.7(a) that $[m_3] = 0$ in $HH^*(UH(\mathcal{L}))$.

Let $\mathrm{Alt}^*\colon HH^*(UH(\mathcal{L})) \to H^*_{\mathrm{CE}}(H(\mathcal{L}),(UH(\mathcal{L}))^{\mathrm{ad}})$ be the cohomology-induced map of the cochain map Alt introduced in Lemma 4.1, and let

$$j^*\colon H^*_{\mathrm{CE}}(H(\mathcal{L}),H(\mathcal{L})) \to H^*_{\mathrm{CE}}(H(\mathcal{L}),(UH(\mathcal{L}))^{\mathrm{ad}})$$

be the cochain map induced by the inclusion $H(\mathcal{L}) \hookrightarrow (UH(\mathcal{L}))^{\mathrm{ad}}$. We have by Theorem 2.1(d) that $\mathrm{Alt}^*[m_3] = j^*[l_3]$. Since $[m_3] = 0$, it follows that $j^*[l_3] = 0$. By Proposition 4.2, j^* is injective, and hence $[l_3] = 0$ in $H^*_{\mathrm{CE}}(H(\mathcal{L}),H(\mathcal{L}))$.

Since $[l_3] = 0$, it follows by Proposition 3.7(b), that there exists a quasi-isomorphism

$$\alpha\colon (H(L),0,l_2,l_3,\dots) \to (H(L),0,l_2,0,l'_4,l'_5,\dots)$$

Applying Theorem 2.1(c) to

$$U^{\mathrm{Bar}}(H(L),0,l_2,0,l'_4,l'_5,\dots) \quad \text{and} \quad U^{\mathrm{Bar}}(H(L),0,l_2,0,\dots),$$

we get that

$$U^{\mathrm{Bar}}(H(L),0,l_2,0,l'_4,l'_5,\dots) = (H^*(UL),0,m_2,0,m'_4,m'_5,\dots).$$

Note that $(H(UL),0,m_2,0,m'_4,m'_5,\dots)$ is A_∞–quasi-isomorphic to $(H(UL),0,m_2)$, since $U^{\mathrm{Bar}}(\alpha)$ is a quasi-isomorphism (by Theorem 2.1(b)) connecting

$$(H(UL),0,m_2,m_3,\dots) \quad \text{and} \quad (H(UL),0,m_2,0,m'_4,m'_5,\dots).$$

Again, by Proposition 3.7(a) it follows that $[m'_4] = 0$ in $HH^*(UH(\mathcal{L}))$. The same reasoning as before will give us that $[l'_4] = 0$ in $C^*_{\mathrm{CE}}(H(\mathcal{L}))$.

Continuing this process will yield a sequence

$$[l_3],\ [l_4'], \dots, [l_n^{(n-3)}], \dots$$

of vanishing Chevalley–Eilenberg cohomology classes. By Theorem 3.6, it follows that $(H(L), 0, l_2, l_3, \dots)$ is L_∞–quasi-isomorphic to $(H(L), 0, l_2)$, which is equivalent to the $\mathbf{DGL}_{\Bbbk}$–formality of $\mathcal{L} = (L, l_1, l_2)$. □

5 Proof of Theorem 1.3

We will compare the cohomology theories $H^*_{\mathscr{A}ss}$ and $H^*_{\mathscr{C}om}$, which correspond to the Hochschild and the Harrison cohomologies, respectively, in order to compare the concept of formality in $\mathbf{DGA}_{\Bbbk}$ and $\mathbf{CDGA}_{\Bbbk}$. We will denote the Harrison cochain complex and the Harrison cohomology of a commutative dg algebra A with coefficients in A by $C^*_{\mathrm{Harr}}(A)$ and $\mathrm{Harr}^*(A)$, respectively.

Hochschild and Harrison cohomology

We will start by recalling the notion of shuffle products. A permutation $\sigma \in S_{p+q}$ is called a (p,q)–shuffle if $\sigma(1) < \dots < \sigma(p)$ and $\sigma(p+1) < \dots < \sigma(p+q)$. Let $\mu_{p,q} \in \Bbbk[S_{p+q}]$ be given by

$$\mu_{p,q} = \sum_{(p,q)\text{–shuffles}} \mathrm{sgn}(\sigma)\sigma$$

There is an action of $\Bbbk[S_n]$ on $A^{\otimes n}$ given by

$$\sigma.(a_1 \cdots a_n) = \epsilon(\sigma; a_1, \dots, a_n) a_{\sigma^{-1}(1)} \cdots a_{\sigma^{-1}(n)}$$

for $\sigma \in S_n$, where $\epsilon(\sigma; a_1, \dots, a_n)$ is the Koszul sign obtained when applying σ to $a_1 \cdots a_n$. The shuffle product $\bar{\mu}_{p,q}\colon A^{\otimes p} \otimes A^{\otimes q} \to A^{\otimes(p+q)}$ is given by letting $\mu_{p,q}$ act on $A^{\otimes p} \otimes A^{\otimes q} \cong A^{\otimes(p+q)}$.

We will now see how this is related to Harrison cohomology. We have that

$$C^n_{\mathrm{Harr}}(A) \cong C^n_{\mathscr{C}om}(A) \cong \mathrm{Hom}_{\Bbbk}(\mathscr{C}om^!(n) \otimes_{S_n} A^{\otimes n}, A) \cong \mathrm{Hom}_{\Bbbk}(\mathscr{L}ie(n) \otimes_{S_n} A^{\otimes n}, A).$$

Over a field of characteristic zero, one can show that $\mathrm{Hom}_{\Bbbk}(\mathscr{L}ie(n) \otimes_{S_n} A^{\otimes n}, A)$ is isomorphic to the space of $\Bbbk$–morphisms $A^{\otimes n} \to A$ that vanish on all shuffle products $\bar{\mu}_{k,n-k}\colon A^{\otimes k} \otimes A^{\otimes(n-k)} \to A^{\otimes n}$ (see [14, Sections 1.3.3 and 13.1.7]) In particular that means that there exists an inclusion

$$\iota\colon C^*_{\mathrm{Harr}}(A) \hookrightarrow C^*_{\mathrm{Hoch}}(A) \cong \mathrm{Hom}_{\Bbbk}(A^{\otimes n}, A).$$

This inclusion induces a map $\iota^*\colon \mathrm{Harr}^*(A) \to HH^*(A)$ in the cohomology. Over a field $\Bbbk$ of characteristic zero, Barr [3] showed that ι^* is injective.

Proposition 5.1 [3] *Let $\Bbbk$ be a field of characteristic zero, and let A be a commutative dg algebra over $\Bbbk$. The map $\iota^*\colon \mathrm{Harr}^*(A) \to HH^*(A)$ induced by the inclusion $\iota\colon C^*_{\mathrm{Harr}}(A) \to C^*_{\mathrm{Hoch}}(A)$ is injective.*

We will briefly explain the techniques used in the proof of the proposition above. First, set $\mu_n = \sum_{i=1}^{n-1} \mu_{i,n-i}$. Next, Barr constructed a family of idempotents $\{e_i\}_{i\geq 2}$ with $e_n \in \Bbbk[S_n]$ that satisfies the following conditions:

(i) **Idempotent** $e_n^2 = e_n$.

(ii) e_n is a polynomial in μ_n (without any constant term).

(iii) $e_n \mu_{i,n-i} = \mu_{i,n-i}$ for $1 \leq i \leq n-1$.

Since $e_n \in \Bbbk[S_n]$, it defines an action on $C^n_{\mathrm{Hoch}}(A) = \mathrm{Hom}(A^{\otimes n}, A)$ (by permuting the inputs). This allows us to formulate a fourth condition that $\{e_i\}$ satisfies

(iv) $\partial_{\mathrm{Hoch}} e_n = e_{n+1} \partial_{\mathrm{Hoch}}$, where ∂_{Hoch} is the Hochschild coboundary.

By (ii)–(iii), there is an equality of ideals $(e_n) = (\mu_{1,n-1}, \mu_{2,n-2}, \dots, \mu_{n-1,1})$. In particular, a map $\phi \in C^n_{\mathrm{Hoch}}(A) = \mathrm{Hom}(A^{\otimes n}, A)$ vanishes under the action of e_n if and only if it vanishes on all $\mu_{i,n-i}$ (which is equivalent to $\phi \in C^*_{\mathrm{Harr}}(A)$).

Recall that an endomorphism $\rho\colon V \to V$ gives a decomposition $V = \ker(\rho) \oplus \mathrm{im}(\rho)$. If ρ is an idempotent we have that $\rho(a,b) = (0,b)$. Applying this to e_n (which defines an endomorphism on $C^n_{\mathrm{Hoch}}(A)$), we get that $C^n_{\mathrm{Hoch}}(A) = \ker(e_n) \oplus \mathrm{im}(e_n)$. Since $(e_n) = (\mu_{1,n-1}, \mu_{2,n-2}, \dots, \mu_{n-1,1})$, it follows that $\ker(e_n) = C^n_{\mathrm{Harr}}(A)$. Set $W^n(A) := \mathrm{im}(e_n)$ and $C^n_{\mathrm{Hoch}}(A)$ is then decomposed as

$$C^n_{\mathrm{Hoch}}(A) \cong C^n_{\mathrm{Harr}}(A) \oplus W^n(A). \tag{5-1}$$

In order to show that $\iota^*\colon \mathrm{Harr}^*(A) \to HH^*(A)$ is injective, we have to show that if $x \in C^n_{\mathrm{Harr}}(A) \subset C^n_{\mathrm{Hoch}}(A)$ is a coboundary in $C^*_{\mathrm{Hoch}}(A)$ then it is also a coboundary in $C^*_{\mathrm{Harr}}(A)$. By (5-1) we have that an element of the Harrison subcomplex may be represented by an element of the form $(x,0) \in C^n_{\mathrm{Harr}}(A) \oplus W^n(A) \cong C^n_{\mathrm{Hoch}}(A)$. Assume $(x,0)$ is a coboundary in $C^n_{\mathrm{Hoch}}(A)$, meaning that there is some element $(y_1,y_2) \in C^{n-1}_{\mathrm{Harr}}(A) \oplus W(A) \cong C^{n-1}_{\mathrm{Hoch}}(A)$ such that $\partial_{\mathrm{Hoch}}(y_1,y_2) = (x,0)$. From property (iv) we get the following commutative diagram:

$$\begin{tikzcd}
(y_1, y_2)_{\partial_{\mathrm{Hoch}}} \arrow[r, "e_{n-1}"] \arrow[d] & (0, y_2) \arrow[d, "\partial_{\mathrm{Hoch}}"] \\
(x,0)_{e_n} \arrow[r] & (0,0)
\end{tikzcd}$$

Now we see that $\partial_{\mathrm{Hoch}}(y_1,0) = \partial_{\mathrm{Hoch}}((y_1,y_2) - (0,y_2)) = (x,0)$, which proves that $(x,0)$ is also a boundary in $C^*_{\mathrm{Harr}}(A)$. This proves that ι^* is injective.

The idempotents $\{e_n\}$ cannot be constructed over a field of characteristic $p > 0$ and, over such a field, ι^* is not injective in general (see the example in Section 4 in [3]).

We would like to remark that the overview above is related to the subject of the λ–decomposition of Hochschild homology (see [13, Section 4.5]).

The proof

As mentioned in the introduction, it is obvious that $\mathbf{CDGA}_{\Bbbk}$–formality implies $\mathbf{DGA}_{\Bbbk}$–formality. Hence what is left to show in order to prove Theorem 1.3 is that if a cdga is formal as a dga, then it is also formal as a cdga.

Proof of Theorem 1.3 Let $(C, \bar{m}_1, \bar{m}_2)$ be a cdga that is formal in $\mathbf{DGA}_{\Bbbk}$. Let $\mathcal{H} = (H(C), 0, m_2)$ be the induced commutative graded algebra structure on the cohomology of $(C, \bar{m}_1, \bar{m}_2)$. The homotopy transfer theorem for C_∞–algebras (see Proposition 3.4) gives that there exists a C_∞–algebra $(H(C), 0, m_2, m_3, \dots)$ equipped with a C_∞–quasi-isomorphism

$$(C, \bar{m}_1, \bar{m}_2) \to (H(C), 0, m_2, m_3, \dots).$$

Since C is formal in $\mathbf{DGA}_{\Bbbk}$, there exists an A_∞–quasi-isomorphism

$$(H(C), 0, m_2, m_3, \dots) \to (H(C), 0, m_2).$$

It follows by Proposition 3.7(a) that $[m_3]_{\mathrm{Hoch}} = 0$ in $HH^*(\mathcal{H})$. Since the cohomology map

$$\iota^*\colon \mathrm{Harr}^*(A) \to HH^*(A)$$

induced by the inclusion $\iota\colon C^*_{\mathrm{Harr}}(H(C)) \hookrightarrow C^*_{\mathrm{Hoch}}(H(C))$ is injective (Proposition 5.1), it follows that $[m_3]_{\mathrm{Harr}} = 0$ in $\mathrm{Harr}^*(\mathcal{H})$. Now, by Proposition 3.7(b) it follows that there exists a C_∞–quasi-isomorphism

$$(H(C), 0, m_2, m_3, \dots) \to (H(C), 0, m_2, 0, m_4', m_5', \dots).$$

Applying Proposition 3.7(a) again gives that $[m_4']_{\mathrm{Hoch}} = 0$ in $HH^*(\mathcal{H})$, which in turn gives together with the injectivity of ι^* that $[m_4']_{\mathrm{Harr}} = 0$ in $\mathrm{Harr}^*(\mathcal{H})$. Continuing this process will yield a sequence

$$[m_3]_{\mathrm{Harr}},\ [m_4']_{\mathrm{Harr}}, \dots, [m_n^{(n-3)}]_{\mathrm{Harr}}, \dots$$

of vanishing Harrison cohomology classes in $\mathrm{Harr}^*(\mathcal{H})$. By Theorem 3.6, it follows that $(H(C), 0, m_2, m_3, \dots)$ is C_∞–quasi-isomorphic to $(H(C), 0, m_2)$, which is equivalent to the $\mathbf{CDGA}_{\Bbbk}$–formality of $(C, \bar{m}_1, \bar{m}_2)$. $\square$

Appendix: Some technicalities concerning A_∞–, C_∞– and L_∞–algebras

Given a Koszul operad $\mathscr{P}$ there are many equivalent ways of viewing a $\mathscr{P}_\infty$–algebra structure on a vector space A.

Theorem A.1 [14, Theorem 10.1.13] *Let $\mathscr{P}$ be a Koszul operad. Then a $\mathscr{P}_\infty$–algebra structure on a vector space A is the same thing as coderivation on the cofree $\mathscr{P}^!$–coalgebra on sA, denoted by $\mathbb{V}^*_{\mathscr{P}^!}(sA)$, and a morphism of $\mathscr{P}_\infty$–algebras is the same thing as a morphism of cofree $\mathscr{P}^!$–coalgebras.*

We have that $\mathbb{V}^*_{\mathscr{P}^!}(sA) = \bigoplus_{n\geq 0} \mathbb{V}^n_{\mathscr{P}^!}(sA)$, where $\mathbb{V}^n_{\mathscr{P}^!}(sA) = \mathscr{P}^{!\vee}(n) \otimes_{S_n} (sA)^{\otimes n}$ (and $\mathscr{P}^{!\vee}$ denotes the cooperad obtained by dualizing $\mathscr{P}^!$). We say that an element of $\mathbb{V}^n_{\mathscr{P}^!}(sA)$ is of word-length n.

We will briefly recall the correspondence between $\mathscr{P}_\infty$–algebras and quasifree dg $\mathscr{P}^!$–coalgebras. A $\mathscr{P}^!$–coalgebra differential d on $\mathbb{V}^*_{\mathscr{P}^!}(sA)$ may be decomposed as

$$d = d_0 + d_1 + \cdots,$$

where d_i is the part of d that decreases the word-length by i. The dg coalgebra $(\mathbb{V}^*_{\mathscr{P}^!}(sA), d = d_0 + d_1 + \cdots)$ corresponds to a $\mathscr{P}_\infty$–algebra $(A, b_1, b_2, \ldots)$, where d_i and b_{i+1} encode each other (ie d_i may be constructed from b_{i+1} and vice versa).

Analogously, a morphism of dg $\mathscr{P}^!$–coalgebras $\Psi\colon (\mathbb{V}^*_{\mathscr{P}^!}(sA), d) \to (\mathbb{V}^*_{\mathscr{P}^!}(sA'), d')$ may be decomposed as $\Psi = \Psi_0 + \Psi_1 + \cdots$, where Ψ_i is the part of Ψ that decreases the word-length by i. We have that Ψ corresponds to a $\mathscr{P}_\infty$–quasi-isomorphism $\phi = (\phi_1, \phi_2, \ldots)\colon A \to A'$, where Ψ_i and ϕ_{i+1} encode each other. With this correspondence we have tools to prove some technical results that we need in this paper. The author was inspired by the techniques used in [7, Section 2.72].

Lemma A.2 *Assume $(A, 0, b_2, b_3, \ldots)$ and $(A, 0, b_2)$ are quasi-isomorphic as $\mathscr{P}_\infty$–algebras. Then there exists a $\mathscr{P}_\infty$–algebra quasi-isomorphism*

$$\phi'\colon (A, 0, b_2, b_3, \ldots) \to (A, 0, b_2),$$

where $\phi'_1 = \mathrm{id}_A$.

Proof Let $\phi = (\phi_1, \phi_2, \ldots)\colon (A, 0, b_2, b_3, \ldots) \to (A, 0, b_2)$ be a quasi-isomorphism. Since ϕ is a quasi-isomorphism of minimal $\mathscr{P}_\infty$–algebras, it follows that ϕ is an isomorphism.

We have that ϕ corresponds to a map

$$\Psi = \Psi_0 + \Psi_1 + \cdots\colon (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1 + d_2 + \cdots) \to (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1),$$

where Ψ_i increases the word-length by i and corresponds to ϕ_{i+1}. Moreover, Ψ_0 is a vector space isomorphism (since Ψ is a dg $\mathscr{P}^!$–coalgebra isomorphism).

We show that Ψ_0 commutes with the differential d_1. Since Ψ is a chain map, we have that
$$(\Psi_0 + \Psi_1 + \cdots) \circ (d_1 + d_2 + \cdots) = d_1 \circ (\Psi_0 + \Psi_1 + \cdots).$$

Collecting the terms that decrease the word-length by 1 from both sides of the equality gives that $\Psi_0 d_1 = d_1 \Psi_0$.

Similar techniques give also that Ψ_0 commutes with the comultiplication Δ on $\mathbb{V}^*_{\mathscr{P}^!}(sA)$. Hence, $\Psi_0\colon (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1) \to (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1)$ is a dg $\mathscr{P}^!$–coalgebra automorphism, which has an inverse Ψ_0^{-1}. Now the composition $\Psi' = (\Psi_0^{-1}) \circ (\Psi_0 + \Psi_1 + \cdots)$ will give the desired result. □

Lemma A.3 *Assume that θ is a $\mathscr{P}^!$–coalgebra coderivation on $\mathbb{V}^*_{\mathscr{P}^!}(V)$ of cohomological degree 0 that decreases the word-length by some number $i \geq 1$. Then the map*
$$e^\theta = \mathrm{id} + \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \cdots$$
is a well-defined map of $\mathscr{P}^!$–coalgebras.

Proof The map is well-defined since, for any element $x \in \mathbb{V}^*_{\mathscr{P}^!}(V)$ of word-length k, we have that $\theta^m(x) = 0$ for all $m \geq \lceil \frac{k}{i} \rceil$, so $e^\theta(x)$ will be a finite sum
$$e^\theta(x) = x + \theta(x) + \cdots + \frac{\theta^{m-1}(x)}{(m-1)!}.$$

Now we prove that e^θ is a map of $\mathscr{P}^!$–coalgebras. One can easily prove by induction that
$$\Delta\theta^n = \left(\sum_{p=0}^{n} \binom{n}{p} \theta^{n-p} \otimes \theta^p\right) \circ \Delta.$$

Thus
$$\begin{aligned}
\Delta \circ e^\theta &= \left(\sum_{n=0}^{\infty} \frac{1}{n!} \sum_{p=0}^{n} \binom{n}{p} \theta^{n-p} \otimes \theta^p\right) \circ \Delta \\
&= \left(\sum_{n=0}^{\infty} \sum_{p=0}^{n} \frac{\theta^{n-p}}{(n-p)!} \otimes \frac{\theta^p}{p!}\right) \circ \Delta \\
&= (e^\theta \otimes e^\theta) \circ \Delta.
\end{aligned}$$
□

Remark A.4 The map e^θ is an automorphism with inverse $e^{-\theta}$.

Lemma A.5 *Assume $(A, 0, m_2, 0, \ldots, 0, m_n, m_{n+1}, \ldots)$ and $(A, 0, m_2)$ are quasi-isomorphic as $\mathscr{P}_\infty$–algebras. Then there exists a map*

$$\phi' \colon (A, 0, m_2, 0, \dots, 0, m_n, m_{n+1}, \dots) \to (A, 0, m_2)$$

such that $\phi'_1 = \mathrm{id}_A$ *and* $\phi'_i = 0$ *for* $2 \le i \le n-2$.

Proof We prove the lemma by induction on n. For $n=3$ the assertion is true by Lemma A.2. Assume the assertion is true for $n-1$ with $n \ge 4$. Then we have that there exists a quasi-isomorphism

$$\phi = (\mathrm{id}, 0, \dots, 0, \phi_{n-2}, \phi_{n-1}, \dots) \colon (A, 0, m_2, 0, \dots, 0, m_n, m_{n+1}, \dots) \to (A, 0, m_2),$$

which corresponds to a $\mathscr{P}^!$–coalgebra map

$$\Psi = \mathrm{id} + \Psi_{n-3} + \Psi_{n-2} + \cdots \colon (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1 + d_{n-1} + d_n + \cdots) \to (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1).$$

Considering the equality $(\Psi \otimes \Psi) \circ \Delta = \Delta \circ \Psi$ and collecting the terms that decrease the word-length by $n-3$ gives that $(\mathrm{id} \otimes \Psi_{n-3} + \Psi_{n-3} \otimes \mathrm{id}) \circ \Delta = \Delta \circ \Psi_{n-3}$. That means that $\pm\Psi_{n-3}$ is a coderivation of $\mathbb{V}^*_{\mathscr{P}^!}(sA)$ and therefore $e^{\pm\Psi_{n-3}} \colon \mathbb{V}^*_{\mathscr{P}^!}(sA) \to \mathbb{V}^*_{\mathscr{P}^!}(sA)$ is a $\mathscr{P}^!$–coalgebra automorphism.

Considering the equality $\Psi \circ (d_1 + d_{n-1} + d_n + \cdots) = d_1 \circ \Psi$ and collecting the terms that decrease the word-length by $n-2$ gives that $\Psi_{n-3} \circ d_1 = d_1 \circ \Psi_{n-3}$, ie that $\pm\Psi_{n-3}$ commutes with the differential d_1. Hence, $e^{\pm\Psi_{n-3}}$ commutes with d_1 and therefore $e^{\pm\Psi_{n-3}} \colon (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1) \to (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1)$ is a dg $\mathscr{P}^!$–coalgebra automorphism.

We consider the composition

$$\Psi' = e^{-\Psi_{n-3}} \circ \Psi \colon (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1 + d_{n-1} + d_n + \cdots) \to (\mathbb{V}^*_{\mathscr{P}^!}(sA), d_1).$$

We have that

$$\begin{aligned}\Psi' &= e^{-\Psi_{n-3}} \circ (\mathrm{id} + \Psi_{n-3} + \Psi_{n-2} + \cdots) \\ &= \left(\mathrm{id} - \Psi_{n-3} + \frac{\Psi_{n-3}^2}{2!} - \cdots\right) \circ (\mathrm{id} + \Psi_{n-3} + \Psi_{n-2} + \cdots) \\ &= \mathrm{id} + (\text{terms that increase the word-length by } \ge n-2).\end{aligned}$$

Hence, Ψ' is of the form $\Psi' = \mathrm{id} + \Psi'_{n-2} + \Psi'_{n-1} + \cdots$, where Ψ'_i decreases the word-length by i and will therefore correspond to a $\mathscr{P}_\infty$–algebra quasi-isomorphism $\phi' \colon (A, 0, m_2, 0, \dots, 0, m_n, m_{n+1}, \dots) \to (A, 0, m_2)$ that satisfies the property given in the lemma. □

References

[1] **M Aubry**, **J-M Lemaire**, *Homotopies d'algèbres de Lie et de leurs algèbres enveloppantes*, from "Algebraic topology: rational homotopy" (Y Félix, editor), Lecture Notes in Math. 1318, Springer (1988) 26–30 MR

[2] **V Baranovsky**, *A universal enveloping for L_∞–algebras*, Math. Res. Lett. 15 (2008) 1073–1089 MR

[3] **M Barr**, *Harrison homology, Hochschild homology and triples*, J. Algebra 8 (1968) 314–323 MR

[4] **A Berglund**, *Homological perturbation theory for algebras over operads*, Algebr. Geom. Topol. 14 (2014) 2511–2548 MR

[5] **P Deligne**, **P Griffiths**, **J Morgan**, **D Sullivan**, *Real homotopy theory of Kähler manifolds*, Invent. Math. 29 (1975) 245–274 MR

[6] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Mathematics 205, Springer (2001) MR

[7] **Y Félix**, **J Oprea**, **D Tanré**, *Algebraic models in geometry*, Oxford Graduate Texts in Mathematics 17, Oxford Univ. Press (2008) MR

[8] **W M Goldman**, **J J Millson**, *The deformation theory of representations of fundamental groups of compact Kähler manifolds*, Bull. Amer. Math. Soc. 18 (1988) 153–158 MR

[9] **S Halperin**, **J Stasheff**, *Obstructions to homotopy equivalences*, Adv. in Math. 32 (1979) 233–279 MR

[10] **T Kadeishvili**, *Cohomology C_∞–algebra and rational homotopy type*, from "Algebraic topology: old and new" (M Golasiński, Y Rudyak, P Salvatore, N Saveliev, N Wahl, editors), Banach Center Publ. 85, Polish Acad. Sci. Inst. Math., Warsaw (2009) 225–240 MR

[11] **B Keller**, *Introduction to A–infinity algebras and modules*, Homology Homotopy Appl. 3 (2001) 1–35 MR

[12] **M Kontsevich**, *Deformation quantization of Poisson manifolds*, Lett. Math. Phys. 66 (2003) 157–216 MR

[13] **J-L Loday**, *Cyclic homology*, 2nd edition, Grundl. Math. Wissen. 301, Springer (1998) MR

[14] **J-L Loday**, **B Vallette**, *Algebraic operads*, Grundl. Math. Wissen. 346, Springer (2012) MR

[15] **M Manetti**, *On some formality criteria for DG–Lie algebras*, J. Algebra 438 (2015) 90–118 MR

[16] **M Markl**, **S Shnider**, **J Stasheff**, *Operads in algebra, topology and physics*, Mathematical Surveys and Monographs 96, Amer. Math. Soc., Providence, RI (2002) MR

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

[18] **D Quillen**, *Rational homotopy theory*, Ann. of Math. 90 (1969) 205–295 MR

Department of Mathematics, Stockholm University
SE-106 91 Stockholm, Sweden

bashar@math.su.se

Received: 10 October 2016 Revised: 1 February 2017

Algebraic & Geometric Topology 17 (2017) 2543–2564

A note on cobordisms of algebraic knots

József Bodnár
Daniele Celoria
Marco Golla

We use Heegaard Floer homology to study smooth cobordisms of algebraic knots and complex deformations of cusp singularities of curves. The main tool will be the concordance invariant ν^+: we study its behaviour with respect to connected sums, providing an explicit formula in the case of L–space knots and proving subadditivity in general.

14B05, 14B07, 57M25; 57M27, 57R58

1 Introduction

A *cobordism* between two knots K and K' in S^3 is a smoothly and properly embedded surface $F \subset S^3 \times [0,1]$, with $\partial F = K \times \{0\} \cup K' \times \{1\}$. Carving along an arc connecting the two boundary components of F, one produces a slice surface for the connected sum $\overline{K} \# K'$, where $\overline{K}$ is the mirror of K. Two knots are *concordant* if there is a genus-0 cobordism between them; this is an equivalence relation, and the connected sum endows the quotient $\mathcal{C}$ of the set of knots with a group operation; $\mathcal{C}$ is therefore called the (smooth) *concordance group*. A knot is *smoothly slice* if it is concordant to the unknot.

Litherland [14] used Tristram–Levine signatures to show that torus knots are linearly independent in $\mathcal{C}$. In fact, Tristram–Levine signatures provide a lower bound for the slice genus of knots. Sharp lower bounds for the slice genus of torus knots are provided by the invariants τ in Heegaard Floer homology — see Ozsváth and Szabó [21] — and s in Khovanov homology; see Rasmussen [25].

More recently, Ozsváth, Stipsicz and Szabó [19] defined the concordance invariant Υ; Livingston and Van Cott [15] used Υ to improve on the bounds given by signatures along some families of connected sums of torus knots.

In this note we consider *algebraic knots*, ie links of irreducible curve singularities (*cusps*), and more generally *L–space knots*. Given two algebraic knots K and L, we give lower bounds on the genus of a cobordism between them by using the concordance

Published: 3 August 2017 DOI: 10.2140/agt.2017.17.2543

invariant ν^+ defined by Hom and Wu [12]. This is computed in terms of the semigroups of the two corresponding singularities, Γ_K and Γ_L, and the corresponding enumerating functions $\Gamma_K(\cdot)$ and $\Gamma_L(\cdot)$.

Theorem 1.1 *If K and L are algebraic knots with enumerating functions $\Gamma_K(\cdot)$ and $\Gamma_L(\cdot)$, respectively, then*

$$\nu^+(K \# \overline{L}) = \max\bigl\{g(K) - g(L) + \max_{n\geq 0}\{\Gamma_L(n) - \Gamma_K(n)\}, 0\bigr\}.$$

In Section 2.3 we define an appropriate enumerating function for L–space knots; Theorem 3.1 below mimics the statement above in this more general setting, and directly implies Theorem 1.1; the key of the definition and of the proofs is the reduced Floer complex defined by Krcatovich [13].

As an application of Theorem 1.1, we give a different proof of a result of Gorsky and Némethi [10] on the semicontinuity of the semigroup of an algebraic knot under deformations of singularities in the cuspidal case. A similar result was obtained by Borodzik and Livingston [7] under stronger assumptions (see Section 4 for details).

Theorem 1.2 *Assume there exists a deformation of an irreducible plane curve singularity with semigroup Γ_K to an irreducible plane curve singularity with semigroup Γ_L. Then for each nonnegative integer n*

$$\#(\Gamma_K \cap [0,n)) \leq \#(\Gamma_L \cap [0,n)).$$

In fact, there is an analogue of Theorem 1.2 when the deformation has multiple (not necessarily irreducible) singularities; see Theorem 5.2 below for a precise statement. As an immediate corollary, we obtain that the multiplicity decreases under deformations. More precisely, we have the following:

Corollary 1.3 *Let K and L be two links of irreducible singularities as above, and $m(K)$ and $m(L)$ denote their multiplicities. Then $m(L) \leq m(K)$.*

It is worth noting that the multiplicity of an irreducible singularity can be interpreted topologically as the *braid index* of the knot, ie the minimal number of strands among all braids whose closure is the given knot.

Finally, we turn to proving some properties of the function ν^+. The first one reflects analogous properties for other invariants (signatures, τ, s, etc.) and gives lower bounds for the unknotting number and related concordance invariants (see Section 5 below).

Theorem 1.4 *If K_+ is obtained from K_- by changing a negative crossing into a positive one, then*

$$\nu^+(K_-) \leq \nu^+(K_+) \leq \nu^+(K_-) + 1.$$

Theorem 1.5 *The function ν^+ is subadditive. Namely, for any two knots $K, L \subset S^3$,*

$$\nu^+(K \# L) \le \nu^+(K) + \nu^+(L).$$

As an application, we consider some concordance invariants, also studied by Owens and Strle [18]. Recall that the *concordance unknotting number* $u_c(K)$ of a knot K is the minimum of unknotting numbers among all knots that are concordant to K; the *slicing number* $u_s(K)$ of K is the minimal number of crossing changes needed to turn K into a slice knot; finally, the 4*–ball crossing number* $c^*(K)$ is the minimal number of double points of an immersed disc in the 4–ball whose boundary is K. It is quite remarkable that there are knots for which these quantities disagree [18].

Proposition 1.6 *The unknotting number, concordance unknotting number, slicing number and 4–ball crossing number of K are all bounded below by $\nu^+(K) + \nu^+(\bar{K})$.*

1.1 Organisation of the paper

In Section 2 we recall some facts about Heegaard Floer correction terms and reduced knot Floer complex. In Section 3 we prove Theorem 1.1 as a corollary of Theorem 3.1, and in Section 4 we prove Theorem 1.2. In Section 5 we study cobordisms between arbitrary knots and prove Theorem 1.4 and Proposition 1.6; in Section 6 we prove Theorem 1.5. Finally, in Section 7 we study some concrete examples.

Acknowledgements We would like to thank Paolo Aceto, Maciej Borodzik, Matt Hedden, and Kouki Sato for interesting conversations, Maciej Borodzik for providing us with some computational tools, Peter Feller for pointing out Corollary 1.3 and the anonymous referee for useful comments. Bodnár has been supported by the ERC grant LDTBud at MTA Alfréd Rényi Institute of Mathematics. Celoria has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 674978). Golla was partially supported by the PRIN–MIUR research project 2010–11 *Varietà reali e complesse: geometria, topologia e analisi armonica* and by the FIRB research project *Topologia e geometria di varietà in bassa dimensione*.

2 Singularities and Heegaard Floer homology

2.1 Links of curve singularities

In what follows, K and L will be two *algebraic knots*. We will recall briefly what this means and also what invariants can be associated with such knots. For further details, we refer to [8; 9; 26].

Assume $F \in \mathbb{C}[x, y]$ is an *irreducible* polynomial which defines an *isolated irreducible plane curve singularity*. This means that $F(0,0) = 0$ and, in a sufficiently small neighbourhood $B_\varepsilon = \{|x|^2 + |y|^2 \leq \varepsilon^2\}$ for some $\varepsilon > 0$ of the origin, $\partial_1 F(x, y) = \partial_2 F(x, y) = 0$ holds if and only if $(x, y) = (0, 0)$. The *link* of the singularity is the zero set of F intersected with a sphere of sufficiently small radius: $K = \{F(x, y) = 0\} \cap \partial B_\varepsilon$. Since F is irreducible, K is a knot, rather than a link, in the 3–sphere ∂B_ε. A knot is called *algebraic* if its isotopy type arises in the above described way. All algebraic knots are *iterated torus knots*, ie they arise by iteratively cabling a torus knot.

The zero set of every isolated irreducible plane curve singularity admits a local parametrisation, ie there exist $x(t), y(t) \in \mathbb{C}[\![t]\!]$ such that $F(x(t), y(t)) \equiv 0$ and $t \mapsto (x(t), y(t))$ is a bijection for $|t| < \eta \ll 1$ to a neighbourhood of the origin in the zero set of F. Consider the set of integers

$$\Gamma_K = \left\{ \operatorname{ord}_t G(x(t), y(t)) \mid G \in \mathbb{C}[\![x, y]\!],\ F \text{ does not divide } G \right\}.$$

It can be seen easily that Γ_K is an additive semigroup. It depends only on the local topological type of the singularity; therefore, it can be seen as an invariant of the isotopy type of the knot K. We will say that Γ_K is the *semigroup of the algebraic knot* K.

We denote by $\mathbb{N} = \{0, 1, \dots\}$ the set of nonnegative integers. The semigroup Γ_K is a cofinite set in $\mathbb{N}$; in fact, $|\mathbb{N} \setminus \Gamma_K| = \delta_K < \infty$ and the greatest element not in Γ_K is $2\delta_K - 1$. The number δ_K is called the *δ–invariant* of the singularity. It is well-known that δ_K is the Seifert genus of K: $\delta_K = g(K)$.

We also write $\Gamma_K(n)$ for the n^{th} element of Γ_K with respect to the standard ordering of $\mathbb{N}$, with the convention that $\Gamma_K(0) = 0$. The function $\Gamma_K(\cdot)$ will be called the *enumerating function* of Γ_K.

2.2 Heegaard Floer and concordance invariants

Heegaard Floer homology is a family of invariants of 3–manifolds introduced by Ozsváth and Szabó [22]; in this paper we are concerned with the "minus" version over the field $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$ with two elements. It associates to a rational homology sphere Y equipped with a spin^c structure $\mathfrak{t}$ a $\mathbb{Q}$–graded $\mathbb{F}[U]$–module $\mathrm{HF}^-(Y, \mathfrak{t})$; the action of U decreases the degree by 2.

The group $\mathrm{HF}^-(Y, \mathfrak{t})$ further splits as a direct sum of $\mathbb{F}[U]$–modules $\mathbb{F}[U] \oplus \mathrm{HF}^-_{\mathrm{red}}(Y, \mathfrak{t})$. We call $\mathbb{F}[U]$ the *tower* of $\mathrm{HF}^-(Y, \mathfrak{t})$. The degree of the element $1 \in \mathbb{F}[U]$ is called the *correction term* of $(Y, \mathfrak{t})$, and it is denoted by[1] $d(Y, \mathfrak{t})$.

[1] Note that our definition of $d(Y, \mathfrak{t})$ would differ by 2 from the original definition of [20]; however, it is more convenient for our purposes to use a shifted grading in HF^-.

When Y is obtained as an integral surgery along a knot K in S^3, one can recover the correction terms of Y in terms of a family of invariants introduced by Rasmussen [24] and then further studied by Ni and Wu [16] and Hom and Wu [12]. We call these invariants $\{V_i(K)\}_{i\geq 0}$, adopting Ni and Wu's notation instead of Rasmussen's — who used $h_i(K)$ — as this seems to have become more standard.

Recall that there is an indexing of spin^c structures on $S^3_n(K)$, as defined in [23, Section 2.4]: $S^3_n(K)$ is the boundary of the surgery handlebody $W_n(K)$ obtained by attaching a single 2–handle with framing n along $K \subset \partial B^4$. Notice that, if we orient K, there is a well-defined generator $[F]$ of $H_2(W_n(K);\mathbb{Z})$ obtained by capping off a Seifert surface of K with the core of the 2–handle. The spin^c structure $\mathfrak{t}_k$ on $S^3_n(K)$ is defined as the restriction of a spin^c structure $\mathfrak{s}$ on $W_n(K)$ such that

$$\langle c_1(\mathfrak{s}), [F]\rangle \equiv n + 2k \pmod{2n}. \tag{1}$$

Theorem 2.1 [24; 16] *The sequence $\{V_i(K)\}_{i\geq 0}$ takes values in $\mathbb{N}$ and is eventually 0. Moreover, $V_i(K) - 1 \leq V_{i+1}(K) \leq V_i(K)$ for every i.*

With the spin^c labelling defined in (1) above, for every integer n we have

$$d(S^3_n(K), \mathfrak{t}_i) = -2\max\{V_i(K), V_{n-i}(K)\} + \frac{(n-2i)^2 - n}{4n}. \tag{2}$$

Definition 2.2 [12] The minimal index i such that $V_i(K) = 0$ is called $\nu^+(K)$.

2.3 Reduced knot Floer homology

Krcatovich [13] introduced the *reduced knot Floer complex* $\underline{\mathrm{CFK}}^-(K)$ associated to a knot K in S^3. This complex is graded by the *Maslov grading* and filtered by the *Alexander grading*; the differential decreases the Maslov grading by 1 and respects the Alexander filtration.

Without going into technical details, for which we refer to [13], any knot Floer complex $\mathrm{CFK}^-(K)$ can be recursively simplified until the differential on the graded object associated to the Alexander filtration becomes trivial (while the differential on the filtered complex is, in general, nontrivial). Moreover, $\underline{\mathrm{CFK}}^-(K)$ still retains an $\mathbb{F}[U]$–module structure.

The power of Krcatovich's approach relies in the application to connected sums; if we need to compute $\mathrm{CFK}^-(K_1 \# K_2) \cong \mathrm{CFK}^-(K_1) \otimes_{\mathbb{F}[U]} \mathrm{CFK}^-(K_2)$ we can first reduce $\mathrm{CFK}^-(K_1)$, and then take the tensor product $\underline{\mathrm{CFK}}^-(K_1) \otimes_{\mathbb{F}[U]} \mathrm{CFK}^-(K_2)$.

This is particularly effective when dealing with L–*space knots*, ie knots that have a positive integral surgery Y such that $\mathrm{HF}^-(Y,\mathfrak{t}) = \mathbb{F}[U]$ for every spin^c structure $\mathfrak{t}$ on Y. Notice that all algebraic knots are L–space knots [11, Theorem 1.8].

In this case, $\underline{\mathrm{CFK}}^-(K)$ is isomorphic to $\mathbb{F}[U]$ as an $\mathbb{F}[U]$–module, and it has at most one generator in each Alexander degree. If we call x the homogeneous generator of $\underline{\mathrm{CFK}}^-(K)$ as an $\mathbb{F}[U]$–module, then $\underline{\mathrm{CFK}}^-(K) = \mathbb{F}[U]x$, and $\{U^n x\}_{n\geq 0}$ is a homogeneous basis of $\underline{\mathrm{CFK}}^-(K)$.

We denote by $\Gamma_K(n)$ the quantity $g(K) - A(U^n\cdot x)$, where A is the Alexander degree, and we call $\Gamma_K(\cdot)$ the *enumerating function* of K. As observed by Borodzik and Livingston [6, Section 4], when K is an algebraic knot, the function $\Gamma_K(\cdot)$ coincides with the enumerating function of the semigroup associated to K as defined above. Accordingly, for an arbitrary L–space knot K, we define the *semigroup* of K as the image of Γ_K.

Example 2.3 In general, $\Gamma_K(\cdot)$ is not the enumerating function of a semigroup; to this end, consider the pretzel knot $K = P(-2,3,7) = 12n_{242}$. K is an L–space knot with Alexander polynomial $t^{-5} - t^{-4} + t^{-2} - t^{-1} + 1 - t + t^2 - t^4 + t^5$, hence the function $\Gamma_K(\cdot)$ takes values $0, 3, 5, 7, 8, 10, 11, 12, \dots$. Since 3 is in the image of $\Gamma_K(\cdot)$ but 6 is not, $\Gamma_K(\cdot)$ is not the enumerating function of a semigroup.

2.4 An example

We are going to show an application of the reduced knot Floer complex in a concrete case. Consider the knot $K = T_{3,7} \# \overline{T_{4,5}}$. The genera, signatures and υ–function [19] of $T_{3,7}$ and $T_{4,5}$ all agree: $g(T_{3,7}) = g(T_{4,5}) = 6$, $\sigma(T_{3,7}) = \sigma(T_{4,5}) = 8$, and $\upsilon(T_{3,7}) = \upsilon(T_{4,5}) = -4$. It follows that $\tau(K) = s(K) = \sigma(K) = \upsilon(K) = 0$. However, we can show the following:

Proposition 2.4 *The knot K satisfies $\nu^+(K) = \nu^+(\overline{K}) = 1$.*

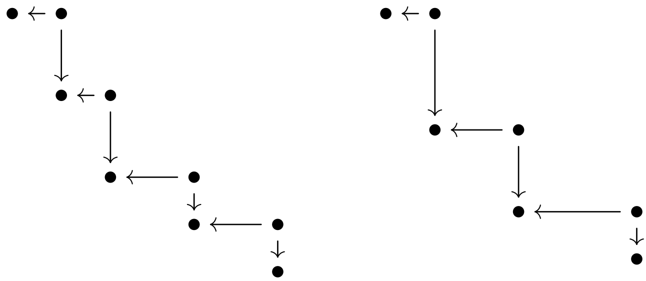
Proof We need to compute a Floer complex of $T_{3,7}$, $T_{4,5}$ and their mirrors, as well as the reduced Floer complex of $T_{3,7}$ and $T_{4,5}$. Let $K_1 = T_{3,7}$ and $K_2 = T_{4,5}$.

For an L–space knot L, and in particular for every positive torus knot, each of the knot Floer complexes $\mathrm{CFK}^-(L)$ and $\mathrm{CFK}^-(\overline{L})$ is determined by a *staircase complex*; the staircase for $\overline{L}$ is obtained by reflecting the one for L across the diagonal of the second and fourth quarters, and switching the direction of all arrows. For example, when $L = T_{2,3}$ the two staircases are:

$$\begin{array}{ccc} \mathrm{CFK}^-(L) & \qquad & \mathrm{CFK}^-(\overline{L}) \\ \begin{array}{ccc} \bullet & \leftarrow & \bullet \\ & & \downarrow \\ & & \bullet \end{array} & & \begin{array}{ccc} \bullet & & \\ \downarrow & & \\ \bullet & \leftarrow & \bullet \end{array} \end{array}$$

In the case of K_1 and K_2, we have:

$\mathrm{CFK}^-(K_1)$ $\qquad$ $\mathrm{CFK}^-(K_2)$



The reduced complex $\underline{\mathrm{CFK}}^-(K_1)$, on the other hand, has a single generator in each of the following bidegrees $(-i,j)$ (where $-i$ records the U–power and j records the Alexander grading):

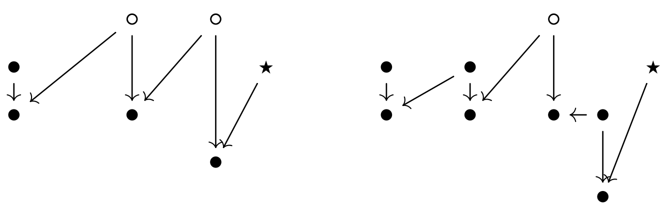
$$(0,6),\quad (-1,3),\quad (-2,0),\quad (-3,-1),\quad (-4,-3),\quad (-5,-4),\quad (-6-n,-6-n),\quad n\geq 0.$$

The reduced complex $\underline{\mathrm{CFK}}^-(K_2)$ has a generator in each of the bidegrees

$$(0,6),\quad (-1,2),\quad (-2,1),\quad (-3,-2),\quad (-4,-3),\quad (-5,-4),\quad (-6-n,-6-n),\quad n\geq 0.$$

In both cases, the U–action carries a generator with i–coordinate k to one with i–coordinate $k-1$. Taking the tensor product over $\mathbb{F}[U]$, one gets twisted staircases as follows, with a generator in bidegree $(0,0)$ (marked with a $\star$):

$\underline{\mathrm{CFK}}^-(K_1)\otimes \mathrm{CFK}^-(\overline{K}_2)$ $\qquad$ $\underline{\mathrm{CFK}}^-(K_2)\otimes \mathrm{CFK}^-(\overline{K}_1)$



The generators marked with a $\circ$ exhibit the fact that $V_0(K_1\#\overline{K}_2)$ and $V_0(K_2\#\overline{K}_1)$ are both strictly positive (see [13, Section 4] for details). $\square$

3 Computing the invariant

In this section we are going to prove a version of Theorem 1.1 for L–space knots. Given an integer x we denote by $(x)_+$ the quantity $(x)_+=\max\{0,x\}$.

Theorem 3.1 *Let K and L be two L–space knots with enumerating functions $\Gamma_K(\cdot), \Gamma_L(\cdot)\colon \mathbb{N} \to \mathbb{N}$. Then*

$$\nu^+(K \# \overline{L}) = \big(g(K) - g(L) + \max_{n\geq 0}\{\Gamma_L(n) - \Gamma_K(n)\}\big)_+.$$

Notice that, since algebraic knots are L–space knots, Theorem 1.1 is an immediate corollary. Theorem 3.1 will in turn be a consequence of the following proposition:

Proposition 3.2 *In the notation of Theorem 3.1, let $\{0 = a_1 < \cdots < a_d = g(L)\}$ be the image of the function $n \mapsto \Gamma_L(n) - n$, and define $a'_k = g(L) - a_{d+1-k}$ for $k = 1, \dots, d$. Then*

$$\nu^+(K \# \overline{L}) = \big(g(K) - g(L) + \max_{1\leq k\leq d}\{a_k + a'_k - \Gamma_K(a'_k)\}\big)_+.$$

Proof Let $\delta_K = g(K)$, $\delta_L = g(L)$. Consider the complex $\underline{\mathrm{CFK}}^-(K) \otimes_{\mathbb{F}[U]} \mathrm{CFK}^-(\overline{L})$ that computes the knot Floer homology of $K \# \overline{L}$. Recall that the function $\Gamma_K(\cdot)$ describes the reduced Floer complex: $\underline{\mathrm{CFK}}^-(K)$ has a generator x_k in each bidegree $(-k, \delta_K - \Gamma_K(k))$. Moreover, $U \cdot x_k = x_{k+1}$.

As observed by Krcatovich [13, Section 4], the sequences $\{a_k\}$ and $\{a'_k\}$ determine a "twisted staircase" knot Floer complex $\mathrm{CFK}^-(\overline{L})$ for $\overline{L}$: the generator of the tower $\mathbb{F}[U]$ in $\mathrm{HFK}^-(\overline{L})$ is represented by the sum of d generators $U^{a'_1} y_1, \dots, U^{a'_d} y_d$, where y_k sits in bidegree $(0, a'_k + a_k - \delta_L)$. In more graphical terms, a_k will be the Alexander grading of $U^{a_k} y_k$, ie its j–coordinate, and $-a'_k$ will be its i–coordinate.

The tensor product $\underline{\mathrm{CFK}}^-(K) \otimes \mathrm{CFK}^-(\overline{L})$ has a staircase in Maslov grading 0 generated by the chain $z = \sum_{k=1}^d x_0 \otimes U^{a'_k} y_k$. Notice that $x_0 \otimes U^{a'_k} y_k = x_{a'_k} \otimes y_k$ sits in Alexander degree $A(x_{a'_k}) + A(y_k) = \delta_K - \Gamma(a'_k) + a_k + a'_k - \delta_L$. Therefore, the maximal Alexander degree in the chain z is precisely $M = \delta_K - \delta_L + \max\{a_k + a'_k - \Gamma_K(a_k)\}$, and we claim that if $M \geq 0$, then $\nu^+(K \# \overline{L}) = M$.

We let A^-_k be the filtration sublevel of $C_\# = \underline{\mathrm{CFK}}^-(K) \otimes \mathrm{CFK}^-(\overline{L})$ defined by $j \leq k$, ie generated by all elements with Alexander filtration level at most k.

If $M \leq k$, the entire staircase is contained in the subcomplex A^-_k. That is, the inclusion $A^-_k \to C_\#$ induces a surjection onto the tower, hence $\nu^+(K \# \overline{L}) \leq k$. In particular, if $M \leq 0$, then $\nu^+(K \# \overline{L}) = 0 = (M)_+$.

If $M > 0$, for each $k < M$ the complex A^-_k misses at least one of the generators of the chain; this implies that the inclusion $A^-_k \to \mathrm{CFK}^-(K)$ does not induce a surjection onto the tower. It follows that $V_k(K \# \overline{L}) > 0$. Hence, by definition of ν^+, we have $\nu^+(K \# \overline{L}) = M = (M)_+$, as desired. $\square$

Proof of Theorem 3.1 As remarked above, the values of a_k and a'_k determine the positions of the generators in the staircase. By the symmetry of the Alexander polynomial (and hence of the staircases), $\Gamma_L(a'_k) - a'_k = a_k$ for each k (compare with [13, Section 4]).

Moreover, for any $a'_k \le n < a'_{k+1}$, we have $\Gamma_L(n) - n = a_k$, and for every $a'_d \le n$ we have $\Gamma_L(n) - n = a_d$. Furthermore, as $\Gamma_K(\cdot)$ is strictly increasing, $n \mapsto \Gamma_K(n) - n$ is nondecreasing; therefore, for any $a'_k \le n < a'_{k+1}$ we have $\Gamma_K(a'_j) - a'_j \le \Gamma_K(n) - n$, so

$$\begin{aligned} a_k + a'_k - \Gamma_K(a'_k) &= (\Gamma_L(a'_k) - a'_k) - (\Gamma_K(a'_k) - a'_k) \\ &= \max_{a'_k \le n < a'_{k+1}} \{(\Gamma_L(n) - n) - (\Gamma_K(n) - n)\} \\ &= \max_{a'_k \le n < a'_{k+1}} \{\Gamma_L(n) - \Gamma_K(n)\}. \end{aligned}$$

□

Remark The same argument shows that, for every $m \le V_0(K \# \bar{L})$,

$$\min\{\, i \mid V_i(K \# \bar{L}) = m \,\} = \big(g(K) - g(L) + \max_{n \ge 0}\{\Gamma_L(n) - \Gamma_K(n+m)\}\big)_+,$$

thus allowing one to compute all correction terms of $K \# \bar{L}$ from the enumerating functions of K and L.

4 Semicontinuity of the semigroups

In this section we prove Theorem 1.2 about the deformations of plane curve singularities. We note here that our Theorem 1.2 differs slightly from both of the results mentioned in the introduction: it reproves [10, Proposition 4.5.1] in the special case when both the central and the perturbed singularity are irreducible, but (in the spirit of [7]) using only smooth topological (not analytic) methods; however, we do not restrict ourselves to δ–constant deformations, as opposed to [7, Theorem 2.16].

In the context of deformations, inequalities which hold for certain invariants are usually referred to as *semicontinuity* of that particular invariant. Our result can be viewed as the semicontinuity of the semigroups (resembling the spectrum semicontinuity; see also [7, Section 3.1.B]).

For a brief introduction to the topic of deformations, we follow mainly [7, Section 1.5] and adapt the notions and definitions from there. By a *deformation* of a singularity with link K we mean an analytic family $\{F_s\}$ of polynomials parametrised by $|s| < 1$ such that there exists a ball $B \subset \mathbb{C}^2$ with the following properties:

- the only singular point of F_0 inside B is at the origin;

- $\{F_s = 0\}$ intersects ∂B transversely and $\{F_s = 0\} \cap \partial B$ is isotopic to K for every $|s| < 1$;
- the zero set of F_s has only isolated singular points in B for every $|s| < 1$;
- all the singular points of F_s inside B are irreducible for every $|s| < 1$;
- all fibres F_s with $s \neq 0$ have the same collection of local topological type of singularities.

For simplicity, we also assume that there is only one singular point of F_s inside B for each s. If such an analytic family of polynomials $\{F_s\}$ exists, we say that the singularity of F_0 at the origin *has a deformation* to the singularity of $F_{1/2}$.

Consider now a sufficiently small ball B_2 around the singular point of $F_{1/2}$ such that $\{F_{1/2} = 0\} \cap \partial B_2$ is isotopic to L, the link of the perturbed singular point. Then $V = \{F_{1/2} = 0\} \cap \overline{B} \setminus B_2$ is a genus-g cobordism between K and L, where $g = g(K) - g(L)$. By a slight abuse of notation, we also say that L is a deformation of K.

Let K and L be two L–space knots, with corresponding semigroups Γ_K and Γ_L, respectively. We define the *semigroup counting functions* $R_K, R_L\colon \mathbb{N} \to \mathbb{N}$ as $R_K(n) = \#[0, n) \cap \Gamma_K$ and $R_L(n) = \#[0, n) \cap \Gamma_L$. For simplicity, we allow n to run on negative numbers as well: if $n < 0$, then we define $R_K(n) = R_L(n) = 0$. In this section, we will assume that $g(K) = \delta_K \geq \delta_L = g(L)$.

Proposition 4.1 *Assume there is a genus-g cobordism between two L–space knots K and L. Then for any $a \in \mathbb{Z}$ we have*

$$R_K(a + \delta_K) \leq R_L(a + \delta_L + g).$$

Proof Since ν^+ is a lower bound for the cobordism genus, by Theorem 1.1 for any $m \in \mathbb{N}$ we have

$$\delta_K - \delta_L + \Gamma_L(m) - \Gamma_K(m) \leq g,$$

equivalently,

$$\Gamma_L(m) - \delta_L - g \leq \Gamma_K(m) - \delta_K.$$

Notice that since $\Gamma_K(m) = a$ implies $R_K(a) = m$, and the largest a for which $R_K(a) = m$ is $a = \Gamma_K(m)$ (and analogously for Γ_L), the above inequality can be interpreted as

$$R_K(a + \delta_K) \leq R_L(a + \delta_L + g). \qquad \square$$

The proposition above should be compared with [7, Theorem 2.14]. There, Borodzik and Livingston introduced the concept of *positively self-intersecting concordance*, and their result is the counterpart of Proposition 4.1 above: their assumption is on the double point count of the positively self-intersecting concordance, while ours is on the cobordism genus. The former is related to the 4–ball crossing number considered in Proposition 1.6.

The assumption in [7] allowed Borodzik and Livingston to treat δ–constant deformations (because irreducible singularities can be perturbed to transverse intersections). However, equipped with Proposition 4.1, we can prove the semigroup semicontinuity even if the deformation is not δ–constant (but assuming that there is only one singularity in the perturbed curve $\{F_{1/2} = 0\}$).

Recall that, with the definition of the function R in place, Theorem 1.2 asserts that if L is a deformation of K then $R_K(n) \leq R_L(n)$ for each nonnegative integer n.

Proof of Theorem 1.2 Apply Proposition 4.1 with $a = n - \delta_K$ and recall that $g = \delta_K - \delta_L$ in this case. □

Remark In [7, Section 3], the example of torus knots $T_{6,7}$ and $T_{4,9}$ was extensively studied. The semigroup semicontinuity proved in Theorem 1.2 obstructs the existence of a deformation between the corresponding singularities. Since the difference of the δ–invariants is 3, a deformation from $T_{6,7}$ to $T_{4,9}$ would produce a genus-3 cobordism between the two knots. However, the bound coming from ν^+ is 4 (compare with [7, Remark 3.1]).

We now turn to proving Corollary 1.3, ie that the braid index/multiplicity is nonincreasing under deformations.

Proof of Corollary 1.3 The multiplicity $m(L)$ of the singularity whose link is L is the minimal positive element in Γ_L. In particular, $R_L(m(L)) = 2$, and $R_L(m) = 1$ for $0 \leq m < m(L)$; symmetrically, $R_K(m) \geq 2$ for every $m \geq m(K)$. Let us apply Theorem 1.2 with $n = m(L)$; we obtain $R_K(m(L)) \leq R_L(m(L)) = 2$, hence $m(L) \leq m(K)$, as desired. □

5 Bounds on the slice genus and concordance unknotting number

Recall that $\nu^+(K) \leq g_*(K)$ for every knot K; as outlined in the introduction, this shows that $\nu^+(K \# \overline{L})$ gives a lower bound on the genus of cobordisms between K and L. Notice that $\nu^+(L \# \overline{K})$ gives a bound, too, and the two bounds are often different.

We now state a preliminary lemma that we will use to prove Theorem 1.4, ie that trading a negative crossing for a positive one does not decrease ν^+, nor does it increase it by more than 1.

Lemma 5.1 *If there is a genus-g cobordism between two knots K and L then, for each $m \geq 0$,*

$$V_{m+g}(K) \leq V_m(L).$$

As a consequence, $\nu^+(K) \leq \nu^+(L) + g$.

Before proving the lemma, we observe some of its consequences. Most notably, it allows us to generalise Theorem 1.2 to the case of more than one irreducible singularity (both in the central and deformed fibre).

Remark In the case of algebraic knots, the lemma is equivalent to Proposition 4.1. Indeed, using the symmetry property of the semigroup, one has that $R_K(a+\delta_K) = R_K(\delta_K - a) + a$ and $R_L(a+\delta_L+g) = R_L(\delta_L - a - g) + a + g$ for every integer a. Using these substitutions in both sides of the statement of Proposition 4.1, we obtain

$$R_K(\delta_K - a) \leq R_L(\delta_L - a - g) + g.$$

If we now set $a = -g-m$, we get

$$R_K(\delta_K + m + g) \leq R_L(\delta_L + m) + g,$$

and by [3, Equation (5.1)] we have $R_K(\delta_K + m + g) = V_{m+g}(K) + m + g$ and $R_L(\delta_L + m) = V_m(L) + m$, thus proving the equivalence of the two statements.

Since the proof of Theorem 1.2 relies on Proposition 4.1, which is in turn equivalent to Lemma 5.1, we can use the latter to generalise its statement. In order to do so, we introduce the concept of the infimum convolution of two functions [6; 4]: given $R_1, R_2\colon \mathbb{N} \to \mathbb{N}$ bounded below, we define the *infimum convolution* of R_1 and R_2 as

$$(R_1 \diamond R_2)(n) := \min_{i+j=n} R_1(i) + R_2(j).$$

Theorem 5.2 *Let $\{F_s\}_{|s|<1}$ be a deformation of F_0, and suppose that F_0 has only irreducible singularities $K_1, \dots, K_a$, while $F_{1/2}$ has irreducible singularities $L_1, \dots, L_b$ (and possibly other reducible singularities). Then, for each nonnegative integer n,*

$$(R_{K_1} \diamond \cdots \diamond R_{K_a})(n) \leq (R_{L_1} \diamond \cdots \diamond R_{L_b})(n).$$

Proof (sketch) Similarly as how we argued in Section 4, it is easy to show that a deformation gives rise to a cobordism Σ_0 from the link $K_1 \cup \cdots \cup K_a$ in the disjoint union of a copies of S^3 to the link $L_1 \cup \cdots \cup L_b$ in the disjoint union of b copies of S^3, living in S^4 with $a+b$ open balls removed. This cobordism will be singular if there are reducible singularities in $F_{1/2}$.

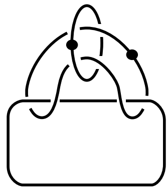


Figure 1: The Borromean knot $K_{B,1}$. The Borromean knot $K_{B,g}$ is the connected sum of g copies of $K_{B,1}$.

We resolve all singularities of Σ_0, replacing each of them with the Milnor fibre of the corresponding reducible singularity and obtain a smooth cobordism, Σ_1; note that the difference $g(\Sigma_1)-g(\Sigma_0)$ is the sum of all δ–invariants of the reducible singularities of $F_{1/2}$.

We can now carve paths along the cobordism Σ_1 connecting all the boundary components containing a K_i and all boundary components containing an L_j, thus obtaining a smooth cobordism Σ from $K=K_1\#\cdots\# K_a$ to $L=L_1\#\cdots\# L_b$. Note that this does not change the genus, ie $g(\Sigma)=g(\Sigma_1)$.

Similarly as in the irreducible case, we have $g(\Sigma)=g(K)-g(L)$; using [6, Theorems 5.4 and 5.6] we know that $V_i(K)+i=(R_{K_1}\diamond\cdots\diamond R_{K_a})(g(K)+i)$ and $V_i(L)+i=(R_{L_1}\diamond\cdots\diamond R_{L_b})(g(L)+i)$.

The statement now follows from Lemma 5.1 and the remark following the lemma, as in the proof of Theorem 1.2. □

Proof of Lemma 5.1 Consider the 4–manifold W obtained by attaching a 4–dimensional 2–handle to $S^3\times[0,1]$ along $L\times\{1\}\subset S^3\times\{1\}$, with framing $n\geq 2\nu^+(L)$.

The cobordism is a genus-g embedded surface F in $S^3\times[0,1]$, whose boundary components are $K\times\{0\}$ and $L\times\{1\}$. Capping off the latter boundary component in W, and taking the cone over $(S^3\times\{0\},K)$, we obtain a singular genus-g surface $\widehat{F}\subset W'=W\cup B^4$, whose only singularity is a cone over K.

As argued in [3, Section 4; 5, Theorem 3.1], the boundary ∂N of a regular neighbourhood N of $\widehat{F}$ in W' is diffeomorphic to the 3–manifold Y_n obtained as n–surgery along the connected sum of K and the Borromean knot $K_{B,g}$ in $\#^{2g}(S^2\times S^1)$. It follows that $Z=-(W'\setminus N)$ can be viewed as a cobordism from $S^3_n(L)$ to Y_n.

We can view N as the surgery cobordism from $\#^{2g}(S^2\times S^1)$ to Y_n, filled with a 1–handlebody; since the class of $[\widehat{F}]$ generates both $H_2(N)$ and $H_2(-W')$, we obtain that the restriction of any spin^c structure on $-W'$ to Z induces an isomorphism between (torsion) spin^c structure on its two boundary components that respects the surgery-induced labelling. Moreover, we also obtain that $b_2^+(Z)=0$.

The 3–manifold Y_n has standard HF^∞ [3; 5], and its bottom-most correction terms have been computed in [3, Proposition 4.4; 5, Theorem 6.10]:

$$d_b(Y_n, \mathfrak{t}_m) = \min_{0\le k\le g}\{2k - g - 2V_{m+g-2k}(K)\} - \frac{n-(2m-n)^2}{4n}.$$

We observe that, choosing $k=0$ in the minimum, we obtain the inequality

$$d_b(Y_n, \mathfrak{t}_m) \le -g - 2V_{m+g}(K) - \frac{n-(2m-n)^2}{4n}.$$

Applying the last inequality and [2, Theorem 4.1] to Z, seen as a negative semidefinite cobordism from $S^3_n(L)$ to Y_n, we get

$$d(S^3_n(L), \mathfrak{t}_m) \le d_b(Y_n, \mathfrak{t}_m) + g,$$

from which we get

$$-2V_m(L) \le -g - 2V_{m+g}(K) + g \iff V_{m+g}(K) \le V_m(L).$$

The last part of the statement now follows from the observation that $V_{\nu^+(L)+g}(K) \le V_{\nu^+(L)}(L) = 0$, hence $\nu^+(K) \le \nu^+(L) + g$, as desired. □

We are now in position to prove Theorem 1.4, which asserts that, if K_+ and K_- differ at a single crossing (which is positive for K_+ and negative for K_-), then $\nu^+(K_-) \le \nu^+(K_+) \le \nu^+(K_-) + 1$.

Proof of Theorem 1.4 The inequality $\nu^+(K_-) \le \nu^+(K_+)$ readily follows from [4, Theorem 6.1]: the latter states that for each nonnegative integer n we have $V_n(K_-) \le V_n(K_+)$. Applying the inequality with $n = \nu^+(K_+)$, we obtain $V_{\nu^+(K_+)}(K_-) \le 0$, hence $\nu^+(K_-) \le \nu^+(K_+)$, as desired.

The inequality $\nu^+(K_+) \le \nu^+(K_-) + 1$ follows from Lemma 5.1 above; in fact, there is a genus-1 cobordism from K_- to K_+ obtained by smoothing the double point of the regular homotopy associated with the crossing change, and the previous lemma concludes the proof. □

Remark In fact, the second inequality follows from [4, Theorem 6.1] as well: Borodzik and Hedden prove that, in the notation of the proposition, $V_{m+1}(K_+) \le V_m(K_-)$, and the claim about ν^+ follows as in the proof of Lemma 5.1. However, Lemma 5.1 is stronger than [4, Theorem 6.1], and we think it might be of independent interest.

We now turn to applications to other, more geometrically defined, concordance invariants, and we prove Proposition 1.6.

Proof of Proposition 1.6 We need at least $\nu^+(K)$ negative crossing changes and at least $\nu^+(\overline{K})$ positive crossing changes to turn K into a knot K_0 such that $\nu^+(K_0) = \nu^+(\overline{K}_0) = 0$. In particular, we need to change at least $\nu^+(K) + \nu^+(\overline{K})$ crossings to make K slice, hence $u_s(K) \geq \nu^+(K) + \nu^+(\overline{K})$.

As for the concordance unknotting number, one simply observes that $\nu^+(K)$ and $\nu^+(\overline{K})$ are concordance invariants, hence every knot in the same concordance class of K has unknotting number at least $\nu^+(K) + \nu^+(\overline{K})$.

Finally, [18, Proposition 2.1] asserts that every immersed concordance can be factored into two concordances and a sequence of crossing changes. That is, given an immersed concordance from K to the unknot with c double points, there exist knots K_0 and K_1 such that K_0 is slice, K_1 is concordant to K, and there is a sequence of c crossing changes from K_0 to K_1; from the proposition above, it follows that $c \geq \nu^+(K_0 \# K_1) + \nu^+(\overline{K_0 \# K_1}) = \nu^+(K) + \nu^+(\overline{K})$. □

6 Subadditivity of ν^+

The goal of this section is to prove Theorem 1.5. We start with a preliminary proposition. In the course of the proof, we will make use of twisted correction terms, as defined in [2]. These are a generalisation of ordinary and bottom-most correction terms to arbitrary 3–manifolds; specifically, given a torsion spin^c structure $\mathfrak{t}$ on a 3–manifold Y, there is an associated rational number $\underline{d}(Y, \mathfrak{t})$, which is a rational homology cobordism invariant of the pair $(Y, \mathfrak{t})$.

When Y is a rational homology sphere, $\underline{d}(Y, \mathfrak{t}) = d(Y, \mathfrak{t})$. If, on the other hand, Y is obtained as 0–surgery along a knot in S^3, equipped with its unique torsion spin^c structure $\mathfrak{t}$, then $\underline{d}(Y, \mathfrak{t}) = d_b(Y, \mathfrak{t})$ (see [2, Section 3.3]).

Moreover, much like d_b, the twisted correction term $\underline{d}$ behaves well under negative semidefinite cobordisms (see [2, Section 4]).

Proposition 6.1 *For any two knots $K, L \subset S^3$ and any two nonnegative integers m and n, we have*

$$V_{m+n}(K \# L) \leq V_m(K) + V_n(L).$$

Proof Consider the surgery diagrams in Figure 2 and Figure 3, representing a closed 4–manifold X and a 4–dimensional cobordism W from $-S^3_{2(m+n)}(K \# L)$ to $-(S^3_{2m}(K) \# S^3_{2n}(L))$. One should be careful with orientation reversals here; in particular, notice that in Figure 3 we represent the cobordism $\overline{W}$ obtained by turning W upside down.

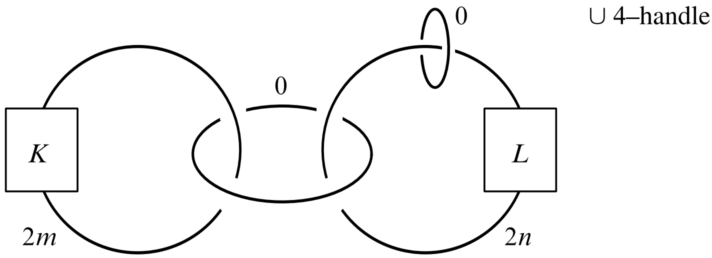


Figure 2: The surgery diagram for the closed 4–manifold X

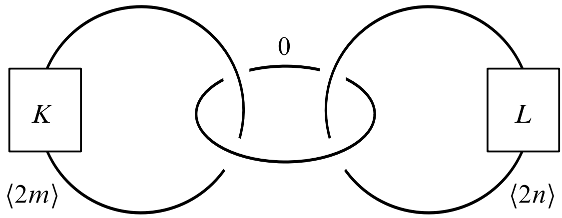


Figure 3: The surgery diagram for the upside-down cobordism $\overline{W}$ from $S^3_{2m}(K) \# S^3_{2n}(L)$ to $S^3_{2(m+n)}(K \# L)$. The coefficients in brackets represent the negative boundary $\partial_- W$.

As observed by Owens and Strle [17], if $m, n > 0$, then $W \subset X$ is a negative definite cobordism from $S^3_{2m}(K) \# S^3_{2n}(L)$ to $S^3_{2(m+n)}(K \# L)$ with $H_2(W;\mathbb{Z}) = \mathbb{Z}$ and $\chi(W) = 1$. When $m = 0$ or $n = 0$, W has signature $\sigma(W) = 0$; therefore, regardless of positivity of m and n, W is negative semidefinite. Moreover, W is obtained from $\partial_- W$ by attaching a single 2–handle, and this does not decrease the first Betti number of the boundary. It follows that we are in the right setup to apply [2, Theorem 4.1].

The 4–manifold X is even; since 0 is a characteristic vector, it is the first Chern class of a spin^c structure $\mathfrak{s}_0$ on X. The spin^c structure $\mathfrak{s}_0$ restricts to the spin^c structure on W with trivial first Chern class, hence $c_1(\mathfrak{s}_0)^2 = 0$.

Notice also that $X \setminus W$ is the disjoint union of two 4–manifolds: one is the boundary connected sum of the surgery handlebodies for $S^3_{2m}(K)$ and $S^3_{2n}(L)$, and the other is the surgery handlebody for $S^3_{2(m+n)}(K \# L)$ with the reversed orientation. In particular, labelling of the restriction of $\mathfrak{s}_0$ onto the two boundary components of W is determined by the evaluation of $c_1(\mathfrak{s}_0)$ on the generators of the second cohomology of the two pieces [23, Section 2.4].

With the chosen convention for labelling spin^c structures (1), since $c_1(\mathfrak{s}_0) = 0$, the spin^c structure $\mathfrak{s}_0$ restricts to $\mathfrak{t}_m$ on $S^3_{2m}(K)$, to $\mathfrak{t}_n$ on $S^3_{2n}(L)$, and to $\mathfrak{t}_{m+n}$ on $S^3_{2(m+n)}(K)$.

We observe here that

$$\underline{d}(-S^3_0(\overline{K}),\mathfrak{t}_0)+\tfrac{1}{2}b_1(-S^3_0(\overline{K}))=2V_0(K),$$

and the same holds for L and $K\#L$ (compare with [20, Proposition 4.12] and [2, Example 3.9]). When $m>0$, however,

$$\underline{d}(-S^3_{2m}(\overline{K}),\mathfrak{t}_m)+\tfrac{1}{2}b_1(-S^3_{2m}(\overline{K}))=d(-S^3_{2m}(\overline{K}),\mathfrak{t}_m)=\tfrac{1}{4}+2V_m(K),$$

and analogous formulae hold for L and $K\#L$.

We now apply additivity of $\underline{d}$ [2, Proposition 3.7] and [2, Theorem 4.1] to W to obtain the inequality

$$\begin{aligned}(3)\quad & b_2^-(W)+4d\big(-(S^3_{2m}(\overline{K})\#S^3_{2n}(\overline{L})),\mathfrak{t}_m\#\mathfrak{t}_n\big)+2b_1(-(S^3_{2m}(\overline{K})\#S^3_{2n}(\overline{L})))\\ &\le 4\underline{d}(-S^3_{2m}(\overline{K}),\mathfrak{t}_m)+2b_1(-S^3_{2m}(\overline{K}))+4\underline{d}(-S^3_{2n}(\overline{L}),\mathfrak{t}_n)+2b_1(-S^3_{2n}(\overline{L})).\end{aligned}$$

When m and n are both positive, (3) becomes

$$1+1+8V_{m+n}(K\#L)\le 1+8V_m(K)+1+8V_n(L).$$

When exactly one among m and n vanishes, say $m=0$, (3) turns into

$$1+8V_n(K\#L)\le 8V_0(K)+1+8V_n(L).$$

Finally, when $m=n=0$, (3) reads

$$8V_0(K\#L)\le 8V_0(K)+8V_0(L).$$

In all cases, we have proved that $V_{m+n}(K\#L)\le V_m(K)+V_n(L)$, as desired. □

We are now ready to prove Theorem 1.5, ie that ν^+ is subadditive.

Proof of Theorem 1.5 This now follows from Proposition 6.1 by setting $m=\nu^+(K)$ and $n=\nu^+(L)$. In fact, since $V_m(K)=V_n(L)=0$,

$$V_{m+n}(K\#L)\le V_m(K)+V_n(L)=0;$$

that is, $\nu^+(K\#L)\le m+n=\nu^+(K)+\nu^+(L)$. □

7 Examples

In this section we study a 3–parameter family of pairs of torus knots on which the lower bound given by ν^+ is sharp. We first start with a 1–parameter subfamily that we study in some detail, and we then turn to the whole family. The techniques used here are inspired by Baader's "scissor equivalence" [1].

Example 7.1 We are going to present an example in which the bound provided by ν^+ on the genus of a cobordism between torus knots is stronger than the ones given by the Tristram–Levine signature function, τ, s and Υ, and moreover it is sharp.

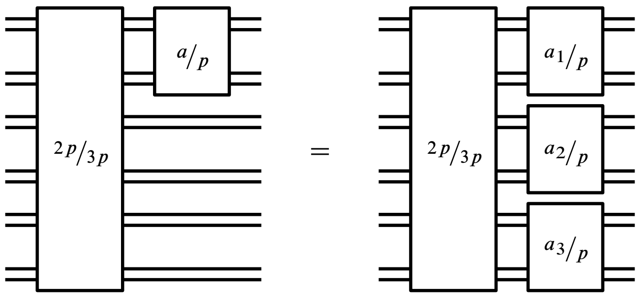


Figure 4: The knot $K_{a,p}$; the boxes indicate the number of full twists. Equality holds whenever $a = a_1 + a_2 + a_3$.

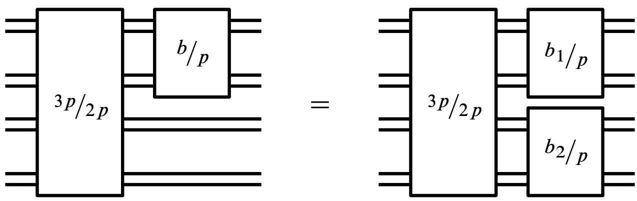


Figure 5: The knot $K'_{b,p}$; the boxes indicate the number of full twists. Equality holds whenever $b = b_1 + b_2$.

Define the two families of links $K_{a,p}$ and $K'_{b,p}$ as the closure of the braids pictured in Figures 4 and 5. Notice that $K_{a,p}$ and $K'_{b,p}$ are (p, s)–cables of the trefoil for some s, and that they are knots if and only if $\gcd(a, p) = 1$ and $\gcd(b, p) = 1$, respectively. Moreover, $K_{a,p}$ is the product of $2p(3p-1) + a(p-1)$ positive generators of the braid group on $3p$ strands, hence its closure represents a transverse knot in the standard contact 3–sphere with self-linking number $6p^2 + (a-5)p - a$. Since for closures of positive braids the self-linking number agrees with the Seifert genus, we can compute the cabling parameter $s = 6p + a$. In conclusion, we have shown that $K_{a,p}$ is the $(p, 6p + a)$–cable of $T_{2,3}$.

The same argument applies to $K'_{b,p}$, the self-linking number computation yields $6p^2 + (b-5)p - b$, hence showing that $K'_{b,p}$ is the $(p, 6p + b)$–cable of $T_{2,3}$. In particular $K_{a,p}$ and $K'_{b,p}$ are isotopic if and only if $a = b$.

Now consider the knots $K_{12,p} = K'_{12,p}$. Denote by σ_i the i^{th} elementary generator of the braid group, and, whenever $i < j$, denote by $\sigma_{i,j}$ the product $\sigma_i \sigma_{i+1} \cdots \sigma_j$.

Setting $a_1 = a_2 = a_3 = 4$ in the right-hand side of Figure 4 exhibits $K_{12,p}$ as the closure of the braid

$$\sigma_{1,3p-1}^{2p} \cdot \overbrace{\underbrace{\sigma_{1,p-1}^{4}}_{\Sigma_1^4} \cdot \underbrace{\sigma_{p+1,2p-1}^{4}}_{\Sigma_2^4} \cdot \underbrace{\sigma_{2p+1,3p-1}^{4}}_{\Sigma_3^4}}^{\Pi} .$$

The three elements Σ_1, Σ_2 and Σ_3 commute, hence $\Pi = (\Sigma_1\Sigma_2\Sigma_3)^4$. Now, notice that $\Sigma_1\sigma_p\Sigma_2\sigma_{2p}\Sigma_3 = \sigma_{1,3p-1}$. Since adding a generator σ_i corresponds to attaching a band between two strands, we produce a cobordism built out of 8 bands from $K_{12,p}$ to $T_{2p+4,3p}$; if p is coprime with 6, both ends of the cobordism are connected, and its genus is 4.

An analogous argument, setting $b_1 = b_2 = 6$ in the right-hand side of Figure 5 produces a 6–band, genus-3 cobordism from $K'_{12,p}$ to $T_{2p,3p+6}$ whenever p is coprime with 6.

Suppose now that $p \equiv 5 \pmod 6$ and $p \geq 11$. Gluing the two cobordisms above yields a genus-7 cobordism between $K = T_{2p+4,3p}$ and $L = T_{2p,3p+6}$.

Applying Proposition 3.2 above we obtain a sharp bound on the slice genus; in fact, in the same notation as in Proposition 3.2, we have

- $2\delta_K = 2g(K) = 6p^2 + 7p - 3$ and $2\delta_L = 2g(L) = 6p^2 + 7p - 5$;
- $\Gamma_K(2) = 3p$ and $\Gamma_L(2) = 3p + 6$;
- $\Gamma_K(3) = 4p + 8$ while $\Gamma_L(3) = 4p$.

It follows that

$$\nu^+(K \# \overline{L}) \geq 1 + \Gamma_L(2) - \Gamma_K(2) = 7,$$
$$\nu^+(L \# \overline{K}) \geq -1 + \Gamma_K(3) - \Gamma_L(3) = 7.$$

A direct computation using [19, Theorem 1.15] shows that for $p = 11,\ 17,\ 23,\ 29$ the bound given by Υ is 3, the one given by the Tristram–Levine signatures is either 2 or 5, and the one given by τ and s is 1.

Moreover, we need at least 7 positive and 7 negative crossing changes to turn K into L, hence their Gordian distance is at least 14. Additionally, suppose that we have a factorisation of the cobordism above into genus-1 cobordisms, and suppose that one of these cobordisms goes from K_1 to K_2. Then both $\nu^+(K_2) = \nu^+(K_1) - 1$ and $\nu^+(\overline{K}_2) = \nu^+(\overline{K}_1) - 1$.

Example 7.2 We can promote the family above to a family parametrised by suitable triples of integers (p, q, r) as follows: Instead of considering the $(p, 6p + 12)$–cable of the trefoil $K_{12,p} = K'_{12,p}$ we can consider the $(p, qr(p + 2))$–cable $K^p_{q,r}$ of $T := T_{q,r}$. The first condition we impose on the triple (p, q, r) is that $q < r$ and $\gcd(q, r) = 1$.

By looking at $K^p_{q,r}$ as a cable of T seen as the closure of an r–braid, we can glue $2q\cdot(r-1)$ bands to $K^p_{q,r}$ and obtain $K=T_{q(p+2),rp}$. Call $x_1=q(p+2)$ and $x_2=rp$ the two generators of the semigroup Γ_K.

By viewing $K^p_{q,r}$ as a cable of T seen as the closure a q–braid instead, we see that we can glue $2r\cdot(q-1)$ bands to $K^p_{q,r}$ and obtain $L=T_{qp,r(p+2)}$. Call $y_1=qp$ and $y_2=r(p+2)$ the two generators of the semigroup Γ_L.

If $\gcd(p,2qr)=\gcd(p+2,2qr)=1$, both K and L have one component, ie they are torus knots; eg both equalities hold if $p\equiv -1 \pmod{2qr}$. Moreover, $\delta_K-\delta_L=g(K)-g(L)=r-q$, and above we produced a cobordism of genus $2qr-q-r$ between K and L, made of $4qr-2q-2r$ bands. Hence, $\nu^+(K\#\overline{L}), \nu^+(L\#\overline{K})\le 2qr-q-r$.

Choose p sufficiently large; it is elementary to check that if $p\ge 2qr-1$ then, for $n_1=\delta_T+q-1$ and $n_2=\delta_T+r-1$, we have

$$\begin{aligned}
\Gamma_T(n_1)&=(q-1)r, & \Gamma_T(n_2)&=(r-1)q,\\
\Gamma_K(n_1)&=(q-1)x_2=(q-1)rp, & \Gamma_K(n_2)&=(r-1)x_1=(r-1)q(p+2),\\
\Gamma_L(n_1)&=(q-1)y_2=(q-1)r(p+2), & \Gamma_L(n_2)&=(r-1)y_1=(r-1)qp.
\end{aligned}$$

If we set $n=n_1$ in Theorem 1.1 we obtain

$$\nu^+(K\#\overline{L})\ge \delta_K-\delta_L+\Gamma_L(n_1)-\Gamma_K(n_1)=2qr-q-r.$$

Reversing the roles of K and L and setting $n=n_2$ yields

$$\nu^+(L\#\overline{K})\ge \delta_L-\delta_K+\Gamma_K(n_2)-\Gamma_L(n_2)=2qr-q-r.$$

The lower bound for the genus given by ν^+ is in this case is tight, as the upper and lower bounds match, and moreover the Gordian distance between K and L is at least $4qr-2q-2r$.

References

[1] **S Baader**, *Scissor equivalence for torus links*, Bull. Lond. Math. Soc. 44 (2012) 1068–1078 MR

[2] **S Behrens**, **M Golla**, *Heegaard Floer correction terms, with a twist*, preprint (2015) arXiv To appear in Quantum Topology

[3] **J Bodnár**, **D Celoria**, **M Golla**, *Cuspidal curves and Heegaard Floer homology*, Proc. Lond. Math. Soc. 112 (2016) 512–548 MR

[4] **M Borodzik**, **M Hedden**, *The Υ function of L–space knots is a Legendre transform*, preprint (2015) arXiv

[5] **M Borodzik**, **M Hedden**, **C Livingston**, *Plane algebraic curves of arbitrary genus via Heegaard Floer homology*, Comment. Math. Helv. 92 (2017) 215–256 MR

[6] **M Borodzik**, **C Livingston**, *Heegaard Floer homology and rational cuspidal curves*, Forum Math. Sigma 2 (2014) art. id. e28 MR

[7] **M Borodzik**, **C Livingston**, *Semigroups, d–invariants and deformations of cuspidal singular points of plane curves*, J. Lond. Math. Soc. 93 (2016) 439–463 MR

[8] **E Brieskorn**, **H Knörrer**, *Plane algebraic curves*, Birkhäuser, Basel (1986) MR

[9] **D Eisenbud**, **W Neumann**, *Three-dimensional link theory and invariants of plane curve singularities*, Annals of Mathematics Studies 110, Princeton Univ. Press (1985) MR

[10] **E Gorsky**, **A Némethi**, *Lattice and Heegaard Floer homologies of algebraic links*, Int. Math. Res. Not. 2015 (2015) 12737–12780 MR

[11] **M Hedden**, *On knot Floer homology and cabling, II*, Int. Math. Res. Not. 2009 (2009) 2248–2274 MR

[12] **J Hom**, **Z Wu**, *Four-ball genus bounds and a refinement of the Ozváth–Szabó tau invariant*, J. Symplectic Geom. 14 (2016) 305–323 MR

[13] **D Krcatovich**, *The reduced knot Floer complex*, Topology Appl. 194 (2015) 171–201 MR

[14] **R A Litherland**, *Signatures of iterated torus knots*, from "Topology of low-dimensional manifolds" (R A Fenn, editor), Lecture Notes in Math. 722, Springer (1979) 71–84 MR

[15] **C Livingston**, **C A Van Cott**, *The four-genus of connected sums of torus knots*, preprint (2015) arXiv

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

[17] **B Owens**, **S Strle**, *Dehn surgeries and negative-definite four-manifolds*, Selecta Math. 18 (2012) 839–854 MR

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

[19] **P Ozsváth**, **A Stipsicz**, **Z Szabó**, *Concordance homomorphisms from knot Floer homology*, preprint (2014) arXiv

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

[21] **P Ozsváth**, **Z Szabó**, *Knot Floer homology and the four-ball genus*, Geom. Topol. 7 (2003) 615–639 MR

[22] **P Ozsváth**, **Z Szabó**, *Holomorphic disks and topological invariants for closed three-manifolds*, Ann. of Math. 159 (2004) 1027–1158 MR

[23] **P S Ozsváth**, **Z Szabó**, *Knot Floer homology and integer surgeries*, Algebr. Geom. Topol. 8 (2008) 101–153 MR

[24] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) MR Available at `http://search.proquest.com/docview/305332635`

[25] **J Rasmussen**, *Khovanov homology and the slice genus*, Invent. Math. 182 (2010) 419–447 MR

[26] **C T C Wall**, *Singular points of plane curves*, London Mathematical Society Student Texts 63, Cambridge Univ. Press (2004) MR

Mathematics Department, Stony Brook University
Stony Brook, NY 11794-3651, United States

Mathematical Institute, University of Oxford
Oxford, OX2 6GG, United Kingdom

Department of Mathematics, Uppsala University
SE-751 06 Uppsala, Sweden

`jozef.bodnar@stonybrook.edu`, `daniele.celoria@maths.ox.ac.uk`,
`marco.golla@math.uu.se`

Received: 23 November 2016 Revised: 23 January 2017

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in AGT is available on the AGT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 17 Issue 4 (pages 1917–2564) 2017