Algebraic & Geometric Topology

Volume 24 (2024)

Issue 5 (pages 2389–2970)

msp

Algebraic & Geometric Topology

msp.org/agt

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

The subscription price for 2024 is US $705/year for the electronic version, and $1040/year (+$70, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Algebraic & Geometric Topology (ISSN 1472-2747 printed, 1472-2739 electronic) is published 9 times per year and continuously online, by Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840. Periodical rate postage paid at Oakland, CA 94615-9651, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

Algebraic & Geometric Topology **24**:5 (2024) 2389–2449
DOI: 10.2140/agt.2024.24.2389
Published: 19 August 2024

Formal contact categories

BENJAMIN COOPER

To each oriented surface Σ, we associate a differential graded category $\mathcal{K}o(\Sigma)$. The homotopy category $\mathrm{Ho}(\mathcal{K}o(\Sigma))$ is a triangulated category which satisfies properties akin to those of the contact categories studied by K Honda. These categories are also related to the algebraic contact categories of Y Tian and to the bordered sutured categories of R Zarev.

53D10; 18G55

1 Introduction

Our purpose here is to associate a differential graded category $\mathcal{K}o(\Sigma)$ to each oriented surface Σ. This category is used to study comparison problems between the categories associated to surfaces by Seiberg–Witten-type manifold invariants. For example, we prove that the categories associated to the disk $(D^2, 2n)$ with $2n$ marked points by each theory are equivalent, and there is a functorial relationship between the categories associated to a surfaces with boundary when they can be defined.

1.1 The unicity of Floer-type invariants of 3–manifolds

In [42; 43] P Ozsváth and Z Szabó introduced invariants of 3–manifolds known as the Heegaard–Floer homologies. Depending upon the setting of a parameter U, there are homology groups: $HF^-_*(M)$, $HF^+_*(M)$ and $HF^\infty_*(M)$ which fit into a long exact sequence:

$$\cdots \to HF^-_*(M) \to HF^\infty_*(M) \to HF^+_*(M) \to \cdots . \tag{1-1}$$

When $U = 0$, there are simpler invariants $\widehat{HF}_*(M)$. The Heegaard–Floer theory has had a profound effect on the study of 3–manifolds and 4–manifolds; see Juhász [23]. This is in part because it was originally conceived of as a way to obtain information in the Seiberg–Witten invariants; see Donaldson [7], Kronheimer and Mrowka [28] and Witten [61]. The relationship between the Heegaard–Floer homology theory and the Seiberg–Witten Floer homology was recently articulated by two independent groups of researchers: Ç Kutluhan, Y-J Lee and C H Taubes [29; 30; 31; 32; 33] and V Colin, P Ghiggini and K Honda [3; 4; 5]. Both teams built upon the earlier work of Taubes [49; 50; 51; 52; 53], which identified the Seiberg–Witten Floer homologies $\widehat{HM}_*(M)$ with the embedded contact homology $ECH_*(M)$ due to M Hutchings [20] and Hutchings and Taubes [21; 22]:

$$\Omega\colon ECH_*(M) \xrightarrow{\sim} \widehat{HM}_*(M).$$

Using the embedded contact homology as an intermediary, both groups completed the diagram

$$\begin{array}{ccc} & & HF^-_*(M) \\ & \nearrow & \downarrow \\ ECH_*(M) & & \\ & \searrow^{\Omega} & \\ & & \widehat{HM}_*(M) \end{array}$$

in a fashion which preserved essential properties of the three homology theories. In particular, the maps defined respect decompositions with respect to $\mathrm{Spin}^{\mathbb{C}}$ structures, carry invariants of contact structures to invariants of contact structures, preserve the long exact sequence (1-1) and support reductions to the simpler $U = 0$ theory:

$$\widehat{ECH}_*(M) \cong \widehat{HF}_*(M) \cong \widetilde{HM}_*(M). \tag{1-2}$$

Intuitively, each component in the equation above corresponds to a codimension-1 piece of a 4–dimensional topological field theory. It is evident that such a theory satisfies the following properties. In codimension 1, a topological field theory associates a chain complex $C(M)$ to each oriented 3–manifold M. The homology of this chain complex $H_*(C(M))$ is an invariant of the diffeomorphism type of the 3–manifold. In codimension 2, a topological field theory associates a differential graded category $\mathcal{C}(\Sigma)$ to each oriented surface Σ. The derived category $D(\mathcal{C}(\Sigma))$ of right $\mathcal{C}(\Sigma)$–modules (see Keller [25; 26]) is an invariant of the diffeomorphism type of the surface, and reversing the orientation of the surface produces the opposite dg category:

$$\mathcal{C}(\overline{\Sigma}) \cong \mathcal{C}(\Sigma)^{\mathrm{op}}.$$

For each 3–manifold X with boundary $\partial X = \Sigma$, there is a right $\mathcal{C}(\Sigma)$–module X_*. When a 3–manifold M is split along a surface $M = X \cup_\Sigma Y$, the invariant $C(M)$ corresponding to M is quasi-isomorphic to the tensor product,

$$C(M) \xrightarrow{\sim} X_* \otimes^{\mathbb{L}}_{\mathcal{C}(\Sigma)} (Y_*)^{\mathrm{op}},$$

of the modules associated to each piece. If the identifications made by (1-2) result from an equivalence between topological field theories, then the codimension-2 extensions of these topological field theories must be equivalent as well.

Question 1.1 *Is there an equivalence between codimension-*2 *extensions of Seiberg–Witten Floer, Heegaard–Floer and embedded contact homology*?

We study the simpler question of establishing a relationship between the categories associated to oriented surfaces Σ by Heegaard–Floer theory and contact topology.

The Heegaard–Floer homology $\widehat{HF}^*(M)$ was extended to surfaces and 3–manifolds with boundary, in the manner described above, by Ozsváth, R Lipshitz and D Thurston [36]. The theory was further developed by R Zarev [62; 63]. In particular, when an oriented surface Σ sports a handle decomposition, determined by combinatorial data $\mathcal{Z}$ called an *arc parametrization*, there is a dg category $\mathcal{A}(-\mathcal{Z})$ which is associated to the surface Σ. The Morita homotopy class of the corresponding categories of dg modules are independent of the handle decomposition $\mathcal{Z}$.

On the contact side, Honda has conjectured the existence of a family of triangulated categories $\mathcal{C}\mathrm{o}(\Sigma)$ associated to oriented surfaces Σ called *contact categories* [15]. These categories might function as part of a codimension-2 component of the embedded contact homology. The morphisms of contact categories are isotopy classes of tight contact structures on a thickened surface $\Sigma \times [0,1]$. Maps in $\mathcal{C}\mathrm{o}(\Sigma)$ are composed by gluing $\Sigma \times [0,1]$ to $\Sigma \times [0,1]$ and rescaling. The contact categories $\mathcal{C}\mathrm{o}(\Sigma)$ are conjectured to contain distinguished triangles associated to special contact structures called bypass moves. Unfortunately, this construction is not yet available in its full generality. For disks and annuli, algebraic analogues of these categories were introduced and studied by Y Tian [54; 55].

1.2 Summary of main results

We associate a $\mathbb{Z}/2$–linear dg category $\mathcal{K}\mathrm{o}(\Sigma)$ to each oriented surface Σ. This category satisfies a universal property which guarantees the existence of a unique map to a dg enhancement of any contact category $\mathcal{C}\mathrm{o}(\Sigma)$, when it exists.

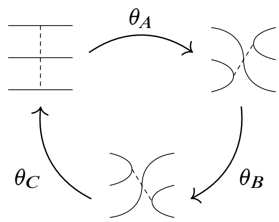
Universal property 1.2 *Suppose $\mathfrak{X}$ is a pretriangulated dg category for which there are choices of maps $\theta\colon \gamma \to \gamma'$ corresponding to bypass moves between dividing sets $\gamma, \gamma' \subset \Sigma$, and these maps satisfy four properties*:

(1) *Bypass moves are cycles.*

(2) *Trivial bypass moves are equal to the identity.*

(3) *Disjoint bypass moves commute.*

(4) *Associated to each bypass move is an exact triangle of the form*



Then there is a unique map $\mathcal{K}\mathrm{o}(\Sigma) \to \mathfrak{X}$ in the homotopy category of differential graded categories. See Section 3 for details.

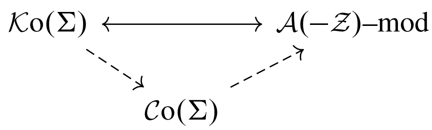
Section 2 contains algebraic background necessary to produce and study $\mathcal{K}\mathrm{o}(\Sigma)$. The definition of pretriangulated hull and a review of Drinfeld–Toën localization construction for dg categories is included. A variation of this localization construction is introduced and related to the standard localization.

Section 3 contains a discussion of surface topology needed for the main construction. The construction of the formal contact categories $\mathcal{K}\mathrm{o}(\Sigma)$ follows immediately by combining these topological considerations and the localization construction introduced in Section 2. The remainder of the paper is dedicated to the study of formal contact categories.

In Section 4, we check that the categories satisfy several elementary properties which were outlined by Honda. In particular, Corollary 4.10 shows that nontrivial boundary conditions are necessary for Giroux's tightness criterion to be satisfied. Theorem 4.14 shows that when such boundary conditions are present, the triangulated structure allows one to simplify the category by writing dividing sets which do not interact with the boundary in terms of those which do, up to homotopy equivalence. In Section 4.5, formal contact categories $\mathcal{K}\mathrm{o}(\Sigma)$ are split into a product of two isomorphic copies of a subcategory $\mathcal{K}\mathrm{o}_+(\Sigma)$, called the positive half of the formal contact category.

In Section 5, Theorem 5.2 shows that the mapping class group $\Gamma(\Sigma)$ of Σ acts naturally on the category $\mathcal{K}\mathrm{o}(\Sigma)$. Theorem 5.11 shows that when the surface Σ supports a handle decomposition, determined by an arc parametrization $\mathcal{Z}$, this produces a collection of generators $\mathfrak{Z}(\mathcal{Z})$ for the category $\mathcal{K}\mathrm{o}(\Sigma)$. After proving the second statement above, in Section 5.4 we study additive invariants of $\mathcal{K}\mathrm{o}_+(\Sigma_{g,1}, 2)$.

The remainder of the paper is dedicated to an investigation of the comparison problem between two codimension-2 extensions: contact categories and Heegaard–Floer categories. The strategy pursued is illustrated by the diagram



When a reasonable candidate for the geometric contact category $\mathcal{C}\mathrm{o}(\Sigma)$ exists, the dashed lines should be taken to be solid.

In Section 6, we study the relationship between three categories associated to the disk $(D^2, 2n)$ with $2n$ points fixed along its boundary. In [54], Tian constructed a candidate $\mathcal{Y}_n$ for $\mathcal{C}\mathrm{o}(D^2, 2n)$, and we introduce an arc parametrization $\mathcal{M}_n$ of the disk $(D^2, 2n)$ which gives a dg category $\mathcal{A}(-\mathcal{M}_n)$ associated to the Heegaard–Floer package; see Zarev [62]. The main result of this section is to prove that the three dg categories are Morita equivalent:

$$\mathcal{C}\mathrm{o}(D^2, 2n) \cong \mathcal{K}\mathrm{o}_+(D^2, 2n) \cong \mathcal{A}(-\mathcal{M}_n). \tag{1-3}$$

The category $\mathcal{A}(-\mathcal{M}_n)$ is a k–linear category because the differential d is always equal to zero. There are several other instances in which categories with this property can be associated to surfaces. In Section 7, we show that functors from these categories to the homotopy categories of the appropriate formal contact categories can be defined.

Section 8 applies the universal property, discussed above, in a much broader context. The section begins with a discussion of the relationship between the formal contact categories $\mathcal{K}\mathrm{o}(\Sigma)$ and the contact categories $\mathcal{C}\mathrm{o}(\Sigma)$. The main theorem leverages the universal property to construct a map

$$\mathcal{K}\mathrm{o}_+(\Sigma) \to \mathcal{A}(-\mathcal{Z})\text{–mod}$$

in the homotopy category of dg categories from the formal contact category associated to Σ to the Heegaard–Floer category associated to Σ, when Σ is parametrized by $\mathcal{Z}$, for any oriented surface Σ with sufficient boundary conditions.

Acknowledgments

The construction of contact categories was inspired by the ideas of K Honda, Y Tian and K Walker [15; 54; 55; 59]. I would especially like to thank Tian for his helpful correspondence and the Simons Center for facilitating our discussion, and also Y Huang, R Lipshitz, A Manion and I Petkova for several helpful emails, and my colleagues C Frohman, A Kaloti, K Kawamuro and R Kinser for their cordiality and their conversation.

In more detail, the author's involvement in this subject began because of the mention of a contact TQFT in Walker's 2006 notes [59, Section 9.4]. These were discussed at length with J Roberts between 2006 and 2009. Several years later, the author spoke to Walker about the possibility of fitting Heegaard–Floer theory into his framework in [59]. In 2014 the author found Tian's papers [54; 55] and a recording of Honda at MSRI discussing his ideas [15]. This project began after several conversations with Kaloti.

The author would like to thank the referees for their readings and very helpful feedback.

The author would also like to thank the organizers and participants of the *Categorifications of quantum groups*, *representations and knot invariants* session at the AMS–EMS–SPM meeting in June 2015, where some of these results were announced.

After this paper was posted, a few papers with complementary results have appeared; see [18; 39].

2 Algebraic constructions

In this section, a discussion of localizations follows a review of pretriangulated hulls. Section 2.2 reviews the standard localization procedure for dg categories. Section 2.3 introduces a form of localization which creates formal extensions among objects in a dg category: rather than creating homotopy equivalences amongst objects, this *Postnikov* localization introduces distinguished triangles. In Section 2.4, properties of Postnikov localizations are discussed.

Most of the materials in this section are standard. Some review is found in the appendix. A review of differential graded categories can be found in [26; 57] or [10, Section 1]; consult [46; 47; 56] for technical details. The language of model categories is reviewed in [37, Section A.2]; more details can be found in [19; 45].

2.1 Pretriangulated hull

This section contains a brief discussion of pretriangulated hulls of dg categories. The key ideas were introduced in [2, Section 4]; see also [1; 8].

Definition 2.1 [8, Section 2.4] If $\mathcal{C}$ is a dg category then there exists a dg category $\mathcal{C}^{\mathrm{pretr}}$, called the *pretriangulated hull* of $\mathcal{C}$. The objects of $\mathcal{C}^{\mathrm{pretr}}$ are one-sided twisted complexes, ie formal expressions

$$x = \Big(\bigoplus_{i=1}^{n} x_i[r_i], p\Big) \quad \text{such that } dp + p^2 = 0$$

where $n \geq 0$, $x_i \in \mathrm{Ob}(\mathcal{C}) \cup \{0\}$ and $r_i \in \mathbb{Z}$. The map $p = (p_{i,j})$ is a matrix such that $|p_{i,j}| = 1$ and

$$p_{i,j} = \begin{cases} x_i[r_i] \to x_j[r_j] & \text{if } j > i, \\ 0 & \text{if } j \leq i. \end{cases}$$

If $x, x' \in \mathrm{Ob}(\mathcal{C}^{\mathrm{pretr}})$ with $x = \big(\bigoplus_{i=1}^{n} x_i[r_i], p\big)$ and $x' = \big(\bigoplus_{i=1}^{n} x'_i[r'_i], p'\big)$, then $\mathrm{Hom}(x, x')$ consists of matrices $f = (f_{i,j})$ for $f_{i,j} \in \mathrm{Hom}^{r'_j - r_i}(x_i, x'_j)$, the composition is given by matrix multiplication and the differential $d : \mathrm{Hom}(x, x') \to \mathrm{Hom}(x, x')$ is determined by the formula

$$(df)_{i,j} = (df)_{i,j} + (p'f)_{i,j} - (-1)^{|f_{i,j}|}(fp)_{i,j}.$$

Remark 2.2 [8, Section 2.4] If $x, y \in \mathrm{Ob}(\mathcal{C})$ and $f : x \to y$ is a closed map of degree zero, then the cone of f exists in $\mathcal{C}^{\mathrm{pretr}}$ by construction: $C(f) = (x \oplus y[-1], p) \in \mathrm{Ob}(\mathcal{C})$ where $p_{1,2} = f$ and $p_{1,1} = p_{2,1} = p_{2,2} = 0$. The objects in $\mathcal{C}^{\mathrm{pretr}}$ can be obtained by iterated applications of the cone construction.

A referee notes that the construction in Remark 2.2 is sometimes called a "cocone."

By construction, the pretriangulated dg category $\mathcal{C}^{\mathrm{pretr}}$ associated to a k–linear category $\mathcal{C}$ factors through its additive closure $\mathrm{Mat}(\mathcal{C})$:

$$\mathrm{Mat}(\mathcal{C})^{\mathrm{pretr}} \cong \mathcal{C}^{\mathrm{pretr}}.$$

(Or set $p=0$ in Definition 2.1.) The canonical inclusion $\mathcal{C}\hookrightarrow\mathcal{C}^{\mathrm{pretr}}$ is fully faithful. A dg category $\mathcal{C}$ is *pretriangulated* when the functor $\mathrm{Ho}(\mathcal{C})\to\mathrm{Ho}(\mathcal{C}^{\mathrm{pretr}})$ induced by inclusion between the associated homotopy categories is an equivalence of categories. The *category of pretriangulated dg categories* will be denoted by $\mathrm{dgcat}_k^{\mathrm{pretr}}$.

Unfamiliar readers may wish to recall that $\mathrm{Ob}(\mathcal{C}\amalg\mathcal{D}) := \mathrm{Ob}(\mathcal{C})\sqcup\mathrm{Ob}(\mathcal{D})$ and

$$\mathrm{Hom}_{\mathcal{C}\amalg\mathcal{D}}(x,y) := \begin{cases}\mathrm{Hom}_{\mathcal{C}}(x,y) & \text{if } x,y\in\mathrm{Ob}(\mathcal{C}),\\ \mathrm{Hom}_{\mathcal{D}}(x,y) & \text{if } x,y\in\mathrm{Ob}(\mathcal{D}),\\ 0 & \text{otherwise.}\end{cases}$$

The proposition below shows how the pretriangulated hull operation distributes over coproducts of dg categories. This is a $p\neq 0$ generalization of the analogous statement about additive closures. It will be used in Theorem 4.4.

Proposition 2.3 *If $\mathcal{C}$ and $\mathcal{D}$ are k–linear then* $(\mathcal{C}\amalg\mathcal{D})^{\mathrm{pretr}}\cong\mathcal{C}^{\mathrm{pretr}}\amalg\mathcal{D}^{\mathrm{pretr}}$.

Proof Since there are no nonzero maps between $\mathcal{C}$ and $\mathcal{D}$, thought of as subcategories of $\mathcal{C}\amalg\mathcal{D}$, a twisted complex $\left(\bigoplus_{i=1}^n x_i[r_i],p\right)\in(\mathcal{C}\amalg\mathcal{D})^{\mathrm{pretr}}$ splits into a direct sum of twisted complexes in $\mathcal{C}^{\mathrm{pretr}}$ and $\mathcal{D}^{\mathrm{pretr}}$. Likewise, matrices $(f_{i,j})$ of maps between twisted complexes in $(\mathcal{C}\amalg\mathcal{D})^{\mathrm{pretr}}$ consist of blocks. It follows that there are functors $\pi_{\mathcal{C}}\colon(\mathcal{C}\amalg\mathcal{D})^{\mathrm{pretr}}\to\mathcal{C}^{\mathrm{pretr}}$ and $\pi_{\mathcal{D}}\colon(\mathcal{C}\amalg\mathcal{D})^{\mathrm{pretr}}\to\mathcal{D}^{\mathrm{pretr}}$ which satisfy the universal property of the product. □

The following proposition is well known; see [1, Section 1.5].

Many of the constructions to follow in this section use ideas which are touched on in the appendix.

Proposition 2.4 *The pretriangulated hull* $-^{\mathrm{pretr}}\colon\mathrm{dgcat}_k\to\mathrm{dgcat}_k^{\mathrm{pretr}}$ *is left adjoint to the forgetful functor*:

$$\mathrm{Hom}_{\mathrm{dgcat}_k^{\mathrm{pretr}}}(\mathcal{C}^{\mathrm{pretr}},\mathcal{D})\cong\mathrm{Hom}_{\mathrm{dgcat}_k}(\mathcal{C},\mathrm{Forget}(\mathcal{D})).$$

If $f:\mathcal{C}\xrightarrow{\sim}\mathcal{D}$ is a quasiequivalence then $f^{\mathrm{pretr}}\colon\mathcal{C}^{\mathrm{pretr}}\to\mathcal{D}^{\mathrm{pretr}}$ is a quasiequivalence of dg categories.

The category Hqe is a localization of dgcat_k in which quasiequivalences between dg categories are isomorphisms. The Morita homotopy category Hmo is a localization of the homotopy category Hqe of dg categories in which derived equivalences are isomorphisms. In Hmo, the homotopy idempotent completion $\mathcal{C}^{\mathrm{perf}}$ of the pretriangulated hull $\mathcal{C}^{\mathrm{pretr}}$ is fibrant replacement; see [46].

2.2 Inverting maps in dg categories

This section contains a brief review of the localization construction for dg categories. Many authors have studied this problem; see [8; 25; 26; 48; 56, Section 8.2].

Definition 2.5 The symbol I will be used to denote the dg category freely generated by a cycle $f:1\to 2$ of degree 0, and I' will be used to denote the dg category freely generated by cycles $f:1\to 2$ and $g:2\to 1$ of degree 0:

$$I = 1\xrightarrow{f}2 \quad\text{and}\quad I' = 1\rightleftarrows 2.$$

The symbol $\bar{I}$ denotes the dg category with a unique degree 0 isomorphism $f : 1 \xrightarrow{\sim} 2$ with $df = 0$. There are canonical inclusions

$$\kappa : I \hookrightarrow \bar{I} \quad \text{and} \quad \kappa' : I' \hookrightarrow \bar{I}.$$

These maps are determined by the assignments $\kappa(f) = f$, $\kappa'(f) = f$ and $\kappa'(g) = f^{-1}$.

Definition 2.6 Suppose that $\mathcal{C}$ is a dg category and $R : \coprod_{r \in \mathcal{R}} I \to \mathcal{C}$ is a dg functor. Then the *localization of* $\mathcal{C}$ *with respect to* R is a dg functor

$$P : \mathcal{C} \to L_R\mathcal{C}$$

which satisfies:

(1) The pullback map $P^* : \mathrm{Hom}_{\mathrm{Hqe}}(L_R\mathcal{C}, \mathcal{X}) \to \mathrm{Hom}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X})$ is injective.

(2) The image of P^* consists of maps $f : \mathcal{C} \to \mathcal{X}$ for which there is a map α making the diagram below commute:

$$\begin{array}{ccc} \coprod_{r \in \mathcal{R}} \mathrm{Ho}(I) & \xrightarrow{\mathrm{Ho}(R^* f)} & \mathrm{Ho}(\mathcal{X}) \\ {\scriptstyle \mathrm{Ho}(\kappa)}\downarrow & \overset{\alpha}{\nearrow} & \\ \coprod_{r \in \mathcal{R}} \mathrm{Ho}(\bar{I}) & & \end{array}$$

The image $\mathrm{im}(P^*)$ may be denoted by $\mathrm{Hom}^{I}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X})$.

Corollary 8.8 in [56] shows that for any dg category $\mathcal{C}$ and any functor $R : \coprod_{r \in \mathcal{R}} I \to \mathcal{C}$, there exists a functor $P : \mathcal{C} \to L_R\mathcal{C}$ in the homotopy category Hqe of dg categories which satisfies the two properties in Definition 2.6. The functor $P : \mathcal{C} \to L_R\mathcal{C}$ is defined to be the homotopy pushout

$$\begin{array}{ccc} \coprod_{r \in \mathcal{R}} I & \xrightarrow{R} & \mathcal{C} \\ {\scriptstyle \kappa}\downarrow & & \downarrow{\scriptstyle P} \\ \coprod_{r \in \mathcal{R}} \bar{I} & \longrightarrow & L_R\mathcal{C} \end{array}$$

When the category $\mathcal{C}$ is cofibrant, this homotopy pushout

$$L_R\mathcal{C} = \Big(\coprod_{r \in \mathcal{R}} \bar{I} \Big) \amalg^{\mathbb{L}}_{R} \mathcal{C}$$

can be computed by replacing the inclusion $\kappa : I \hookrightarrow \bar{I}$ by a well-known cofibration $I \hookrightarrow \tilde{I}$. The dg category $\tilde{I}$ appears in Drinfeld, where it is denoted by $\mathcal{K}$ [8, Section 3.7.1].

Definition 2.7 The category $\tilde{I}$ has two objects: 1 and 2. Its maps are generated by the elements $f \in \mathrm{Hom}^0_{\tilde{I}}(1, 2)$, $g \in \mathrm{Hom}^0_{\tilde{I}}(2, 1)$, $h_{1,1} \in \mathrm{Hom}^{-1}_{\tilde{I}}(1, 1)$, $h_{2,2} \in \mathrm{Hom}^{-1}_{\tilde{I}}(2, 2)$ and $h_{1,2} \in \mathrm{Hom}^{-2}_{\tilde{I}}(1, 2)$:

$$h_{1,1} \circlearrowright 1 \underset{g}{\overset{f, h_{1,2}}{\rightleftarrows}} 2 \circlearrowleft h_{2,2}$$

The differential is determined by the Leibniz rule together with the equations

$$df = 0, \quad dg = 0, \quad dh_{1,1} = gf - 1_1, \quad dh_{2,2} = fg - 1_2 \quad \text{and} \quad dh_{1,2} = h_{2,2}f - fh_{1,1},$$

and the maps are subject to no relations.

Remark 2.8 In Definition 2.6, the category I and the map $\kappa\colon I \hookrightarrow \bar{I}$ can be replaced by the category I' and the map $\kappa'\colon I' \hookrightarrow \bar{I}$. Suppose that $R\colon I \to \mathcal{C}$ and a candidate $R(f)^{-1}$ for the inverse of the map $R(f)$ already exists in the category $\mathcal{C}$. Then R can be extended to a functor $R'\colon I' \to \mathcal{C}$ such that $R'(f) = R(f)$ and $R'(g) = R(f)^{-1}$ and there is an analogous localization

$$P : \mathcal{C} \to L_{R'}\mathcal{C} \quad \text{where } L_{R'}\mathcal{C} = \tilde{I} \amalg^{\mathbb{L}}_{R'} \mathcal{C}.$$

2.3 Postnikov localization

A variation of the localization procedure discussed in the previous section is introduced. This *Postnikov localization* introduces distinguished triangles rather than homotopy equivalences. In particular, given a sequence

$$1 \to 2 \to 3 \to 1$$

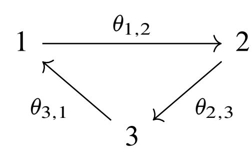
of maps S in a dg category $\mathcal{C}$, there is a dg category $L_S\mathcal{C}$ in which this sequence forms a distinguished triangle.

The dg categories considered in this section are $\mathbb{Z}/2$–graded for simplicity. The equivalences discussed below commute with the forgetful functor to the ungraded setting introduced in Section 2.5. On the other hand, $\mathbb{Z}$–graded lifts determined by grading conventions for distinguished triangles can be found in [10, Section 2.4.1]. See also [57, Section 4.3].

Historically, Postnikov systems appear in the study of triangulated categories [12]. The name Postnikov may be attached to that construction because it is a generalization of the Postnikov decomposition of topological spaces to algebraic triangulated categories.

First we introduce a dg category D' which corepresents triangles; see (2-1) and Proposition 2.11. Then Definition 2.13 introduces dg categories $\overline{D}$ and $\widetilde{D}$ which corepresent distinguished triangles. A dg functor $\kappa\colon D' \hookrightarrow \widetilde{D}$ will be used to construct the Postnikov localization in Definition 2.15.

Definition 2.9 The symbol D' will be used to denote the dg category freely generated by cycles $\theta_{1,2}\colon 1 \to 2$, $\theta_{2,3}\colon 2 \to 3$ and $\theta_{3,1}\colon 3 \to 1$:



The degrees are determined by $|\theta_{1,2}| = 1$, $|\theta_{2,3}| = 1$ and $|\theta_{3,1}| = 1$.

Since a dg functor $f: D' \to \mathcal{C}$ is uniquely determined by where it maps the generators in the definition above, there is a bijection between the set of such functors and (symmetric) triangles in $\mathcal{C}$:

$$\mathrm{Hom}_{\mathrm{dgcat}_k}(D', \mathcal{C}) \xrightarrow{\sim} \{\text{symmetric triangles in } \mathcal{C}\}. \tag{2-1}$$

Definition 2.10 If $f, g: D' \to \mathcal{C}$ are two triangles in $\mathcal{C}$, then f *is isomorphic to* g when $\mathrm{Ho}(f) \cong \mathrm{Ho}(g)$ as objects in the functor category $\mathrm{Hom}(\mathrm{Ho}(D'), \mathrm{Ho}(\mathcal{C}))$.

The proposition below states that in the homotopy category Hqe of dg categories, the left-hand side of (2-1) is in canonical bijection with isomorphism classes of triangles.

Proposition 2.11 [10, Proposition 2.4.7] *For any dg category $\mathcal{C}$, there is a one-to-one correspondence between homotopy classes of functors $f: D' \to \mathcal{C}$ and isomorphism classes of triangles in $\mathcal{C}$:*

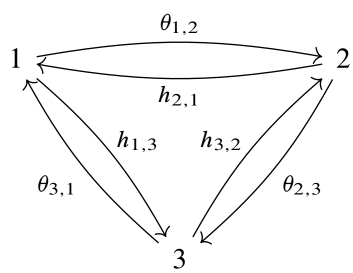
$$\mathrm{Hom}_{\mathrm{Hqe}}(D', \mathcal{C}) \leftrightarrow \{\textit{symmetric triangles in } \mathcal{C}\}/\mathrm{iso}.$$

Just as isomorphisms are distinguished types of maps, distinguished triangles are distinguished types of triangles. A distinguished triangle is a recipe for constructing one of its objects in terms of the other two.

Definition 2.12 If S is a symmetric triangle $1 \xrightarrow{\theta_{1,2}} 2 \xrightarrow{\theta_{2,3}} 3 \xrightarrow{\theta_{3,1}} 1$ in a dg category $\mathcal{C}$, then S is a *distinguished triangle* if and only if S is isomorphic to the distinguished triangle S' given by $1 \xrightarrow{\theta_{1,2}} 2 \to C(\theta_{1,2}) \to 1$ in the homotopy category of $\mathcal{C}^{\mathrm{pretr}}$.

In keeping with Section 2.2, the distinguished property of triangles is formulated as a lifting problem. An innocuous-looking dg category $\overline{D}$ which corepresents distinguished triangles and a quasiequivalent cofibrant replacement $\widetilde{D} \xrightarrow{\sim} \overline{D}$ are introduced below.

Definition 2.13 [10, Section 2.4.1] The dg category $\overline{D}$ consists of objects $\mathrm{Ob}(\overline{D}) = \{1, 2, 3\}$. The maps are generated by cycles: $\theta_{1,2}: 1 \to 2$, $\theta_{2,3}: 2 \to 3$ and $\theta_{3,1}: 3 \to 1$ of degree 1, and homotopies $h_{2,1}: 2 \to 1$, $h_{3,2}: 3 \to 2$ and $h_{1,3}: 1 \to 3$ of degree 1,



with $dh_{2,1} = \theta_{3,1}\theta_{2,3}$, $dh_{3,2} = \theta_{1,2}\theta_{3,1}$ and $dh_{1,3} = \theta_{2,3}\theta_{1,2}$ and the relations

$$\theta_{2,3}h_{3,2} + h_{1,3}\theta_{3,1} = 1_3, \quad \theta_{1,2}h_{2,1} + h_{3,2}\theta_{2,3} = 1_2 \quad \text{and} \quad \theta_{3,1}h_{1,3} + h_{2,1}\theta_{1,2} = 1_1.$$

The dg category $\widetilde{D}$ consists of objects $\mathrm{Ob}(\widetilde{D}) = \{1, 2, 3\}$. The maps $\theta_{i,j}\colon i \to j$ in this category are clockwise-oriented paths between vertices, from i to j, in the triangular graph featured in Definition 2.9. The differential is zero on paths of length zero or one; when $\theta_{i,i}$ is a cycle, a path of topological degree one (a loop),

$$d\theta_{i,i} = 1_i - \sum_k \theta_{k,i}\theta_{i,k},$$

otherwise $d\theta_{i,j}$ is the sum over compositions of all possible factorizations of the path:

$$d\theta_{i,j} = \sum_k \theta_{k,i}\theta_{j,k}.$$

The projection $p\colon \widetilde{D} \to \overline{D}$ given by mapping cycles of length 1 to their respective θ–maps is a quasiequivalence [10, Proposition 2.4.13]. In the other direction, there is an inclusion $\kappa'\colon D' \hookrightarrow \widetilde{D}$ given by sending the θ–maps to their respective length-1 cycles. There is also an inclusion $\kappa'\colon D' \hookrightarrow \overline{D}$ given by the same formula. A $\mathbb{Z}$–graded analogue of $\widetilde{D}$ is discussed in [27]. This dg category is the cobar–bar construction on the partially wrapped Fukaya category of the disk with three stops [41].

The proposition below states that the dg category $\widetilde{D}$ corepresents distinguished triangles and satisfies the key properties necessary for the localization construction.

Proposition 2.14 [10, Proposition 2.4.14] (1) *For any dg category* $\mathcal{C}$, *the set of homotopy classes of dg functors from* $\widetilde{D}$ *to* $\mathcal{C}$ *is in bijection with the set of isomorphism classes of distinguished triangles in* $\mathcal{C}$:

$$\mathrm{Hom}_{\mathrm{Hqe}}(\widetilde{D}, \mathcal{C}) = \{1 \xrightarrow{\theta_{1,2}} 2 \xrightarrow{\theta_{2,3}} C(\theta_{1,2}) \to 1\}/\mathrm{iso}.$$

(2) *The image of the pullback induced by the map* κ' *appearing in Definition 2.13 coincides with the subset of triangles which are distinguished*:

$$(\kappa')^*\colon \mathrm{Hom}_{\mathrm{Hqe}}(\widetilde{D}, \mathcal{C}) \to \mathrm{Hom}_{\mathrm{Hqe}}(D', \mathcal{C}).$$

(3) *The set* $\mathrm{Hom}_{\mathrm{Hqe}}(\widetilde{D}, \mathcal{C})$ *is equal to the set of maps* $f \in \mathrm{Hom}_{\mathrm{Hqe}}(D', \mathcal{C})$ *for which there is a map* $\alpha\colon \mathrm{Ho}(\widetilde{D}) \to \mathrm{Ho}(\mathcal{C})$ *such that* $\mathrm{Ho}(f) = \alpha \circ \mathrm{Ho}(\kappa')$.

We are now ready to discuss a generalization of the localization procedure presented earlier in Section 2.2. Instead of inverting maps in the associated homotopy category, this new operation creates distinguished triangles in the associated homotopy category.

Definition 2.15 Suppose that $\mathcal{C}$ is a dg category and $S\colon \coprod_{s\in\mathcal{S}} D' \to \mathcal{C}$ is a dg functor. Then the *Postnikov localization of* $\mathcal{C}$ *with respect to* S is a dg functor

$$Q\colon \mathcal{C} \to L_S\mathcal{C}$$

such that for any dg category $\mathcal{X}$ the following properties are satisfied:

(1) The pullback map $Q^*\colon \mathrm{Hom}_{\mathrm{Hqe}}(L_S\mathcal{C}, \mathcal{X}) \to \mathrm{Hom}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X})$ is injective.

(2) The set of maps $\mathrm{Hom}_{\mathrm{Hqe}}(L_S\mathcal{C}, \mathcal{X})$ in the image of Q^* is equal to the set of maps $f \in \mathrm{Hom}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X})$ such that there is a map α making the diagram below commute:

$$\begin{array}{ccc} \coprod_{s\in\mathcal{S}} \mathrm{Ho}(D') & \xrightarrow{\mathrm{Ho}(f\circ S)} & \mathrm{Ho}(\mathcal{X}) \\ {\scriptstyle\kappa'}\downarrow & \overset{\alpha}{\nearrow} & \\ \coprod_{s\in\mathcal{S}} \mathrm{Ho}(\widetilde{D}) & & \end{array}$$

The image $\mathrm{im}(Q^*)$ may also be denoted by $\mathrm{Hom}^T_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X})$.

Recall from above that a functor from $S: D' \to \mathcal{C}$ is determined by the choice of cycles $f: 1 \to 2$, $g: 2 \to 3$ and $h: 3 \to 1$. The Postnikov localization $L_S\mathcal{C}$ associated to the functor S requires that the sequence

$$1 \xrightarrow{f} 2 \xrightarrow{g} 3 \xrightarrow{h} 1$$

is a distinguished triangle in the sense of Definition 2.12. The category $L_S\mathcal{C}$ is uniquely determined up to homotopy by the property that a functor $f: \mathcal{C} \to \mathcal{X}$ factors through $Q: \mathcal{C} \to L_S\mathcal{C}$ in Hqe when it maps triangles in the image of S to distinguished triangles in the homotopy category $\mathrm{Ho}(\mathcal{X})$ of $\mathcal{X}$.

When $\mathcal{C}$ is a cofibrant dg category, the category $L_S\mathcal{C}$ is a pushout, obtained by gluing a copy of $\widetilde{D}$ along the subcategory determined by the image of a functor S. If $\mathcal{C}$ is not cofibrant then $L_S\mathcal{C}$ is a homotopy pushout: the pushout of a cofibrant replacement $\widetilde{\mathcal{C}} \xrightarrow{\sim} \mathcal{C}$ of $\mathcal{C}$ [37, Section A.2.4.4].

The next proposition states that Postnikov localizations always exist.

Proposition 2.16 *For any dg category $\mathcal{C}$ and any collection $S: \coprod_{s\in\mathcal{S}} D' \to \mathcal{C}$, there is a Postnikov localization $Q: \mathcal{C} \to L_S\mathcal{C}$ in* Hqe.

Proof It follows from Proposition 2.14 that the functor $\kappa': D' \to \widetilde{D}$ is a Postnikov localization in the sense of Definition 2.15. Therefore, any coproduct of inclusions $\coprod_{s\in\mathcal{S}} D' \to \coprod_{s\in\mathcal{S}} \widetilde{D}$ is a Postnikov localization. For any dg category $\mathcal{C}$, the localization $Q: \mathcal{C} \to L_S\mathcal{C}$ is given by the homotopy pushout:

$$\begin{array}{ccc} \coprod_{s\in\mathcal{S}} D' & \xrightarrow{S} & \mathcal{C} \\ \downarrow & & \downarrow{\scriptstyle Q} \\ \coprod_{s\in\mathcal{S}} \widetilde{D} & \longrightarrow & L_S\mathcal{C} \end{array}$$

That $L_S\mathcal{C}$ is a Postnikov localization follows Definition 2.15 and properties of homotopy pushouts [19]. □

2.4 Properties of Postnikov localization

In this section, we explore properties of the Postnikov localization procedure, establish a relationship between it and the ordinary localization of dg categories, and introduce an analogue of Heller's lemma which facilitates the computation of additive invariants such as the Grothendieck group.

Triangle insertion The appendix reviews relevant concepts such quasifully faithful embedding.

The proposition below assures us that, after having added a triangle, it persists in the pretriangulated hull.

Proposition 2.17 *Suppose that $S\colon D' \to \mathcal{C}$, $Q\colon \mathcal{C} \to L_S\mathcal{C}$ and $R\colon L_S\mathcal{C} \to \mathcal{X}$ is a quasifully faithful embedding of the Postnikov localization of $\mathcal{C}$ into a pretriangulated category $\mathcal{X}$. If $f = RQS(1 \to 2)$ and $c = RQS(3)$ then c is isomorphic to the cone $C(f)$ of f in the homotopy category of $\mathcal{X}$:*

$$c \cong C(f) \quad \textit{in } \mathrm{Ho}(\mathcal{X}).$$

Proof For the sake of notation, everything to follow takes place inside of the category $\mathrm{Ho}(\mathcal{X})$. By the triangulated category axiom TR3, there is a map h in $\mathcal{X}$ which yields a map $(1, 1, h, 1)$ from the triangle $S(1) \to S(2) \to S(3) \to S(1)$ to the triangle $S(1) \to S(2) \to C(f) \to S(1)$. For all $x \in \mathcal{X}$, both triangles determine long exact sequences after applying the functor $\mathrm{Hom}(x, -)$. By the five lemma, $h_*\colon \mathrm{Hom}^*(x, c) \to \mathrm{Hom}^*(x, C(f))$ is an isomorphism. Therefore Yoneda's lemma implies the result. □

Decategorification of localizations For references concerning short exact sequences of dg categories see [26, Section 4.6].

Lemma 2.18 *Suppose $S\colon D' \to \mathcal{C}$ is a triangle, $\theta_{1,2} = S(1 \to 2)$ and $c = S(3)$ in a dg category $\mathcal{C}$. Then S is isomorphic to a distinguished triangle if and only if the double cone complex $K = C(C(\theta_{1,2}) \xrightarrow{\tilde{\theta}_{2,3}} c)$ is contractible, where $\tilde{\theta}_{2,3}$ is the extension of the map $\theta_{2,3}\colon S(2) \to c$ to the cone $C(\theta_{1,2})$.*

Proof If S is distinguished, then the triangle $S(1) \to S(2) \to S(3) \to S(1)$ is isomorphic to $1 \to 2 \to C(\theta_{1,2}) \to 1$ in the homotopy category via the map $(1, 1, \tilde{\theta}_{2,3}, 1[1])$, so $\tilde{\theta}_{2,3}$ is a homotopy equivalence and $C(\tilde{\theta}_{2,3})$ is contractible. Conversely, $C(\tilde{\theta}_{2,3}) \simeq 0$ implies $\tilde{\theta}_{2,3}$ is a homotopy equivalence and the map above determines an equivalence of triangles. □

Recall that if $a \in \mathrm{Ob}(\mathcal{C})$, then Drinfeld's dg quotient $\mathcal{C}/\langle a\rangle$ can be formed by adding a homotopy h which satisfies $dh = 1_a$ to a cofibrant replacement of $\mathcal{C}$; see [8]. This makes the object contractible in the homotopy category of the Drinfeld quotient. (This can be reformulated as a homotopy pushout [48, Theorem 4.0.1].)

The proposition below constructs a short exact sequence of dg categories by relating the Postnikov localization $L_S\mathcal{C}$ of a dg category $\mathcal{C}$ to a Drinfeld quotient $\mathcal{C}/\langle K\rangle$. The subcategory $\langle K\rangle$ is generated by the object K in Lemma 2.18.

Proposition 2.19 *Suppose that $S\colon D' \to \mathcal{C}$ is a triangle, $f = S(1 \to 2)$ and $c = S(3)$ in a dg category $\mathcal{C}$. Then there is a short exact sequence of dg categories*

$$\langle K\rangle \to \mathcal{C} \to L_S(\mathcal{C})$$

in the Morita homotopy category Hmo, *where $\langle K\rangle$ is the dg category determined by the cone $K = C(C(f) \to c)$ of the natural map from the cone on f to c in $\mathcal{C}^{\mathrm{pretr}}$.*

Proof First assume that K is represented by an object in $\mathcal{C}$. By Definition 2.15, the Postnikov localization $L_S\mathcal{C}$ satisfies the universal property

$$\mathrm{Hom}_{\mathrm{Hqe}}(L_S\mathcal{C}, \mathcal{X}) \xrightarrow{\sim} \mathrm{Hom}^{T}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X}), \tag{2-2}$$

the set of homotopy classes of functors from $L_S\mathcal{C}$ to any dg category $\mathcal{X}$ is in bijection with the set of homotopy classes of functors $f: \mathcal{C} \to \mathcal{X}$ which map $\mathrm{im}(S)$ to distinguished triangles in the homotopy categories: $\mathrm{Ho}(f): \mathrm{Ho}(\mathcal{C}) \to \mathrm{Ho}(\mathcal{X})$. By the lemma above, the condition that $\mathrm{Ho}(fS): D' \to \mathrm{Ho}(\mathcal{X})$ maps to a distinguished triangle is equivalent to the condition that a certain double cone complex K is contractible. If $\tilde{\theta}_{2,3}: C(\theta_{1,2}) \to 3$ is given by $\tilde{\theta}_{2,3} = (0, \theta_{2,3})$, then set $K = C(\tilde{\theta}_{2,3})$ so that

$$K = C(\tilde{\theta}_{2,3}) = (1[2] \oplus 2[1] \oplus 3, d_K) \quad \text{where } d_K = \begin{pmatrix} d_1 & \theta_{1,2} & \\ & -d_2 & \theta_{2,3} \\ & & d_3 \end{pmatrix}$$

is contractible in $\mathcal{X}$. So there is a bijection of sets

$$\mathrm{Hom}^{T}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X}) \xrightarrow{\sim} \mathrm{Hom}^{\langle K\rangle}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X}), \tag{2-3}$$

where $\mathrm{Hom}^{\langle K\rangle}(\mathcal{C}, \mathcal{X})$ is the set of maps $f: \mathcal{C} \to \mathcal{X}$ which send K to a contractible object in $\mathcal{X}$. Then

$$\mathrm{Hom}_{\mathrm{Hqe}}(\mathcal{C}/\langle K\rangle, \mathcal{X}) \xrightarrow{\sim} \mathrm{Hom}^{\langle K\rangle}_{\mathrm{Hqe}}(\mathcal{C}, \mathcal{X}). \tag{2-4}$$

See [48, Theorem 4.0.1]. The maps in (2-2), (2-3) and (2-4) combine to show that the Postnikov localization satisfies the same universal property as the Drinfeld quotient. Therefore, $\mathcal{C}/\langle K\rangle$ and $L_S\mathcal{C}$ are isomorphic in Hqe. Associated to any such Drinfeld quotient, there is a short exact sequence

$$\langle K\rangle \hookrightarrow \mathcal{C} \to \mathcal{C}/\langle K\rangle$$

in the Morita homotopy category Hmo [48, Remark 4.0.2]. Since Hmo is a quotient of Hqe, the isomorphism $\mathcal{C}/\langle K\rangle \cong L_S\mathcal{C}$ in Hqe implies the isomorphism $\mathcal{C}/\langle K\rangle \cong L_S\mathcal{C}$ in Hmo, and there is a short exact sequence of dg categories

$$\langle K\rangle \hookrightarrow \mathcal{C} \to L_S\mathcal{C}.$$

Now suppose that K is *not* representable by an object in $\mathcal{C}$. In the Morita homotopy category Hmo, the fibrant replacement $\mathcal{C}^{\mathrm{perf}}$ of $\mathcal{C}$ is the category of perfect modules over $\mathcal{C}$ — an idempotent completion of the pretriangulated hull. The object K is representable in $\mathcal{C}^{\mathrm{perf}}$ (see Remark 2.2), and so, by the argument above, there is a short exact sequence

$$\langle K\rangle \to \mathcal{C}^{\mathrm{perf}} \to L_S(\mathcal{C}^{\mathrm{perf}}).$$

In the homotopy category of any model category, every object $\mathcal{C}$ is isomorphic to its fibrant replacement $\beta: \mathcal{C} \xrightarrow{\sim} \mathcal{C}^{\mathrm{perf}}$. Since cofibrations in Hmo and Hqe are identical, a homotopy pushout in Hqe is a homotopy pushout in Hmo. The map β determines an equivalence of pushout diagrams from $\widetilde{D} \leftarrow \coprod_s D' \to \mathcal{C}$ to $\widetilde{D} \leftarrow \coprod_s D' \to \mathcal{C}^{\mathrm{perf}}$, from which it follows that the map $L_S\beta: L_S\mathcal{C} \to L_S(\mathcal{C}^{\mathrm{perf}})$ is an isomorphism in Hmo.

There is a commuting diagram extending the right-hand side of the short exact sequence in which both of the vertical maps are isomorphisms in Hmo:

$$\begin{array}{ccc} \mathcal{C} & \longrightarrow & L_S\mathcal{C} \\ {\scriptstyle\beta}\big\downarrow & & \big\downarrow{\scriptstyle L_S\beta} \\ \mathcal{C}^{\mathrm{perf}} & \longrightarrow & L_S(\mathcal{C}^{\mathrm{perf}}) \end{array}$$

So there is a short exact sequence: $E \to \mathcal{C} \to L_S\mathcal{C}$ where E is a dg category Morita equivalent to $\langle K\rangle$. □

A short exact sequence of dg categories in Hmo induces a long exact sequence among additive invariants of dg categories [26; 46]. The corollary below is the first part of the long exact sequence associated to Hochschild homology.

Corollary 2.20 *Suppose that S, $\langle K\rangle$ and $\mathcal{C}$ are as in the proposition above. Then there is an exact sequence of abelian groups*

$$HH_0(\langle K\rangle) \to HH_0(\mathcal{C}) \to HH_0(L_S(\mathcal{C})) \to 0$$

A Postnikov localization as a module In this section, we explain how Postnikov localizations inherit the structure of a module category over $\mathrm{End}(\widetilde{D})$ in Hmo.

If $\mathcal{C} \cong L_S\mathcal{X}$ is a Postnikov localization of a dg category $\mathcal{X}$, then the map $\iota\colon \coprod_{s\in\mathcal{S}} \widetilde{D} \to \mathcal{C}$ from the proof of Proposition 2.16 yields a map $\iota^{\mathrm{pretr}}\colon (\coprod_{s\in\mathcal{S}} \widetilde{D})^{\mathrm{pretr}} \to \mathcal{C}^{\mathrm{pretr}}$. Therefore, by Proposition 2.3 there is a map $\hat{\iota}^{\mathrm{pretr}}\colon \prod_{s\in\mathcal{S}} \widetilde{D}^{\mathrm{pretr}} \to \mathcal{C}^{\mathrm{pretr}}$. The pullback of the map $\hat{\iota}^{\mathrm{pretr}}$ along the diagonal map $\Delta_{\mathcal{S}}\colon \widetilde{D}^{\mathrm{pretr}} \to \prod_{s\in\mathcal{S}} \widetilde{D}^{\mathrm{pretr}}$ is a functor $j : \widetilde{D}^{\mathrm{pretr}} \to \mathcal{C}^{\mathrm{pretr}}$. The map j determines an action of $\mathrm{End}(\widetilde{D}^{\mathrm{pretr}})$ on $\mathcal{C}^{\mathrm{pretr}}$:

$$\begin{array}{ccc} \widetilde{D}^{\mathrm{pretr}} & \xrightarrow{\;j\;} & \mathcal{C}^{\mathrm{pretr}} \\ {\scriptstyle g}\big\downarrow & & \big\downarrow{\scriptstyle \bar{g}} \\ \widetilde{D}^{\mathrm{pretr}} & \xrightarrow{\;j\;} & \mathcal{C}^{\mathrm{pretr}} \end{array}$$

The universal property in Definition 2.15 gives us a lift $\bar{g}$ of $j \circ g$ for each $g \in \mathrm{End}(\widetilde{D}^{\mathrm{pretr}})$, and uniqueness of lifts implies that lifts commute with compositions.

2.5 Ungraded dg categories

The main body of the paper will use the trivial grading; a more sophisticated G–grading will be introduced at a later time [6]. Here we require k to be a field of characteristic 2.

There is a category $\mathrm{Kom}_k^{\mathrm{un}}$ of ungraded chain complexes. In more detail, An *ungraded chain complex* is a k–vector space C and a differential $d_C : C \to C$ which satisfies $d_C^2 = 0$. A map $f : C \to D$ of ungraded chain complexes is a map of vector spaces. If $\mathrm{Hom}(C, D)$ denotes the vector space of such maps from C to D, then there is an associative composition and for each C there is an identity map $1_C : C \to C$. This determines the category $\mathrm{Kom}_k^{\mathrm{un}}$.

The monoidal structure in $\mathrm{Kom}_k^{\mathrm{un}}$ is the tensor product; the differential is defined by

$$d_{C\otimes D}(x\otimes y)=d_C x\otimes y+x\otimes d_D y.$$

If $f\in \mathrm{Hom}(C,D)$ then the formula $df=fd_C-d_D f$ defines a differential which makes $(\mathrm{Hom}(C,D),d)$ an ungraded chain complex, and $\mathrm{Kom}_k^{\mathrm{un}}$ is a category which is enriched over itself.

If $\mathrm{Kom}_k^{\mathbb{Z}/2}$ denotes the dg category of $\mathbb{Z}/2$–graded chain complexes then there is an adjunction

$$\iota\colon \mathrm{Kom}_k^{\mathrm{un}}\leftrightarrow \mathrm{Kom}_k^{\mathbb{Z}/2}:\rho$$

in which ι maps (C,d) to the chain complex $(C_n,d_n)_{n\in\mathbb{Z}/2}$ where $C_n=C$ and $d_n=d$ for each $n\in\mathbb{Z}/2$. If $(C_n,d_n)_{n\in\mathbb{Z}/2}$ is a chain complex then $C=\bigoplus_n C_n$ and $d=\sum_n d_n$ determine a forgetful functor $\rho\colon \mathrm{Kom}_k^{\mathbb{Z}/2}\to \mathrm{Kom}_k^{\mathrm{un}}$.

An *ungraded dg category* $\mathcal{C}$ is a category which is enriched over $\mathrm{Kom}_k^{\mathrm{un}}$. The adjunction above induces an adjunction between the category $\mathrm{dgcat}_k^{\mathrm{un}}$ of ungraded dg categories and the category $\mathrm{dgcat}_k^{\mathbb{Z}/2}$ of $\mathbb{Z}/2$–graded categories. This extends to a Quillen adjunction which induces model structures corresponding to Hqe and Hmo on $\mathrm{dgcat}_k^{\mathrm{un}}$; for analogous details see [9, Section 5.1].

3 Formal contact categories

In this section, a contact category $\mathcal{K}o(\Sigma)$ is associated to each oriented surface Σ. The remainder of the paper will assume that k is a field of characteristic 2 and use the trivial grading.

3.1 Bypass moves

In what follows, surfaces will always be pointed in the sense defined below.

Definition 3.1 A *pointed surface* Σ is a compact connected surface Σ in which the connected components of the boundary have been ordered and each boundary component $\partial_i\Sigma$ contains a marked point $z_i\in\partial_i\Sigma$:

$$\partial\Sigma=\partial_1\Sigma\cup\cdots\cup\partial_n\Sigma,\quad z=\{z_1,\dots,z_n\}\quad\text{and}\quad z_i\in\partial_i\Sigma.$$

Every closed surface is canonically pointed.

A pointed oriented surface Σ in which a collection of points $m\subset\partial\Sigma$ satisfies the conditions

$$m\cap z=\varnothing\quad\text{and}\quad |m|\in 2\mathbb{Z}_+$$

will be denoted by (Σ,m). We write $m=\bigcup_i m_i$ where $m_i\subset\partial_i\Sigma$. Often notation will be abused and m will be used to denote both the set m and the cardinality $|m|$.

An orientation on a pointed surface Σ induces an orientation of each boundary component. The points $m_i\subset\partial_i\Sigma$ inherit an ordering by starting from the basepoint $z_i\in\partial_i\Sigma$ and traversing the boundary circle in this direction. Combining the order on each $m_i\subset\partial_i\Sigma$ with the ordering of the boundary components $\{\partial_1\Sigma,\partial_2\Sigma,\dots,\partial_n\Sigma\}$ produces a total ordering on the set m.

Recall that an arc γ is properly embedded in a pointed surface when $\partial\gamma \subset \partial\Sigma \setminus z$ and $\mathrm{int}(\gamma) \cap \partial\Sigma = \varnothing$. Arcs γ are required to intersect the boundary transversely.

Definition 3.2 Let Σ be a pointed orientable surface possibly with boundary. Then a properly embedded collection of smooth curves and arcs γ on Σ is a *multicurve*.

If γ is a multicurve on (Σ, m) then we require that the set $\gamma \cap \partial\Sigma$ coincides with the points m chosen on the boundary $\partial\Sigma$.

Definition 3.3 A nonempty multicurve γ is said to be a *dividing set on the surface* Σ when there are disjoint subsurfaces R_+ and R_- of Σ such that

$$\Sigma \setminus \gamma = R_+ \cup R_- \quad \text{and as sets} \quad \gamma = \partial R_+ \setminus \partial\Sigma = \partial R_- \setminus \partial\Sigma.$$

If Σ is a surface with boundary, then we require that the intersection number $i(\gamma, \partial\Sigma)$ is a positive even integer. In particular, when Σ has boundary we *require* that $m \geq 2$.

The subsets R_+ and R_- of Σ are the *positive region* and the *negative region* of γ on Σ, respectively. These regions may be labeled by $+$ and $-$ signs in illustrations.

If a multicurve γ is a dividing set, then for each boundary component $\partial_i\Sigma$, the number of points $\gamma \cap \partial_i\Sigma$ must be even.

Definition 3.4 For any dividing set γ on Σ, there is a *dual dividing set* $\gamma^\vee$ on Σ that is obtained by exchanging the positive and negative regions.

The *equator* $\ell = \{(x, y) : y = 0\} \subset D^2 = \{x \in \mathbb{R}^2 : |x| < 1\}$ of a disk is the line formed by the x–axis in the standard embedding $D^2 \subset \mathbb{R}^2$. The equator ℓ divides the disk D^2 into two *half-disks*, a bottom B and a top T:

$$D^2 = B \cup T \quad \text{and} \quad B \cap T = \ell.$$

The boundary ∂T of the top half-disk T consists of the equator ℓ and the northern hemisphere $\nu \subset \partial D^2$ of the boundary circle:

$$\partial T = \ell \cup \nu.$$

Definition 3.5 Suppose that γ is a dividing set on an oriented surface Σ. Then a *bypass disk on* γ is a smoothly embedded oriented half-disk $(T, \ell) \subset (\Sigma \times [0, 1], \Sigma \times 0)$ which satisfies the following properties:

(1) The equatorial arc ℓ intersects γ at exactly three points, a, b and c, such that

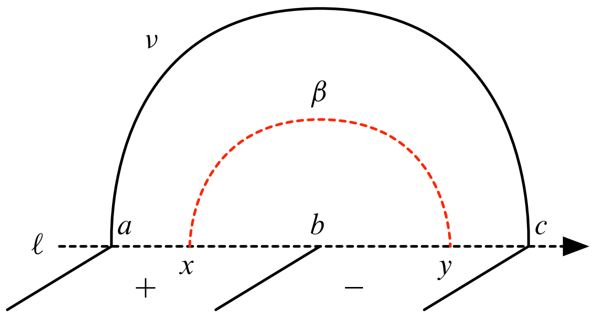
$$\ell = [a, b] \cup [b, c] \quad \text{and} \quad a < b < c,$$

where the order of the points is induced by the orientation.

(2) The boundary points of the arcs ℓ and ν are the points a and c.

A *dividing set* β of a bypass disk T is a properly embedded arc starting at a point x between a and b and ending at a point y between b and c.

Definition 3.5 is illustrated below:



This picture shows a bypass disk T embedded in a thickened surface $\Sigma \times [0,1]$. The boundary of the half-disk consists of the dashed equatorial arc ℓ and the boundary of the northern hemisphere ν. The dashed curve β is the dividing set for the bypass disk. The three straight lines at the bottom are part of a dividing set γ on the surface Σ. The labels a, b and c indicate the intersection points of the arc ℓ with the dividing set γ. The orientation of T is determined by fixing the direction of the equator ℓ and using the standard orientation along the normal axis. The equator ℓ is drawn beyond the boundary of T for aesthetic reasons.

Remark 3.6 If $\Sigma \subset (M,\xi)$ is a convex surface in a contact 3–manifold, then ξ determines a dividing set γ on Σ. A bypass disk T, embedded into a regular neighborhood of Σ, determines an operation on the dividing set called *bypass attachment* that changes the dividing set and the contact structure in a well-understood way [16]. These operations generate the contact structures on $M = \Sigma \times [0,1]$ in a sense which has been made precise by Honda [17, Lemma 3.10 (isotopy discretization)].

If Σ is an oriented surface then the space $\Sigma \times [0,1]$ will be always be oriented by appending the vertical direction to the orientation of Σ.

Definition 3.7 A bypass disk (T,ℓ) in $\Sigma \times [0,1]$ determines the product orientation on $\Sigma \times [0,1]$. In more detail, if ℓ represents the direction of the equator and n is the direction of the disk normal to the surface, then the three vectors $(\ell, \ell \times n, n)$ determine this orientation of $\Sigma \times [0,1]$. If the orientation induced by T agrees with that of $\Sigma \times [0,1]$ then the bypass disk is said to be *orientation preserving*; otherwise it is *orientation reversing*.

Definition 3.8 (bypass move) Suppose that γ is a dividing set on an oriented surface Σ, T is a bypass disk on γ and $N(T)$ is a regular neighborhood of the half-disk $T \subset \Sigma \times [0,1]$. The boundary $\partial N(T)$ contains two copies of the half-disk T, which we will call *faces*. Each face, being a parallel copy of the half-disk T, contains a collection of points

$$a < x < b < y < c$$

ordered along an equator ℓ, a dividing set β and a northern hemisphere ν. Moreover, there are three line segments γ_a, γ_b and γ_c from γ, on either side, meeting the points a, b and c, respectively. The face in the $\ell \times n$ direction of $T \times \{\frac{1}{2}\} \subset T \times [0,1]$ is called the *positive face*; the other face is the *negative face*.

There is a dividing set η on the surface $\Sigma' = \partial(\Sigma \cup N(T))$, which is constructed by regluing the segments of γ according to the prescription below:

(1) If T is orientation preserving, then on the positive face attach γ_b to the point x of β and attach γ_c to the point y of β, and on the negative face attach γ_a to the point x and attach γ_b to the point y.

(2) If T is not orientation preserving, then on the positive face attach γ_a to the point x of β and attach γ_b to the point y of β, and on the negative face attach γ_b to the point x and attach γ_c to the point y.

(3) Attach the curve γ_a on the latter face to the curve γ_c on the former face by an interval that crosses over the $\nu \times [0,1] \subset \partial N(T)$ boundary component along the diagonal.

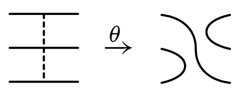
After smoothing the corners, the surface Σ' is diffeomorphic to Σ by a diffeomorphism ψ which is isotopic to the identity. If $\gamma' = \psi(\eta)$ then the *bypass move* $\theta\colon \gamma \to \gamma'$ is the tuple

$$\gamma \xrightarrow{\theta} \gamma' = (T, \gamma, \gamma')$$

given by the bypass disk T, the dividing set γ and the curve γ' determined by the operation described above.

Remark 3.9 The definition of the bypass move requires a choice of smoothing. We fix one choice and use it consistently. Any two such choices will produce equivalent categories.

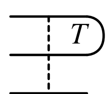
The picture below shows the orientation-preserving bypass move defined above. On the left-hand side, the dividing set γ consists of three horizontal lines and the equator ℓ of the bypass disk T is indicated by the vertical line. The rest of the bypass disk T is assumed to come out of the page. The positive and negative regions on the right are determined by the positive and negative regions on the left.



In the contact category, bypass moves are required to be orientation preserving. Since the orientation of a bypass disk T is determined by the direction of the equator, we will always choose orientations which are compatible with the ambient orientation of the surface. So it is not necessary to denote the orientation in most illustrations.

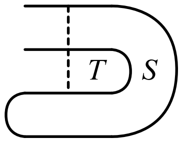
Special types of bypass moves The two special types of bypass moves isolated below correspond precisely to the relations (1) and (2) in Definition 3.15.

Definition 3.10 A bypass move $\theta\colon \gamma \to \gamma'$ is *capped* when either the subset $[a,b]$ or the subset $[b,c]$ of the associated equator ℓ is the equator ρ of an embedded half-disk $(T, T \setminus \rho) \to (\Sigma, \gamma)$ which does not intersect the equator at any other point.



Intercardinal directions will be used to locate caps. For instance, a bypass featuring a cap T in its northeastern corner is pictured above.

Example 3.11 The picture below contains one cap T in the southeastern corner. The half-disk labeled S is not a cap because it intersects the equator twice.



Capped bypass moves are the least interesting bypass moves since, depending upon where the cap is found, a capped bypass must be either nullhomotopic or equal to the identity map in the formal contact category.

Definition 3.12 Two distinct bypass moves $\theta\colon \gamma \to \gamma'$ and $\theta'\colon \gamma \to \gamma''$ are *disjoint*, up to isotopy with endpoints fixed in the dividing set, when the equators of their bypass disks have geometric intersection number zero.

If a collection of bypass moves $\{\theta_i\}_{1\leq i\leq n}$ on a dividing set γ is pairwise disjoint, then performing the moves in any order produces the same result: γ'. So the union

$$\coprod_{i=1}^{n} \theta_i : \gamma \to \gamma'$$

may be viewed as a kind of bypass combo move.

Isotopy of curves and disks

Definition 3.13 If γ and γ' are dividing sets on a surface Σ then they are *isotopic*, $\gamma \simeq \gamma'$, when they are isotopic as multicurves on Σ. If Σ is a pointed surface then the isotopy is required to fix the basepoints $z \subset \partial\Sigma$. If (Σ, m) is a surface with points m on each boundary component then the isotopy is required to fix the points at which the dividing sets attach to each boundary component.

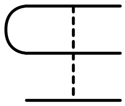
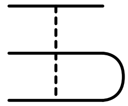
Two bypass moves $\theta = (T, \gamma, \gamma')$ and $\theta' = (S, \delta, \delta')$ are *isotopic*, $\theta \simeq \theta'$, when the graph $\gamma \cup \ell$ is isotopic to $\delta \cup \rho$ where ℓ and ρ are equators of T and S, respectively.

Remark 3.14 If Σ is realized as a convex surface in the 3–manifold $M = \Sigma \times [0, 1]$ and the dividing sets γ and γ' corresponding to two contact structures ξ and ξ' are isotopic, then ξ and ξ' are contactomorphic [16]. Since our motivation is to produce a category in which morphisms behave like contact structures up to contactomorphism, isotopic dividing sets are identified in Definition 3.15.

3.2 The contact category

Definition 3.15 The *preformal contact category* Pre–$\mathcal{K}$o(Σ) is the ungraded k–linear category with objects corresponding to isotopy classes of dividing sets on Σ and maps generated by isotopy classes of orientation-preserving bypass moves subject to the following relations:

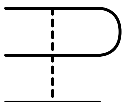
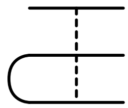
(1) If θ is a capped bypass move then $\theta = 1$ when the cap can be found in the northwest or southeast:

 $= 1$ and  $= 1$.

(2) If θ and θ' are disjoint bypass moves then the maps that they determine commute:

$$\theta\theta' = \theta \amalg \theta' = \theta'\theta.$$

These relations are required for the formal contact category, defined below, to have any bearing on contact geometry; see Remark 3.14. In Section 4.3, we will show that the first relation implies that $\theta = 0$ in the associated homotopy category when the corresponding bypass is capped in the northeast or the southwest:

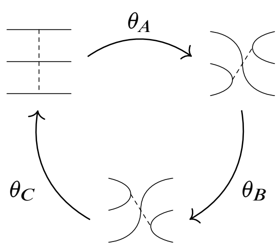
 $= 0$ and  $= 0$.

The next proposition shows that every bypass move determines a triple of composable morphisms. This determines a functor from the category D' in Definition 2.9 to the category $\text{Pre–}\mathcal{K}\text{o}(\Sigma)$. This proposition is due to Honda and K Walker; see [15; 59].

Proposition 3.16 *For each oriented surface Σ and each dividing set γ on Σ, each bypass move θ on γ determines a functor $\tilde{\theta}\colon D' \to \text{Pre–}\mathcal{K}\text{o}(\Sigma)$.*

Proof Set $\gamma_A = \gamma$ and $\theta_A = \theta$. By definition, a bypass move $\theta_A = (T_A, \gamma_A, \gamma_B)$ is locally modeled on a bypass disk T_A in $\Sigma \times [0, 1]$ which intersects γ_A in three points. There is a bypass disk T_B on the dividing set γ_B which results from the bypass move θ_A. The disk T_B determines a bypass move $\theta_B = (T_B, \gamma_B, \gamma_C)$, and there is a bypass disk T_C on the dividing set γ_C. The disk T_C determines a bypass move $\theta_C = (T_C, \gamma_C, \gamma_A)$; the result of the bypass T_C is the original dividing set $\gamma = \gamma_A$. These choices are unique up to isotopy. □

The construction above is illustrated below. Each of the arrows in the diagram is a bypass move. The solid lines represent dividing sets on the surface Σ and the dashed lines represent the equators of bypass disks:



The icon at the source of a given arrow represents a dividing set γ on the surface Σ. The icon at the target of the arrow represents the dividing set obtained by performing the bypass move with equator given by the dashed line in the source.

The proposition above allows us to associate a functor $\tilde{\theta}\colon D' \to \text{Pre–}\mathcal{K}\text{o}(\Sigma)$ to each bypass move $\theta\colon \gamma \to \gamma'$ between dividing sets on Σ. Composing the coproduct $\coprod_\theta \tilde{\theta}\colon \coprod_\theta D' \to \coprod_\theta \text{Pre–}\mathcal{K}\text{o}(\Sigma)$ of all such functors with the fold map $\coprod_\theta \text{Pre–}\mathcal{K}\text{o}(\Sigma) \to \text{Pre–}\mathcal{K}\text{o}(\Sigma)$ yields the functor

$$\Xi\colon \coprod_\theta D' \to \text{Pre–}\mathcal{K}\text{o}(\Sigma). \tag{3-1}$$

Definition 3.17 The *formal contact category* $\mathcal{K}\text{o}(\Sigma)$ is the pretriangulated hull of the Postnikov localization of the preformal contact category $\text{Pre–}\mathcal{K}\text{o}(\Sigma)$ along the functor Ξ above:

$$\mathcal{K}\text{o}(\Sigma) = L_\Xi \text{Pre–}\mathcal{K}\text{o}(\Sigma)^{\text{pretr}}$$

By Proposition 2.17, the bypass triangles introduced by the Postnikov localization remain distinguished triangles in the homotopy category of the hull. The formal contact category $\mathcal{K}\text{o}(\Sigma)$ is the universal pretriangulated category generated by bypass moves, containing bypass triangles and satisfying the relations (1) and (2).

Conjecture 3.18 *A cofibrant–fibrant replacement for* $\mathcal{K}\text{o}(\Sigma)$ *can be constructed without homotopy pushouts. Note that, before relations* (1) *and* (2) *are applied to the preformal contact category*

$$\text{Pre–}\mathcal{K}\text{o}(\Sigma) = \text{Pre–Pre-}\mathcal{K}\text{o}(\Sigma)/\langle(1),(2)\rangle,$$

the "prepreformal contact category" is freely generated by bypass moves. Any freely generated category is cofibrant as it can be obtained by a series of pushouts along generating cofibrations in Hqe. *One can then adjoin copies of Drinfeld's category* $\tilde{I}$ *via pushout and copies of a resolution for the symmetric algebra for each instance of relations* (1) *and* (2), *respectively. The result is cofibrant in* Hqe, *so the homotopy pushout which underlies the Postnikov localization in Definition 3.17 is now an ordinary pushout and the result of this pushout is both cofibrant and fibrant in* Hqe. *The idempotent completion* $L_\Xi \text{Pre–}\mathcal{K}\text{o}(\Sigma)^{\text{perf}}$ *of* $\mathcal{K}\text{o}(\Sigma)$ *is cofibrant and fibrant in the Morita category* Hmo.

4 Elementary properties of contact categories

In this section, many of the properties which should hold for the contact categories [15] are shown to hold for the formal contact categories. The formal contact category associated to a surface decomposes into a product of formal contact categories with fixed Euler invariant. The category with Euler invariant n is equivalent to the category with Euler invariant $-n$. Reversing the orientation of the surface is equivalent to forming the opposite category. A dividing set featuring a homotopically trivial curve is contractible and dividing sets featuring regions which are disconnected from the boundary are shown to be homotopy equivalent to convolutions of dividing sets which are connected to the boundary.

4.1 Decompositions of contact categories

The contact categories $\mathcal{K}\mathrm{o}(\Sigma)$ consist of noninteracting subcategories $\mathcal{K}\mathrm{o}^n(\Sigma, m)$. Each subcategory is determined by fixing some points m on each boundary component and the Euler number $n = \mathfrak{e}(\gamma)$ of the dividing sets γ on Σ.

Euler decomposition If $(\Sigma \times [0,1), \xi)$ is a contact 3–manifold and $e(\xi)$ is the Euler class of ξ, then the Euler number of ξ is $\mathfrak{e}(\xi) = \langle e(\xi), [\Sigma]\rangle$. This number can be computed from the dividing set $\gamma \subset \Sigma$.

Definition 4.1 If γ is a dividing set on an orientable surface Σ then the *Euler number* $\mathfrak{e}(\gamma)$ of γ is the Euler characteristic of the positive region minus the Euler characteristic of the negative region:

$$\mathfrak{e}(\gamma) = \chi(R_+) - \chi(R_-).$$

The proposition below shows that this is a reasonable thing to consider.

Proposition 4.2 *The Euler number satisfies the following properties*:

(1) *If two dividing sets are isotopic then the corresponding Euler numbers are equal*:

$$\gamma \simeq \gamma' \text{ implies that } \mathfrak{e}(\gamma) = \mathfrak{e}(\gamma').$$

(2) *If $\theta\colon \gamma \to \gamma'$ is a bypass move then the Euler numbers of γ and γ' must be equal.*

Proof The first statement follows from the observation that $\gamma \simeq \gamma'$ implies that $R_+ \simeq R'_+$ and $R_- \simeq R'_-$. The second statement follows from computing each Euler characteristic as a union of the region in which the bypass move is performed and its complement. Suppose that $B \subset \Sigma$ is a small ball containing the bypass moves. If $X_\pm = R_\pm \setminus B$ and $Y_\pm = R_\pm \cap B$, then $Y_\pm$ is homeomorphic to the disjoint union of two disks and $X_\pm \cap Y_\pm$ is homeomorphic to the disjoint union of three intervals. See the illustration following Definition 3.8. □

Remark 4.3 If γ is a dividing set on a surface $(\Sigma_{g,1}, 2)$ of genus g with one boundary component and two points on the boundary, then $\chi(R_+ \cap R_-) = 1$ because γ consists of a disjoint union of circles and one interval connecting the two points which are fixed on the boundary. So $2 - 2g = \chi(R_+) + \chi(R_-)$. If $\mathfrak{e}(\gamma) = 2(g-k)$ then $\chi(R_+) = 1 - k$ and $\chi(R_-) = 1 - l$, where $k + l = 2g$ for $0 \le k \le 2g$.

Since the preformal contact category Pre–$\mathcal{K}\mathrm{o}(\Sigma, m)$ in Definition 3.15 is generated by bypass moves, the proposition above is equivalent to the statement that the Euler number yields a well-defined map $\mathfrak{e}\colon \mathrm{Ob}(\text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m)) \to \mathbb{Z}$ which determines a decomposition

$$\text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m) \cong \coprod_{n \in \mathbb{Z}} \text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m)$$

in which Pre–$\mathcal{K}\mathrm{o}^n(\Sigma, m)$ is the full subcategory of Pre–$\mathcal{K}\mathrm{o}(\Sigma, m)$ such that $\mathfrak{e}(\gamma) = n$ for all $\gamma \in \mathrm{Ob}(\text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m))$. The theorem below shows that this decomposition extends to the formal contact category $\mathcal{K}\mathrm{o}(\Sigma, m)$.

Theorem 4.4 *The formal contact category $\mathcal{K}\mathrm{o}(\Sigma, m)$ splits into a product of categories $\mathcal{K}\mathrm{o}^n(\Sigma, m)$:*

$$\mathcal{K}\mathrm{o}(\Sigma, m) \cong \prod_{n\in\mathbb{Z}} \mathcal{K}\mathrm{o}^n(\Sigma, m).$$

Here $\mathcal{K}\mathrm{o}^n(\Sigma, m)$ is the full subcategory of $\mathcal{K}\mathrm{o}(\Sigma, m)$ with objects that satisfy $\mathfrak{e}(\gamma) = n$.

Proof By the proposition above, $\Xi: \coprod D' \to \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m)$ splits into a union $\Xi = \coprod_n \Xi_n$ where $\Xi_n: \coprod D' \to \text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m)$ corresponds to the bypass triangles contained in $\text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m)$. The localization functor $Q: \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m) \to L_\Xi \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m)$ splits into a union of localizations:

$$\text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m) \cong \coprod_n \text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m) \to L_\Xi \coprod_n \text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m) \cong \coprod_n L_{\Xi_n} \text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m).$$

The theorem follows from Proposition 2.3. □

4.2 Dualities of contact categories

Two forms of duality are introduced, corresponding to switching the labelings of the regions and the ambient orientation of the surface.

Euler duality Definition 3.4 introduced an operation $\gamma \mapsto \gamma^\vee$ on dividing sets which exchanged the positive and negative regions: $R_+ \leftrightarrow R_-$. This reverses the sign of the Euler number: $\mathfrak{e}(\gamma^\vee) = -\mathfrak{e}(\gamma)$. Here this operation is extended to an involution

$$-^\vee: \mathcal{K}\mathrm{o}(\Sigma, m) \to \mathcal{K}\mathrm{o}(\Sigma, m)$$

of the formal contact category which exchanges $\mathcal{K}\mathrm{o}^n(\Sigma, m)$ and $\mathcal{K}\mathrm{o}^{-n}(\Sigma, m)$ from Theorem 4.4.

Proposition 4.5 *The Euler duality map on dividing sets: $-^\vee: \mathrm{Ob}(\text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m)) \to \mathrm{Ob}(\text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m))$ extends to an involution of dg categories:*

$$-^\vee: \mathcal{K}\mathrm{o}^n(\Sigma, m) \to \mathcal{K}\mathrm{o}^{-n}(\Sigma, m) \quad \textit{and} \quad (-^\vee)^\vee \cong 1.$$

Proof If γ is a dividing set on Σ, then for any bypass move $\theta: \gamma \to \gamma'$ the positive and negative regions of γ determine positive and negative regions of γ'; see the illustration after Definition 3.8. Therefore, on the generators θ of $\text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m)$:

$$\theta: \gamma \to \gamma' \quad \mapsto \quad \theta^\vee: \gamma^\vee \to \gamma'^\vee.$$

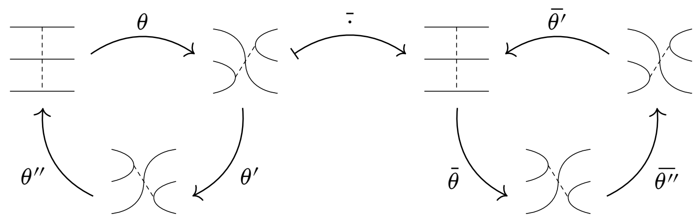
This extends to an involution of $\text{Pre–}\mathcal{K}\mathrm{o}(\Sigma, m)$ which takes triangles to triangles and so descends to a functor $-^\vee: \mathcal{K}\mathrm{o}^n(\Sigma, m) \to \mathcal{K}\mathrm{o}^{-n}(\Sigma, m)$. The uniqueness of this extension implies the relation $(-^\vee)^\vee \cong 1$. The map $-^\vee$ is an equivalence as it is its own inverse. □

Orientation reversal The formal contact category $\mathcal{K}\mathrm{o}(\bar{\Sigma})$ of a surface with reversed orientation is identified with the opposite formal contact category $\mathcal{K}\mathrm{o}(\Sigma)^{\mathrm{op}}$ of the surface.

Proposition 4.6 *There is an equivalence of formal contact categories*

$$\mathcal{K}\mathrm{o}^n(\Sigma, m)^{\mathrm{op}} \xrightarrow{\sim} \mathcal{K}\mathrm{o}^n(\bar{\Sigma}, m).$$

Proof It is a consequence Definition 3.8 that reversing the orientation of the surface is equivalent to reversing the orientation of each bypass half-disk or equator. It suffices to analyze the correspondence between bypass triangles. In the eyeglass-shaped diagram below, reversing the orientation of each bypass disk $\theta \mapsto \bar{\theta}$ in a triangle fixes the source and changes the sink of each map:



Reversing the arrows on the left-hand side of the diagram produces the bypass triangle for $\mathcal{K}\mathrm{o}^n(\Sigma, m)^{\mathrm{op}}$. The assignment $\gamma \mapsto \gamma$ on objects and $\theta^{\mathrm{op}} \mapsto \bar{\theta}'$ on maps determines a functor

$$\bar{\cdot}\colon \text{Pre–}\mathcal{K}\mathrm{o}^n(\Sigma, m)^{\mathrm{op}} \to \text{Pre–}\mathcal{K}\mathrm{o}^n(\bar{\Sigma}, m)$$

because it preserves the cap relations and disjoint unions. Moreover, the relation $\theta^{\mathrm{op}} \mapsto \bar{\theta}'$ implies that $(\theta')^{\mathrm{op}} \mapsto \bar{\theta}''$ and $(\theta'')^{\mathrm{op}} \mapsto \bar{\theta}$, so that triangles are mapped to triangles and the functor $\bar{\cdot}$ descends to a map between formal contact categories. By applying the same construction to the surface after reversing its orientation again, one obtains an inverse functor, and so the functor $\bar{\cdot}$, introduced above, is an isomorphism of formal contact categories. □

4.3 Relations for overtwisted contact structures

A theorem of E Giroux [13] states that a contact structure on $\Sigma \times [0,1]$, when $\Sigma \neq S^2$, is overtwisted if and only if its dividing set contains no homotopically trivial closed curves. When $\Sigma = S^2$, a contact structure is overtwisted if and only if the dividing set contains any two such curves. Corollary 4.10 states that Giroux's criterion is satisfied for surfaces with boundary. The surface Σ is assumed to be connected in this section.

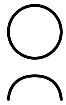
The lemma below shows that the local relations can be applied to parts of more complicated dividing sets.

Lemma 4.7 (local relations) *Suppose that R and Σ are orientable surfaces and $R \subset \Sigma$. Then a distinguished triangle in* $\mathrm{Ho}(\mathcal{K}\mathrm{o}(R))$ *yields a distinguished triangle in* $\mathrm{Ho}(\mathcal{K}\mathrm{o}(\Sigma))$.

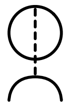
Proof The embedding $R \subset \Sigma$ determines a functor $i\colon \text{Pre–}\mathcal{K}\mathrm{o}(R) \hookrightarrow \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma)$. A bypass triangle $\tilde{\theta}\colon D' \to \text{Pre–}\mathcal{K}\mathrm{o}(R)$ determines a bypass triangle $D' \to \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma)$ after composing with i. □

Definition 4.8 If γ is a dividing set then we write $S^1 \subset \gamma$ when γ contains a homotopically trivial closed curve. All such curves are isotopic when Σ is connected. If γ contains any collection of $n \in \mathbb{Z}_+$ such curves then we write $nS^1 \subset \gamma$.

Proposition 4.9 *The object represented by the dividing set pictured below is contractible*:

 $\cong 0$

Proof The formal contact category $\mathrm{Ho}(\mathcal{K}\mathrm{o}(D^2, 2))$ associated to the disk D^2 with two boundary points contains a bypass move with equator indicated by the dashed line below:



All of the objects in the distinguished triangle associated to the bypass move are isotopic, and the first relation in Definition 3.15 implies two out of three of the maps are the identity. □

Corollary 4.10 (1) *If* Σ *is a surface with boundary*, *then for all dividing sets* γ *on* Σ,

$$S^1 \subset \gamma \textit{ implies } \gamma \cong 0 \textit{ in } \mathrm{Ho}(\mathcal{K}\mathrm{o}(\Sigma)).$$

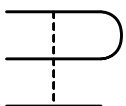
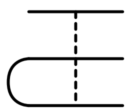
(2) *If* Σ *is a closed surface then for all dividing sets* γ *on* Σ,

$$S^1 \subset \gamma \textit{ and } \gamma \neq S^1 \textit{ implies } \gamma \cong 0 \textit{ in } \mathrm{Ho}(\mathcal{K}\mathrm{o}(\Sigma)).$$

Proof The proposition above applies to surfaces with boundary as they are required to contain properly embedded arcs. □

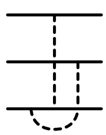
Without further complicating the main construction, this corollary appears to be optimal: bypass moves do not imply that $S^1 \cong 0$ in the disk category $\mathrm{Ho}(\mathcal{K}\mathrm{o}(D^2, 0))$, and any such proof would contradict Giroux's theorem for $\Sigma = S^2$.

Corollary 4.11 *The relation in Proposition 4.9 implies that a bypass move is zero in the homotopy category when it is capped in either the northeast or southwest*:

 $= 0$ *and*  $= 0.$

Proof The dividing set γ' resulting from either bypass move $\theta\colon \gamma \to \gamma'$ must contain a homotopically trivial curve. So the isomorphism $\gamma' \cong 0$ is obtained by applying Lemma 4.7 and Proposition 4.9. This implies the relation $\theta = 0$ in the homotopy category of the formal contact category. □

Remark 4.12 Two consecutive bypass moves occurring in a bypass triangle are disjoint:



The second bypass is capped when it is performed before the first, so the commutativity of disjoint bypasses and the corollary above suffice to imply that compositions of consecutive bypass moves must be zero in the homotopy category.

4.4 Dividing sets containing disconnected regions are convolutions

Suppose γ is a dividing set on a surface Σ with boundary and $\Sigma \setminus \gamma$ contains a connected component B which is disjoint from the boundary of Σ. We will show that γ is homotopy equivalent to an iterated cone construction on dividing sets which do not contain a region such as B.

Definition 4.13 A multicurve γ on a surface Σ with boundary is *boundary disconnected* when there is a connected component B of $\Sigma \setminus \gamma$ which does not touch the boundary:

$$B \subset \Sigma \setminus \gamma \quad \text{and} \quad B \cap \partial\Sigma = \varnothing$$

A dividing set γ is *boundary connected* when it is not boundary disconnected.

Theorem 4.14 *In the homotopy category of the formal contact category $\mathcal{K}\mathrm{o}(\Sigma, m)$ associated to a surface (Σ, m) with boundary, every boundary-disconnected dividing set γ is isomorphic to an iterated extension of dividing sets γ_i which are boundary connected.*

Proof Observe that boundary-disconnected regions can be nested. For example, an annulus can be placed within the annulus illustrated below. For the purpose of this argument, the amount of nesting $n(\gamma)$ is defined to be

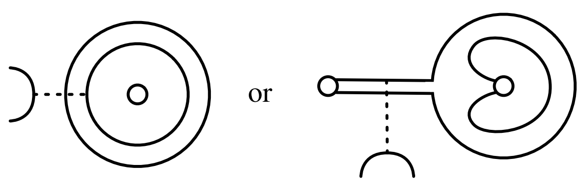
$$n(\gamma) := \max_B \min_a |a \cap \gamma|,$$

where $a\colon (I, \{0\}, \{1\}) \to (B, \partial\Sigma, \mathrm{int}(B))$ is an arc from the boundary $\partial\Sigma$ to an interior point of a connected component $B \subset \Sigma \setminus \gamma$.

The proof is by induction on the amount of nesting in boundary-disconnected regions. Fix a dividing set γ. If $n(\gamma) = 0$ and there are no boundary-disconnected regions then there is nothing to show. So assume that the statement of the theorem holds for all γ with $n(\gamma) = N$ and suppose $n(\gamma) = N + 1$.

There are innermost disconnected regions B and arcs $a\colon I \to B$ in Σ which satisfy $|a \cap \gamma| = N + 1$. Fix such a disconnected region B.

If this disconnected region is a disk then γ is isomorphic to zero because $|m| \geq 2$ by Proposition 4.9. If γ is a dividing set on a surface with boundary and $\Sigma \setminus \gamma$ contains an annulus or a punctured torus component, then there are bypass moves



respectively. The first picture above shows two concentric homotopically nontrivial circles in the annulus $(S^1 \times [0,1], 2)$. In the second picture above, the two small circles are identified by folding the page to form a torus with one boundary component $(T^2 \setminus D^2, 2)$. In either case, the triangle associated to the indicated bypass move results in two dividing sets which connect B to either the boundary, when $n(\gamma) = 1$, or a region outside of B, when $n(\gamma) > 1$. In either case this lowers $n(\gamma)$ by 1.

In general, the innermost region B is an orientable surface with boundary. Any such surface is obtained by attaching 1–handles to the boundary components of a disjoint union of punctured tori $\Sigma_{1,1}$ and annuli $\Sigma_{0,2}$. If B has genus g and $n+1$ boundary components, then B is abstractly homeomorphic to g copies of $\Sigma_{1,1}$ and n copies of $\Sigma_{0,2}$ glued together in this fashion. In particular, there is a 1–handle H which, when cut along its cocore I, produces a disjoint union of surfaces with lower genus or number of boundary components. There is an interval ℓ in Σ which is obtained by connecting I to a point on the boundary of the region outside of B (which is not in ∂B itself). By construction, this interval ℓ intersects γ at three points. The bypass move θ determined by ℓ determines a distinguished triangle

$$\gamma \xrightarrow{\theta} \gamma' \to \gamma'' \to \gamma[1]$$

with objects γ' and γ'' that must contain disconnected regions B and B'' with lower genus or number of boundary components. This procedure can be iterated until the result contains only annuli and tori, to which one applies the bypasses in the previous paragraph.

Applying the procedure in the previous two paragraphs to each innermost disconnected region expresses the result as an iterated extension of dividing sets for which $n(\gamma) < N+1$. It follows by induction that γ can be further expressed as an iterated extension of dividing sets, for which $n(\gamma) = 0$, which are boundary connected. □

4.5 The positive half of the contact category

The decomposition of the formal contact category introduced by the proposition below will clarify our discussion later.

Proposition 4.15 *The formal contact category $\mathcal{K}o(\Sigma, m)$ associated to a surface with boundary splits into a product of two pieces*,

$$\mathcal{K}o(\Sigma, m) \cong \mathcal{K}o_+(\Sigma, m) \times \mathcal{K}o_-(\Sigma, m),$$

supported on the dividing sets $\gamma \in \mathcal{K}o(\Sigma, m)$, in which the basepoint $z_1 \in \partial_1 \Sigma$ is contained in a positive or negative region, *respectively*.

Proof If two dividing sets γ and γ' are isotopic, then the signs of the regions containing the basepoint must be equal. If $\theta\colon \gamma \to \gamma'$ is a bypass move then it cannot change the sign of the region containing the basepoint z_1. The rest of the proof follows along the same lines of the proof of Theorem 4.4. □

By Proposition 4.5, the two pieces found in the decomposition above are equivalent:

$$-^{\vee} : \mathcal{K}\mathrm{o}^{n}_{+}(\Sigma, m) \xrightarrow{\sim} \mathcal{K}\mathrm{o}^{-n}_{-}(\Sigma, m).$$

In Corollary 5.3, moving the basepoint z_1 to an adjacent region is shown to yield an equivalence $r : \mathcal{K}\mathrm{o}^{n}_{+}(\Sigma, m) \xrightarrow{\sim} \mathcal{K}\mathrm{o}^{n}_{-}(\Sigma, m)$. By composing the two maps we obtain an equivalence

$$\mathcal{K}\mathrm{o}^{n}_{+}(\Sigma, m) \xrightarrow{\sim} \mathcal{K}\mathrm{o}^{-n}_{+}(\Sigma, m).$$

See also Proposition 6.15.

5 Symmetries and generators of contact categories

The mapping class group of the surface Σ is shown to act naturally on the formal contact category $\mathcal{K}\mathrm{o}(\Sigma)$. After introducing arc diagrams and parametrizations of surfaces by arc diagrams, each parametrization of Σ by an arc diagram is shown to yield a system of generators for the formal contact category. Section 5.4 contains a discussion of decategorification.

5.1 The mapping class group action

In this section, we show that the mapping class group $\Gamma(\Sigma)$ acts naturally on $\mathcal{K}\mathrm{o}(\Sigma)$.

Definition 5.1 Suppose that Σ is an oriented surface. Then the mapping class group $\Gamma(\Sigma)$ is the group of connected components of the group of orientation-preserving and boundary-fixing diffeomorphisms:

$$\Gamma(\Sigma) = \pi_0 \operatorname{Diff}^{+}(\Sigma, \partial\Sigma).$$

Recall that an action of a group G on a dg category $\mathcal{C}$ is a homomorphism from G to the group $\operatorname{Aut}(\mathcal{C}) \subset \operatorname{End}_{\mathrm{Hmo}}(\mathcal{C})$ of derived equivalences.

Theorem 5.2 *The mapping class group $\Gamma(\Sigma)$ acts naturally on the formal contact category $\mathcal{K}\mathrm{o}(\Sigma)$.*

Proof The proof occurs in two steps: first we construct a natural $\Gamma(\Sigma)$–action on the preformal contact category Pre–$\mathcal{K}\mathrm{o}(\Sigma)$, and second this group action is extended to the formal contact category $\mathcal{K}\mathrm{o}(\Sigma)$.

A diffeomorphism class $g \in \Gamma(\Sigma)$ determines a functor $f_g : \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma) \to \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma)$ that is defined by its action on dividing sets and bypass moves. If γ is an isotopy class of dividing set on Σ then there is a unique isotopy class of dividing set $g\gamma$, and if $\theta = (T, \gamma, \gamma')$ is a bypass move then there is a unique bypass disk gT and associated bypass move $g\theta = (gT, g\gamma, g\gamma')$. Since the category Pre–$\mathcal{K}\mathrm{o}(\Sigma)$ is generated by bypass moves and the assignment $\theta \mapsto g\theta$ preserves disjointness of bypass moves and caps of bypass moves, there is a functor

$$f_g : \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma) \to \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma) \quad \text{such that } f_g(\gamma) = g\gamma \text{ and } f_g(\theta) = g\theta.$$

Both the composition law $f_{gg'} = f_g \circ f_{g'}$ and naturality follow directly from the definition. In particular, since the identity diffeomorphism $1 \in \Gamma(\Sigma)$ fixes both dividing sets and bypass moves, the functor f_1 is the identity functor $1_{\text{Pre–}\mathcal{K}\mathrm{o}(\Sigma)}$.

Suppose that $f_g : \text{Pre–}\mathcal{K}\text{o}(\Sigma) \to \text{Pre–}\mathcal{K}\text{o}(\Sigma)$ is a functor occurring in the construction above. Composing with the localization functor $Q : \text{Pre–}\mathcal{K}\text{o}(\Sigma) \to L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma)$ from (3-1) yields a functor $\text{Pre–}\mathcal{K}\text{o}(\Sigma) \to L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma)$. By Definition 2.15, the image of

$$Q^* : \text{Hom}_{\text{Hqe}}(L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma), L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma)) \to \text{Hom}_{\text{Hqe}}(\text{Pre–}\mathcal{K}\text{o}(\Sigma), L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma))$$

is the subset of functors $f : \text{Pre–}\mathcal{K}\text{o}(\Sigma) \to L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma)$ whose restriction to a bypass triangle extends to a distinguished triangle in the localization $L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma)$.

If $\tilde{\theta} : D' \to \text{Pre–}\mathcal{K}\text{o}(\Sigma)$ is the bypass triangle

$$\gamma \xrightarrow{\theta} \gamma' \xrightarrow{\theta'} \gamma'' \xrightarrow{\theta''} \gamma[1]$$

associated to a bypass move $\theta = (T, \gamma, \gamma')$ on Σ by Proposition 3.16, then $f_g(\theta) = (gT, g\gamma, g\gamma')$ and $f_g(\tilde{\theta})$ corresponds to the bypass triangle

$$g\gamma \xrightarrow{g\theta} g\gamma' \xrightarrow{g\theta'} g\gamma'' \xrightarrow{g\theta''} g\gamma[1].$$

Since the criteria of Definition 2.15 are satisfied, there is a unique lift of the functor $Q \circ f_g$ to a functor $\tilde{f}_g : L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma) \to L_\Xi\text{Pre–}\mathcal{K}\text{o}(\Sigma)$. By Proposition 2.4, there is an induced functor between the associated pretriangulated hulls:

$$h_g : \mathcal{K}\text{o}(\Sigma) \to \mathcal{K}\text{o}(\Sigma) \quad \text{where } h_g = \tilde{f}_g^{\text{pretr}}.$$

Uniqueness of the lift and functoriality of $-^{\text{pretr}}$ imply that the stated group action is obtained. □

The same argument as above allows us to define an automorphism r which moves the first basepoint across the first adjacent boundary point. The corollary below records the existence of this map.

Corollary 5.3 *There is a distinguished automorphism r of $\mathcal{K}\text{o}(\Sigma, m)$ which moves the first basepoint $z_1 \in \partial_1\Sigma$ on the first boundary component over the nearest boundary point in the direction of the orientation.*

The functor r induces functors $r : \mathcal{K}\text{o}^n_\pm(\Sigma, m) \to \mathcal{K}\text{o}^n_\mp(\Sigma, m)$ with respect to the decomposition of $\mathcal{K}\text{o}^n(\Sigma, m)$ found in Proposition 4.15. See also Proposition 6.15.

5.2 Arc diagrams

An arc diagram is a combinatorial way to record a handle decomposition of a surface. The definitions below are due to Zarev [62] and constitute generalizations of ideas which were used by Lipshitz, Ozsváth and Thurston [36, Section 3.2].

Definition 5.4 An *arc diagram* $\mathcal{Z}$ consists of three things:

(1) an ordered collection $Z = \{\mathcal{Z}_1, \dots, \mathcal{Z}_l\}$ of l oriented line segments,

(2) a set $\boldsymbol{a} = \{a_1, \dots, a_{2k}\}$ of distinct points in the line segments Z, and

(3) a two-to-one function $M : \boldsymbol{a} \to \{1, \dots, k\}$ called the *matching*.

In order to apply to any version of the bordered Heegaard–Floer package, this data is required to be *nondegenerate*: after performing surgery on Z at each 0–sphere $M^{-1}(j)$ for $1 \le j \le k$, the resulting 1–manifold must have no closed components.

The set of points $\boldsymbol{a}$ receives a total ordering from the order on the set Z and the orientations of the line segments. The numbers l and k are allowed to be zero. Each arc diagram $\mathcal{Z}$ determines a surface $F(\mathcal{Z})$.

Definition 5.5 The *surface* $F(\mathcal{Z})$ associated to an arc diagram $\mathcal{Z}$ is given by thickening each line segment $\mathcal{Z}_i$ to $\mathcal{Z}_i \times [0,1]$ for $1 \le i \le l$ and attaching oriented 1–handles $D^1 \times D^1$ along the normal bundles of the 0–spheres $M^{-1}(j) \times \{0\}$ for $1 \le j \le k$. The surface $F(\mathcal{Z})$ is oriented by extending the orientation of the line segment $\mathcal{Z}_1$ and its positive normal.

Remark 5.6 One can regard $\mathcal{Z}_i$ as part of the boundary of $\mathcal{Z}_i \times [0,1]$. In Definition 5.10, an arc parametrization will be used to construct dividing sets $\mathfrak{z}_C \in \mathcal{K}\mathrm{o}_+(F(\mathcal{Z}))$ in which the positive regions correspond to the handles of $\mathcal{Z}$. In particular, $\mathcal{Z}_i$, when regarded as part of the boundary, will always be contained in a positive region of $\mathfrak{z}_C \in \mathcal{K}\mathrm{o}_+(F(\mathcal{Z}))$ (and a negative region of $z_C \in \mathcal{K}\mathrm{o}_-(F(\mathcal{Z}))$.

Recall that the points m on a pointed oriented surface (Σ, m) are also ordered by the ordering of the boundary components, and the order on each boundary component is obtained by starting from each basepoint and traveling in the direction of the orientation induced on the boundary.

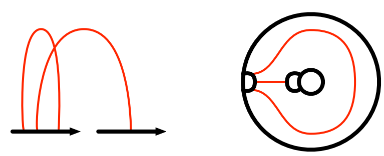
Definition 5.7 Suppose that $m \subset \partial\Sigma$ is the set of *sutures* or points fixed along the boundary of Σ. An *arc parametrization* $(\mathcal{Z}, \varphi_{\mathcal{Z}})$ of a pointed oriented surface (Σ, m) is an arc diagram $\mathcal{Z}$ and a proper orientation-preserving diffeomorphism

$$\varphi_{\mathcal{Z}} \colon \Bigl(F(\mathcal{Z}), \bigcup_{i=1}^{l} \partial \mathcal{Z}_i\Bigr) \to (\Sigma, m)$$

which preserves the total order on the points $\boldsymbol{a}$ and m.

Remark 5.8 An arc parametrization identifies $\bigcup_{i=1}^{l} \partial \mathcal{Z}_i$ with m. The sets m and $\boldsymbol{a}$ play different roles, but under this identification pairs in m partition the points of $\boldsymbol{a}$.

Example 5.9 The annulus $(S^1 \times [0,1], (2,2))$ with two points fixed on each boundary component is parametrized by the arc diagram $\mathcal{Z}$ pictured on the left:



This picture contains two oriented lines $Z = \{\mathcal{Z}_1, \mathcal{Z}_2\}$ and four points $\boldsymbol{a} = \{x, x', y, y'\}$ with $\mathcal{Z}_1 = xyx'$ and $\mathcal{Z}_2 = y'$. The matching function $M \colon \boldsymbol{a} \to \{1,2\}$ is determined by the assignments $M(x) = 1 = M(x')$ and $M(y) = 2 = M(y')$. The picture on the right shows the surface $F(\mathcal{Z})$ associated to $\mathcal{Z}$.

5.3 Generators from arc diagrams

In this section, we show that a parametrization $\mathcal{P} = (\mathcal{Z}, \varphi_{\mathcal{Z}})$ of a pointed oriented surface (Σ, m) determines a canonical collection $\mathfrak{Z}(\mathcal{Z})$ of generators for the associated contact category $\mathcal{K}\mathrm{o}(\Sigma, m)$. This material is motivated by a reading of Zarev [63].

Definition 5.10 Suppose that a pointed oriented surface (Σ, m) is parametrized by an arc diagram $\mathcal{Z}$. Then for each subset $C \subset \{1, \dots, k\}$ of matched pairs, there is an *elementary dividing set*

$$\mathfrak{z}_C = \partial R_C \quad \text{on } \Sigma,$$

where $R_C \subset \Sigma$ is the union of a thickening of the core of each 1–handle indexed by C with the collection of thickened oriented arcs $\mathcal{Z}_i \times [0,1]$. The region R_C is the positive region of $\mathfrak{z}_C$ and its complement $\Sigma \setminus R_C$ is the negative region of $\mathfrak{z}_C$.

An elementary dividing set may be also be called a *positive elementary dividing set*. The *set of positive elementary dividing sets* will be denoted by $\mathfrak{Z}_+(\mathcal{Z})$. The *set of negative elementary dividing sets* $\mathfrak{Z}_-(\mathcal{Z}) = \mathfrak{Z}_+(\mathcal{Z})^\vee$ is obtained by reversing the positive and negative regions. The *set of elementary dividing sets* is the union

$$\mathfrak{Z}(\mathcal{Z}) = \mathfrak{Z}_+(\mathcal{Z}) \cup \mathfrak{Z}_-(\mathcal{Z}).$$

Theorem 5.11 *Suppose (Σ, m) is a pointed oriented surface with boundary and (Σ, m) is parametrized by an arc diagram $\mathcal{Z}$. Then the elementary dividing sets $\mathfrak{Z}(\mathcal{Z})$ classically generate the contact category $\mathcal{K}\mathrm{o}(\Sigma, m)$: any dividing set γ is homotopy equivalent to an iterated extension of dividing sets $\mathfrak{z} \in \mathfrak{Z}(\mathcal{Z})$.*

Proof Suppose that γ is a dividing set on Σ. We will show that γ can be expressed in terms of elementary dividing sets. The proof will be divided into a number of steps.

First By Theorem 4.14 we can assume that γ is boundary connected.

Second Here we simplify γ within the 1–handles of $F(\mathcal{Z})$.

Let $\{c_1, \dots, c_k\}$ be the set of cocores of 1–handles of $F(\mathcal{Z})$. If c_i is a cocore of a 1–handle in $F(\mathcal{Z})$ and the intersection number $|\gamma \cap c_i|$ is greater than 2, then there is a bypass disk with equator parallel to c_i with associated bypass triangle $\gamma \to \gamma' \xrightarrow{\theta_B} \gamma'' \to \gamma[1]$ with $|\gamma' \cap c_i|, |\gamma'' \cap c_i| < |\gamma \cap c_i|$. So γ is isomorphic to a cone

$$\gamma \cong C(\theta_B) \quad \text{such that } |\gamma' \cap c_i|, |\gamma'' \cap c_i| < |\gamma \cap c_i|.$$

Since γ bounds an orientable surface contained within the 1–handle, $|\gamma \cap c_i|$ is even. In more detail, γ bounds $R \subset \Sigma \setminus \gamma$ so $R \cap c_i$ is a disjoint union of intervals. Since the cardinality of the boundary of an interval is 2, $\gamma \cap c_i = \partial(R \cap c_i)$ is even.

Therefore, after iterating this procedure some number of times, we can assume that

$$|\gamma \cap c_i| = 0 \quad \text{or} \quad |\gamma \cap c_i| = 2 \quad \text{for } 1 \le i \le k. \tag{5-1}$$

If the intersection number is 0 then the i^{th} 1–handle is *unoccupied*, and if the number is 2 then the i^{th} 1–handle is *occupied*.

Third Here we simplify γ within the 0–handles of $F(\mathcal{Z})$.

After removing the cocores from the surface, one obtains a disjoint union of disks

$$F(\mathcal{Z}) \setminus \{c_1, \ldots, c_k\} = \coprod_{i=1}^{l} D_i^2. \tag{5-2}$$

The positive regions of a dividing set γ produced by the second step intersects the boundary of each such disk along intervals where occupied 1–handles are attached and the endpoints of the oriented line segment $\mathcal{Z}_i \times [0,1] \subset \partial D_i^2$.

Let us formalize the situation which we will simplify in the remainder of the proof. Suppose R is a positive region bounded by γ, and D_i is a disk from (5-2). Then R is *disconnected in* D_i if $R \cap \partial D_i \neq \varnothing$ and $(R \cap D_i) \cap \mathcal{Z}_i \times [0,1] = \varnothing$. A region R is *disconnected* if R is disconnected in D_i for some disk D_i in (5-2).

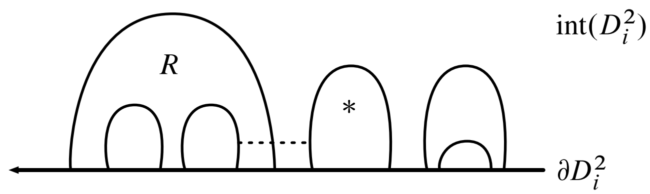
A dividing set γ is elementary if and only if there is one positive region in each disk. So in order to express γ produced by step two in terms of elementary dividing sets, we must reduce the number of disconnected regions. (This is just a version of Theorem 4.14 with the boundary components $\mathcal{Z}_i \subset \partial\Sigma$ treated separately.)

Let $R_1, \ldots, R_N$ be the positive regions of γ which are disconnected. Our complexity function is

$$n(\gamma) := \sum_{i=1}^{N} \sum_{j=1}^{l} |\pi_0(R_i \cap \partial D_j)| \in \mathbb{Z}_{\geq 0},$$

the total number of 1–handles occupied by the disconnected regions. Notice that if $N > 0$ then there exists an R such that $R \cap \partial D_i \neq \varnothing$ and so $n(\gamma) > 0$. On the other hand, if $n(\gamma) = 0$ then there are no disconnected regions and $N = 0$.

We claim that any γ which satisfies (5-1) with $n(\gamma) > 0$ can be expressed as a twisted complex in dividing sets γ' which satisfy $n(\gamma') = 0$. Suppose $n(\gamma) > 0$. Then there is a disk D_i which contains a disconnected region. Let $*$ be the positive region which contains $\mathcal{Z}_i \times [0,1] \subset D_i$. Now follow the orientation around D_i to the region R disconnected in D_i which is adjacent to $*$ and consider the bypass move illustrated below:



This results in a triangle $\gamma \to \gamma' \to \gamma''$ for which $n(\gamma'), n(\gamma'') < n(\gamma)$.

Lastly, our dividing sets may still contain some positive regions which do not intersect the boundary of any disk. Such regions can be removed with Theorem 4.14. □

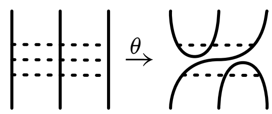
Corollary 5.12 *When a pointed oriented surface Σ is parametrized by an arc diagram $\mathcal{Z}$, the positive half of the formal contact category $\mathcal{K}\mathrm{o}_+(\Sigma)$ is generated by the positive elementary dividing sets $\mathfrak{Z}_+(\mathcal{Z})$.*

5.4 Decategorification

In this section, we prove a variety of structural properties and conjecture a decategorification statement for the formal contact category.

Proposition 5.13 *A single bypass $\theta\colon \gamma \to \gamma$ which takes γ to γ is capped.*

Proof One can make a small perturbation a (or b) above (or below) the equator ℓ of the bypass θ, as pictured on the left-hand side below. The bypasses associated to a (or b) are isotopic to θ.



Now by assumption the right-hand side, or the result of performing θ, is isotopic to the left-hand side. This isotopy takes the caps pictured on the right-hand side to caps of the bypasses on the left-hand side, so a and b are capped. But a and b arose as perturbations of θ, so θ is capped. □

Proposition 5.14 *Let Σ be a surface with boundary together with a parametrization $(\mathcal{Z}, \varphi_{\mathcal{Z}})$. There is a surjective map*

$$\epsilon\colon \mathbb{F}_2\langle \mathrm{Ob}(\mathcal{K}\mathrm{o}_+(\Sigma))\rangle \to \Lambda^* H_1(F(\mathcal{Z}), F(\partial\mathcal{Z}); \mathbb{F}_2),$$

where $F(\partial\mathcal{Z}) := \bigcup_i \mathcal{Z}_i \subset \partial F(\mathcal{Z})$. This map satisfies the following property: if

$$\gamma \to \gamma' \to \gamma''$$

is a bypass triangle then $\epsilon(\gamma'') = \epsilon(\gamma) + \epsilon(\gamma')$.

Proof A dividing set $\gamma \subset \Sigma$ determines a collection of positive regions: if $\Sigma \setminus \gamma = \bigsqcup_{i\in I} R_i$ then the set of positive regions is given by $\mathcal{R} := \{i \in I : R_i \text{ is positive}\}$. For each such region $R \in \mathcal{R}$, let $\partial_+ R := \partial R \cap F(\partial\mathcal{Z})$; the pair $(R, \partial_+ R)$ gives an inclusion

$$i_R\colon (R, \partial_+ R) \to (F(\mathcal{Z}), F(\partial\mathcal{Z})).$$

Let $n_R := \dim H_1(R, \partial_+ R; \mathbb{F}_2)$, so that $\Lambda^{n_R} H_1(R, \partial_+ R; \mathbb{F}_2)$ is 1–dimensional and there is a unique choice of nonzero vector $v_R \in \Lambda^{n_R} H_1(R, \partial_+ R; \mathbb{F}_2)$. Now tensoring gives a map

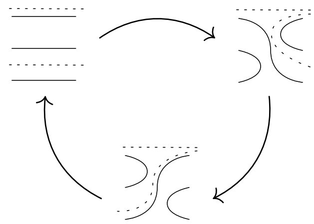
$$\hat{\imath}\colon \bigotimes_{R\in\mathcal{R}} \Lambda^{n_R} H_1(R, \partial_+ R; \mathbb{F}_2) \xrightarrow{\bar{\imath}} \bigotimes_{R\in\mathcal{R}} \Lambda^{n_R} H_1(F(\mathcal{Z}), F(\partial\mathcal{Z}); \mathbb{F}_2) \hookrightarrow \Lambda^* H_1(F(\mathcal{Z}), F(\partial\mathcal{Z}); \mathbb{F}_2),$$

where $\bar{\imath} := \bigotimes_{R\in\mathcal{R}} \wedge^{n_R} (i_R)_*$ and the last map is a composition of wedge products. The map ϵ is defined as

$$\epsilon(\gamma) := \hat{\imath}\Big(\bigwedge_{R\in\mathcal{R}} v_R\Big).$$

The 1–handles in $F(\mathcal{Z})$ span $H_1(F(\mathcal{Z}), F(\partial\mathcal{Z}); \mathbb{F}_2)$. If C corresponds to a subset of 1–handles, then by construction $\epsilon(\mathfrak{z}_C)$ is the wedge product of these classes in $\Lambda^* H_1(F(\mathcal{Z}), F(\partial\mathcal{Z}); \mathbb{F}_2)$. Since wedge products of 1–handles span the exterior algebra, ϵ is onto.

Additivity of ϵ can be observed by examining how the bypass moves affect elements in the first homology.



In the picture above, the dashed arcs represent (local) choices of generators in a positive region. If the $\epsilon(\gamma) = A \wedge C$ and $\epsilon(\gamma') = B \wedge C$ are the wedge products of arcs depicted on the left and right, respectively, then $\epsilon(\gamma'') = (A + B) \wedge C$. The other possible cases are handled similarly. □

Corollary 5.15 *For any bypass $\theta\colon \mathfrak{z}_C \to \mathfrak{z}_{C'}$ between elementary dividing sets, the third dividing set γ'' in the associated bypass triangle,*

$$\mathfrak{z}_C \xrightarrow{\theta} \mathfrak{z}_{C'} \to \gamma'', \tag{5-3}$$

is not an elementary dividing set.

Proof As above, elementary dividing sets $\mathfrak{z}_C$ determine basis vectors for $\Lambda^* H_1(\Sigma, \partial\Sigma; \mathbb{F}_2)$ in a canonical way. Since $\epsilon(\gamma'')$ in (5-3) must be a sum of the vectors determined by $\mathfrak{z}_C$ and $\mathfrak{z}_{C'}$ in this correspondence, it cannot be an elementary generator. □

Conjecture 5.16 *For any parametrization $\mathcal{Z}$ of Σ, there is a map $\bar{\epsilon}$, induced by ϵ, which is an isomorphism, as in the following diagram:*

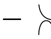
$$\begin{array}{ccc} \mathbb{F}_2\langle \mathrm{Ob}(\mathcal{K}\mathrm{o}_+(\Sigma))\rangle & & \\ \Big\downarrow \pi & \searrow^{\epsilon} & \\ K_0(\mathcal{K}\mathrm{o}_+(\Sigma)) & \xrightarrow{\bar{\epsilon}} & \Lambda^* H_1(F(\mathcal{Z}), F(\partial\mathcal{Z}); \mathbb{F}_2) \end{array}$$

Here π is the quotient map found in the definition of K_0.

Relation to work of J Murakami and O Viro The representation theory of the quantum group $U_q(\mathfrak{sl}_2)$ at $q^4 = 1$ determines a degenerate instance of the Chern–Simons topological field theory that has been related to the Alexander polynomial [40; 58]. The Jones–Wenzl projector $p_3 \in \mathrm{End}_{U_q(\mathfrak{sl}_2)}(V^{\otimes 3})$ takes the form

$$p_3 = ||| - \frac{d}{d^2-1}(\asymp| + |\asymp) + \frac{1}{d^2-1}(\asymp\!\asymp + \asymp\!\asymp),$$

where $d = q + q^{-1}$. Taking $q = \sqrt{-1}$ gives $d = 0$ and $d^2 - 1 = -1$. This eliminates the middle term above, leaving the bypass triangle

$p_3 = |||$ −  − .

Since the right-hand side should be zero, there is only a relationship between the contact geometry and representation theory after reducing by the Goodman–Wenzl ideal $\langle p_3 \rangle$ [14].

6 Comparison between categories associated to disks

In this section, we show that the categories associated to the disk $(D^2, 2n)$ with $2n$ points by the Heegaard–Floer theory $\mathcal{A}(D^2, 2n)$, the contact topology $\mathcal{C}o(D^2, 2n)$ and the formal contact construction, are Morita equivalent:

$$\mathcal{A}(D^2, 2n) \cong \mathcal{C}o(D^2, 2n) \cong \mathcal{K}o_+(D^2, 2n).$$

This is accomplished by choosing an arc parametrization $\mathcal{M}_n$ of the disk $(D^2, 2n)$ so that the associated Heegaard–Floer category $\mathcal{A}(D^2, 2n) \cong \mathcal{A}(-\mathcal{M}_n)$ has the same quiver presentation as the algebraic contact category $\mathcal{C}o(D^2, 2n) \cong \mathcal{Y}_n$ studied by Tian. This equivalence is combined with Theorem 5.11 to show that both categories are Morita equivalent to the positive half of the formal contact category $\mathcal{K}o_+(D^2, 2n)$. In this section $n \geq 2$.

6.1 The Heegaard–Floer categories associated to a disk

In this section, an arc diagram $\mathcal{M}_n$ and an arc parametrization of the disk $(D^2, 2n)$ with $2n$ marked points by $\mathcal{M}_n$ are introduced. The bordered sutured Floer theory developed by Zarev associates a dg category $\mathcal{A}(\mathcal{M}_n)$ to this parametrization. In Section 6.3, we will find that this category is the same as Tian's quiver algebra $\mathcal{R}_n$.

The disk will be oriented in the opposite direction of later sections. In this way the boundary of the disk is oriented clockwise. When viewed from above, as in the illustration below, each interval $\mathcal{Z}_i \subset \partial D$ has a well-defined left direction (counterclockwise) and right direction (clockwise). This terminology is used by the definition below.

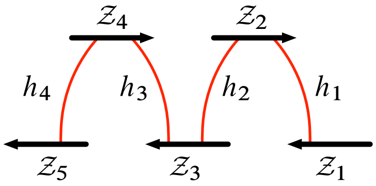
Definition 6.1 The *zigzag arc diagram* $\mathcal{M}_n$ is defined inductively as follows:

(1) The arc diagram $\mathcal{M}_2$ consists of two lines $Z = \{\mathcal{Z}_1, \mathcal{Z}_2\}$ and two points $\boldsymbol{a} = \{a_1, a_1'\}$, where $a_1 \in \mathcal{Z}_1$, $a_1' \in \mathcal{Z}_2$ and $M(a_1) = M(a_1')$.

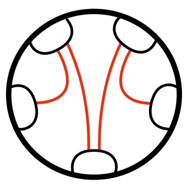
(2) If n is odd then $\mathcal{M}_n$ is obtained from $\mathcal{M}_{n-1}$ by adding a new line $\mathcal{Z}_n$, containing the point a_{n-1}, to the right of the line $\mathcal{Z}_{n-2}$ and adding the point a_{n-1}' to the line $\mathcal{Z}_{n-1}$ immediately to the left of a_{n-2}'.

(3) If n is even then $\mathcal{M}_n$ is obtained from $\mathcal{M}_{n-1}$ by adding a new line $\mathcal{Z}_n$, containing the point a_{n-1}', to the left of $\mathcal{Z}_{n-2}$ and then adding the point a_{n-1} to $\mathcal{Z}_{n-1}$ to the right of the point a_{n-2}.

If we imagine the line segments $\{\mathcal{Z}_i\}_{i=1}^n$ to be embedded sequentially along the real line $\mathbb{R}$, then an orientation on each line segment is induced by choosing an orientation of $\mathbb{R}$; they all point either to the left or to the right. The name zigzag becomes clear after rearranging the line segments into a zigzag pattern:



The arc diagram for $\mathcal{M}_5$ is pictured above. The line labeled $\mathcal{Z}_i$ is the i^{th} line segment in the construction from Definition 6.1. The lines h_i connect the matched pairs $\{a_i, a_i'\}$. If the illustration above is understood to specify an embedding of the arc diagram into the plane, then thickening each of the components produces the parametrization of the disk $(D^2, 2\cdot 5)$ with 10 points, pictured below:



Giving the plane the standard $\langle x, y\rangle$ orientation induces an orientation on $(D^2, 2n)$ in which the boundary is oriented clockwise.

The proposition below may be clear to readers who are more familiar with the algebras involved.

Proposition 6.2 *The dg category $\mathcal{A}(-\mathcal{M}_n)$ has trivial differential $d = 0$.*

Proof This follows from the definition of the differential. In more detail, by construction, as an algebra with idempotents, the dg category $\mathcal{A}(\mathcal{M}_n)$ is a subalgebra of a tensor product of copies of strands algebras $\mathcal{A}(1)$ and $\mathcal{A}(2)$. Neither of these algebras have differentials. Any tensor product of algebras without differentials does not have a differential. Any subalgebra of an algebra without differential does not have a differential; see also [62, Proposition 9.2]. □

Without a differential, the dg category $\mathcal{A}(-\mathcal{M}_n)$ is a category. The definition below comes from [62, Section 2.3]. It is summarized in Definition 6.3.

First note that the idempotents in this construction correspond to the objects of the category $\mathcal{A}(-\mathcal{M}_n)$. The idempotents are indexed by a choice of a subset

$$S \subset \{h_1, \dots, h_{n-1}\}$$

of the 1–handles which identify matched pairs in the arc diagram $\mathcal{M}_n$ [62, Definition 2.5].

In the definition of $\mathcal{M}_n$ above, there are n line segments $\{\mathcal{Z}_1, \dots, \mathcal{Z}_n\}$. On the segment $\mathcal{Z}_1$, there is only one point a_1. If n is even then $\mathcal{Z}_n$ contains only one point a'_{n-1}. If n is odd then $\mathcal{Z}_n$ contains only the point a_{n-1}. The line segment $\mathcal{Z}_k \in \{\mathcal{Z}_2, \dots, \mathcal{Z}_{n-1}\}$ contains the two points

$$a'_k a'_{k-1} \text{ for } k \text{ even} \quad \text{or} \quad a_k a_{k+1} \text{ for } k \text{ odd.} \tag{6-1}$$

Since the algebra $\mathcal{A}(1)$ only contains the identity element, the nonidentity elements in the parts of $\mathcal{A}(-\mathcal{M}_n) \subset \mathcal{A}(1) \otimes \mathcal{A}(2)^{\otimes n-2} \otimes \mathcal{A}(1)$ correspond to the $\mathcal{A}(2)$–tensor factors. Each such factor contains a Reeb chord $\rho_{k,k+1}$ or $\rho_{k+1,k}$. If the line segment contains the points $a'_{k+1}a'_k$ then the Reeb chord $\rho_{k,k+1}$ connects $\rho^-_{k,k+1} = a'_k$ to $\rho^+_{k+1,k} = a'_{k+1}$. If the line segment contains the points $a_k a_{k+1}$ then the Reeb chord $\rho_{k+1,k}$ connects $\rho^-_{k+1,k} = a_{k+1}$ to $\rho^+_{k+1,k} = a_k$. Since the k^{th} 1–handle h_k corresponds to the matching of the pair a_k and a'_k, the Reeb chords $\rho_{k,k+1}$ and $\rho_{k+1,k}$ correspond to maps

$$\rho_{k,k+1}\colon h_k \to h_{k+1} \quad \text{and} \quad \rho_{k+1,k}\colon h_{k+1} \to h_k \tag{6-2}$$

Translating (6-1) into the language of (6-2) tells us when such maps can be found in the category $\mathcal{A}(-\mathcal{M}_n)$. If n is even then there are maps

$$h_{n-1} \xrightarrow{\rho_{n-1,n-2}} h_{n-2} \xleftarrow{\rho_{n-3,n-2}} h_{n-3} \to \cdots \leftarrow h_3 \xrightarrow{\rho_{3,2}} h_2 \xleftarrow{\rho_{1,2}} h_1$$

and if n is odd then there are maps

$$h_{n-1} \xleftarrow{\rho_{n-2,n-1}} h_{n-2} \xrightarrow{\rho_{n-2,n-3}} h_{n-3} \leftarrow \cdots \leftarrow h_3 \xrightarrow{\rho_{3,2}} h_2 \xleftarrow{\rho_{1,2}} h_1.$$

Increasing the number n by one has the effect of adding one new Reeb chord.

The generators of the full category $\mathcal{A}(-\mathcal{M}_n)$ are obtained by extending each Reeb chord by identity in all possible ways [62, Definition 2.9]. In more detail, if $S = h_{i_1} h_{i_2} \cdots h_{i_j} \cdots h_{i_{k-1}} h_{i_k}$ is a subset of 1–handles which have been ordered so that $i_j < i_{j+1}$, then there is a generator

$$h_{i_1} h_{i_2} \cdots h_{i_j} \cdots h_{i_{k-1}} h_{i_k} \to h_{i_1} h_{i_2} \cdots h_{i_j \pm 1} \cdots h_{i_{k-1}} h_{i_k} \tag{6-3}$$

in $\mathcal{A}(-\mathcal{M}_n)$ when there is a Reeb chord $\rho_{i_j, i_j \pm 1}\colon h_{i_j} \to h_{i_j \pm 1}$ as above and the 1–handle $h_{i_{j \pm 1}}$ isn't contained in set S:

$$i_{j \pm 1} \notin \{i_1, i_2, \dots, i_k\}.$$

None of the relations satisfied by the strands algebras apply in our context because the Reeb chords are contained in independent strands algebras $\mathcal{A}(2)$ of order two. There is only one relevant family of relations, stemming from the observation that maps applied to independent tensor factors commute:

$$\begin{array}{ccccc} & & \cdots h_{i_j \pm 1} \cdots h_{i_l} \cdots & & \\ & \nearrow & & \searrow & \\ \cdots h_{i_j} \cdots h_{i_l} \cdots & & & & \cdots h_{i_j \pm 1} \cdots h_{i_l \pm 1} \cdots \\ & \searrow & & \nearrow & \\ & & \cdots h_{i_j} \cdots h_{i_l \pm 1} \cdots & & \end{array} \tag{6-4}$$

Said differently, whenever generators can be applied out-of-order to form a square, as pictured above, this square must commute.

The definition below summarizes the discussion above.

Definition 6.3 $\mathcal{A}(-\mathcal{M}_n)$ is the dg category with $d = 0$. The objects

$$\mathrm{Ob}(\mathcal{A}(-\mathcal{M}_n) = \{S : S \subset \{h_1, \dots, h_{n-1}\}\}$$

are subsets of the set of arcs in Definition 6.1. We write $S = \prod_{h_{i_k} \in S} h_{i_k}$ for any $S \in \mathrm{Ob}(\mathcal{A}(-\mathcal{M}_n))$. The category $\mathcal{A}(-\mathcal{M}_n)$ is generated by maps of the form (6-3) subject to relations in (6-4).

The examples below will be compared to Examples 6.9 and 6.10 in Section 6.3.

Example 6.4 The structure of $\mathcal{A}(-\mathcal{M}_3)$ can be pictured in the following way:

$$\varnothing \qquad h_1 \xrightarrow{\rho_{1,2}} h_2 \qquad h_1 h_2$$

Example 6.5 The structure of $\mathcal{A}(-\mathcal{M}_4)$ is illustrated by the diagram below:

$$\varnothing \qquad \begin{array}{ccc} & & h_2h_3 \\ & \nearrow^{\rho_{1,2}} & \\ h_1h_3 & & \\ & \searrow_{\rho_{3,2}} & \\ & & h_1h_2 \end{array} \qquad \begin{array}{ccc} h_1 & & \\ & \searrow^{\rho_{1,2}} & \\ & & h_2 \\ & \nearrow_{\rho_{3,2}} & \\ h_3 & & \end{array} \qquad h_1h_2h_3$$

Remark 6.6 Bordered sutured theory usually associates different algebras to different parametrizations of a surface. The categories of modules associated to these algebras are equivalent. In this sense, the algebras associated to surfaces are Morita equivalent; see the appendix. In order to understand why this is the case, consider that the mapping cylinder 3–manifolds associated to a diffeomorphism between parametrizations and its inverse determine a pair of bimodules [62, Section 8]. Product with a bimodule determines a functor between modules over algebras. The composition of functors gives the bimodule associated to the identity, which is algebraically the identity [62, Section 8.6]. See also [63].

In particular, there is an arc parametrization $\mathcal{W}_n$ [62, Example 9.1] for which there is an isomorphism of dg categories $\mathcal{A}(\mathcal{W}_n) \cong \mathcal{A}(n-1)^{\mathrm{op}}$ [62, Proposition 9.1], where $\mathcal{A}(n-1)$ is the strands algebra [36, Section 3.1]. Therefore $\mathcal{A}(-\mathcal{M}_n) \cong \mathcal{A}(n-1)^{\mathrm{op}}$ in Hmo.

6.2 The contact category associated to a disk

Here we introduce the category $\mathcal{Y}_n$ that Tian associates to the disk with $2n$ boundary points [54]. We will not discuss gradings.

6.2.1 Indexing multicurves with nil-Temperley–Lieb notation Monomials in the nil-Temperley–Lieb algebra will be used to denote multicurves $\gamma \subset (D^2, 2n)$ in the disk. In particular, multicurves determined by monomials $e_{i_1} e_{i_2} \cdots e_{i_k}$, which have been ordered so as to satisfy $i_1 < i_2 < \cdots < i_k$, correspond to the objects in Tian's construction; see Definition 6.8.

Definition 6.7 The *nil-Temperley–Lieb algebra* $\mathcal{N}_n$ is the k–algebra on generators e_i for $1 \leq i < n$, subject to the relations

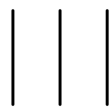
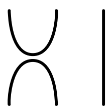
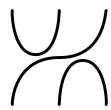
(1) $e_i^2 = 0$ for $1 \leq i < n$,
(2) $e_i e_j = e_j e_i$ for $|i-j| > 2$, and
(3) $e_i e_{i\pm 1} e_i = e_i$.

If the ground ring k is changed to $\mathbb{Z}[q, q^{-1}]$ and the first relation is changed from $e_i^2 = 0$ to $e_i^2 = q + q^{-1}$, then the algebra $\mathcal{N}_n$ introduced above becomes the well-known Temperley–Lieb algebra $\mathcal{TL}_n$; see [24].

The relationship between the Temperley–Lieb algebra and the planar algebra of multicurves extends to the nil-variant $\mathcal{N}_n$ introduced above. There is a basis for the algebra $\mathcal{N}_n$ consisting of monomials which is in one-to-one correspondence with isotopy classes of boundary-connected multicurves in a pointed oriented disk $(D^2, 2n)$. This can be seen after each generator e_i is identified with a multicurve $\gamma(e_i)$,

$$e_i \mapsto \gamma(e_i).$$

If the disk is pictured so that the first n points are situated on the top of the disk and the last n points are situated on the bottom of the disk, then all of the strands of $\gamma(e_i)$ are vertical except for two which connect the i^{th} and $(i+1)^{\text{st}}$ points in each collection. The products, $\gamma(e_i e_j) = \gamma(e_i)\gamma(e_j)$, of generators correspond to vertically stacking the multicurves. For instance, when $n = 3$ we have the following pictures:

$\gamma(1) =$  , $\gamma(e_1) =$  or $\gamma(e_1 e_2) =$  .

In the image of the map γ, the second and third relations in Definition 6.7 correspond to isotopy, and the first relation implies that any multicurve containing a homotopically trivial component is zero.

This observation can be used to construct a set map γ from the monomials of the nil-Temperley–Lieb algebra $\mathcal{N}_n$ to positive dividing sets on $(D^2, 2n)$. Since all of the defining relations for $\mathcal{N}_n$ preserve monomiality, the product of monomials is a monomial and each monomial $x \in \mathcal{N}_n$ corresponds to a multicurve $\gamma(x)$. After signing the regions of $D^2 \setminus \gamma(x)$, this determines a dividing set on the disk. Knowledge of the map γ is assumed throughout the next section.

6.2.2 Tian's disk category Tian's category $\mathcal{Y}_n$ is introduced by the sequence of definitions below. The construction presented here is equivalent to the original [54]. However, we will use the algebra $\mathcal{N}_n$ to express the presentation in more familiar notation.

Definition 6.8 The quiver $\mathcal{Q}_n$ has vertices $V := \{S : S = \{i_1 < i_2 < \cdots < i_k : 1 \leq i_j < n \text{ for } j = 1, \ldots, k\}\}$ and edges

$$E(S,T) := \begin{cases} \{\theta_p\} & \text{if } |T| = |S| + 2 \text{ and } T = S \cup \{p, p+1\}, \\ \varnothing & \text{otherwise.} \end{cases}$$

In more detail, the vertices S of the quiver $\mathcal{Q}_n$ are the ordered monomials

$$e_S = e_{i_1} e_{i_2} \cdots e_{i_k} \in \mathcal{N}_n \quad \text{where } S = \{i_1 < i_2 < \cdots < i_k\}$$

and $1 \le i_j < n$ for $j = 1, \dots, k$ in the nil-Temperley–Lieb algebra. There is an edge $\theta_p \colon e_S \to e_T$ from e_S to e_T when the set T can be obtained from the set S by adjoining the disjoint subset $\{p, p+1\}$.

Before introducing the category $\mathcal{Y}_n$ we illustrate this definition:

Example 6.9 When $n = 3$, the quiver $\mathcal{Q}_3$ assumes a rather unassuming form:

$$e_1 \qquad 1 \xrightarrow{\theta_1} e_1 e_2 \qquad e_2$$

Example 6.10 When $n = 4$, the quiver $\mathcal{Q}_4$ is more complicated:

$$\begin{array}{ccccccc} & & & e_1e_2 & e_1 & & \\ e_1e_3 & & 1 \begin{array}{c} \overset{\theta_1}{\nearrow} \\ \underset{\theta_2}{\searrow} \end{array} & & & \begin{array}{c} \overset{\theta_2}{\searrow} \\ \underset{\theta_1}{\nearrow} \end{array} e_1e_2e_3 & e_2 \\ & & & e_2e_3 & e_3 & & \end{array}$$

Each arrow $\theta_p \colon e_S \to e_T$ corresponds to a bypass move $\gamma(e_S) \to \gamma(e_T)$ between the multicurves $\gamma(e_S)$ and $\gamma(e_T)$, involving the p^{th} and $(p+1)^{\text{st}}$ regions in the disk; see (6-5).

The disk category $\mathcal{R}_n$ is the category generated by the graph $\mathcal{Q}_n$, modulo the relation that compositions of disjoint bypass moves commute.

Definition 6.11 The *disk category* $\mathcal{R}_n$ is the k–linear category generated by the graph $\mathcal{Q}_n$ subject to the relations

$$\theta_p \theta_q = \theta_q \theta_p \quad \text{for each pair of arrows } \theta_p\theta_q, \theta_q\theta_p \colon e_S \to e_T \text{ in } \mathcal{Q}_n.$$

The disk category $\mathcal{R}_n$ can be viewed as a dg category with $d = 0$. Recall the notion of a pretriangulated hull from Section 2.1.

Definition 6.12 The category $\mathcal{Y}_n$ associated to the disk $(D^2, 2n)$ is the pretriangulated hull of the disk category $\mathcal{R}_n$:

$$\mathcal{Y}_n = \mathcal{R}_n^{\text{pretr}}.$$

6.3 Relationship between the contact category and the Heegaard–Floer category

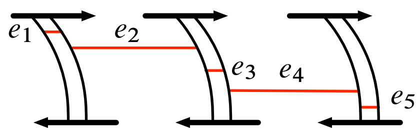
Here we show that the category $\mathcal{A}(-\mathcal{M}_n)$ found in Section 6.1 is isomorphic to Tian's disk category $\mathcal{R}_n$ from Section 6.2.2.

Theorem 6.13
$$\mathcal{R}_n \xrightarrow{\sim} \mathcal{A}(-\mathcal{M}_n).$$

Proof The similarities between Examples 6.4 and 6.9 and Examples 6.5 and 6.10 are suggestive. We will discuss the case when n is even; the case when n is odd is similar. We first give a bijective correspondence between the objects in either category. After this, the generators in either category are related to one another by representing each by geometric bypass moves.

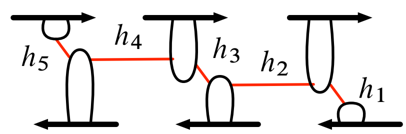
There is a one-to-one correspondence between the objects in each category. Recall that for $\mathcal{R}_n$, the objects are $\mathrm{Ob}(\mathcal{R}_n) = V(\mathcal{Q}_n) = \{e_S : S = \{i_1 < i_2 < \cdots < i_k\}\}$, which correspond to multicurves in the disk determined by the product $e_S = e_{i_1} \cdots e_{i_k}$ in the nil-Temperley–Lieb algebra. For $\mathcal{A}(-\mathcal{M}_n)$, the objects are $\mathrm{Ob}(\mathcal{A}(-\mathcal{M}_n)) = \{S : S \subset \{h_1, h_2, \ldots, h_{n-1}\}\}$, which correspond to a selection of 1–handles in the zigzag diagram. The maps in the next two paragraphs are constructed using these two topological interpretations for S.

First we construct a map $\Phi\colon \mathrm{Ob}(\mathcal{R}_n) \to \mathrm{Ob}(\mathcal{A}(-\mathcal{M}_n))$. In this correspondence, the identity diagram $1 \in \mathcal{N}_n$ corresponds to selecting all of the odd 1–handles, $\Phi(1) = h_1 h_3 \cdots h_{n-1}$. Suppose that $e_S = e_{i_1} e_{i_2} \cdots e_{i_k} \in \mathcal{N}_n$ is an ordered monomial. Then to construct the selection of 1–handles in $\mathrm{Ob}(\mathcal{A}(-\mathcal{M}_n))$ associated to e_S we perform surgery on this identity surface $h_1 h_3 \cdots h_{n-1}$ along the arcs pictured below for each e_{i_k} appearing in e_S:



After performing this surgery, there is a uniquely determined set $S \subset \{h_1, \ldots, h_{n-1}\}$ of 1–handles in the arc diagram $\mathcal{M}_n$ corresponding to this surface; this is the map from ordered monomials to subsets S of the set of 1–handles.

Now we construct an inverse map $\Psi\colon \mathrm{Ob}(\mathcal{A}(-\mathcal{M}_n)) \to \mathrm{Ob}(\mathcal{R}_n)$. The empty set of 1–handles $\varnothing$ corresponds to the product of the odd generators, $\Psi(\varnothing) = e_1 e_3 \cdots e_{n-1}$. If $h_{i_1} h_{i_2} \cdots h_{i_k}$ is an arbitrary selection of 1–handles, then gluing each 1–handle h_{i_j} into the picture below, in the indicated fashion, uniquely determines a multicurve associated to a positive monomial:



The maps introduced above are inverse. There is a bijection between the objects of either category. Observe that performing the odd e_i surgeries in the first illustration above produces the picture below it. From this observation the following two rules can be deduced:

(1) If i is odd then the effect of choosing or not choosing e_i corresponds to removing or adding h_{n-i}.

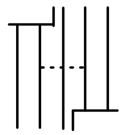
(2) If i is even then the effect of choosing or not choosing e_i corresponds to adding or removing h_{n-i}.

Here it is in algebraic notation:

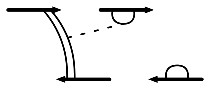
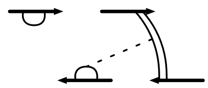
$$\Phi(e_{i_1} e_{i_2} \cdots e_{i_k}) = \{h_{n-s} : \exists r, s = i_r \text{ and } s \text{ even}\} \cup \{h_{n-s} : \forall r, s \neq i_r \text{ and } s \text{ odd}\},$$
$$\Psi(\{h_{i_1}, h_{i_2}, \ldots, h_{i_k}\}) = \{e_{n-s} : \exists r, s = i_r \text{ and } s \text{ even}\} \cup \{e_{n-s} : \forall r, s \neq i_r \text{ and } s \text{ odd}\}.$$

The variable r is restricted to the relevant subset of indices and the subscripts of a word $e_{i_1} e_{i_2} \cdots e_{i_k}$ are placed in order so as to coincide with conventions. These rules determine a bijection.

If $w, w' \in \mathcal{N}_n$ are ordered monomials, then an arrow $\theta_p\colon ww' \to we_pe_{p+1}w'$ in the graph $\mathcal{Q}_n$ corresponds to the bypass move $\theta_p\colon \gamma(ww') \to \gamma(we_pe_{p+1}w')$ pictured below:

(6-5) $\theta_p =$ 

For example, after a rotation the only arrow in the quiver $\mathcal{Q}_3$ corresponds to the bypass illustrated before Definition 3.12. Conversely, the basic Reeb chords $\rho_{k,k+1}\colon h_k \to h_{k+1}$ and $\rho_{k+2,k+1}\colon h_{k+2} \to h_{k+1}$ from Section 6.1 correspond to the pictures

(6-6)  and 

so that the two combinatorial notions perform the same function between multicurves in the correspondence between the objects.

There are no relations in either category besides the commutativity of (6-4) and Definition 6.11. □

6.4 Relationship between the disk category and the formal contact category

In this section, we will construct a Morita equivalence between the Heegaard–Floer category $\mathcal{A}(-\mathcal{M}_n)$ considered in Section 6.1 and the formal contact category $\mathcal{K}\mathrm{o}_+(D^2, 2n)$.

The discussion in prior sections suffices to define a functor

$$\mu\colon \mathcal{A}(-\mathcal{M}_n) \to \mathcal{K}\mathrm{o}_+(D^2, 2n).$$

To each collection of 1–handles $C = h_{i_1}h_{i_2}\cdots h_{i_k}$ we associate the elementary generator

$$\mathfrak{z}_C \in \mathrm{Ob}(\text{Pre–}\mathcal{K}\mathrm{o}_+(D^2, 2n)).$$

The basic Reeb chords correspond to the bypass moves pictured in (6-6). Composing this functor with the quotient map $Q\colon \text{Pre–}\mathcal{K}\mathrm{o}_+(D^2, 2n) \to \mathcal{K}\mathrm{o}_+(D^2, 2n)$ yields μ above.

Theorem 6.14 *The functor $\mu\colon \mathcal{A}(-\mathcal{M}_n) \to \mathcal{K}\mathrm{o}_+(D^2, 2n)$ determines a Morita equivalence.*

The proof of the theorem will use the fact that if $\mathcal{A}$ and $\mathcal{C}$ are small dg categories then $\mathcal{A}$ is Morita equivalent to $\mathcal{C}$ when $\mathcal{C}$ is quasiequivalent to a full dg subcategory $\mathcal{B}$ of the category of $\mathcal{A}$ whose objects form a set of small generators. This is a special case of a more general statement [25, Theorem 8.2].

Proof Using Theorem 5.11, it suffices to check that for each pair of collections of 1–handles C and C',

$$\mu_{C,C'}\colon \mathrm{Hom}_{\mathcal{A}(-\mathcal{M}_n)}(C, C') \to \mathrm{Hom}_{\mathcal{K}\mathrm{o}_+(D^2,2n)}(\mathfrak{z}_C, \mathfrak{z}_{C'})$$

is a quasi-isomorphism Since the trivial bypasses must bound caps and are removed by Definition 3.15(1), the only bypasses $\mathfrak{z}_C \to \mathfrak{z}'_C$ between elementary generators are those that appear in (6-6). These bypasses and their compositions are the cycles in $\text{Pre–}\mathcal{K}\mathrm{o}_+(D^2, 2n)$. It suffices to show that they remain cycles in the quotient.

The remainder follows from the commutativity of pushouts,

$$L_S L_{S'}\mathcal{C} \cong L_{S\amalg S'}\mathcal{C} \cong L_{S'} L_S\mathcal{C},$$

and the observation that the maps $Q_{C,C'}\colon \mathrm{Hom}_{\mathcal{C}}(C,C') \to \mathrm{Hom}_{L_S\mathcal{C}}(C,C')$ are quasi-isomorphisms for any single Postnikov localization. This can be seen by identifying a single Postnikov localization as an instance of Drinfeld localization under the Yoneda embedding; see Proposition 2.19. The Drinfeld localization modifies the homological structure of the morphisms by adding a single map h which is a boundary $dh = 1_K$, where K is as in the proof of Proposition 2.19. This makes any cycle to or from K into a boundary, but does not create any other boundaries. Since K is not an elementary generator $\mathfrak{z}_C$ for some C, the result follows. □

6.5 Dualities

Our discussion concludes with some mention of dualities. In Examples 6.9 and 6.10, duality is found in the lateral symmetry of the graph $\mathcal{Q}_n$. If $[n]$ denotes an ordered set $\{1 < 2 < \cdots < n\}$, then the assignment

$$e_S^y = e_{[n]\setminus S}$$

determines a contravariant involution:

$$-^y\colon \mathcal{Y}_n^{\mathrm{op}} \to \mathcal{Y}_n.$$

In $\mathcal{Y}_n$ there are no signed regions and the lateral symmetry is contravariant. So the functor $-^y$ cannot directly correspond to a functor, such as $-^\vee$, between formal contact categories. The proposition records the correct formulation. The proof is left to the reader.

Proposition 6.15 *The diagram*

$$\begin{array}{ccc} \mathcal{Y}_n^{\mathrm{op}} & \xrightarrow{\mathrm{i}_+^{\mathrm{op}}} & \mathcal{K}\mathrm{o}_+(D^2,2n)^{\mathrm{op}} \\ {\scriptstyle -^y}\downarrow & & \downarrow{\scriptstyle \alpha} \\ \mathcal{Y}_n & \xrightarrow{\mathrm{i}_+} & \mathcal{K}\mathrm{o}_+(D^2,2n) \end{array}$$

commutes, where the functor $\alpha = (-)^\vee \circ \overline{(-)} \circ (r^{-1})^{\mathrm{op}}$ is the composition of three equivalences: r is the element of the mapping class group which rotates the basepoint z by one region clockwise (Corollary 5.3), $\overline{(-)}$ reverses the orientation of the disk (Proposition 4.6) and $(-)^\vee$ changes the signs of the regions (Proposition 4.5).

7 Linear bordered Heegaard–Floer categories

Within the framework of the bordered Heegaard–Floer theory, a differential graded category $\mathcal{A}(\mathcal{Z})$ is associated to each arc diagram $\mathcal{Z}$. For some choices this category satisfies $d = 0$ and it is possible to write down a quiver presentation. In this section, these categories are related to the corresponding formal contact categories. We define functors

$$\mathcal{A}(-\mathcal{Z}_{0,n}, 1-n) \xrightarrow{\sigma_n} \mathrm{Ho}(\mathcal{K}\mathrm{o}_+^{2n-4}(\Sigma_{0,n}, n\cdot 2)) \quad\text{and}\quad \mathcal{A}(-\mathcal{Z}_{g,1}, 2g-1) \xrightarrow{\tau_g} \mathrm{Ho}(\mathcal{K}\mathrm{o}_+^{2g-2}(\Sigma_{g,1}, 2)),$$

where $\mathcal{Z}_{0,n}$ and $\mathcal{Z}_{g,1}$ are arc diagrams which parametrize surfaces $\Sigma_{0,n}$ and $\Sigma_{g,1}$ of genus zero with n boundary components, and of genus g with one boundary component, respectively. We fix two points on every boundary component and require that $n > 1$ and $g > 0$.

The bordered algebras studied in this section are the "one moving strand" algebras corresponding to the second-largest weight; see [62, Section 2], [34, Section 2] or [36, Section 3].

7.1 A surface $\Sigma_{0,n}$ of genus 0 with several boundary components

When n disks are removed from the 2–sphere,

$$\Sigma_{0,n} = S^2 \setminus \coprod_{i=1}^{n} D^2 \quad \text{for } n > 1,$$

and two points are fixed on each of its boundary components, the resulting surface can be parametrized by the arc diagram $\mathcal{Z}_{0,n}$ found in the definition below.

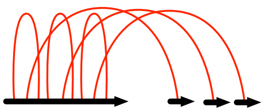
Definition 7.1 The arc diagram $\mathcal{Z}_{0,n}$ consists of n oriented line segments $Z = \{\mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_n\}$. On the first line segment $\mathcal{Z}_1$ there are $3n-3$ points, and there is one point on each of the remaining line segments $\{\mathcal{Z}_2 \dots \mathcal{Z}_n\}$:

$$\mathcal{Z}_1 = a_1 b_1 a_1' a_2 b_2 a_2' \cdots a_{n-1} b_{n-1} a_{n-1}' \qquad \text{and} \qquad \mathcal{Z}_i = b_{i-1}' \quad \text{for } 2 \le i \le n.$$

The set of points is given by $\boldsymbol{a} = \{a_i, a_i', b_i, b_i' : 1 \le i < n\}$. The line $\mathcal{Z}_1$ is oriented so that the subscripts of the points increase in value. The matching function is determined by the rules $M(a_i) = M(a_i')$ and $M(b_i) = M(b_i')$.

The annulus $\Sigma_{0,2}$ and its parametrization by $\mathcal{Z}_{0,2}$ are pictured in Example 5.9.

Example 7.2 When $n = 4$, the definition above is illustrated by the picture below:

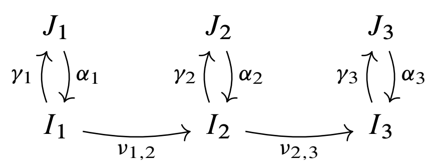


Definition 7.3 The category $\mathcal{A}(-\mathcal{Z}_{0,n}, 1-n)$ associated to the arc diagram $\mathcal{Z}_{0,n}$ is the k–linear category determined by a quiver with vertices I_i and J_i corresponding to the pairs $\{a_i, a_i'\}$ and $\{b_i, b_i'\}$ for $1 \le i < n$, respectively. There are arrows $\alpha_i : I_i \to J_i$, $\gamma_i : J_i \to I_i$ and $\nu_{i,i+1} : I_i \to I_{i+1}$ subject to the relations

(1) $\alpha_i \gamma_i = 0 \colon J_i \to J_i$, and

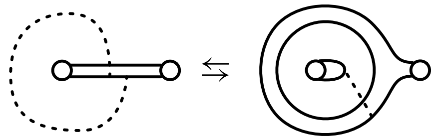
(2) $\nu_{i+1,i+2} \nu_{i,i+1} = 0 \colon I_i \to I_{i+2}$.

Example 7.4 We illustrate quiver underlying the category $\mathcal{A}(-\mathcal{Z}_{0,4}, -2)$ in the definition above:



The construction of the functor $\sigma_n\colon \mathcal{A}(\mathcal{Z}_{0,n}, 1-n) \to \mathrm{Ho}(\mathcal{K}\mathrm{o}_+^{2n-4}(\Sigma_{0,n}, n\cdot 2))$ will occur in two stages. First note that the parametrization of $\Sigma_{0,n}$ by the arc diagram allows us to associate to each object I_i or J_i for $1 \le i < n$ a dividing set contained in an annulus. In fact, Theorem 5.11 states that these dividing sets generate the contact category. In each annulus we will describe bypass moves corresponding to the arrows $\alpha_i\colon I_i \to J_i$ and $\gamma_i\colon J_i \to I_i$. We will check that these bypass moves satisfy the first collection of relations in the definition above. After this has been done, bypass moves corresponding to the lateral arrows $\nu_{i,i+1}\colon I_i \to I_{i+1}$ will be introduced and shown to satisfy the second collection of relations.

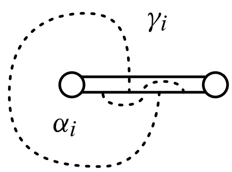
Step 1 For each annulus, the dividing sets J_i and I_i, and the bypass moves corresponding to the maps $\gamma_i\colon J_i \to I_i$ and $\alpha_i\colon I_i \to J_i$ can be depicted by the curves



The dividing set associated to J_i is featured on the left-hand side and the dividing set associated to I_i is shown on the right-hand side. The map γ_i runs from left to right and the map α_i runs from right to left. The equators of γ_i and α_i are determined by the dashed lines in the dividing sets corresponding to J_i and I_i, respectively.

Proposition 7.5 *The relation $\alpha_i\gamma_i = 0$ holds in the formal contact category* $\mathrm{Ho}(\mathcal{K}\mathrm{o}_+(S^1\times[0,1],(2,2)))$.

Proof The map $\alpha_i\gamma_i\colon J_i \to J_i$ is a composition of two disjoint bypass moves. This is illustrated below:



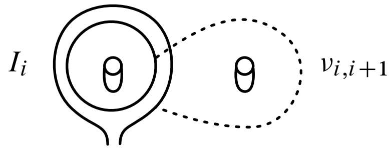
Relation (2) in the definition of the formal contact category (Definition 3.17) implies that applying the two bypass moves in either order must commute:

$$\begin{array}{ccc} J_i & \xrightarrow{\gamma_i} & I_i \\ \big\downarrow & & \big\downarrow{\scriptstyle \alpha_i} \\ 0 & \longrightarrow & J_i \end{array}$$

But performing the bypass move α_i before the bypass move γ_i must be zero since α_i is capped. $\square$

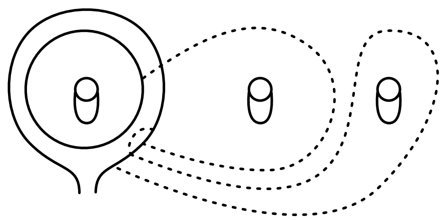
The same argument shows that one of the terms in the commutative diagram associated to the other composition $\gamma_i\alpha_i$ is a capped bypass equivalent to the identity.

Step 2 As pictured above, the idempotents I_i correspond to the boundaries of regular neighborhoods of loops about each boundary component of $\Sigma_{0,n}$. Here we think of $\Sigma_{0,n}$ as a subset of the plane $D^2 \setminus \coprod_{i=1}^{n-1} D^2 \subset \mathbb{R}^2$ with $n-1$ disks removed from its interior. The arc parametrization orders the boundary components and the associated idempotents. When two of them are adjacent, I_i and I_{i+1}, there is a bypass move $\nu_{i,i+1}\colon I_i \to I_{i+1}$ determined by the equator of the bypass disk in the illustration below:



Proposition 7.6 *In the formal contact category* $\mathrm{Ho}(\mathcal{K}\mathrm{o}_+(\Sigma_{0,n}, n\cdot 2))$, *the relation* $\nu_{i+1,i+2}\nu_{i,i+1} = 0$ *holds.*

Proof The proof is analogous to the proof of Proposition 7.5. The bypass moves representing $\nu_{i,i+1}$ and $\nu_{i+1,i+2}$ are disjoint. Considering them simultaneously produces the visual aid below:



The curve on the far right represents the equator of the bypass $\nu_{i+1,i+2}$. Since this bypass move is capped, the composition factors through zero. □

Tian's annulus As in Section 6, in Tian's work [55, Section 2.2] the category associated to an annulus with two points on each boundary component is the pretriangulated hull on the free k–linear category associated to a quiver with five vertices: I, E, F and EF. The dividing sets associated to E and F are Euler dual and are neither the source nor the target of any nontrivial edges. There are two dividing sets I and EF generating the subcategory with Euler number zero via maps $\gamma\colon I \to EF$ and $\alpha\colon EF \to I$ which are required to satisfy the relation

$$\alpha\gamma = 0.$$

This description is summarized by the illustration below:

$$F \qquad \gamma\colon I \rightleftarrows EF :\alpha \qquad E$$

The quiver in the center is precisely $\mathcal{A}(-\mathcal{Z}_{0,2}, 0)$ above.

Remark 7.7 It is natural to ask about surfaces $\Sigma_{0,n}$ with $n > 2$. There are presently two constructions in the literature. In [55], the category associated to $\Sigma_{0,n}$ is a bordered Heegaard–Floer category by definition. Precisely the same can be said for the categories considered by Petkova and V Vértesi [44]. While the former chooses an arc parametrization which yields a heart encoding contact geometry, the latter chooses an arc parametrization which yields an extension [60] of the strands algebra [36]. In both cases the arc parametrizations are *degenerate*, so the bordered Heegaard–Floer construction does not suffice to imply an equivalence between the two, and the materials here do not necessarily apply.

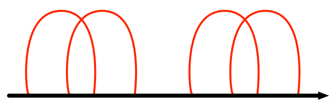
7.2 A surface $\Sigma_{g,1}$ of genus g with one boundary component

Definition 7.8 The arc diagram $\mathcal{Z}_{g,1}$ consists of $4g$ points $\boldsymbol{a} = \{a_i, a'_i, b_i, b'_i : 1 \le i \le g\}$ on one line segment $Z = \{\mathcal{Z}_1\}$,

$$\mathcal{Z}_1 = a_1 b_1 a'_1 b'_1 a_2 b_2 a'_2 b'_2 \cdots a_g b_g a'_g b'_g,$$

which is oriented so that the indices above are increasing. The matching function is determined by the rules $M(a_i) = M(a'_i)$ and $M(b_i) = M(b'_i)$ for $1 \le i \le g$.

Example 7.9 The arc diagram $\mathcal{Z}_{2,1}$ is illustrated below:



Definition 7.10 The category $\mathcal{A}(-\mathcal{Z}_{g,1}, 2g-1)$ associated to the arc diagram $\mathcal{Z}_{g,1}$ is the k–linear category determined by a quiver with vertices: I_i and J_i corresponding to the pairs $\{a_i, a'_i\}$ and $\{b_i, b'_i\}$ for $1 \le i \le g$, respectively. There are arrows

$$\alpha_i, \beta_i : I_i \to J_i, \quad \gamma_i : J_i \to I_i \quad \text{and} \quad \eta_{i,i+1} \colon I_i \to J_{i+1},$$

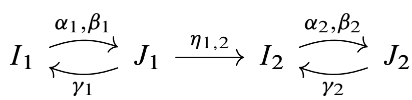
the compositions of which satisfy the relations

(1) $\alpha_i \gamma_i = 0 \colon J_i \to J_i$ and $\gamma_i \beta_i = 0 \colon I_i \to I_i$, and

(2) $\eta_{i,i+1} \alpha_i = 0 \colon I_i \to I_{i+1}$ and $\beta_{i+1} \eta_{i,i+1} = 0 \colon J_i \to J_{i+1}$.

Note that $\eta_{i,i+1} \colon I_i \to J_{i+1}$ is not the same as $\nu_{i,i+1} \colon I_i \to I_{i+1}$ in the previous section.

Example 7.11 The quiver underlying the construction of the category $\mathcal{A}(-\mathcal{Z}_{2,1}, 3)$ is illustrated below:

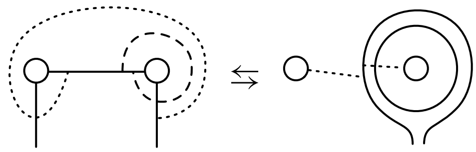


The construction of the functor $\tau_g : \mathcal{A}(-\mathcal{Z}_{g,1}, 2g-1) \to \mathcal{K}\mathrm{o}^{2g-2}(\Sigma_{g,1}, 2)$ will occur in two stages.

First note that the parametrization of $\Sigma_{g,1}$ by the arc diagram allows us to associate to each i, for $1 \le i \le g$, a pair of dividing sets I_i and J_i contained in a torus $\Sigma_{1,1} \subset \Sigma_{g,1}$ with one boundary component. In fact, Theorem 5.11 states that these dividing sets generate the category. In each torus, we will describe bypass moves corresponding to the arrows $\alpha_i, \beta_i : I_i \to J_i$ and $\gamma_i : J_i \to I_i$, and check that these bypass moves satisfy the first collection of relations in the definition above.

After this has been done, bypass moves corresponding to the lateral arrows $\eta_{i,i+1}$ will be introduced and shown to satisfy the second collection of relations.

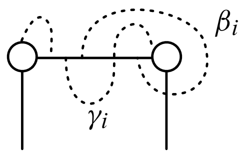
Step 1 For each torus, the dividing sets I_i and J_i, and the bypass moves corresponding to the maps $\alpha_i, \beta_i : I_i \to J_i$ and $\gamma_i : J_i \to I_i$, can be depicted by the following curves:



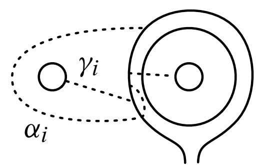
On either side of the arrows in the picture above, the two small circles are identified by folding the page to form the surface $(T^2 \setminus D^2, 2)$. The dividing set associated to I_i is featured on the left-hand side and the dividing set associated to J_i is featured on the right-hand side. The maps α_i and β_i run from left to right and the map γ_i runs from right to left. The equator of the map α_i is dotted and the equator of β_i is dashed.

Proposition 7.12 *In the formal contact category* $\mathrm{Ho}(\mathcal{K}o_+(\Sigma_{1,1}, 2))$, *the relations* $\beta_i \gamma_i = 0$ *and* $\gamma_i \alpha_i = 0$ *hold.*

Proof The logic is analogous to the proof of Proposition 7.5. The map $\beta_i \gamma_i$ is a composition of two disjoint bypass moves. When performed in the opposite order the bypass γ_i is capped, implying that the composition $\beta_i \gamma_i$ factors through zero. This is illustrated below:

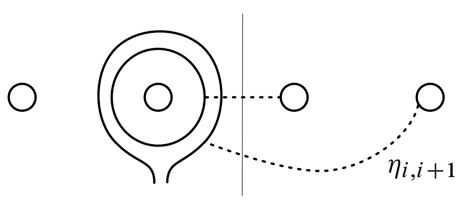


The map $\gamma_i \alpha_i$ is a composition of two disjoint bypass moves. When performed in the opposite order the bypass α_i is capped, implying that the composition $\beta_i \gamma_i$ factors through zero. This illustrated below:



□

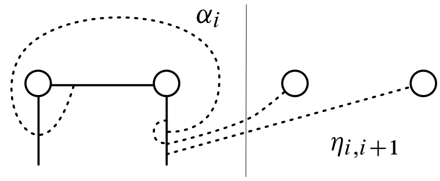
Step 2 As pictured above, the idempotents I_i correspond to the boundaries of regular neighborhoods of loops about the first 1–handle and the idempotents J_i to the boundaries of regular neighborhoods of loops about the second 1–handle in the i^{th} torus $\Sigma_{1,1} \subset \Sigma_{g,1}$. The tori $\Sigma_{1,1}$ are ordered by the arc parametrization and, when two tori are adjacent, there is a bypass move $\eta_{i,i+1} : J_i \to I_{i+1}$ from the dividing set about the second 1–handle of the first torus to the dividing set about the first 1–handle of the second torus. The map $\eta_{i,i+1}$ is pictured below:



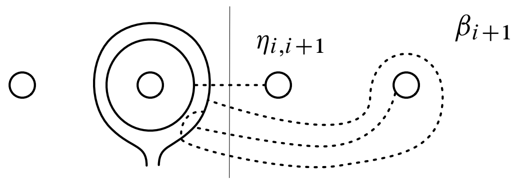
Here the first two and the second two smaller circles are connected by annuli $S^1 \times [0,1]$ to form the k^{th} and $(k+1)^{\text{st}}$ tori $\Sigma_{1,1} \subset \Sigma_{g,1}$.

Proposition 7.13 *The relations*: $\eta_{i,i+1}\alpha_i = 0\colon I_i \to I_{i+1}$ *and* $\beta_{i+1}\eta_{i,i+1} = 0\colon J_i \to J_{i+1}$ *hold in the formal contact category* $\mathrm{Ho}(\mathcal{K}\mathrm{o}_+(\Sigma_{g,1}, 2))$.

Proof The logic is analogous to the proof of Proposition 7.12. The map $\eta_{i,i+1}\alpha_i$ is a composition of two disjoint bypass moves. When performed in the opposite order the bypass $\eta_{i,i+1}$ is capped, implying that the composition factors through zero. This is illustrated below:



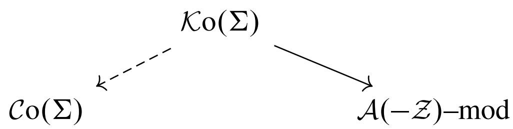
The map $\beta_{i+1}\eta_{i,i+1}$ is a composition of two disjoint bypass moves. When performed in the opposite order the bypass β_{i+1} is capped, implying that the composition factors through zero. This is illustrated below:



□

8 Comparison to geometric categories

One of the appealing qualities of the formal contact category $\mathcal{K}\mathrm{o}(\Sigma)$ is that it has a universal property with respect to other dg categories, by construction. Although there is no underlying Floer theory or contact geometry, this property allows us compare $\mathcal{K}\mathrm{o}(\Sigma)$ to other constructions which stem from observations involving either. In this section, we will discuss why the universal property of $\mathcal{K}\mathrm{o}(\Sigma)$ implies the existence of maps



in the homotopy category of dg categories which relate contact categories $\mathcal{C}\mathrm{o}(\Sigma)$ with the corresponding component of the bordered Heegaard–Floer theory. See Sections 8.1 and 8.2 for precise statements.

8.1 Relation to the contact category

Much of the material in this paper was inspired by Honda's proposed *contact category* $\mathcal{C}\mathrm{o}(\Sigma)$ [15]. Although a full account of this construction is in preparation, in this section a modest comparison is drawn between the formal and geometric contact categories.

The morphisms in the contact category $\mathcal{C}\mathrm{o}(\Sigma)$ are tight contact structures on $\Sigma \times [0,1]$. More precisely, $\mathcal{C}\mathrm{o}(\Sigma)$ is the additivization [38, Section 1.1.2.1] of a category with objects given by dividing sets γ on the surface Σ and morphisms $\theta\colon \gamma \to \gamma'$ given by contactomorphism classes of contact structures on $\Sigma \times [0,1]$, which induce γ and γ' on $\partial\Sigma \times [0,1]$, subject to the relation that an overtwisted contact structure is zero. The composition is induced by the pullback of contact plane fields along the rescaling diffeomorphism: $\Sigma \times [0,1] \xrightarrow{\sim} \Sigma \times [0,1] \cup_\Sigma \Sigma \times [0,1]$.

The contact category $\mathcal{C}\mathrm{o}(\Sigma)$ plainly exists. The maps in the contact category $\mathcal{C}\mathrm{o}(\Sigma)$ are generated by bypass moves between dividing sets [17, Lemma 3.10 (isotopy discretization)]. Since the bypass moves satisfy the elementary relations (1) and (2) in Definition 3.15, there is a functor: $\sigma\colon \text{Pre–}\mathcal{K}\mathrm{o}(\Sigma) \to \mathcal{C}\mathrm{o}(\Sigma)$. When (Σ, m) is a surface with boundary then the discussion in Section 4.3 suggests that these categories are very closely related.

For the purposes of comparison, we must make the nontrivial assumption below:

Assumption 8.1 *The contact category $\mathcal{C}\mathrm{o}(\Sigma)$ has pretriangulated dg enhancement $\mathcal{C}\mathrm{o}^{\mathrm{dg}}(\Sigma)$ in which bypass triangles are distinguished triangles.*

If this assumption is correct then there is a canonical lift

$$\tilde{\sigma}\colon \mathcal{K}\mathrm{o}(\Sigma) \to \mathcal{C}\mathrm{o}^{\mathrm{dg}}(\Sigma)$$

of the dg functor σ to a functor from the formal contact category to the dg category $\mathcal{C}\mathrm{o}^{\mathrm{dg}}(\Sigma)$.

Remark 8.2 In the formal contact category $\mathcal{K}\mathrm{o}(\Sigma)$, the bypass "an" involving the annulus in the proof of Theorem 4.14 determines a distinguished triangle

$$\gamma \xrightarrow{\mathrm{an}} \gamma' \xrightarrow{\mathrm{an}'} \gamma'' \xrightarrow{\mathrm{an}''} \gamma[1].$$

The map an' is not necessarily zero. However, in the geometric setting $\mathrm{an}' = 0$, making the convolution $\gamma \simeq C(\mathrm{an}')$ isomorphic to a direct sum (Tian, personal communication, 2015). As $\sigma(\mathrm{an}') = 0$, it is possible to view $\mathrm{Ho}(\mathcal{K}\mathrm{o}(\Sigma))$ as a deformation.

8.2 Relation to the bordered sutured Floer categories

In this section, we construct a functor $\widetilde{\mathcal{K}\mathrm{o}}_+(\Sigma, m) \to \mathcal{A}(-\mathcal{Z})$–mod from a cofibrant replacement of the positive part of the formal contact category to the category of left dg modules over an arc algebra of an arc diagram $\mathcal{Z}$ that parametrizes Σ. Assume that (Σ, m) has at least one boundary component, and every boundary component $\partial_i\Sigma$ contains a positive even number of points m_i. The ground ring k of $\mathcal{K}\mathrm{o}_+(\Sigma, m)$ is fixed to be the field $\mathbb{F}_2$. We will not discuss gradings here. The cofibrant replacement is a slightly larger but quasiequivalent category; see Conjecture 3.18. In particular, there is a functor $\mathcal{K}\mathrm{o}_+(\Sigma, m) \to \mathcal{A}(-\mathcal{Z})$–mod in Hqe.

If γ is a dividing set on Σ then Zarev associates a bordered sutured manifold [62, Section 3.2] called the *cap* W_γ to γ. The cap W_γ is the 3–manifold $\Sigma \times [0,1]$ in which the surface $\Sigma \times \{0\}$ is parametrized by the arc diagram $\mathcal{Z}$, the *sutures* m are the m boundary points, the dividing set γ appears on $\Sigma \times \{1\}$ and the two sides are connected by straight lines segments in $\partial\Sigma \times [0,1]$:

$$W_\gamma = (\Sigma \times [0,1], \gamma \times \{1\} \cup \Lambda \times [0,1], (-\Sigma \times \{0\}, -\Lambda \times \{0\})).$$

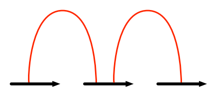
For details concerning this definition consult [63, Definition 2.5].

Associated to each bordered sutured manifold Y, there is a Heegaard diagram $H(Y)$ [62, Section 4]. Associated to each Heegaard diagram $H(Y)$, there is a left dg $\mathcal{A}(-\mathcal{Z})$–module $\widehat{\mathrm{BSD}}(Y)$ [62, Section 7.3]. Notation for the module does not include the intermediate Heegaard diagram because the homotopy type of the module is independent of this choice.

If γ is a dividing set on Σ such that the basepoint z_1 is contained in the positive region $R_+ \subset \Sigma \setminus \gamma$, then γ determines an object $\gamma \in \mathrm{Ob}(\mathcal{K}o_+(\Sigma, m))$. To each such γ we associate the left dg module $\widehat{\mathrm{BSD}}(\gamma) = \widehat{\mathrm{BSD}}(W_\gamma)$ associated to the cap for some choice of Heegaard diagram:

$$\gamma \mapsto \widehat{\mathrm{BSD}}(\gamma) \quad \text{where } \widehat{\mathrm{BSD}}(\gamma) = \widehat{\mathrm{BSD}}(W_\gamma). \tag{8-1}$$

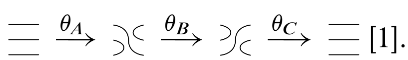
The disk $(D^2, 6)$ can be parametrized by an arc diagram $\mathcal{W}_3$ pictured below:



The diagram $\mathcal{W}_3$ consists of three oriented line segments $Z = \{\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3\}$ containing the points $\{a\}$, $\{a' < b\}$ and $\{b'\}$, respectively. The matching function M is determined by $M(a) = M(a')$ and $M(b) = M(b')$.

As discussed in Proposition 3.16, the three important dividing sets γ_A, γ_B and γ_C in $(D^2, 6)$ can be connected by three bypass moves

$$\gamma_A \xrightarrow{\theta_A} \gamma_B \xrightarrow{\theta_B} \gamma_C \xrightarrow{\theta_C} \gamma_C[1] \quad \text{or}$$



(The signs of the regions are fixed by requiring that the region containing the basepoint is positive.) Associated to these three dividing sets, there are three left $\mathcal{A}(-\mathcal{W}_3)$–modules $\widehat{\mathrm{BSD}}(\gamma_A)$, $\widehat{\mathrm{BSD}}(\gamma_B)$ and $\widehat{\mathrm{BSD}}(\gamma_C)$ corresponding to the bordered sutured diagrams given by the caps W_{γ_A}, W_{γ_B} and W_{γ_A}.

In [11, Section 6.2], the authors J B Etnyre, D S Vela-Vick and Zarev made a fundamental computation: after choosing Heegaard diagrams for the caps W_{γ_A}, W_{γ_B} and W_{γ_C}, they found that there are chain maps $\phi_A \colon \widehat{\mathrm{BSD}}(\gamma_A) \to \widehat{\mathrm{BSD}}(\gamma_B)$, $\phi_B \colon \widehat{\mathrm{BSD}}(\gamma_B) \to \widehat{\mathrm{BSD}}(\gamma_C)$ and $\phi_C \colon \widehat{\mathrm{BSD}}(\gamma_C) \to \widehat{\mathrm{BSD}}(\gamma_A)$ such that

$$\widehat{\mathrm{BSD}}(\gamma_A) \xrightarrow{\phi_A} \widehat{\mathrm{BSD}}(\gamma_B) \xrightarrow{\phi_B} \widehat{\mathrm{BSD}}(\gamma_C) \xrightarrow{\phi_C} \widehat{\mathrm{BSD}}(\gamma_A)[1]$$

is a distinguished triangle. They show explicitly that

(1) $\widehat{\mathrm{BSD}}(\gamma_A) = C(\phi_B)$,

(2) ϕ_A is projection, and

(3) ϕ_C is inclusion.

(Alternatively, this follows from Section 6 and the Morita invariance of the category associated to the disk by bordered sutured Floer theory.) Our functor is defined using the pairing theorem to extend the assignments $\theta_A \mapsto \phi_A$, $\theta_B \mapsto \phi_B$ and $\theta_C \mapsto \phi_C$ to all of the other bypass moves between dividing sets.

Throughout the remainder of this section we will make repeated use of the pairing theorem. Suppose that γ is a dividing set on Σ and the first basepoint z_1 is contained in a positive region. Then if $D = (D^2, 2m) \subset \Sigma$ is an embedded disk with $2m$ points on the boundary such that $\gamma^\circ = \gamma \setminus (D \cap \gamma)$ is a dividing set on $\Sigma \setminus D$, then the pairing theorem [62, Theorem 8.7] gives a homotopy equivalence

$$\widehat{\mathrm{BSD}}(\gamma) \xrightarrow{\sim} \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes \widehat{\mathrm{BSD}}(\gamma \cap D) \quad \text{where } \gamma = \gamma^\circ \cup_{\gamma \cap \partial D} (\gamma \cap D),$$

and where $\gamma \cap \partial D = 2m$, $\widehat{\mathrm{BSD}}(\gamma) = \widehat{\mathrm{BSD}}(W_\gamma)$ is the left dg $\mathcal{A}(-\mathcal{Z})$–module assigned to the dividing set γ, $\widehat{\mathrm{BSDA}}(\gamma^\circ) = \widehat{\mathrm{BSDA}}(W_{\gamma^\circ})$ is a left $\mathcal{A}(-\mathcal{Z})$–module and right A_∞ $\mathcal{A}(-\mathcal{W}_3)$–module, $\widehat{\mathrm{BSD}}(\gamma \cap D) = \widehat{\mathrm{BSD}}(W_{\gamma \cap D})$ is the left $\mathcal{A}(-\mathcal{W}_3)$–module determined by γ in the interior of the disk D, and the box product $\boxtimes$ is an analogue of the derived tensor product; see [36, Section 2.4].

Definition 8.3 If $\theta\colon \gamma \to \eta$ is a bypass move then the map $\theta_*\colon \widehat{\mathrm{BSD}}(\gamma) \to \widehat{\mathrm{BSD}}(\eta)$ of dg modules associated to θ is determined by the commutative diagram

$$\begin{array}{ccc} \widehat{\mathrm{BSD}}(\gamma) & \xrightarrow{\sim} & \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes \widehat{\mathrm{BSD}}(\gamma_A) \\ \Big\downarrow{\scriptstyle \theta_*} & & \Big\downarrow{\scriptstyle 1\boxtimes\phi_A} \\ \widehat{\mathrm{BSD}}(\eta) & \xrightarrow{\sim} & \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes \widehat{\mathrm{BSD}}(\gamma_B) \end{array}$$

where $\gamma^\circ = \gamma \setminus D$, introduced above, denotes the dividing set minus the region containing the equator of the bypass disk associated to θ.

In order for the maps chosen above to yield a functor from the preformal contact category, we must check that relations (1) and (2) in Definition 3.15 above are satisfied. Since these relations hold up to homotopy in the category $\mathcal{A}(-\mathcal{Z})$–mod, this determines a functor from the cofibrant replacement of the preformal contact category. Lastly we will show that this functor factors through the Postnikov localization introduced by Proposition 3.16.

Relation (1) If θ is capped in the northwest or southeast then relation, (1) must hold up to homotopy by the invariance of the bordered sutured theory [62, Section 7].

In more detail, suppose that $\theta\colon \gamma \to \eta$ is a bypass move and D is a neighborhood of the equator of the underlying bypass disk. Then when there is a cap, the region D can be enlarged to a region $\widetilde{D}$ which contains the cap disk in Σ. Two applications of the pairing theorem give

$$\begin{array}{ccccc} \widehat{\mathrm{BSD}}(\gamma) & \xrightarrow{\sim} & \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes \widehat{\mathrm{BSD}}(\gamma_A) & \xrightarrow{\sim} & \widehat{\mathrm{BSDA}}(\tilde{\gamma}^\circ) \boxtimes \widehat{\mathrm{BSD}}(\tilde{\gamma}_A) \\ \Big\downarrow{\scriptstyle \theta_*} & & \Big\downarrow{\scriptstyle 1\boxtimes\phi_A} & & \Big\downarrow{\scriptstyle 1\boxtimes\tilde{\phi}_A} \\ \widehat{\mathrm{BSD}}(\eta) & \xrightarrow{\sim} & \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes \widehat{\mathrm{BSD}}(\gamma_B) & \xrightarrow{\sim} & \widehat{\mathrm{BSDA}}(\tilde{\gamma}^\circ) \boxtimes \widehat{\mathrm{BSD}}(\tilde{\gamma}_B) \end{array}$$

where $\gamma^\circ = \gamma \setminus D$ and $\tilde{\gamma}^\circ = \gamma \setminus \widetilde{D}$. The dividing sets $\tilde{\gamma}_A$ and $\tilde{\gamma}_B$, on the right-hand side above, are identical when the cap is either northwestern or southeastern. They are both represented by the same Heegaard diagram and the map $\tilde{\phi}_A$ is the identity. It follows that θ_* is homotopic to the identity.

Relation (2) In order to see that disjoint bypass moves $\theta \amalg \theta'\colon \gamma \to \eta$ commute, we must cut the dividing set γ along the two disjointly embedded disks corresponding to neighborhoods of the equators of our bypass moves to form $\gamma^{\circ\circ} = \gamma \setminus (D \amalg D')$. The arc algebra associated to a disjoint union splits, $\varphi\colon \mathcal{A}(-(\mathcal{W}_3 \amalg \mathcal{W}_3)) \xrightarrow{\sim} \mathcal{A}(-\mathcal{W}_3) \otimes_k \mathcal{A}(-\mathcal{W}_3)$, the module $\widehat{\mathrm{BSD}}(\gamma_A) \otimes_k \widehat{\mathrm{BSD}}(\gamma_A)$ appears in the pairing theorem,

$$\widehat{\mathrm{BSD}}(\gamma) \xrightarrow{\sim} \widehat{\mathrm{BSDA}}(\gamma^{\circ\circ}) \boxtimes [\widehat{\mathrm{BSD}}(\gamma_A) \otimes_k \widehat{\mathrm{BSD}}(\gamma_A)],$$

and the disjoint union of Heegaard diagrams splits as a tensor product compatible with the isomorphism φ above. Under this identification, the maps θ_* and θ'_* induced by θ and θ' correspond to different tensor factors and must commute by the standard algebraic fact that

$$(1_{\gamma^{\circ\circ}} \boxtimes [1_A \otimes \theta'_*])(1_{\gamma^{\circ\circ}} \boxtimes [\theta_* \otimes 1_A]) = (1_{\gamma^{\circ\circ}} \boxtimes [\theta_* \otimes 1_A])(1_{\gamma^{\circ\circ}} \boxtimes [1_A \otimes \theta'_*]),$$

where $1_{\gamma^{\circ\circ}}$ and 1_A are used to denote the identity maps $1_{\widehat{\mathrm{BSDA}}(\gamma^{\circ\circ})}$ and $1_{\widehat{\mathrm{BSD}}(\gamma_A)}$, respectively.

Triangles Finally, it is necessary to see that the objects and the maps assigned by (8-1) and Definition 8.3 factor through the Postnikov localization constructed in Proposition 3.16.

These choices form distinguished triangles because

$$\widehat{\mathrm{BSD}}(\gamma) = \widehat{\mathrm{BSD}}(\gamma \cup \gamma_A) \simeq \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes \widehat{\mathrm{BSD}}(\gamma_A) \simeq \widehat{\mathrm{BSDA}}(\gamma^\circ) \boxtimes C(\phi_B) \simeq C(1_{\widehat{\mathrm{BSDA}}(\gamma^\circ)} \boxtimes \phi_B) \simeq C(\theta'_*),$$

where the last equivalence corresponds to the commutative diagram in Definition 8.3 after rotating the triangle. An analogue of this argument appears in [35, Theorem 4.1].

Appendix Dg categories

This section contains some materials about dg categories and the model structures. All of the definitions are from the literature. More information about differential graded categories can be found in [26; 57] or [10, Section 1]; consult [46; 47; 56] for technical details. The language of model categories is reviewed in [37, Section A.2]; more details can be found in [19; 45].

Definition A.1 A *dg category* $\mathcal{C}$ *over* $\mathcal{A}$ is a category enriched in the monoidal category of chain complexes,

$$\mathrm{Hom}_{\mathcal{C}}(x, y) \in \mathrm{Kom}_k(\mathcal{A}) \quad \text{for all } x, y \in \mathrm{Ob}(\mathcal{C}),$$

such that composition in $\mathcal{C}$ is a map in $\mathrm{Kom}_k(\mathcal{A})$. A *functor* $f : \mathcal{C} \to \mathcal{D}$ between two such dg categories is required to consist of maps in $\mathrm{Kom}_k(\mathcal{A})$:

$$f_{x,y} : \mathrm{Hom}_{\mathcal{C}}(x, y) \to \mathrm{Hom}_{\mathcal{D}}(f(x), f(y)) \in \mathrm{Kom}_k(\mathcal{A}). \tag{A-2}$$

A dg functor $f : \mathcal{C} \to \mathcal{D}$ is *fully faithful* when, for any pair $x, y \in \mathrm{Ob}(\mathcal{C})$, the map $f_{x,y}$ in (A-2) is an isomorphism of chain complexes. If the homology $H^*(f_{x,y})$ induces an isomorphism for all pairs, then $f_{x,y}$ is called *quasifully faithful*. A functor $f : \mathcal{C} \to \mathcal{D}$ is a *quasi-isomorphism* of dg categories when $H^*(f) \colon H^*(\mathcal{C}) \to H^*(\mathcal{D})$ induces an equivalence of graded k–linear categories.

Example A.2 The category of chain complexes $\mathrm{Kom}_k(\mathcal{A})$ is a subcategory $\mathrm{Kom}_k(\mathcal{A}) \subset \mathrm{Kom}_k^*(\mathcal{A})$ of a dg category. The objects of $\mathrm{Kom}_k^*(\mathcal{A})$ are the chain complexes $(C, \partial_C) \in \mathrm{Kom}_k(\mathcal{A})$. The maps are now given by the chain complex $(\mathrm{Hom}^*((C, \partial_C), (D, \partial_D)), \delta)$, where

$$\mathrm{Hom}^n((C, \partial_C), (D, \partial_D)) := \prod_{m \in \mathbb{Z}} \mathrm{Hom}(C^m, D^{n+m}),$$

with differential $\delta(f) := d_D f + (-1)^{n+1} f d_C$ for f of degree n.

When $\mathcal{A}$ is Vect_k, the category of dg categories will be denoted by dgcat_k. Important for this paper is a sequence of localizations obtained by different model category structures on dgcat_k:

$$\mathrm{dgcat}_k \xrightarrow{(1)} \mathrm{Hqe} \xrightarrow{(2)} \mathrm{Hmo}. \tag{A-3}$$

Hqe The first category $\mathrm{Hqe} := \mathrm{dgcat}_k[W^{-1}]$ is obtained by requiring quasi-isomorphisms $f \in W$ to be isomorphisms. In this model structure, cofibrations are determined by the left lifting property with respect to fibrations, and fibrations are dg functors $f : \mathcal{C} \to \mathcal{D}$ for which $f_{x,y}$ in (A-2) are surjective and

- for $x \in \mathrm{Ob}(\mathcal{C})$ and any homotopy equivalence $\beta \colon f(x) \to y$ in $\mathcal{D}$ there is a homotopy equivalence $\alpha \colon x \to z$ in $\mathcal{C}$ such that $f(\alpha) = \beta$.

The initial object is the empty category $\varnothing$ with no objects and the final object 0 is the zero dg category consisting of one object with no endomorphisms. In Hqe nontrivial dg categories are fibrant, and cofibrant resolutions are can be obtained from cobar–bar construction.

Modules For any dg category there are associated categories of modules over that dg category.

A right dg module M over a dg category $\mathcal{C}$ is a dg functor $\mathcal{C}^{\mathrm{op}} \to \mathrm{Kom}_k^*(\mathrm{Vect}_k)$. The dg category of such functors will be denoted by $\mathrm{Mod}_{\mathcal{C}}$. The homology $H^*(M) \colon \mathcal{C}^{\mathrm{op}} \to \mathrm{Vect}_k^{\mathbb{Z}}$ of a dg module M is the functor $c \mapsto H^*(M(c))$ taking values in graded vector spaces. A *quasi-isomorphism* $g \colon M \to N$ of dg modules is a map inducing an isomorphism between their respective homologies. The derived category $D(\mathcal{C})$ of dg modules over a dg category $\mathcal{C}$ is obtained by inverting the quasi-isomorphisms Q:

$$D(\mathcal{C}) := \mathrm{Mod}_{\mathcal{C}}[Q^{-1}].$$

This is a triangulated category [25]. If $f : \mathcal{C} \to \mathcal{D}$ is a dg functor then there is a pushforward functor $f_! \colon \mathrm{Mod}_{\mathcal{C}} \to \mathrm{Mod}_{\mathcal{D}}$ which is left adjoint to the pullback $f^* \colon \mathrm{Mod}_{\mathcal{D}} \to \mathrm{Mod}_{\mathcal{C}}$. These functors induce functors between derived categories

$$f_! \colon D(\mathcal{C}) \leftrightarrow D(\mathcal{D}) : f^*.$$

A dg functor $f : \mathcal{C} \to \mathcal{D}$ is a *Morita equivalence* when $f^* : D(\mathcal{D}) \to D(\mathcal{C})$ is an equivalence of triangulated categories.

Hmo The category Hmo is obtained by inverting Morita equivalences M:

$$\mathrm{Hmo} := \mathrm{Hqe}[M^{-1}].$$

The category Hmo is pointed: the dg category 1 consisting of a single object and a single morphism is both initial and terminal. The cofibrant objects of Hmo and Hqe remain the same. Fibrant objects become pretriangulated dg categories, as discussed in the next paragraph.

There is a full subcategory $\mathcal{C}^{\mathrm{perf}} \subset \mathrm{Mod}_{\mathcal{C}}$ consisting of modules M which are compact in $D(\mathcal{C})$. Since representable modules are compact, the Yoneda embedding factors through the subcategory of perfect modules, giving a dg functor

$$\gamma\colon \mathcal{C} \to \mathcal{C}^{\mathrm{perf}}.$$

A dg category $\mathcal{C}$ is called *perfect* when γ is a quasiequivalence. A dg category $\mathcal{C}$ in Hmo is fibrant if and only if it is perfect. So γ is fibrant replacement. An explicit model for $\mathcal{C}^{\mathrm{perf}}$ is given by the idempotent completion of the category of one-sided twisted complexes over $\mathcal{C}$ [8, Section 2.4].

Maps Toën's theorem shows that maps $\mathcal{C} \to \mathcal{D}$ in Hqe and are given by bimodules $\mathcal{C} \otimes \mathcal{D}^{\mathrm{op}} \to \mathrm{Kom}_k^*(\mathrm{Vect}_k)$ satisfying certain cofibrancy and representability conditions [56]. If $\mathcal{D}$ is fibrant then these are also the maps in Hmo. Dg functors described above define maps in each of these settings.

Constructions in Hqe vs Hmo If $\mathcal{C} \to \mathcal{D}$ and $\mathcal{C} \to \mathcal{E}$ are in Hqe, then the homotopy pushout $\mathcal{D} \sqcup_{\mathcal{C}}^h \mathcal{E}$ can be constructed by using the coproduct of dg categories on the associated pushout of cofibrant replacements. Since cofibrant objects in Hqe and Hmo agree, the quotient Hqe $\to$ Hmo commutes with homotopy pushout.

Since all of our localization constructions are homotopy pushouts, they are indifferent to the distinction between Hqe and Hmo in the manner described above.

List of symbols

After Section 2, dg categories are ungraded over a field of characteristic 2. The homotopy category of dg categories $\mathrm{Ho}(\mathrm{dgcat}_k)$ over k will be denoted by Hqe or Hmo when the equivalence relation is quasiequivalence or Morita equivalence, respectively. All surfaces denoted by Σ are connected unless otherwise mentioned. $\Sigma_{g,n}$ is the orientable surface of genus g with n boundary components.

$-^\vee$	Proposition 4.5
$-^{\mathrm{op}}$	opposite category
$\boldsymbol{a}$	points $\{a_1, \dots, a_{2k}\}$ in arc diagram, Definition 5.4
a_k, a_k'	points in an arc diagram, Definition 5.4
$\mathcal{A}(\mathcal{Z})$	arc algebra [36; 62]
B	bottom of D^2

B	$B \subset \Sigma$, Definition 4.13
$\widehat{\mathrm{BSD}}(\gamma)$	(8-1)
$\mathcal{C}$	dg category, after Section 2 ungraded; see Section 2.5
c_i	cocore of 1–handle
$\mathcal{C}o(\Sigma)$	geometric contact category or Tian algebraic contact category
d	differential, $d^2 = 0$
dgcat_k	category of dg categories [8; 57]
$D', \overline{D}, \widetilde{D}$	Definitions 2.9 and 2.13
D^2	unit disk
e_k	generator of $\mathcal{N}_n$, Definition 6.7
$\mathfrak{e}(\gamma)$	Definition 4.1
$F(\mathcal{Z})$	surface of arc diagram, Definition 5.5
$F(\partial \mathcal{Z})$	Proposition 5.14
γ	dividing set, Definition 3.3
$\gamma(\epsilon_i)$	dividing set associated to e_i, Section 6.2.1
$\gamma_A, \gamma_B, \gamma_C$	bypass triangle, Proposition 3.16 and Section 8.2
$\gamma^\vee$	dual dividing set, Definition 3.4 and Proposition 4.5
$\left(\bigoplus_{i=1}^n \gamma_i, p\right)$	convolution of dividing sets, Definition 2.1
$\Gamma(\Sigma)$	mapping class group, Section 5.1
h_k	1–handle in $F(\mathcal{M}_n)$
$\mathrm{Ho}(\mathcal{C})$	$[\mathcal{C}]$ or $H^0(\mathcal{C})$ [57]
Hom^I	Definition 2.6
Hom^T	Definition 2.15
$\mathrm{Hom}^{\langle K \rangle}$	Proposition 2.19
Hmo	Morita homotopy category [46]
Hqe	homotopy category [47]
$i(x, y)$	geometric intersection number
$\mathrm{int}(X)$	interior of X
$I, I', \bar{I}$	Definitions 2.5 and 2.7
k	ground field; after Section 2 $\mathrm{char}(k) = 2$
κ, κ'	Definitions 2.5 and 2.13
$\langle K \rangle$	Proposition 2.19
$K_0(\mathcal{C})$	Grothendieck group [46]
$\mathcal{K}o(\Sigma)$	Definition 3.17
$\mathcal{K}o^n(\Sigma, m)$	Theorem 4.4
$\mathcal{K}o^n_\pm(\Sigma, m)$	$\pm$–halves of $\mathcal{K}o^n(\Sigma, m)$, Section 4.5
$L_R\mathcal{C}$	Definition 2.5
$L_S\mathcal{C}$	Proposition 2.16

m	boundary points $m \subset \partial\Sigma$, Section 3.1
M	a matching $M: \boldsymbol{a} \to \{1, \dots, k\}$ in arc diagram, Definition 5.4
μ	Section 6.3
$\mathcal{M}_n$	zigzag diagram for $(D^2, 2n)$, Definition 6.1
$\mathrm{Mat}(\mathcal{C})$	the additive closure, Section 2.1
$\mathcal{N}_n$	nil-Temperley–Lieb algebra, Definition 6.7
$\mathbb{N}$	$\mathbb{N} = \{0\} \cup \mathbb{Z}_+$
nS^1	Definition 4.8
$N(T)$	neighborhood of disk, Definition 3.8
Pre–$\mathcal{K}$o(Σ)	Definition 3.15
Pre–Pre-$\mathcal{K}$o(Σ)	Conjecture 3.18
$\mathcal{Q}_n$	Tian quiver, Definition 6.8
r	basepoint automorphism, Corollary 5.3
$\rho_{k,k\pm 1}$	(6-2)
$\mathcal{R}_n$	Tian disk category, Definition 6.11
$R_\pm$	positive and negative regions, Definition 3.3
S^2	the 2–sphere
$\Sigma_{g,n}$	connected surface of genus g with n boundary components
(Σ, m)	pointed oriented surface, Definition 3.1
$\bar{\Sigma}$	orientation reversal, Proposition 4.6
$\partial_i \Sigma$	i^{th} boundary component of Σ, Definition 3.1
$\theta: \gamma \to \gamma'$	bypass move, Definition 3.8
$\theta_{i,j}$	Definitions 2.9 and 2.13
T	top of D^2
(T, γ, γ')	bypass attachment, Definition 3.8
W_γ	cap associated to γ [63, Definition 2.5], Section 8.2
Ξ	(3-1) and Definition 3.17
$\mathcal{Y}_n$	$\mathcal{R}_n^{\text{pretr}}$, Definition 6.12
z	basepoints $z = \{z_1, \dots, z_n\}$ for $z_i \in \partial_i \Sigma$, Definition 3.1
$\mathfrak{z}_C$	$z_C \in \mathfrak{Z}(\mathcal{Z})$, Definition 5.10
$\mathbb{Z}_+$	$\{1, 2, 3, \dots\} \subset \mathbb{Z}$
$\mathbb{Z}/2$	$\mathbb{Z}/2\mathbb{Z}$
Z	ordered set of lines, Definition 5.4
$\mathcal{Z}$	arc diagram, Definition 5.4
$\mathcal{Z}_i$	arc in arc diagram, Definition 5.4
$\mathcal{Z}_{0,n}$	arc diagram for $\Sigma_{0,n}$, Definition 7.1
$\mathcal{Z}_{g,1}$	arc diagram for $\Sigma_{g,1}$, Definition 7.8
$\mathfrak{Z}(\mathcal{Z})$	set of elementary dividing sets, Definition 5.10

References

[1] **A Beilinson**, **V Vologodsky**, *A DG guide to Voevodsky's motives*, Geom. Funct. Anal. 17 (2008) 1709–1787 MR Zbl

[2] **A I Bondal**, **M M Kapranov**, *Enhanced triangulated categories*, Mat. Sb. 181 (1990) 669–683 MR Zbl In Russian; translated in Math. USSR-Sb. 70 (1991) 93–107

[3] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions*, *I*, preprint (2012) arXiv 1208.1074

[4] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions*, *II*, preprint (2012) arXiv 1208.1077

[5] **V Colin**, **P Ghiggini**, **K Honda**, *The equivalence of Heegaard Floer homology and embedded contact homology*, *III*: *From hat to plus*, preprint (2012) arXiv 1208.1526

[6] **B Cooper**, *Formal contact categories*, *II*, in preparation

[7] **S K Donaldson**, *The Seiberg–Witten equations and* 4*–manifold topology*, Bull. Amer. Math. Soc. 33 (1996) 45–70 MR Zbl

[8] **V Drinfeld**, *DG quotients of DG categories*, J. Algebra 272 (2004) 643–691 MR Zbl

[9] **T Dyckerhoff**, *Compact generators in categories of matrix factorizations*, Duke Math. J. 159 (2011) 223–274 MR Zbl

[10] **T Dyckerhoff**, **M Kapranov**, *Triangulated surfaces in triangulated categories*, J. Eur. Math. Soc. 20 (2018) 1473–1524 MR Zbl

[11] **J B Etnyre**, **D S Vela-Vick**, **R Zarev**, *Sutured Floer homology and invariants of Legendrian and transverse knots*, Geom. Topol. 21 (2017) 1469–1582 MR Zbl

[12] **S I Gelfand**, **Y I Manin**, *Methods of homological algebra*, Springer (1996) MR Zbl

[13] **E Giroux**, *Structures de contact sur les variétés fibrées en cercles audessus d'une surface*, Comment. Math. Helv. 76 (2001) 218–262 MR Zbl

[14] **F M Goodman**, **H Wenzl**, *Ideals in Temperley–Lieb category* (2003) Zbl Appendix to M H Freedman, *A magnetic model with a possible Chern–Simons phase*, Comm. Math. Phys. 234 (2003) 129–183

[15] **K Honda**, *Contact structures*, *Heegaard Floer homology and triangulated categories*, in preparation

[16] **K Honda**, *On the classification of tight contact structures, I*, Geom. Topol. 4 (2000) 309–368 MR Zbl

[17] **K Honda**, *Gluing tight contact structures*, Duke Math. J. 115 (2002) 435–478 MR Zbl

[18] **K Honda**, **Y Tian**, *Contact categories of disks*, J. Symplectic Geom. 20 (2022) 665–759 MR Zbl

[19] **M Hovey**, *Model categories*, Math. Surv. Monogr. 63, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[20] **M Hutchings**, *An index inequality for embedded pseudoholomorphic curves in symplectizations*, J. Eur. Math. Soc. 4 (2002) 313–361 MR Zbl

[21] **M Hutchings**, **C H Taubes**, *Gluing pseudoholomorphic curves along branched covered cylinders, I*, J. Symplectic Geom. 5 (2007) 43–137 MR Zbl

[22] **M Hutchings**, **C H Taubes**, *Gluing pseudoholomorphic curves along branched covered cylinders, II*, J. Symplectic Geom. 7 (2009) 29–133 MR Zbl

[23] **A Juhász**, *A survey of Heegaard Floer homology*, from "New ideas in low dimensional topology" (L H Kauffman, V O Manturov, editors), Ser. Knots Everything 56, World Sci., Hackensack, NJ (2015) 237–296 MR Zbl

[24] **L H Kauffman**, **S L Lins**, *Temperley–Lieb recoupling theory and invariants of* 3*–manifolds*, Ann. of Math. Stud. 134, Princeton Univ. Press (1994) MR Zbl

[25] **B Keller**, *Deriving DG categories*, Ann. Sci. École Norm. Sup. 27 (1994) 63–102 MR Zbl

[26] **B Keller**, *On differential graded categories*, from "International Congress of Mathematicians, II" (M Sanz-Solé, J Soria, J L Varona, J Verdera, editors), Eur. Math. Soc., Zürich (2006) 151–190 MR Zbl

[27] **M Kontsevich**, *Symplectic geometry of homological algebra*, preprint (2009) Available at `http://www.ihes.fr/~maxim/TEXTS/Symplectic_AT2009.pdf`

[28] **P Kronheimer**, **T Mrowka**, *Monopoles and three-manifolds*, New Math. Monogr. 10, Cambridge Univ. Press (2007) MR Zbl

[29] **Ç Kutluhan**, **Y-J Lee**, **C H Taubes**, HF = HM*, I: Heegaard Floer homology and Seiberg–Witten Floer homology*, Geom. Topol. 24 (2020) 2829–2854 MR Zbl

[30] **Ç Kutluhan**, **Y-J Lee**, **C H Taubes**, HF = HM*, II: Reeb orbits and holomorphic curves for the ech/Heegaard Floer correspondence*, Geom. Topol. 24 (2020) 2855–3012 MR Zbl

[31] **Ç Kutluhan**, **Y-J Lee**, **C H Taubes**, HF = HM*, III: Holomorphic curves and the differential for the ech/Heegaard Floer correspondence*, Geom. Topol. 24 (2020) 3013–3218 MR Zbl

[32] **Ç Kutluhan**, **Y-J Lee**, **C H Taubes**, HF = HM*, IV: The Seiberg–Witten Floer homology and ech correspondence*, Geom. Topol. 24 (2020) 3219–3469 MR Zbl

[33] **Ç Kutluhan**, **Y-J Lee**, **C H Taubes**, HF = HM*, V: Seiberg–Witten Floer homology and handle additions*, Geom. Topol. 24 (2020) 3471–3748 MR Zbl

[34] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *A faithful linear-categorical action of the mapping class group of a surface with boundary*, J. Eur. Math. Soc. 15 (2013) 1279–1307 MR Zbl

[35] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Bordered Floer homology and the spectral sequence of a branched double cover, I*, J. Topol. 7 (2014) 1155–1199 MR Zbl

[36] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Bordered Heegaard Floer homology*, Mem. Amer. Math. Soc. 1216, Amer. Math. Soc., Providence, RI (2018) MR Zbl

[37] **J Lurie**, *Higher topos theory*, Ann. of Math. Stud. 170, Princeton Univ. Press (2009) MR Zbl

[38] **J Lurie**, *Higher algebra*, preprint (2017) Available at `https://url.msp.org/Lurie-HA`

[39] **D V Mathews**, *Strand algebras and contact categories*, Geom. Topol. 23 (2019) 637–683 MR Zbl

[40] **J Murakami**, *The multi-variable Alexander polynomial and a one-parameter family of representations of* $\mathfrak{U}_q(\mathfrak{sl}(2,\mathbb{C}))$ *at* $q^2=-1$, from "Quantum groups" (P P Kulish, editor), Lecture Notes in Math. 1510, Springer (1992) 350–353 MR Zbl

[41] **D Nadler**, *Cyclic symmetries of* A_n*–quiver representations*, Adv. Math. 269 (2015) 346–363 MR Zbl

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

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

[44] **I Petkova**, **V Vértesi**, *Combinatorial tangle Floer homology*, Geom. Topol. 20 (2016) 3219–3332 MR Zbl

[45] **D G Quillen**, *Homotopical algebra*, Lecture Notes in Math. 43, Springer (1967) MR Zbl

[46] **G Tabuada**, *Invariants additifs de DG-catégories*, Int. Math. Res. Not. 2005 (2005) 3309–3339 MR Zbl

[47] **G Tabuada**, *Une structure de catégorie de modèles de Quillen sur la catégorie des DG–catégories*, C. R. Math. Acad. Sci. Paris 340 (2005) 15–19 MR Zbl

[48] **G Tabuada**, *On Drinfeld's DG quotient*, J. Algebra 323 (2010) 1226–1240 MR Zbl

[49] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, I*, Geom. Topol. 14 (2010) 2497–2581 MR Zbl

[50] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, II*, Geom. Topol. 14 (2010) 2583–2720 MR Zbl

[51] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, III*, Geom. Topol. 14 (2010) 2721–2817 MR Zbl

[52] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, IV*, Geom. Topol. 14 (2010) 2819–2960 MR Zbl

[53] **C H Taubes**, *Embedded contact homology and Seiberg–Witten Floer cohomology, V*, Geom. Topol. 14 (2010) 2961–3000 MR

[54] **Y Tian**, *A categorification of* $\mathrm{U}_q(\mathfrak{sl}(1|1))$ *as an algebra*, preprint (2012) arXiv 1210.5680

[55] **Y Tian**, *A categorification of* $\mathrm{U}_T(\mathfrak{sl}(1|1))$ *and its tensor product representations*, Geom. Topol. 18 (2014) 1635–1717 MR Zbl

[56] **B Toën**, *The homotopy theory of DG–categories and derived Morita theory*, Invent. Math. 167 (2007) 615–667 MR Zbl

[57] **B Toën**, *Lectures on DG–categories*, from "Topics in algebraic and topological K–theory" (G Cortiñas, editor), Lecture Notes in Math. 2008, Springer (2011) 243–302 MR Zbl

[58] **O Y Viro**, *Quantum relatives of the Alexander polynomial*, Algebra i Analiz 18 (2006) 63–157 MR Zbl In Russian; translated in St. Petersburg Math. J. 18 (2007) 391–457

[59] **K Walker**, *TQFTs*, incomplete notes (2006) Available at `https://canyon23.net/math/tc.pdf`

[60] **B Webster**, *Tensor product algebras, Grassmannians and Khovanov homology*, from "Physics and mathematics of link homology" (S Gukov, M Khovanov, J Walcher, editors), Contemp. Math. 680, Amer. Math. Soc., Providence, RI (2016) 23–58 MR Zbl

[61] **E Witten**, *Monopoles and four-manifolds*, Math. Res. Lett. 1 (1994) 769–796 MR Zbl

[62] **R Zarev**, *Bordered Floer homology for sutured manifolds*, preprint (2009) arXiv 0908.1106

[63] **R Zarev**, *Joining and gluing sutured Floer homology*, preprint (2010) arXiv 1010.3496

Department of Mathematics, University of Iowa
Iowa City, IA, United States

`ben-cooper@uiowa.edu`

Received: 20 July 2017 Revised: 3 April 2023

Algebraic & Geometric Topology **24**:5 (2024) 2451–2508
DOI: 10.2140/agt.2024.24.2451
Published: 19 August 2024

Comparison of period coordinates and Teichmüller distances

IAN FRANKEL

We show that when two unit-area quadratic differentials are ϵ–close with respect to good systems of period coordinates and lie over a compact subset K of the moduli space of Riemann surfaces $\mathcal{M}_{g,n}$, then their underlying Riemann surfaces are $C\epsilon^{\alpha}$–close in the Teichmüller metric. Here, α depends only on the genus g and the number of marked points, while C depends on K.

30F60

1 Introduction and statement of main result

1.1 Preliminaries

The moduli space of compact Riemann surfaces of genus g with n unlabeled marked points (or n deleted points), $\mathcal{M}_{g,n}$, is a complex variety and orbifold of dimension $3g-3+n$, with each point of $\mathcal{M}_{g,n}$ representing a biholomorphism class a of compact genus Riemann surface with n points deleted (or with n marked points). The topology of $\mathcal{M}_{g,n}$ is induced by the Kobayashi metric. (The Kobayashi pseudometric on a complex analytic X space is the largest pseudometric for which all holomorphic maps of the hyperbolic disk into X are nonexpanding, and it is a nondegenerate metric in all cases we will consider.) Teichmüller constructed a metric whereby a K–quasiconformal homeomorphism (see Section 3) between X and Y exists if and only if the distance between $(X; p_1, \dots, p_n)$ and $(Y; q_1, \dots, q_n)$ is at most K. Royden [1971] showed that this is the same as the Kobayashi metric. The orbifold universal cover of $\mathcal{M}_{g,n}$, the *Teichmüller space* $\mathcal{T}_{g,n}$, is a contractible complex manifold of dimension $3g-3+n$ when $3g-3+n>0$. In the same paper, Royden showed that the biholomorphism group of $\mathcal{T}_{g,n}$ is naturally isomorphic to the Teichmüller modular group, or mapping class group $\mathrm{Mod}(S_{g,n})$, of the surface $S_{g,n}$, and the quotient by this action is $\mathcal{M}_{g,n}$.

Even though the Teichmüller–Kobayashi metric on $\mathcal{T}_{g,n}$ is not Riemannian, it has a geodesic flow, which is usually described as a dynamical system on the unit *cotangent bundle*. For $X \in \mathcal{T}_{g,n}$, the cotangent space to $\mathcal{T}_{g,n}$ at X consists of those meromorphic sections of the tensor square of the cotangent bundle to the complete Riemann surface corresponding to X for which the only poles are simple and occur at the marked points of X. We refer to such sections as *quadratic differentials*. We will write $\mathrm{QD}(\mathcal{T}_{g,n})$ to denote the space of nonzero quadratic differentials on surfaces of genus g with n marked points.

Many analogies have been made between the geodesic flow for this metric and the geodesic flow on a closed negatively curved Riemannian manifold. Euclidean geometry is the main tool in the study of the Teichmüller geodesic flow; we describe briefly the connection here. Given a nonzero quadratic differential q on X, there is an associated flat (Gaussian curvature = 0) metric on X with a finite number of cone-type singularities, which we will call the *q–metric*. The metric can be defined locally by taking the integral of the holomorphic 1–form $\sqrt{q}$ to give an isometric chart to $\mathbb{C}$, ie

$$z \mapsto \int_{z_0}^{z} \sqrt{q}$$

gives an isometry between a neighborhood of $z_0 \in X$ with the q–metric and an open set in $\mathbb{C} = \mathbb{R}^2$ with the standard Euclidean metric, for each point z_0 where q is holomorphic and nonvanishing.

Singularities of the q–metric occur at points where such charts cannot be defined — at zeros and poles of q. At these points, we have cone-type singularities with cone angle $(n+2)\pi$ at each zero of order $n \geq -1$. (If we allowed q to have poles of higher order, they would be an infinite distance away, so the metric cannot extend.) It should be noted that, since $\sqrt{q}$ is only defined up to sign, these charts to $\mathbb{C}$ are unique up to the group $\{z \mapsto C \pm z : C \in \mathbb{C}\}$. We will refer to a surface equipped with such a metric as a *half-translation surface*. We note that it is also possible to take a surface with charts and transition maps lying in $\{z \mapsto C \pm z : C \in \mathbb{C}\}$, and give it a metric for which those charts are isometric and recover a complex structure and a quadratic differential which is $dz \otimes dz$ in local coordinates for the system of charts. If, in addition, the resulting metric space can be completed by adding only finitely many points with cone-type singularities, the complex structure extends uniquely to the completion and the quadratic differential extends meromorphically. We also establish the following convention:

Convention 1.1 Given $(X; p_1, \dots, p_n; q)$ a Riemann surface with n marked points $p_1, \dots, p_n$ and a nonzero meromorphic quadratic differential q on X whose only poles are simple and occur at a subset of $\{p_1, \dots, p_n\}$, the set of *singularities* of $(X; p_1, \dots, p_n; q)$ are the zeros of q on X together with the collection of all marked points, regardless of whether or not q has poles at those points.

1.2 Period coordinates and the main theorem

We can integrate $\sqrt{q}$ along q–geodesic arcs $\gamma_i : [0,1] \to X$ chosen so that $\gamma_i((0,1))$ contains no singularities and γ_i is injective on $(0,1)$, but $\gamma(0)$ and $\gamma(1)$ are (not necessarily distinct) singularities. Such arcs are called *saddle connections*. A collection of such integrals forms a local holomorphic coordinate chart for $\mathrm{QD}(\mathcal{T}_{g,n})$ in a neighborhood of (X,q), provided that q has $4g-4+n$ simple zeros and n simple poles. We call such charts *period coordinate* charts. A period coordinate chart can be extended to the boundary of any open precompact subset of $\mathrm{QD}(\mathcal{T}_{g,n})$ on which it is well defined, and embeds into $\mathbb{C}^{6g-6+2n}$ as a convex set. $\mathrm{QD}(\mathcal{T}_{g,n})$ is locally finitely covered by the closures of such sets by Corollary A.7.

If a set $S \subset \mathrm{QD}(\mathcal{T}_{g,n})$ is homeomorphic to a compact convex set with nonempty interior $K \subset \mathbb{C}^{6g-6+2n}$ via a homeomorphism $f : S \to K$ whose restriction to the interior of S is a period coordinate embedding, we say S is a *compact convex period coordinate patch*. In Appendix A we show that $\mathrm{QD}(\mathcal{T}_{g,n})$ is locally finitely covered by compact convex period coordinate patches.

We will define a $\mathrm{Mod}(S_{g,n})$–invariant path metric $d_{\mathrm{Euclidean}}$ (see Definition 3.2) that has the property that every compact convex period coordinate patch is locally bi-Lipschitz to the corresponding subset of $\mathbb{C}^{6g-6+2n}$.

We now state our main theorem:

Theorem 1.2 *Let g and n be nonnegative integers such that $3g-3+n>0$. Let $a_n = 1$ if $n=0$ and 2 if $n>0$. Let K be a compact subset of $\mathcal{T}_{g,n}$. Then for any unit-area quadratic differentials (X_1,q_1) and (X_2,q_2) with in $X_1, X_2 \in K$, there is a constant C_K such that*

$$d_{\mathrm{Teich}}(X_1,X_2) < C_K\, d_{\mathrm{Euclidean}}(q_1,q_2)^{2/[2+a_n(4g-4+n)]}.$$

Remark Theorem 1.2 is true whether we restrict $d_{\mathrm{Euclidean}}$ to unit-area quadratic differentials intrinsically or extrinsically (whether or not we consider paths that leave the space of unit-area differentials). The reason for this is that $\|q\|$, the area of the q–metric, is locally Lipschitz as a function of (X,q) with respect to the metric $d_{\mathrm{Euclidean}}$. Therefore if (X_1,q_1) and (X_2,q_2) are unit area and sufficiently close, the shortest path between them stays near the set of unit-area quadratic differentials. A path in $\mathrm{QD}(\mathcal{T}_{g,n})$ that starts and ends in K can be projected onto the unit-area subspace, and if the path stays sufficiently close to K this projection will only increase its length by a bounded factor.

The real content of the theorem is that it remains valid even when zeros of the quadratic differential are allowed to coincide with each other, or with marked points. If we required K to be a compact subset of the *principal stratum* of quadratic differentials, ie the space of quadratic differentials with n simple poles and $4g-4+n$ simple zeros, then we could replace the Hölder exponent $2/[2+a_n(4g-4+n)]$ in Theorem 1.2 with 1, by finding a common triangulation of (X_1,q_1) and (X_2,q_2) by saddle connections and mapping triangles to triangles.

We further remark that no inequality in the other direction is possible, since two quadratic differentials $q_1 \neq q_2$ may have the same underlying Riemann surface $X = X_1 = X_2$.

The *strongly stable leaves* for the Teichmüller geodesic flow are the collections of unit-area quadratic differentials that, on each chart given by period coordinates, can be obtained from each other by changing only the imaginary parts of period coordinates and changing systems of period coordinates. One might hope for a converse inequality if we make the additional assumption that (X_1, q_1) and (X_2, q_2) are on the same strongly stable leaf or strongly unstable leaf of the Teichmüller geodesic flow, for in such cases X_1 and X_2 are known to be different points in $T_{g,n}$, by the main result of [Hubbard and Masur 1979].

Before we proceed toward the proof, we would like to make a remark about the optimal Hölder exponent. We do not make any claim that the exponent we produce is sharp, but we do claim that, asymptotically, it can only be improved by a factor of 2, at least when $n \geq 1$.

The reason for this is the following example:

Let (X_t, q_t) be a family of quadratic differentials that are all isometric to some (X, q), but on which one marked point moves according to a holomorphic parameter t. Moreover, assume that the quadratic differential q has the form $z^k\, dz^2$ in some local coordinate z on the underlying Riemann surface. The q–metric is then given by $ds^2|z|^k|dz|^2$, and we take the parameter t to be the location of the marked point in the coordinate z, so t varies over a neighborhood of zero.

For t near 0, we have that $C^{-1}|t| \leq d_{\mathrm{Teich}}(X_t, X_0) < C|t|$ since t is a coordinate in a holomorphic coordinate system for Teichmüller space, and the Teichmüller metric is Finsler with respect to this complex structure.

On the other hand, the Euclidean distance between q_0 and q_t is comparable to the q–distance between the points $z = 0$ and $z = t$, which equals $\int_0^{|t|} x^{k/2}\, dx$.

Thus $C^{-1} t^{(2+k)/2} < d_{\mathrm{Euclidean}}(q_0, q_t) < C|t|^{(2+k)/2}$.

(The actually behavior of the Euclidean metric near q_0 is treated more carefully in the appendices, but the intuition is that one of our period coordinates comes from a saddle connection joining 0 to t, and this coordinate accounts for the Euclidean distance from q_0 to q_t.)

The largest possible k is $4g - 4 + n$, so the best possible Hölder exponent is then $2/[4g - 2 + n]$.

Acknowledgments

This research was conducted for the author's PhD thesis at the University of Chicago. The author would like to thank Howard Masur for suggesting this problem. The author would also like to thank Howard Masur, Kasra Rafi, and especially Alex Eskin for helpful conversations, and the referee(s) whose input helped to clarify the exposition.

2 Sketch of proof

The proof is by construction of a quasiconformal map between nearby surfaces. The easiest way to build a quasiconformal map between (X_1, q_1) and (X_2, q_2) is by finding a common triangulation by saddle connections and applying a piecewise linear (PL) map. Then one can simply estimate the dilatation (the best possible constant K such that the map is K–quasiconformal) on each triangle, and in so doing obtain a bound for the dilatation of the entire map.

Unfortunately, this strategy yields poor estimates unless all of the edge lengths and angles of the triangles can be bounded away from zero, and usually can't be used to build maps between surfaces in different strata.

Our solution to this problem is to find a system of disks, which we call *nearly regular right polygons* or NRRPs, that isolate clusters of singularities. We can triangulate the remainder of each of the surfaces, and build a PL map in the complement of these disks. As for the disks themselves, we need to find a quasiconformal extension of the boundary map to the NRRP, and estimate its dilatation. For this we use a Beurling–Ahlfors extension, which requires as input an estimate of the quasisymmetry of the boundary map. This is carried out Section 6.

The boundary maps between disks are fixed since the map is PL outside of the disks. However, we need to understand the uniformization of these disks to the upper half-plane in order to estimate its dilatation. Our disks will be chosen so that, when doubled along the boundary, they have the isometry type of half-translation surfaces coming from conjugation-invariant quadratic differentials on $\widehat{\mathbb{C}}$. Almost all of the work consists of bounding the changes in the locations of the singularities as a function of the changes in period coordinates. Since the quadratic differentials are actually of the form $f(z)\,dz^2$ for some rational function $f \in \mathbb{R}(z)$, it is actually easier to give a lower bound for the changes in period coordinates in terms of the changes of locations of the zeros and poles, for a sequence of perturbations tending to 0. This is most of the work, and it is carried out in Section 5.

The key to finding a lower bound is finding a scale on which perturbations do not have canceling effects on period coordinates. For this we use a partial compactification of strata similar to the compactification used in [Bainbridge et al. 2019] to handle collisions of singularities. For us, the most important organizing information is which collections of singularities are colliding faster than others. The limit of a differential in this compactification is a stable curve with one component for each such cluster, and a meromorphic quadratic differential on each component. A colliding cluster of singularities will correspond to a component with a quadratic differential that has a higher-order pole on ∞. The partial compactification is described in Definition 4.35, and a convergent sequence is defined in Notation 5.3.

When a cluster of colliding singularities moves mostly in the same direction, we show that it essentially behaves as a point, and shadows a perturbation in a lower-dimensional stratum. This is the content of Proposition 5.4. When a cluster of singularities move in what would appear to be canceling directions according to Proposition 5.4, we find a leading-order approximation for how the perturbation affects saddle connections in this cluster. If there are subclusters the leading term may again be zero, but we can break the perturbation into components, at least one of which must be detectable on some cluster. Appendix B is dedicated to showing that these kinds of perturbations do not cancel out. The end result of this argument is the key estimate, Lemma 5.11, which bounds the change in the location of the singularities as a function of the change in the periods. (More precisely, we prove the contrapositive.) We finish Section 5 by converting Lemma 5.11 to the form we need to apply to the specific disks, Lemma 5.12.

3 Teichmüller spaces and quadratic differentials

In this section, we assemble basic facts about Teichmüller spaces and coordinate systems on strata.

3.1 Flat metrics and period coordinates

On a half-translation surface, the slope of a tangent vector at any nonsingular point is well defined; in particular, vertical and horizontal are well-defined notions. Length and area are also well defined in the q–metric, since the change of charts preserves them. It is also evident that if X has the q–metric, then geodesics are polygonal arcs and only change direction at cone points, and when they do they turn by an angle of at least π measured either way around the cone point. If q is given in local coordinates by $(dx + i\,dy)^{\otimes 2}$, then the area of X with the q metric, given by $|q| = \int_X dx \wedge dy$, is the norm of q in the Teichmüller cometric. (Note that this area form does not depend on which square root of q is picked.)

The collection of all quadratic differentials whose associated metrics have finite area is a vector bundle over $\mathcal{T}_{g,n}$ (if we include the zero section), and the set of nonzero quadratic differentials admits a holomorphic stratification whereby each stratum consists of differentials whose metrics have the same number of singularities of each type; two singularities have the same type if they are both marked points and have the same cone angle, or they are both unmarked points and have the same cone angle. Two different strata can have the same underlying Riemann surfaces and quadratic differentials, but differ in which points are marked. The *principal stratum* consists of those quadratic differentials with a simple pole at each of the n marked points and $4g-4+n$ simple zeros. The complement of the principal stratum is an analytic subvariety, so the principal stratum sits inside $\mathrm{QD}(\mathcal{T}_{g,n})$ as a dense set of full measure (with respect to Lebesgue measure class).

Definition 3.1 The *universal half-translation surface* is a surface bundle over $T^*\mathcal{T}_{g,n}$ with fiber S_g, and where the fiber over (X,q) is a copy of $\overline{X}$ equipped with the q–metric.

We will give a more explicit description of the topology of this bundle later in this section.

With our convention, the set of all singularities of q–metrics is an *incidence subvariety* of the universal half-translation surface. Locally they define a collection of sections of the universal half-translation surface, at least in a neighborhood of each point in the principal stratum. If $p_1,\dots,p_r$ are the singularities at a point (X,q) in the principal stratum, any nearby (X',q') will have singularities $p'_1,\dots,p'_r$, and each p'_i varies holomorphically with (X',q') in a bundle over $T^*\mathcal{M}_{g,n}$ whose fibers are the Riemann surfaces represented. It is thus meaningful to talk about points being the same singularity on different surfaces. We may pick homotopy classes of arcs whose endpoints are the singularities, and the integrals $\int_{p_i'}^{p_j'} \sqrt{q'}$ over such arcs will vary holomorphically. When the arcs are represented by saddle connections, we call the integrals $\int_{p_i'}^{p_j'} \sqrt{q'}$ *period coordinates*. In a neighborhood of any point in the principal stratum, some such collection of period coordinates associated to saddle connections will form a holomorphic coordinate chart to $\mathbb{C}^{6g-6+2n}$. On a dense subset of such coordinate charts, we can paste together Euclidean metrics to form a path metric on the moduli space of quadratic differentials, described in the next section.

3.2 The Euclidean metric on $\mathrm{QD}(\mathcal{T}_{g,n})$

A choice of norm on a vector space is equivalent to a choice of the closed unit ball, ie a compact convex set with nonempty interior which is symmetric about the origin. Given a finite collection of norms we can simply take the convex hull of the union of their unit balls to be the unit ball of the pasted metric. The resulting norm is of course equivalent (bi-Lipschitz) to any of the original Euclidean norms at each point. Call this the *union convex hull operation*. It produces the largest norm that is less than or equal to a given set of norms. If we have an arc in a stratum, at each point of which some system of period coordinates has locally constant derivative, we describe a norm which is the speed of the arc at almost every point. For each real $L > 0$, there are locally only finitely many coordinate systems represented by saddle connections of length $\le L$ (see Appendix A) in the following sense: for any point $(X, q) \in \mathrm{QD}(\mathcal{T}_{g,n})$ with $q \neq 0$, there is a neighborhood of (X, q) in which only finitely many such systems exist, *even if* (X, q) *is not in the principal stratum.*

Recall the definition of compact convex period coordinate patch in the discussion immediately before Theorem 1.2. At each point (X, q), we consider all good embeddings of sets containing (X, q) in which (X, q) maps to $\{z : |z| \le \Theta(X, q)\}^n$ for some continuous proper function Θ which is large enough to ensure that the set of such good embeddings is not empty for any (X, q). From Appendix A it is clear such a function exists. For example, 4 times the q–diameter of (X, q) is such a function.

Definition 3.2 Fix a continuous function $\Theta\colon \mathrm{QD}(\mathcal{M}_{g,n}) \to (0, \infty)$ as above. We define the *Euclidean metric* $d_{\mathrm{Euclidean}}$ on the space of half-translation surfaces to be the path metric obtained by applying the union convex hull operation to the norms coming from compact convex period coordinate patches containing (X, q), whose defining saddle connections have length less than $\Theta((X, q))$.

It should be clear from the definition that $d_{\mathrm{Euclidean}}$ is the largest path metric such that, for all (X, q), all compact convex period coordinate patches containing (X, q) that take it into $\{v : |v|_\infty < \Theta(X, q)\}^{6g-6+2n}$ are noncontracting in a neighborhood of (X, q).

Remark The local Lipschitz class of the metric $d_{\mathrm{Euclidean}}$ depends only on the fact that it comes from a locally finite collection of systems of period coordinates. It does not depend on the specific coordinate systems chosen. Theorem 1.2 and any other statements that depend only on the local Lipschitz class remain true if we define $d_{\mathrm{Euclidean}}$ differently, for instance by applying the union convex hull operation on a different locally finite collection of period coordinate systems that is rich enough to cover quadratic differential space.

We note that the Euclidean metric extends to a path metric on the whole of $\mathrm{QD}(\mathcal{T}_{g,n})$, but in the absence of the lower bound for the area, arbitrarily short paths with the Euclidean metric move arbitrarily far in Teichmüller space. For this reason, and for the reason that the Teichmüller flow is usually studied on unit-area differentials, we find it prudent to restrict to the unit-area locus.

Also, when we are referring to a particular stratum or stratum closure, the Euclidean metric is understood to be the *intrinsic* metric on the stratum or stratum closure, rather than the Euclidean metric in the ambient moduli space.

3.3 Teichmüller's metric

The Teichmüller space of Riemann surfaces of genus g with n marked points is the set of marked complex structures on a Riemann surface of genus g with n points deleted. One way to define this more precisely is, given a smooth oriented surface S diffeomorphic to a surface of genus g with n points deleted, to look at all orientation-preserving homeomorphisms of complete finite volume hyperbolic Riemann surfaces into S, subject to the following equivalence relation: for $f_1\colon M_1 \to S$ and $f_2\colon M_2 \to S$, we say (M_1, f_1) and (M_2, f_2) are equivalent if $f_1^{-1} \circ f_2$ is homotopic to a biholomorphism. The set of equivalence classes of pairs (M, f) is the *Teichmüller* space of genus g surfaces with n punctures, which we will call $\mathcal{T}_{g,n}$. When it is convenient, we will sometimes view the punctures as marked points on a compact surface instead of deleted points.

$\mathcal{T}_{g,n}$ carries several metrics; the one we will be concerned with is the *Teichmüller* metric. We will briefly summarize the geometric properties of this metric by characterizing its geodesics. Proofs of these can be found in, for instance, [Hubbard 2006] or [Farb and Margalit 2012].

- Start with any nonzero holomorphic quadratic differential α, ie a nonzero holomorphic section of the tensor square of the cotangent bundle of a Riemann surface. If z is a local coordinate on a Riemann surface, then locally $\alpha = f(z)\,dz^2$, with f holomorphic in z; at each point at which f does not vanish we can find a holomorphic coordinate z for which $\alpha = dz^2$.

- In any simply connected region on such a chart, the level sets of the real and imaginary parts of z give a pair of transverse *measured foliations*, that is to say, foliations equipped with a transverse measure on the local leaf space, and the transverse measure is invariant under transition maps of a system of charts defining the foliation. (See [Fathi et al. 2012, Section 1.3] for a precise construction of the space of measured foliations.) Given any local coordinates $z = x + iy$, the integrals $\int_C |dx|$ and $\int_C |dy|$ are well defined for any smooth contour of integration C. There is also an invariant volume form, and in the case that our surface has punctures we will only consider those quadratic differentials for which the area form $dx \otimes dy$ assigns finite measure. This is equivalent to assuming the quadratic differential extends meromorphically to a compact Riemann surface with at most simple poles at each of finitely many new points.

- Now, fix a Riemann surface M_0 and a pair of transverse measured foliations associated to a holomorphic quadratic differential α with finite area. Then for each $\lambda \in \mathbb{R}$ there is a Riemann surface M_λ with a homeomorphism $f: M_0 \to M_\lambda$, smooth away from the zeros of α, with holomorphic charts such that the pushforwards of $|dx|$ and $|dy|$ are $|e^{\lambda}\,dx|$ and $|e^{-\lambda}\,dy|$, and $\lambda \mapsto M_\lambda$ is a unit-speed geodesic. The map f is called a *Teichmüller map*.

- All geodesics in $\mathcal{T}_{g,n}$ arise in this way, and the length of a geodesic between two points in $\mathcal{T}_{g,n}$ is equal to the distance between those points in the Teichmüller metric. Any two distinct points in $\mathcal{T}_{g,n}$ lie on a unique geodesic.

Suppose $f: U \to V$ is an almost-everywhere differentiable orientation-preserving homeomorphism of bounded domains in $\mathbb{C}$, with first-order distributional derivatives in $L^2(U)$. Then the partial derivatives

$$f_z := \tfrac{1}{2}(f_x - i f_y) \quad \text{and} \quad f_{\bar z} := \tfrac{1}{2}(f_x + i f_y)$$

are defined for almost all z and $|f_{\bar z}| < |f(z)|$ for almost all $z \in U$.

Definition 3.3 If

$$K_{z_0}(f) := \left| \frac{|f_z(z_0)| + |f_{\bar z}(z_0)|}{|f_z(z_0)| - |f_{\bar z}(z_0)|} \right|$$

satisfies $K_{z_0}(f) \le K$ for almost all z_0, with respect to the Lebesgue measure class, then we say that f is K–quasiconformal.

Since the quantity $K_{z_0}(f)$ is invariant under holomorphic changes of coordinates on U and V, we say that a homeomorphism between Riemann surfaces is K–quasiconformal if it is K–quasiconformal with respect to a choice of holomorphic coordinate charts; by the above discussion the choice of charts does not matter. It is a theorem that 1–quasiconformal maps are actually conformal.

The Teichmüller metric also has the following characterization:

Theorem 3.4 (Teichmüller) *If (Y_1, f_1) and (Y_2, f_2) represent two points in $\mathcal{T}_{g,n}$, then their distance is at most $\frac{1}{2}\log(K)$ if and only if there is a K–quasiconformal homeomorphism $g: Y_1 \to Y_2$ such that $f_2 \circ g \circ f_1^{-1}$ is homotopic to the identity on S. Teichmüller maps are the unique dilatation-minimizing maps in their homotopy classes (that is, they are K–quasiconformal for the smallest possible K).*

With this metric, $\mathcal{T}_{g,n}$ is homeomorphic to $\mathbb{R}^{6g-6+2n}$. This manifold admits a complex structure, and $\mathrm{QD}(\mathcal{T}_{g,n})$ is a holomorphic vector bundle over $\mathcal{T}_{g,n}$ with the zero section deleted (as we have defined it). However we will not need to use this complex structure.

In addition to quadratic differentials that we used to characterize the Teichmüller metric, we will sometimes consider quadratic differentials on $\mathbb{C}$ of the form $f(z)\,dz^2$, where f is a polynomial or a rational function with at most one pole in $\mathbb{C}$, which is simple. They extend to $\widehat{\mathbb{C}}$ with a higher-order pole at ∞. The induced flat metrics on $\mathbb{C}$ have infinite area, and the distance to ∞ is infinite in such metrics.

We can use Teichmüller maps to create various bundles over Teichmüller space. We will pick a basepoint $X \in \mathcal{T}_{g,n}$. If q is a quadratic differential over the Riemann surface X, we can associate to q the Teichmüller map ϕ_q that moves X along the geodesic described by q by a distance equal to the area of (X, q). This identifies Teichmüller space with the space of quadratic differentials over X, and we can use coordinates (x, q) to refer to the point $\phi_q(x)$.

We can also trivialize the bundles of quadratic differentials and half-translation surface structures using Teichmüller maps. To do so, we need the following definition from [Fathi et al. 2012] and theorem from [Hubbard and Masur 1979]. In the definition below, our base surface $S_{g,n}$ and all Riemann surfaces are understood to have n points *deleted*.

Definition 3.5 Let q_1 and q_2 be vertical foliations of quadratic differentials on two Riemann surfaces $X_1, X_2 \in \mathcal{T}_{g,n}$, with vertical measured foliations $\mathcal{F}_v(q_1)$ and $\mathcal{F}_v(q_2)$, respectively. For a simple closed curve C on the base surface $S_{g,n}$ with the n points *deleted*, let $i(\mathcal{F}_v(q_j), C)$ denote the infimum of the transverse measure evaluated on curves homotopic to C in the q_j–metric. If $i(\mathcal{F}_v(q_1), C) = i(\mathcal{F}_v(q_2), C)$ for all C, we say that $\mathcal{F}_v(q_1)$ and $\mathcal{F}_v(q_2)$ are equivalent and we write $\mathcal{F}_v(q_1) = \mathcal{F}_v(q_2)$.

Definition 3.6 The set of all equivalence classes of measured foliations is the space $\mathcal{MF}$. Its quotient by the action of $\mathbb{R}^{>0}$ by scalars is *PMF*.

The values of the numbers $i(\mathcal{F}, C)$, where C ranges over all (homotopy classes of) essential simple closed curves, give a weak-$*$ topology on the set of measured foliations, which gives $\mathcal{MF}$ the homeomorphism type of the product of $\mathbb{R}^{6g-6+2n} \setminus \{0\}$; see exposé 6 of [Fathi et al. 2012]. Moreover, if we quotient by the obvious $\mathbb{R}^{>0}$ action by scalars, the result is homeomorphic to a sphere of dimension $6g - 7 + 2n$.

Theorem 3.7 [Hubbard and Masur 1979] *The map from* $\mathrm{QD}(\mathcal{T}_{g,n})$ *to* $\mathcal{T}_{g,n} \times \mathcal{MF}$ *defined by*

$$(X, q) \mapsto (X, \mathcal{F}_v(q))$$

is a homeomorphism.

Theorem 3.8 [Hubbard and Masur 1979] *The map from the set of unit-area quadratic differentials in* $\mathrm{QD}(\mathcal{T}_{g,n})$ *to* $\mathcal{T}_{g,n} \times \mathcal{PMF}$ *defined by*

$$(X, q) \mapsto (X, \mathbb{R}^{>0}\mathcal{F}_v(q))$$

is a homeomorphism.

Corollary 3.9 [Hubbard and Masur 1979] *Pick* $(X, q) \in \mathrm{QD}(\mathcal{T}_{g,n})$. *For each* Y *in* $\mathcal{T}_{g,n}$, *there is a unique* $q' \in T^*_Y\mathcal{T}_{g,n}$ *such that* $\mathcal{F}_v(q) = \mathcal{F}_v(q')$. *Moreover, given any basepoint* $Y \in \mathcal{T}_{g,n}$, *there is a trivialization of the vector bundle* $T^*\,\mathrm{QD}(\mathcal{T}_{g,n})$ *over* $\mathrm{QD}(\mathcal{T}_{g,n})$ *that sends each* (X, q) *to* (X, q'), *where* q' *is the unique quadratic differential on* Y *whose vertical foliation is equivalent to that of* q.

We note that the foliation of $\mathrm{QD}(\mathcal{T}_{g,n})$ whose leaves are the sets of quadratic differentials with equivalent vertical foliations have the following property: if we fix a system of period coordinates, then moving along a leaf of this foliation only changes the imaginary parts of the period coordinates.

Now, we describe a trivialization of another bundle. Let q be a quadratic differential over the basepoint surface X, let q' be a nonzero quadratic differential over X, and let $x \in \overline{X}$, the completion of X. We can use coordinates (x, q, q') to refer to the point $\phi_q(x)$ on the surface $\phi_q(\overline{X})$ equipped with the metric given by the unit-area quadratic differential on $\phi_q(\overline{X})$ whose vertical foliation is equivalent to $\mathcal{F}_v(q')$. The union of all such points (x, q, q') is the *universal half-translation surface of type* $S_{g,n}$. Similarly, we can take the universal cover of this bundle, which results in replacing each fiber with its universal cover in our coordinates (Teichmüller maps lift to universal covers). The coordinates are not particularly important, but the induced manifold topologies will be used for various compactness statements — for instance, we may refer to a sequence of saddle connections on quadratic differentials (X_i, q_i) converging to a saddle connection on (X, q). This means that the convergence is in the Hausdorff metric on the space of compact subsets of the bundle. In the case when the surfaces in question have genus 0 and are explicitly uniformized to $\widehat{\mathbb{C}}$, we have an alternative but topologically equivalent trivialization of the same bundles, so the notion of Hausdorff convergence is the same.

Definition 3.10 Given a quadratic differential (X, q), we may form a branched double cover as follows: Let $(\widetilde{X}, \tilde{q})$ be the metric completion of a metric double cover of the set of nonsingular points of X given by $\{(X, v) : v \text{ is a vertical unit vector with respect to } q\}$. We call this the *orienting double cover* of (X, q), and it admits a degree-2 map to X that is branched over all points where the cone angle is of the form $(2n-1)\pi$ for $n \in \mathbb{N}$.

The orienting double cover is the square of a holomorphic one-form on $\widetilde{X}$, and $\widetilde{X}$ is connected only if q is not the square of a holomorphic one-form. We can see that the orienting double cover comes with a flat metric and a unit vertical vector field at all nonsingular points. Moreover, the map $i : X \to X$ defined by $(x, v) \mapsto (x, -v)$ is an involution fixing the ramification points. If Σ is the set of singular points of X, then $\widetilde{\Sigma}$ is the preimage of the set of singular points in the double cover. Now, on $\widetilde{X}$ we can find a holomorphic 1–form α such that $f^*q = \alpha^{\otimes 2}$, and α is unique up to sign.

3.4 Triangulations and degenerations

Every nonzero quadratic differential admits a cellular decomposition whose 1–cells are saddle connections and whose open 2–cells are isometric to the interiors of triangles in Euclidean space. We shall refer to such decompositions as *triangulations*, even though not every edge has distinct vertices.

There is one triangulation, due to [Masur and Smillie 1991], which can be constructed as follows: If we delete all the poles of a nonzero meromorphic quadratic differential (X, α), we can take the metric universal cover of $X \setminus \{x_1, \dots, x_n\}$ where the x_i are the marked points; let Γ be the deck group of this covering map. This makes sense because on $X \setminus \{x_1, \dots, x_n\}$ the flat metric associated to α is still a path metric. Then complete the resulting space $\widetilde{X}$. (The new points added will be points of cone angle ∞.) Then, for each point in $\widetilde{X}$, we consider the set of singularities and cone points of angle ∞ that are closest to X. If for some point p there are three or more such points distance R_p from p, then there is an embedded Euclidean disk of radius $R > 0$ around p. Since the length of the shortest saddle connection is nonzero, there are only finitely many singularities on the boundary of this disk, and they have a cyclic ordering. For each such p draw the chords that are the boundary of the convex hull of these points in the disk, and label them as 1–cells. We now have a Γ–invariant cell decomposition. Now quotient by Γ. The result is a cell decomposition in which the 0–cells are singularities, the 1–cells are saddle connections, and the 2–cells are convex polygons, which can then be triangulated by diagonals.

Definition 3.11 We refer to the above construction as the *Delaunay triangulation*.

The Delaunay triangulation has a number of useful properties. It is invariant under scaling the quadratic differential by nonzero complex numbers, and for a typical surface it does not involve a choice — if it does, only finitely many choices are possible. (By a typical surface, we mean a full-measure set with respect to a Lebesgue measure, described below.) One of them is that the lengths of the saddle connections are bounded by twice the diameter of the flat metric, and another is that all of the angles of the triangles are bounded away from zero given an upper bound on the diameter of the surface and a lower bound on the length of the shortest saddle connection.

A triangulation by saddle connections, together with the collection of cone angles, imposes relations between the periods of the saddle connections. Given a triangulation, there is some number of periods that determines all of the others.

In [Masur 1982] the following complex manifold structure on strata in $\mathrm{QD}(\mathcal{T}_{g,n})$ was described, using the homology of the orienting double cover:

Definition 3.12 Let (X, q) be a quadratic differential with double cover $(\widetilde{X}, \tilde{q})$, let Σ be the set of singularities of X with inverse image $\widetilde{\Sigma}$, and let ι be the involution of $\widetilde{X}$. Then

$$H_1^{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) := \{\varphi \in H_1(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) : \iota_*(\varphi) = -\varphi\},$$
$$H^1_{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) := \{\alpha \in H^1(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) : \iota^*(\alpha) = -\alpha\}.$$

Similarly, define H_1^{even} and H^1_{even} to be the eigenspaces of ι_* and ι^* with eigenvalue 1.

Since ι^2, ι_*^2 and $(\iota^*)^2$ are the identity map on their domains, we have

$$H_1^{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma}, \mathbb{C}) \oplus H_1^{\mathrm{even}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) = H_1(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C})$$

and

$$H^1_{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) \oplus H^1_{\mathrm{even}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) = H_1(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}).$$

We also have the usual duality between homology and cohomology groups since we are using field coefficients.

Fix a subset S of a stratum of $\mathrm{QD}(\mathcal{T}_{g,n})$ sharing a triangulation T by saddle connections in fixed homotopy classes relative to the singularities; each saddle connection γ on any $(X, q) \in S$ has two lifts γ' and γ'' to $\widetilde{X}$. Fix T as an oriented graph, and assume the edges have names. On the orienting double cover of a (X,q), there is a triangulation $\widetilde{T}$ that maps to T as an oriented graph. For each γ, fix a choice of γ' and γ''. Then $\widetilde{T}$ comes equipped with this choice of names as well. In other words, if we say two surfaces have the same triangulation, then an isomorphism of the two as simplicial complexes determines a name-preserving isomorphism of their orienting double covers as simplicial complexes, which is unique up to composition with the involution. We require that one of the two possible isomorphisms of the double cover preserves the edge names.

Definition 3.13 A triangulation T by saddle connections *degenerates* to a cell decomposition T' if there is a sequence of unit-area half-translation surfaces $\{X_i\}_{i=1}^{\infty}$, all with the same marked triangulation T, that converges in the Hausdorff topology in the universal curve, and each cell of T in the sequence of surfaces X_i converges in the Hausdorff topology to a cell or union of cells of a surface X with the cell decomposition T'. If v is a vertex of T in each X_i, we say that v *limits to* v'. If a collection of vertices all limit to the same $v \in v'$ in X, we say that those vertices *collide*.

Definition 3.14 Let T be a triangulation of S_g, together with a prescribed cone angle for each vertex of T. Assume T has $4g-4+n$ vertices with prescribed cone angle 3π (the zeros of T) and n vertices with cone angle π (the poles of T). We say a simply connected subset A of $\mathrm{QD}(\mathcal{T}_{g,n})$ is T*–convex* if each $(X,q) \in A$ has a triangulation by saddle connections isotopic to T or a degeneration T' of T such that:

- The order of vanishing of q at each vertex v of T' is the sum of the orders of vanishing of all vertices v_k of a sequence (X_i,q_i) triangulated by T and degenerating to v in (X,q). (Given a maximal collection D of vertices joined by degenerate edges of T whose union is connected and contains only nullhomotopic closed curves, the order of vanishing of q at the degenerate singularity is equal to the number of zeros minus the number of poles in D.)

- For each oriented edge γ of T there is a half-plane $H_\gamma \subset \mathbb{C}$ not containing 0 in its interior such that if we vary (X,q) continuously, we always have
$$\int_{\gamma'} \sqrt{\tilde{q}} \in H_\gamma.$$

- For (X,q) let $v(X,q)$ be the vector of periods of saddle connections of (X,q) that belong to T, such that for each edge γ the coordinate is in H_γ. Then for any $t \in (0,1)$ and $(X_1,q_1),(X_2,q_2) \in A$, there is some $(X_3,q_3) \in A$ such that
$$v(X_3,q_3) = t\,v(X_1,q_1) + (1-t)v(X_2,q_2).$$

We will show in Appendix A that there is a locally finite collection of triangulations T_i on $\mathrm{QD}(\mathcal{T}_{g,n})$ such that $\mathrm{QD}(\mathcal{T}_{g,n})$ is a locally finite union of T_i–convex sets, and the sets are invariant under scaling by $\mathbb{R}^{>0}$. The purpose of this definition is to allow us to analyze collisions of singularities while letting them retain their individual identities.

A T–convex subset of a stratum admits an embedding into $H^1_{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{C})$ by integrating a choice of square root of $\tilde{q}$ along relative cycles in $H^1_{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{C})$. We always choose the square root so that

$$\int_{\gamma'}\sqrt{\tilde{q}}\in H_\gamma.$$

The choice of triangulation gives a local trivialization of the vector space $H^1_{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{C})$, so this gives us a system of charts into $\mathbb{C}^n$, where n is the dimension of $H^1_{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{C})$; in fact they form an atlas of charts, and the dimension of the connected component of a stratum is the dimension of $H^1_{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{C})$.

There is a canonical identification

$$H^1_{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{C})=\mathrm{Hom}(H_1^{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{Z}),\mathbb{C})$$

which gives rise to a natural Lebesgue measure on each stratum, by picking a basis for the free $\mathbb{Z}$–module $H_1^{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma};\mathbb{Z})$; we can write any quadratic differential in terms of this basis as an element of $\mathbb{C}^d$ for some d by taking the integral over this basis, and the natural Lebesgue measure on $\mathbb{C}^d=\mathbb{R}^{2d}$ is invariant under the adjoint action by $\mathrm{GL}_n(\mathbb{Z})$. We refer to this measure on any stratum of $\mathrm{QD}(\mathcal{T}_{g,n})$ as the *Masur–Veech* measure. We can obtain a measure on the space of unit-area differentials by a standard cone construction; the coned measure of a set of unit-area differentials is defined to be the measure of the union of their multiples by scalars in $(0,1)$.

It is well known that these measures are finite on the whole stratum, and invariant under the Teichmüller geodesic flow. They were constructed by Masur [1982] and Veech [1986] in order to apply ergodic theory to the study of the Teichmüller flow. It is of use to us that the condition of having a vertical or horizontal saddle connection is Masur–Veech measure 0. If μ and ν are strata and $\nu\subset\bar{\mu}$, then ν is nowhere dense and measure 0 with respect to the Masur–Veech measure on $\bar{\mu}$, in the sense that there are neighborhoods of ν whose intersections with μ have arbitrarily small measure.

4 Quadratic differentials on the sphere

In this section, we collect basic facts about collisions of singularities in quadratic differentials. We show that every cluster of singularities is biholomorphic to a cluster of singularities on a meromorphic quadratic differential on $\widehat{\mathbb{C}}$, and describe a partial compactification of strata in terms of this uniformization.

4.1 Singularities and δ–clusters

Notation 4.1 Let A and B be two positive real quantities that depend on a common variable. We write $A\stackrel{.}{\prec}B$ or $B\stackrel{.}{\succ}A$ if there is a positive constant c such that $A\le cB$. We write $A\stackrel{.}{\asymp}B$ if $A\stackrel{.}{\prec}B\stackrel{.}{\prec}A$.

We will use the following version of Mumford's compactness criterion:

Theorem 4.2 *Let K be a closed set in the moduli space of unit-area quadratic differentials $\mathrm{QD}^1(\mathcal{M}_{g,n})$. For $(X,q) \in K$, let $S(q)$ denote the infimum of the q–lengths of simple closed curves in X which are not nullhomotopic or homotopic to a loop around a puncture. Then $S(q) \dot{\succ} 1$ as (X,q) ranges over K if and only if K is compact.*

This is really a combination of Mumford's original theorem [1971], which was about lengths of the curves in hyperbolic metrics, with any of several other works that can be used to compare q–metrics to hyperbolic metrics. See for instance [Rafi 2007] or [Maskit 1985].

There are two concrete equivalent descriptions of complex manifold structures on strata of $\mathrm{QD}(\mathcal{T}_{0,n})$. Given a quadratic differential, we may assume three of its poles are at 0, 1, and ∞, since it necessarily has four more poles than zeros (counted with multiplicity) and all of the poles are simple. Once this is fixed, it follows that each stratum can be represented locally as a space of differentials of the form $P(z)/(z(z-1)Q(z))\,dz^2$, where the roots of P have prescribed multiplicities, the roots of Q do not repeat and are distinct from 0 and 1, and the number of roots of P of each multiplicity that are also roots of Q is fixed. The locations of the roots of P and Q, together with the ratio of leading coefficients, give local coordinate charts.

It follows easily that, once we restrict to any fixed triangulation, the periods are defined by holomorphic functions of the coefficients on each stratum, or equivalently, by the locations of the singularities in $\mathbb{C} \setminus \{0,1\}$. Periods vary holomorphically with respect to these coordinates. If the locations of singularities vary holomorphically in the plane, we can assume that the endpoints of any saddle connection remain fixed at a and b, and one other singularity remains fixed at ∞ by applying a holomorphic choice of Möbius transformation. Along a saddle connection, $\sqrt{q}$ varies in a 1–parameter family h_t. We can compute the derivative as

$$\frac{d}{dt}\int_a^b h_t(z)\,dz,$$

which we can simply differentiate under the integral over a fixed contour of integration. In fact, even if the endpoints do not remain fixed, differentiation under the integral remains valid when the endpoints are zeros of the differential. If we never allow our singularities to come close to colliding, this is an effective way to estimate the order of magnitude of $d_{\mathrm{Euclidean}}$ as a function of the change of the locations of the singularities. When collections of singularities collide or come close, we can make estimates from two perspectives: one, that they never really collided (we zoom in) or two, that they stay together (we zoom out). To aid us in analyzing these clusters of singularities we make the following definition:

Definition 4.3 Given $0 < \delta < 1$, a metric space X, and a discrete set $S \subset X$, we say that $D \subset S$ is a *δ–cluster* in S if D contains at least two points, and the distance between any two points in D is at most δ times the distance from any point in D to any point in $S \setminus D$.

We will first show that when a subset of the singularities forms a δ–cluster D, and the diameter of D is small compared to the lengths of all nontrivial simple closed curves, there are enough saddle connections connecting singularities in D.

Definition 4.4 Given a graph G and a subset W of the vertices, the *induced* subgraph with respect to W is the maximal subgraph whose vertex set is W.

The Delaunay triangulation has the following property:

Lemma 4.5 *Suppose $(X,(x_1,\dots,x_n),q) \in \mathrm{QD}(\mathcal{T}_{g,n})$ or $\mathrm{QD}(\mathcal{M}_{g,n})$. Let $S(q)$ denote the infimum length of simple closed curves on $X \setminus (x_1,\dots,x_n)$ that are not homotopic to punctures or constant maps, where lengths are taken in the q–metric, and let $\Delta(q)$ be the diameter of X in the q–metric. Then for each $\epsilon > 0$ there is a number $\delta_0 > 0$ such that whenever $S(q)/\Delta(q) > \epsilon$ and $\delta < \delta_0$, for every δ–cluster D in the singularity set of q, the induced subgraph on the vertex set D from the Delaunay triangulation is a connected graph. Moreover, D contains at most one marked point and all cycles in this induced subgraph are nullhomotopic in X.*

Proof We may scale the metric by a real number so that $\Delta(q) = 1$, and we may pick δ to be less than $\frac{1}{8}S(q)$. We may also assume $\delta < \frac{1}{8}$. We note that the distance between two distinct poles is at least half of $S(q)$, so no δ–cluster contains two distinct marked points.

Now suppose that T_1 and T_2 are disjoint subsets of T with no edges between them in the Delaunay triangulation, and $T_1 \cup T_2 = T$. We may consider the completion of the metric universal cover of $X \setminus \{x_1,\dots,x_n\}$. Let v be a saddle connection connecting points that project into T_1 and T_2 whose length is minimal. Indeed, the shortest path from T_1 to T_2 must be a single saddle connection, since if it were a sequence of more than one saddle connection, any intermediate singularity would be closer to the start and end than the start and end are to each other, and must therefore belong to any δ–cluster containing the start and end. Pick a lift of this saddle connection and, starting at its midpoint, move along the locus of points equidistant from its ends until a point is reached that is equidistant from a third singularity. It may be the case that we do not have to move at all. We eventually meet a point p which is equidistant from our two original singularities plus some more, which is the center of a 2–cell in the decomposition into convex polygons, a triangulation of which forms the Delaunay triangulation. Moving around the boundary of the circle in one direction or the other, it must be the case that all singularities we hit are at least as close to one of the ends of the saddle connection we started on the midpoint of as to each other, and therefore they are edges in the Delaunay triangulation (no matter how it is chosen). Thus they are in the original δ–cluster, and the sequence of chords connecting them projects down to sequence of edges in the Delaunay triangulation that connect T_1 and T_2, all of which have both ends in T. □

Notation 4.6 Let (X,q) be a quadratic differential, and let γ be an (oriented) arc such that the endpoints of γ are the only singularities of the q–metric contained in γ. Let $\tilde{\gamma}^1$ and $\tilde{\gamma}^2$ be the two oriented lifts of γ in the orienting double cover $(\widetilde{X},\tilde{q})$.

Lemma 4.7 *Given a quadratic differential $(X, q) \in \mathrm{QD}(\mathcal{T}_{0,n})$ with a simple pole at ∞ and a collection of saddle connections forming a tree B whose vertex set is the set of singularities of (X, q) that are not ∞, let $\gamma_1, \dots, \gamma_r$ be oriented saddle connections in X and let γ_i'' and γ_i' be the lifts of γ_i in the orienting double cover of (X, q). Then $\{[\gamma_i^2] - [\gamma_i^1] : 1 \le i \le r\}$ is a basis for $H_1^{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C})$.*

Proof In $\widetilde{X}$, the inverse image of B is a graph $\widetilde{B}$, and $\widetilde{X} \setminus \widetilde{B}$ is a ramified double cover of $\widehat{\mathbb{C}} \setminus B$ whose only branching is of order 2 over ∞; call the ramification point $\widetilde{\infty}$. This is biholomorphic to a disk. Thus $\widetilde{B}$ is the 1–skeleton of a CW decomposition of $\widetilde{X}$ with a unique 2–cell, which we may call U. The boundary of U is zero (which can be seen since $H^2(X, \mathbb{Z}) \neq 0$ and there is only one 2–cell, or alternatively because its boundary wraps twice around the graph, on opposite sheets, and hence ∂U traverses each 1–cell twice in each direction). So $\widetilde{B} \hookrightarrow X$ induces isomorphisms $H_i(\widetilde{B}; \mathbb{C}) \to H_i(\widetilde{X}; \mathbb{C})$ for $i \le 1$. Since $\widetilde{B}$ is invariant with respect to the involution, the isomorphism on the first homology respects the even–odd decomposition. So all absolute cycles in $H_1^{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C})$ are homologous to cycles in $H_1^{\mathrm{odd}}(\widetilde{B}, \widetilde{\Sigma} \setminus \{\widetilde{\infty}\}; \mathbb{C})$. It is clear that all of the absolute cycles are in the odd part, for both $\widetilde{B}$ and $\widetilde{X}$, since $\widehat{\mathbb{C}}$ and B have no homology in dimension 1.

From the long exact sequences of the pairs $(\widetilde{X}, \widetilde{\Sigma})$ and $(\widetilde{B}, \widetilde{\Sigma} \setminus \{\widetilde{\infty}\})$ we get natural inclusions

$$H_1(\widetilde{X}; \mathbb{C}) \hookrightarrow H_1(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C}) \quad \text{and} \quad H_1(\widetilde{B}; \mathbb{C}) \hookrightarrow H_1(\widetilde{B}, \Sigma \setminus \{\widetilde{\infty}\}; \mathbb{C}).$$

An element of either quotient group is exactly determined by the image of any cycle representing it under the connecting map into $H_0(\widetilde{\Sigma})$ or $H_0(\widetilde{\Sigma} \setminus \{\widetilde{\infty}\})$. If such an element is odd, it must be the case that its boundary is a linear combination terms of the form $\{p_i - \iota(p_i)\}$ for $p_i \in \Sigma$. Since $\widetilde{\infty}$ is invariant under ι, the exact same relative cycle classes are realizable as well. □

Corollary 4.8 *The periods of B form a holomorphic coordinate chart.* □

A similar argument shows that, when $\tilde{\cdot}$ denotes the preimage of $\cdot$ in the orienting double cover:

Corollary 4.9 *If γ is a simple closed curve containing no singularities which bounds a disk $\mathbb{D} \subset X$, and the singularities in D form a set S, then the periods of any tree with vertex set S and all edges saddle connections contained in $\mathbb{D}$ form a basis for $H^1(\widetilde{\mathbb{D}}, \widetilde{S}; \mathbb{C})$ as a vector space over $\mathbb{C}$.*

Lemma 4.10 *Let $(X, q) \in \mathrm{QD}(\mathcal{T}_{0,n})$. If Σ is its singularity set and p is a pole in Σ, then any Delaunay triangulation has a connected induced subgraph with respect to $\Sigma \setminus \{p\}$.*

Remark This property is shared with the L^∞ Delaunay triangulations considered in Appendix A.

Proof There is only one ray of each slope emanating from p, so there cannot be a saddle connection that starts and ends at p. The edges emanating from p are cyclically ordered counterclockwise, and there is an edge connecting each consecutive pair of them, so for any pair of neighbors of p there is a path in the graph avoiding p. Therefore p is not a cut vertex of the 1–skeleton. □

On compact sets, any two holomorphic coordinate systems are bi-Lipschitz, so the main difficulty will be in dealing with degenerations to lower-dimensional strata. Our main object will be to examine periods of clusters of singularities close to 0.

In the sequel, we will often have to deal with a meromorphic quadratic differential q defined on $\mathbb{C}$, and we will often have to speak of distances between points, diameters of sets, etc in two different metrics: the usual Euclidean metric on $\mathbb{C} = \mathbb{R}^2$ (which does not depend on the choice of q) and the singular metric that depends on q.

Notation 4.11 We use $d_{\mathbb{C}}$ and $\operatorname{diam}_{\mathbb{C}}$ to denote the distance and diameter in the usual metric on $\mathbb{C}$, and d_q and diam_q to denote the distance and diameter in the singular metric defined by q. We use $\operatorname{perim}_{\mathbb{C}}$ and perim_q to denote perimeters of regions, and $\operatorname{rds}_{\mathbb{C}}(A)$ and $\operatorname{rds}_q(A)$ will denote the maximum radius of a ball contained entirely in A with respect to $d_{\mathbb{C}}$ and d_q.

We now discuss a certain class of quadratic differential on $\mathbb{C}$ which we use to model clusters of singularities. These differentials will extend meromorphically to $\widehat{\mathbb{C}}$, but will have higher-order poles at ∞.

Definition 4.12 We say q is a *cluster differential* if is a quadratic differential on $X = \mathbb{C}$ which takes one of the following forms: $q = p(z)\,dz^2$, where p is a monic polynomial of degree at least 2 whose roots sum to 0, or $q = p(z)/z\,dz^2$. If p is of the second type, we say that 0 is a *marked point*.

We consider two cluster differentials to be distinct if one has a marked point and the other does not, even though they may have the same underlying quadratic differential on $\mathbb{C}$. That is to say, 0 is *allowed* to be a root of p.

The sets of cluster differentials with a fixed degree and number of marked points are complex manifolds, with the coefficients of the polynomial as coordinates. The complex dimension of the space of cluster differentials is $\deg(p) - 1$ if the differentials are of the form $p(z)$ and there is no marked point, and $\deg(p)$ if there is a marked point.

Spaces of cluster differentials are also stratified by the numbers and types of singularities, and the periods of any spanning tree of saddle connections forms a local holomorphic coordinate system.

In fact, the periods of the saddle connections in the tree and the cone angles determine the q–metric entirely, since the metric has constant curvature 0 away from the tree. This determines $(\mathbb{C}, q)$ as a metric space, from which the conformal structure can be recovered; this determines the differential and the locations of its singularities up to a Möbius transformation fixing ∞. Since we know that p is monic and we know which direction is vertical, we can recover p up to multiplying all of its singularities by a root of unity. If the singularities have names, then locally there is a unique correct choice. We may write q as $q(z_1, \dots, z_n, w)$ where n is the number of zeros of q and w is the marked point. We are of course constrained to the hyperplane $\sum(z_i) = 0$ if there is no marked point.

If a cluster differential has no poles, then its infinitesimal metric is locally nonpositively curved in the sense of Alexandrov. If a quadratic differential induces a complete Alexandrov nonpositively curved metric on a simply connected Riemann surface, then there is a unique length-minimizing geodesic between every pair of points. In particular, a cluster differential has only finitely many saddle connections if it has no poles. We also see that if we take a branched double cover of a cluster differential branched only over the pole and pull back the differential, we get a cluster differential with no poles and a marked point. All saddle connections on our original cluster differential lift to one of finitely many saddle connections upstairs so there were only finitely many saddle connections on the original cluster differential, and the number of saddle connections is bounded by the number of pairs of singularities in the double cover.

There is an analogous definition of T–convex for cluster differentials: we may say that a set A is Γ*–convex* if Γ is the graph whose vertices are singularities and whose edges are saddle connections, the periods of each saddle connection that is an edge of Γ vary continuously in a half-plane, and we can take convex combinations in period coordinates without leaving A.

Proposition 4.13 *The effect of multiplying the zeros and poles of a cluster differential q by t is to scale all periods by $t^{(m+2)/2}$, where m is the rational function degree of q, and angles between saddle connections meeting at a singularity are preserved.*

Proof We give the proof for $q = p(z)\,dz^2$ without the marked point since the proof with a marked point is identical. The first statement is a straightforward change of variables:

$$\int_{ta}^{tb} [t^m p(z/t)]^{1/2}\,dz = \int_a^b t^{m/2}[p(z)]^{1/2} t\,dz.$$

The two sides are the values of the two periods in question, which clearly differ by $t^{(m+2)/2}$. The only way that the cone angles can differ is by integral multiples of π, but they don't differ at all since we can vary t continuously and cone angles vary continuously with t. □

A similar change of variables tells us the following:

Proposition 4.14 *Fix a stratum $\mathcal{Q}$ of quadratic differential with n singularities in $\mathbb{C}$, and let m be the total number of zeros minus the number of poles (counted with multiplicity). For an n–tuple p of distinct points in $\mathbb{C}$ which can be the singularity set of an element in $\mathcal{Q}$, let q_p be the cluster differential with those singularities (it is assumed that for each coordinate the singularity has a prescribed type). Then for any $v \in \mathbb{C}^n$ and any period P_i of any saddle connection, we have*

$$\frac{d}{dt} P_i(q_{sp+tv})\Big|_{t=0} = s^{m/2}\frac{d}{dt} P_i(q_{p+tv})\Big|_{t=0}.$$

Corollary 4.15 *Let q be a cluster differential of rational function degree m. Let Σ be the set of singularities of q. Then*

$$\operatorname{diam}_q(\Sigma) \stackrel{.}{\asymp} \operatorname{diam}_{\mathbb{C}}(\Sigma)^{(2+m)/2}.$$

Proof By Proposition 4.13 this reduces to the following claim: if $\operatorname{diam}_{\mathbb{C}}(\Sigma) = 1$, then $\operatorname{diam}_q(\Sigma) \stackrel{\cdot}{\asymp} 1$.

We will only deal with the case in which there is a marked point since that is more difficult. What we will do is prove an upper bound on the q–diameter of the $d_{\mathbb{C}}$ ball of radius 1 about the marked point and a lower bound on the q–length of an arc whose endpoints belong to Σ and are distance 1 apart with respect to $d_{\mathbb{C}}$.

Indeed, suppose that the marked point is 0. We will give an upper bound for the q–length of any radius of the unit circle $|z| = 1$, since any two points in Σ can be joined by a path contained in a pair of two such radii. By rotating our coordinate system, we may assume that the radius is the interval $[0, 1] \subset \mathbb{R}$.

If $q = f(z)/z\, dz^2$ and f has degree $m+1$, then for all points $x \in [0, 1]$ we have $|f(x)/x| \le (x+1)^n/x$. Therefore, the q–length of the line segment $[0, 1]$ is bounded by

$$d_q(0,1) \le \int_0^1 x^{-1/2}\, dx.$$

This gives us the upper bound.

For a lower bound, note that a rectifiable arc γ from a to b with $d_{\mathbb{C}}(a,b) = 1$ must travel Euclidean distance at least $\frac{1}{2}$ outside of the following union of $m+2$ disks:

$$\bigcup_{z_i \in \Sigma} \left\{ z : |z - z_i| < \frac{1}{4(m+2)} \right\}.$$

As before, assume that $q = [f(z)/z]\, dz^2$ and $z_1, \dots, z_{m+1}$ are the roots of f. Then

$$\left|\frac{f(z)}{z}\right| \ge \left|\frac{z - z_{m+1}}{z}\right| \prod_{j=1}^{m} |z - z_j| \ge \frac{|z - z_{m+1}|}{1 + |z - z_{m+1}|} \prod_{j=1}^{m} |z - z_j| \ge \frac{1/(4m+8)}{1 + 1/(4m+8)} \frac{1}{(4m+8)^m}.$$

The length of γ is therefore bounded below by

$$\int_\gamma |\sqrt{q}| \ge \frac{1}{2} \left| \frac{1/(4m+8)}{1 + 1/(4m+8)} \frac{1}{(4m+8)^m} \right|^{1/2}. \qquad \square$$

Proposition 4.16 *Let f be a monic polynomial of degree d with $f(z) \neq 0$ whenever $|z| < 1$, and assume $r < 1$. Then there is a constant $C_{r,d}$, depending on r and d but not on f, such that on the disk $\{|z < r|\}$ we have $C_{r,d}^{-1} \le |f(z)/f(0)| \le C_{r,d}$. Moreover, if r is small enough we may take C arbitrarily close to 1.*

Proof For a monic polynomial $z - w$, this follows from the fact that whenever $|w| \ge 1$ and $|z_1|, |z_2| < r$, we have

$$\left| \frac{z_1 - w}{z_2 - w} - 1 \right| \le \frac{|z_1| + |z_2|}{|w| - r} \le \frac{2r}{1 - r}. \qquad \square$$

Notation 4.17 For the purpose of Corollary 4.18, let B_s denote the disk $\{z : z < s\}$.

Corollary 4.18 *Let $r < \frac{1}{4}$. Let $p(z)\, dz^2$ be a cluster differential with a pole of order $m+4$ at ∞ and no singularities in $\mathbb{C}$ outside B_r. Let f and g be polynomials with fixed degrees and no zeros in B_1, and let $q = (f(z)/g(z))\, p(z)\, dz^2$.*

Then

$$d_q(0,\partial B_r) \asymp \left|\frac{f(0)}{g(0)}\right|^{1/2} r^{(2+m)/2}, \qquad \operatorname{rds}_q(B_r) \asymp \left|\frac{f(0)}{g(0)}\right|^{1/2} r^{(2+m)/2},$$

$$\operatorname{diam}_q(B_r) \asymp \left|\frac{f(0)}{g(0)}\right|^{1/2} r^{(2+m)/2}, \quad \operatorname{perim}_q(B_r) \asymp \left|\frac{f(0)}{g(0)}\right|^{1/2} r^{(2+m)/2}.$$

The implied constants depend on m and the degrees of f and g.

Proof For the base case where f and g are both constant, we can apply Corollary 4.15. For other cases, we may simply apply Proposition 4.16. □

Corollary 4.19 *Let q be as in Corollary 4.18. Then for each $C \in (0,1)$ there is a constant $c > 0$ such that if all zeros and poles of p are contained in B_{cR}, any length-minimizing path in the d_q–metric between points in B_{cr} is contained in B_{Cr}.*

Proof By the estimates in Corollary 4.18, if c/C is small enough, then the diameter of B_{cr} is less than the d_q distance from ∂B_{cr} to ∂B_{Cr}. □

The following lemma says that while a faraway singularity may change the size of a δ–cluster, it does little to change the shape:

Lemma 4.20 *Let f, g, and B_s be as in Corollary 4.18. Let $q = p(z)\,dz^2$ be a cluster differential with all singularities in B_r, and let $q' = (f(z)/g(z))p(z)\,dz^2$. Then for each $\epsilon > 0$ there is some $r_0 > 0$ such that whenever $r < r_0$ and $a,b,c,d \in B_r$,*

$$1-\epsilon < \frac{d_q(a,b)d_{q'}(c,d)}{d_q(c,d)d_{q'}(a,b)} < 1+\epsilon.$$

Moreover, if Γ is a tree whose vertices are the singularities of p and whose edges are saddle connections whose lengths with respect to d_q are as small as possible, then there is a tree Γ' whose endpoints are saddle connections of q' and whose edges are homotopic (rel endpoints) to those of Γ whenever $r < r_0$ is sufficiently small, and angles between corresponding edges differ by less than ϵ.

Proof The first claim is immediate from the pointwise estimates (Proposition 4.16), since the metric is scaled pointwise by a near constant.

For the second claim, we need to show that if we start with our embedding of the graph $\Gamma \subset \mathbb{C}$ and modify it by making its edges geodesic with respect to $d_{q'}$, the edges remain saddle connections. Γ is obtained by a greedy algorithm: pick saddle connections one at a time, repeatedly picking the shortest saddle connection that does not join two vertices belonging to the same connected component of the graph with edges already picked. (This is because for any saddle connection e not used, the fundamental cycle of $\Gamma \cup e$ with e must consist of edges no longer than e.) Therefore, whenever an edge e is drawn, it must be the case that there is no singularity of distance less than the length of e away from both ends of e, and there are two equilateral triangles sharing the edge e with no singularities in the interior of either triangle. Also there are no other singularities on e. It follows that the angle formed by e and any previously drawn edge is at least $\frac{1}{3}\pi$.

The estimate of Proposition 4.16 then implies that the lengths and angles of all tangent vectors are scaled by nearly the same constant, so for any points a and b joined by a q–geodesic with no singularities, we have

$$\frac{\int_a^b \sqrt{q'}}{\int_a^b \sqrt{q}} = \sqrt{f(0)/g(0)}[1+o(1)] \quad \text{as } r_0 \to 0.$$

We may scale the entire differential by a constant without changing the geodesics, so we may assume $\sqrt{f(0)/g(0)} = 1$. Now, if we have a map from a region bounded by the union of two equilateral triangles into $\mathbb{R}^2$ sharing an edge, and its derivative is close enough to $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, then the geodesic in $\mathbb{R}^2$ joining the endpoints is contained in the image of the region. □

Proposition 4.21 *Let f and g be polynomials such that g has no repeated roots, and $\deg(f) \geq \deg(g) - 3$. Let $q = (f(z)/g(z))\,dz^2$ be a meromorphic quadratic differential on $\widehat{\mathbb{C}}$ whose poles are all simple, except perhaps for a higher-order pole at ∞. Assume that the roots of f and g are all contained in B_{r_0} for some $r_0 > 0$. Then for each $r > 0$ there is a number R depending only on r, r_0, and the degrees of f and g, such that the shortest path between any points in B_r with respect to the metric d_q is contained in B_R.*

Proof We will assume f and g are monic, since scaling the metric by a constant doesn't change whether or not a path is length-minimizing. Let $m = \deg(f) - \deg(g)$. As $|z| \to \infty$, we have $|f(z)/g(z)|^{1/2}/|z^{m/2}| \to 1$ uniformly for all q satisfying our hypotheses. Integrating this pointwise bound gives us the following:

The distance from the circle of radius r to the circle of radius r^2 has asymptotics

$$\int_r^{r^2} |\sqrt{q}| \sim \int_r^{r^2} |z^{m/2}\,dz| \sim \begin{cases} 2r^{-1/2} & \text{if } m = -3, \\ \log(r) & \text{if } m = -2, \\ 2r^{m+2}/(m+2) & \text{if } m \geq -1. \end{cases}$$

As $r \to \infty$, the ratio between the actual q–distance between ∂B_r and ∂B_{r^2} and the asymptotic value converges to 1 at a rate that depends only on r_0, and the degrees of f and g.

The q–lengths of semicircles on ∂B_r are asymptotic to $\pi r \cdot r^{m/2}$. Thus, the distance between two points on ∂B_r is at most $[1+o(1)]\frac{\pi}{4}$ times the distance from ∂B_r to ∂B_{r^2} and back when $m = -3$, and $o(1)$ times the distance from ∂B_r to ∂B_{r^2} and back for $m \geq -3$. □

We could have used B_{cr} in place of B_{r^2} and gotten the same conclusion in the case $m = -3$, for any sufficiently large C. By performing the change of coordinates $z \mapsto 1/z and\ c = 1/C$, we conclude the following:

Proposition 4.22 *If a quadratic differential q on $\widehat{\mathbb{C}}$ has only simple poles, one occurring at $z = 0$, and no other poles or zeros in the ball B_r, then there is some $c \in (0,1)$, depending only the number of zeros and poles of q of each multiplicity, such that no length-minimizing path from two points outside of B_r passes through B_{cr}.* □

Proposition 4.23 *Fix integers $e_1, \dots, e_k \geq -1$ with $\sum_{i=1}^k e_i \geq -3$. Let $q(\alpha, z_1, \dots, z_k)$ denote the quadratic differential $\left[\prod_{i=1}^k (z - z_i)^{e_i}\right] dz^2$ on $\mathbb{C}$. (It will be meromorphic on the Riemann sphere with a pole at ∞ which is simple if and only if $\sum e_i = -3$.) Assume the set of tuples $(\alpha, z_1, \dots, z_k)$ is restricted to the set where $z_i \neq z_j$ if $e_i = e_j = -1$. Then for all $i, j \in \{1, \dots, k\}$, we have that $d_{q(\alpha, z_1, \dots, z_k)}(z_i, z_j)$ varies continuously with respect to $(\alpha, z_1, \dots, z_k)$.*

Proof Let $\epsilon > 0$. By the estimates of Corollary 4.18, the q–diameter, radius, perimeter, and distance from the boundary to the point z_i of

$$B_{r_i}(z_i) := \{z : |z - z_i| < r_i\}$$

are all $O(r_i^{1/2})$ as $r \to 0$. We pick around all singularities a ball small enough that each has diameter less than ϵ, and also a ball B_∞ about ∞ such that no length-minimizing path between z_i and z_j can enter. We can pick the balls much smaller than the distance between any two points z_i and z_j unless $z_i = z_j$. In this case we pick $r_i = r_j$ in a neighborhood of $(\alpha, z_1, \dots, z_k)$. We can fix $r_1, \dots, r_k$ and α so that each of these balls has diameter, radius, and perimeter less than ϵ in the q–metric in a neighborhood of $(\alpha, z_1, \dots, z_k)$. Then, as $(\beta, y_1, \dots, y_k)$ converges to $(\alpha, z_1, \dots, z_k)$, the sizes of the balls $B_i(z_i)$ still satisfy

$$\operatorname{diam}_{q(\beta, y_1, \dots, y_k)}(B_{r_i}(z_i)) \dot{\prec} \epsilon.$$

The distances $d_{q(\beta, y_1, \dots, y_k)}(B_{r_i}(z_i), B_{r_j}(z_j))$ vary continuously since the metric varies smoothly on $\widehat{\mathbb{C}} \setminus \left(B_\infty \cup \bigcup_{i=1}^k B_{r_i}(z_i)\right)$. Since length-minimizing paths from $B_{r_i}(z_i)$ to $B_{r_j}(z_j)$ stay in this region, it follows that $d_{q(\beta, y_1, \dots, y_k)}(y_i, y_j)$, viewed as a function of $(\beta, y_1, \dots, y_k,)$, can be written as a sum of a continuous function and a function taking values in $[-\epsilon, \epsilon]$. Since ϵ was arbitrary, the proposition follows. □

Definition 4.24 A *Möbius normalization* of $\mathcal{T}_{0,n}$ or $\mathrm{QD}(\mathcal{T}_{0,n})$ is a collection of uniformization maps of the underlying Riemann surfaces to $\widehat{\mathbb{C}}$ such that the points mapped to 0, 1, and ∞ define three continuous sections of the universal curve or universal half-translation surface, and all each of these three sections intersects each fiber at a marked point. A *Möbius normalized* collection A of quadratic differentials on $\mathcal{T}_{0,n}$ is the set of quadratic differentials on $\widehat{\mathbb{C}}$ that pull back to elements of A across the uniformization maps.

Proposition 4.25 *The distance $d_{q(\alpha, z_1, \dots, z_k)}(a, b)$ is jointly continuous in $(\alpha, z_1, \dots, z_k, a, b)$.*

Proof We can just pretend a and b are singularities z_{k+1} and z_{k+2} with $e_{k+1} = e_{k+2} = 0$ and apply Proposition 4.23. □

Corollary 4.26 *Fix a Möbius normalization of* $\mathrm{QD}(\mathcal{T}_{0,n})$ *and a compact set* $K \subset \mathrm{QD}(\mathcal{T}_{0,n})$ *for which* ∞ *is always a pole. Then for any compact subset* $V \subset \mathbb{C}$ *and any* $\epsilon > 0$, *there is a number* $\delta > 0$ *such that for all* $q \in K$, $d_q(x, y) < \epsilon$ *whenever* $x, y \in V$ *and* $d_{\mathbb{C}}(x, y) < \delta$. *The same is true if we allow* K *to vary over a compact set of cluster differentials.*

Proof The set $\{(x, y, q) : x, y \in V \text{ and } d_q(x, y) \geq \epsilon\}$ is compact so $d_{\mathbb{C}}(x, y)$ attains a minimum there, which we can take to be δ. □

Corollary 4.27 *Suppose* $\{t_m\} \to t_\infty$ *and* $\{u_m\} \to u_\infty$ *are convergent sequences in* $\mathbb{C}$ *and* $\{q_m\} \to q_\infty$ *is a Möbius-normalized convergent sequence in* $\mathrm{QD}(\mathcal{T}_{0,n})$ *or a space of cluster differentials. Then any sequence* $\gamma_m : [0, 1] \to \mathbb{C}$ *of* d_{q_m}*–geodesics from* t_m *to* u_m *which are length minimizing has a subsequence that converges uniformly (as* $\mathbb{C}$*–valued functions) to a constant-speed length-minimizing geodesic in* q_m *along a subsequence.*

Proof By Proposition 4.25 it is clear that any subsequential limit of constant-speed length-minimizing geodesics must be a constant-speed length-minimizing geodesic. By a diagonalization argument, we can pass to a subsequence such that $\gamma_m(t)$ converges for all $t \in [0, 1] \cap \mathbb{Q}$. Uniform convergence follows from Corollary 4.26. □

Corollary 4.28 *Suppose that* $q_n \to q_\infty$ *in* $\mathrm{QD}(\mathcal{T}_{g,n})$ *or a space of cluster differentials. Let* $\gamma_n : [0, 1] \to \mathbb{C}$ *be a geodesic arc for the* q_n*–metric, parametrized to be constant speed, such that the* γ_n *converge to a geodesic arc* γ_∞ *in the* q_∞*–metric that does not pass through any singularities. Then the sequence of maps* γ_n *converges to* γ_∞ *in the* C^1 *topology on* $C^1([0, 1])$.

Proof We already know that the lengths converge. It is easy to see that the directions converge as well, since the arcs formed by concatenating the q_∞ geodesic from $\gamma_\infty(0)$ to $\gamma_n(0)$, the γ_n–geodesic from $\gamma_n(a)$ to $\gamma_n(b)$, and the q_∞–geodesic from $\gamma_n(1)$ to $\gamma_\infty(1)$ are homotopic rel endpoints, via homotopies that pass through no singularities, to γ_∞ for all sufficiently large n. The integrands of $\sqrt{q_n}$ converge to $\sqrt{q_\infty}$ on an open set containing all of these arcs. It thus follows that, if we take the correct branch of the square root,

$$\int_{\gamma_n(0)}^{\gamma_n(1)} \sqrt{q_n} \to \int_{\gamma_\infty(0)}^{\gamma_\infty(1)} \sqrt{q_\infty}.$$

This implies that the complex lengths ℓ_n of the segments $\gamma_n([0, 1])$ converge to some ℓ_∞. By Corollary 4.27 and the fact that the tangent vector that maps $\sqrt{q}$ to 1 varies continuously in (q, z) away from singularities, we conclude that $\ell_i / \sqrt{q}$, the lengths and directions of the tangent vectors $\{\gamma_n'(t) : t \in [0, 1]\}$ converge uniformly in $T\mathbb{C}$. □

Fix a space of cluster differentials, or else fix $c > 0$ and a compact subset K of $\mathrm{QD}(\mathcal{T}_{0,n})$, and assume that for each $(X, q) \in K$, we pick a marking such that ∞ is a pole and such that for all singularities p, we have $d_q(p, \infty) > c\Delta(q)$. (Recall that $\Delta(q)$ is the diameter of the q–metric.) This choice is invariant under the group of Möbius transformations fixing ∞. Then we have the following:

Proposition 4.29 *The notions of δ–cluster with respect to $\mathbb{C}$–metric and q–metric are comparable on K in the following sense*:

- *For each sufficiently small $\delta > 0$ there exists $\delta' > 0$, depending only on δ and K, such that whenever S is a δ–cluster of the singularity set of $q \in K$ with respect to the $\mathbb{C}$–metric, S is a δ'–cluster of singularities with respect to the q–metric.*
- *Conversely, for each sufficiently small $\delta > 0$ we can find $\delta' > 0$, depending only on δ and K, such that whenever S is a δ–cluster of the singularity set of $q \in K$ with respect to the q–metric, S is a δ'–cluster of singularities with respect to the $\mathbb{C}$–metric.*
- *The two statements above are true for strata of cluster differentials in place of differentials in $\mathrm{QD}(\mathcal{T}_{0,n})$.*

In other words, we can tell when a collection of singularities is collapsing strictly faster than any proper superset using either metric.

Proof By scale invariance of the ratios of distances in both metrics, we will assume all quadratic differentials in K are of the form $(P(z)/Q(z))\,dz^2$ with P and Q monic of fixed degrees. P is allowed to have multiple roots and Q is not, and P and Q may have roots in common. By applying Möbius translations fixing ∞, we can also assume that the set Σ of singularities of each q is contained in a ball of radius R about 0 but not in a ball of radius r, for some $R > r > 0$. A closed subset of the space of differentials with these restrictions is compact if and only if there is a lower bound on the distance between any two roots of Q; conversely, for any K we can find some R and r constraining the singularities of q for all $q \in K$, after a Möbius transformation.

Let $\{q_n\}_{n=1}^{\infty}$ be a sequence in K, uniformized to $\widehat{\mathbb{C}}$ so as to satisfy the constraints in the above paragraph, and let S_n be a collection of singularities in the q_n–metrics. Then by Corollary 4.18, $\mathrm{diam}_{\mathbb{C}}(S_n) \to 0$ if and only if $\mathrm{diam}_{q_n}(S_n) \to 0$. We may rephrase this as $\mathrm{diam}_{q_n}(S_n)/\Delta(q_n) \to 0$ if for each $\delta > 0$, S_n is eventually contained in a δ–cluster with respect to $d_{\mathbb{C}}$, if and only if for each $\delta > 0$, S_n is eventually contained in a δ–cluster with respect to d_{q_n}.

Now let $S_n \subsetneq S_n'$ be sets of singularities such that S_n' is a δ_n cluster in q_n, and $\mathrm{diam}_{q_n}(S_n) \to 0$. By Lemma 4.20 we can ignore all singularities outside of S_n' and use the above argument to deduce that $\mathrm{diam}_{\mathbb{C}}(S_n)/\mathrm{diam}_{\mathbb{C}}(S_n') \to 0$ if and only if $\mathrm{diam}_{q_n}(S_n)/\mathrm{diam}_{q_n}(S_n') \to 0$.

The proposition then follows by induction on the total number of singularities. □

Definition 4.30 Let $\delta < \frac{1}{10}$. We say that a saddle connection is *internal* to a δ–cluster D in the q–metric if it stays in the $2\,\mathrm{diam}_q(D)$–neighborhood of D (in the q–metric). (In particular, this includes any saddle connection that is the shortest path from one endpoint to the other.)

Definition 4.31 Given a meromorphic quadratic differential $(P(z)/Q(z))\,dz^2$ on $\mathbb{C}$ whose poles are all simple, we say that a δ–cluster of singularities (where δ–cluster is with respect to the flat metric) is *shrunk* if it contains more zeros than poles (counted with multiplicity). Given a shrunk δ–cluster of singularities, we say that the *center* of the δ–cluster is the weighted average of the singularities, where each point is counted with weight equal to the order of vanishing of $P(z)/Q(z)$ at that point (simple poles count as -1.)

These definitions are motivated by the fact that the diameter in the singular q–metric shrinks much faster than the diameter in the nonsingular metric as a cluster converges to a point and all other singularities remain fixed, and all other periods in a good coordinate system are well approximated by the periods of the quadratic differential that is obtained when we replace the δ–cluster with a single singularity at its center, which is a zero of order equal to the sum of the orders of vanishing of P/Q at the singularities being replaced.

The equivalence of complex coordinate systems then follows because a holomorphic local homeomorphism of complex manifolds is a local biholomorphism — that is, the locations of the singularities in $\mathbb{C}$ are holomorphic functions of the periods.

4.2 Modeling collisions of singularities by cluster differentials

We claim that cluster differentials form a model for a neighborhood of any disk without poles:

Proposition 4.32 *If (X, q) is a simply connected Riemann surface equipped with a holomorphic quadratic differential with finitely many zeros and the metric d_q is complete, then (X, q) is determined up to isometry by the isometry type of the geodesic convex hull of its zeros and the exterior angles at the vertices of the convex hull.*

Proof Assume that a ball of radius r about a singularity p contains all singularities. By Alexandrov nonpositive curvature, there is a unique geodesic between any pair of points. Moreover, geodesic rays from a fixed point p have a circular order from their initial direction, and the directions they exit singularities if they enter the same singularity. We can therefore construct all geodesic rays from p, and reconstruct the complement of the geodesic convex hull of the zeros as a union of sectors of disks of infinite radii. □

Proposition 4.33 *If (X, q) is a (noncompact) simply connected Riemann surface with a complete metric d_q coming from a holomorphic quadratic differential q, and q has finitely many zeros and no poles, then (X, q) is a cluster differential.*

Proof Given (X, q) we will construct a sequence of quadratic meromorphic differentials $\{q_n\}$ on $\widehat{\mathbb{C}}$ that converge to a holomorphic quadratic differential q_∞ on every compact subset of $\mathbb{C}$. Consider the balls $B_R(p)$ in (X, q) for some fixed singularity $p \in X$. Now, the circle $c_R(p)$ of radius R about p has geodesic curvature $1/(R-s)$ at z if the geodesic from p to z last changed direction at a cone point distance s from z. In particular, the geodesic curvature of $c_R(p)$ is between $1/R$ and $1/(R-r)$ if $B_r(p)$ contains all singularities of d_q, except at finitely many points where the curvature is undefined, and never changes sign where the curvature is discontinuous.

Step 1 (construction of $\{q_n\}$) There are geodesic tangent lines to $c_R(p)$, and a finite number of these lines are vertical or horizontal. For large R, the number of such lines remains fixed, their points of tangency vary continuously in R, and they alternate vertical and horizontal. If $V(R)$ and $H(R)$ are consecutive vertical and horizontal tangent lines with respect to the circular order, then V intersects H. The convex hull of all such intersection points is a piecewise geodesic arc with right-angles at its singularities, and it bounds a disk D_R. We will take (X_n, q_n), a sequence of quadratic differentials, to be the union of disks D_{R_n} with their Schwarz reflections about the boundary (as metric spaces). Note that this Schwarz reflection gives us a canonical way to extend the vertical and horizontal foliations, and creates cone points of angle π at the corners of ∂D_R. However, we have not yet described the sequence of uniformization maps on X_n that makes q_n convergent on compact sets.

Step 2 (picking coordinates for $\{X_n\}$) We note that if we rescale q_n to have unit area, then the d_q distances between poles remain bounded below. (The area of D_R is a quadratic polynomial in R for large R, and the distances between poles grow linearly.) Thus we can apply Mumford's compactness criterion. By the equivalence of δ–clusters in d_{q_n} and $d_{\mathbb{C}}$, for any sufficiently small positive δ there is an $N < \infty$ such that set of zeros on D_{R_n} forms a δ–cluster in $\mathbb{C}$ for all $n > N$ in both $d_{\mathbb{C}}$ and d_q. However, it is also true that for any sufficiently small $\delta > 0$ no proper subset of the zeros in D_{R_n} forms a δ–cluster, for any sufficiently small δ and large n. Fix any pair of vertical and horizontal tangent lines V and H to have intersection equal to ∞ under choice of uniformizations $X_n \to \widehat{\mathbb{C}}$. Then, there is a sequence of Möbius transformations $T_n(z) = a_n z + b_n$ such that the collection of zeros converges in the configuration space $\mathrm{Conf}_n(\mathbb{C})$, along a subsequence, since ratios of $\mathbb{C}$–distances between zeros in D_{R_n} remain bounded above and below. Moreover, the locations of all other singularities go to ∞.

Step 3 (proving uniform convergence on compact sets) It follows that on $\mathbb{C}$, each q_n takes the form $p_n(z)h_n(z)\,dz^2$, where we may assume (passing to a subsequence if necessary) that $p_n(z)$ has convergent coefficients, and h_n is a rational function whose zeros and poles tend to ∞ uniformly in n. Therefore, there are numbers C_n such that $h_n(z) = C_n(1 + o(1))$ on any ball as $n \to \infty$. Moreover, since the periods of saddle connections converge, h_n must converge uniformly to a *fixed constant* on each compact subset of $\mathbb{C}$. Moreover, judicious choices of a_n and b_n allow us to assume that p_n is monic and its roots sum to 0. □

Corollary 4.34 *Let B be the closed ball of d_q–radius 1 in a quadratic differential (X, q) about a point $p \in X$. If B has compact closure in X and is homeomorphic to a closed disk, and q has no poles except possibly at p, then the interior of B is isometric to a ball in a cluster differential.*

Proof The case of no poles is obvious, and the case of one pole follows from taking a double cover branched only over the pole. □

Definition 4.35 Let μ be a set whose elements correspond to singularities of a quadratic differential in $\mathcal{T}_{g,n}$ or a space of cluster differentials. To each element of μ we associate an integer no less than -1, corresponding to order of vanishing, and a boolean variable which is 1 or 0 for marked or unmarked points, respectively. Let $\mathcal{Q}(\mu)$ be the corresponding stratum in the Teichmüller space of half-translation surfaces.

For each stratum $\mathcal{Q}(\mu)$ of quadratic differential or cluster differential with at least one zero and at least two singularities, there is a partial compactification $\overline{\mathcal{Q}}(\mu)$, each of whose elements consist of the following data, up to an equivalence relation:

- S_ℓ is a collection of subsets of μ of size at least 2, such that each S_ℓ contains at most one singularity corresponding to a marked point. We require that any two subsets in this collection are either nested or disjoint. We also require $S_0 = \mu$ to be one of the subsets. Let m_ℓ be the sum of the orders of vanishing of singularities in S_ℓ (poles count as -1).
- T is a directed tree whose vertices consist of the sets S_ℓ. Let the edges be such that there is a directed path from v_{ℓ_1} to v_{ℓ_2} if and only if $S_{\ell_1} \subset S_{\ell_2}$. For each vertex $S_\ell \neq S_0$, let $\phi(\ell)$ be the unique subset of μ such that there is a directed edge from S_ℓ to $S_{\phi(\ell)}$.
- (X_ℓ, q_ℓ) is a quadratic differential with one singularity for each element p of μ in S_ℓ not in any $S_j \subsetneq S_\ell$, and one singularity for each $S_j \subsetneq S_\ell$. Each p not in any S_ℓ corresponds to the same type of singularity as p in μ, and for each S_j there is a singularity whose order of vanishing is the sum m_j of orders of vanishing of the singularities in S_j, and which is a marked point if and only if S_j contains a marked point.
- We require that (X_0, q_0) have genus g with n marked points. For $\ell \neq 0$ we require that (X_ℓ, q_ℓ) be a cluster differential with a pole of order $m_\ell + 4$ at ∞, with a marked point if and only if S_ℓ has a marked point. We also require that the $d_{\mathbb{C}}$–diameter of the set of singularities of S_ℓ is 1.
- To each oriented saddle connection γ_i of the quadratic differential corresponding to vertex $\phi(S_\ell)$ that starts at the singularity corresponding to S_ℓ, we associate an angle $\theta(\gamma_i) \in \mathbb{R}/2\pi\mathbb{Z}$ such that the direction of γ_i is counterclockwise from the direction of any other such saddle connection γ_k by a (cone) angle of $[\theta(\gamma_i) - \theta(\gamma_k)] \cdot \frac{1}{2}(m_\ell + 2)$.

The equivalence relation is generated by the condition that if $q_\ell = f(z)\,dz^2$ with the collection of angles $\theta(\gamma_k)$, it is equivalent to $Af(e^{i\alpha}z)[d(e^{i\alpha}z)]^2$ and the collection of angles $\theta(\gamma_k) - \operatorname{Re}\alpha$ for $A \in \mathbb{R}$ and $\alpha \in \mathbb{C}$.

This equivalence relation is simply because we can't distinguish isometric cluster differentials under coordinate changes on $\mathbb{C}$ or scaling by real constant.

This is similar to the compactification of strata in [Bainbridge et al. 2019], except that we do not consider quadratic differentials that escape to ∞ in the moduli space of quadratic differentials, and we consider the limiting objects to be different if their components are rotated relative to each other but are less specific with regard to the relative sizes of different components.

We will not need to construct the topology of $\overline{\mathcal{Q}}(\mu)$, though it can be constructed real analytically by repeatedly taking blowups (over $\mathbb{R}$) of the closure of $\mathcal{Q}(\mu)$ in $\mathcal{QD}(\mathcal{T}_{g,n})$ along the loci where singularities collide and taking finite branched covers of the blown up subspace. Such a construction would go beyond our scope here. However, for our purposes it is enough describe when a sequence in $\mathcal{Q}(\mu)$ converges to an element of $\overline{\mathcal{Q}}(\mu)$.

Definition 4.36 Let $\{(Y_N, q_N)\}_{N=1}^{\infty}$ be a sequence of quadratic differentials in $\mathcal{Q}(\mu)$. We say the sequence converges to a given element of $\overline{\mathcal{Q}}(\mu)$ if for every $\delta \in (0,1)$ there exists some $M < \infty$ such that:

- $d_{\text{Euclidean}}((Y_N, q_N), (X_0, q_0)) < \delta$ whenever $N > M$.
- δ–clusters in the set of singularities of (Y_m, q_m) correspond exactly to the sets S_ℓ whenever $M > N$. That is, they have the same number of singularities of each type, same nesting, same graph of inclusions, etc. Call these clusters $S_\ell(N)$. For each ℓ we will refer to the sequence $S_\ell(N)$ as a *vanishing cluster*.
- For each S_ℓ, $\ell \neq 0$, there is a disk $\mathbb{D}_\ell(N) \subset Y_N$ containing only the singularities in $S_\ell(N)$. For some $\lambda_N \in \mathbb{C}^\times$, $(\mathbb{D}_\ell(N), d_{\lambda_N q_N})$ is locally isometric to the disk $\{z : z < 1/\delta\}$ in the q'_ℓ–metric on $\mathbb{C}$, for some cluster differential $(\mathbb{C}, q_\ell^N)$, where $d_{\text{Euclidean}}(X'_\ell, q_\ell^N) < \delta$.
- For each oriented saddle connection γ_k, if we pick a sequence of saddle connections γ_k^N converging to γ_k in the universal half-translation surface containing $q_{\phi(\ell)}$, and the choice of disks and isometries above, the $d_{q_i^N}$ ray in $\mathbb{C}$ that has the same initial location and direction as the image of γ_k, when parametrized by arc length, is of the form $\mathbb{R}(t)e^{i\alpha(t)}$ for $t \in [0,\infty)$ and $|\theta(\gamma_k) - \lim_{t\to\infty} \alpha(t)| < \epsilon \pmod{2\pi/\mathbb{Z}}$.

Proposition 4.37 *Every convergent sequence in* $\mathrm{QD}(\mathcal{T}_{g,n})$ *consisting of elements of* $\mathcal{Q}(\mu)$ *has a subsequence that converges to an element of* $\overline{\mathcal{Q}}(\mu)$.

Proof This is just a simple matter of extracting subsequences, since the space of cluster differentials with $\operatorname{diam}_{\mathbb{C}} = 1$ is compact for each stratum of cluster differentials, as well as the fact that the space of directions θ is compact. □

We remark that there is an equivalent characterization of the convergence to an element of $\overline{\mathcal{Q}}(\mu)$ when the underlying Riemann surface is $\widehat{\mathbb{C}}$ and the differentials are Möbius normalized. In this case, there is a sequence of Möbius transformations $T_i(N)$, all fixing ∞, such that the images of singularities in $S_\ell(N)$ under $T_i(N)$ converge to those of a suitable cluster differential. By Propositions 4.13 and 4.16, and Lemma 4.20, it follows that this sequence of cluster differentials can be rescaled to converge in the Euclidean metric so that the limit is isometric to $(\mathbb{C}, d_{q_i})$. It should also be noted that if q_i has no poles, then for some numbers $\lambda_N \in \mathbb{C}$, $\{\lambda_N (T_N^*)^{-1}(q_N)\, dz^{-2}\}_{N=1}^{\infty}$ is a uniformly convergent sequence of holomorphic functions on each compact subset of $\mathbb{C}$. For each vanishing cluster, let $z_{j,N}$ be a sequence of elements corresponding to a singularity in $S_\ell(N)$. After possibly permuting singularities of the same type, we conclude that there are sequences of numbers $a_N \in \mathbb{R}$ and $b_N \in \mathbb{C}$ such that $T_N(z) = a_N(z) + b_N$, with $a_N \to \infty$, such that the singularities $T_N(z_{j,N})$ all converge to the singularities of the differential q_i. In summary, we have the following:

Proposition 4.38 *Let* $\{q_N\}_{N=1}^{\infty}$ *be a Möbius normalized sequence of quadratic differentials on* $\widehat{\mathbb{C}}$ *in a stratum of cluster differentials or in* $\mathrm{QD}(\mathcal{T}_{0,n})$*, converging to an element of* $\overline{\mathcal{Q}}(\mu)$ *with a pole at* ∞*. If* S_ℓ *is a vanishing cluster, then after permuting singularities* $z_{j,N}$ *of the same type in* $S_\ell(N)$*, there exist Möbius transformations* $T_N(z) = a_N z + b_N$ *for* $a_N \in \mathbb{R}^+$ *and* $b_N \in \mathbb{C}$ *with* $a_N \to \infty$*, such that the singularities* $T_N(z_{j,N})$ *converge to the singularities of the differential* q_i*. There are scalars* $\lambda_N \in \mathbb{C}$ *such* *that $\lambda_N((T_N)^*)^{-1}$ converges uniformly on compact subsets of $\mathbb{C}$ if S_ℓ does not contain a pole, and on compact subsets of $\mathbb{C} \setminus \{0\}$ if S_ℓ contains a pole.* □

Because of the convergence of the rescaled quadratic differentials, we may also conclude the following version of Corollaries 4.27 and 4.28:

Corollary 4.39 *Let λ_N, T_n, q_N, and $S_\ell(N)$ be as in Proposition 4.38. If $\gamma_N : [0,1] \to \mathbb{C}$ is a sequence of constant-speed saddle connections internal to $S_\ell(N)$, then $T_N(\gamma_N)$ has a Hausdorff convergent subsequence. If the limit is a saddle connection in q_ℓ then the convergence is C^1 on compact subsets of $(0,1)$.* □

We conclude this section by collecting several simple facts about cluster differentials, whose proofs we defer to Appendix B. These results are needed for the proof of Lemma 5.11.

Definition 4.40 Let $\delta \in (0,1)$. A *possibly collapsed* δ–cluster is a δ–cluster or a single point.

Proposition 4.41 *Let K be a compact subset of $\mathrm{QD}^1(\mathcal{T}_{g,n})$ and let $(X,q) \in K$ have singularity set Σ. Let $\gamma : [0,1] \to \mathrm{QD}(\mathcal{T}_{g,n})$ be a rectifiable path with respect to the Euclidean metric on $\mathrm{QD}(\mathcal{T}_{g,n})$ starting at $\gamma(0) = (X,q) \in K$. We write $\gamma(t) = (X(t), q(t))$, and we let $\Sigma(t)$ be the singularity set of $\gamma(t)$.*

Let $\Sigma' \subset \Sigma$, let $\Sigma'(t)$ be a finite subset of $X(t)$ that varies continuously in the Hausdorff topology along $\gamma(t)$, and assume that $\Sigma'(0)$ is a possibly collapsed δ–cluster for some $\delta > 1$.

Then for any $\delta_1 \in (0,1)$ there are positive real numbers $C, \delta_2 > 0$, depending only on K and δ_1, such that the following holds: if $\Sigma'(0)$ is a δ_2–cluster of Euclidean length at most $C \operatorname{diam}_q(\Sigma)/\delta_2$ in $\mathrm{QD}(\mathcal{T}_{g,n})$, there is a unique way to choose $\Sigma'(t)$ for all t such that the number of marked points in $\Sigma(t)$ and the number of zeros minus poles (counted with multiplicity) remain constant along γ, even if the cardinality of $\Sigma'(t)$ does not remain constant.

Proposition 4.42 *Let $K \subset \mathrm{QD}(\mathcal{T}_{g,n})$ be compact. Let U be a contractible open subset of $\mathrm{QD}(\mathcal{T}_{g,n})$ with closure in K. Suppose that there is a continuous choice of possibly collapsed δ–cluster in U, that is to say, a continuous function h from U to the space of compact sets in the universal half-translation surface of type g,n (endowed with the topology of the Hausdorff metric) that maps each element $(X,q) \in U$ to a finite collection of singularities of (X,q), such that $h(X,q)$ is a possibly collapsed δ–cluster in the singularity set of (X,q), and moreover the number of marked points and sum of the orders of vanishing of quadratic differentials at points in the image of $h(X,q)$ are constant. (By Proposition 4.41, if δ is small enough and the diameter of U is small enough, then any such function h is uniquely determined by its value at any point in U.)*

Let M be the moduli space of cluster differentials with the same number of marked points and total order of vanishing as $h(X,q)$. Then there is a number δ_K such that if $\delta < \delta_K$ and U and M are given Euclidean metrics in the sense of Definition 3.2, there is a Lipschitz map $F : U \to M$ with the following property: a disk containing all of the singularities of $F(X,q)$ maps isometrically into (X,q) so that singularities of $F(X,q)$ map to $h(X,q)$ and the vertical foliation is preserved. The Lipschitz constant depends only on K.

Proof This is immediate from local finiteness of period coordinate systems coming from bounded length saddle connections, ie from Proposition A.1. □

Lemma 4.43 *Let $\{q_N\}$ be a sequence in $\mathrm{QD}(\mathcal{T}_{0,n})$ that converges to an element of $\overline{\mathcal{Q}}$, and fix a vanishing cluster $\{S_\ell(N)\}$ for the sequence q_N. Assume q_N is given the normalization such that one pole, which is not part of a vanishing cluster, is at ∞, the center of $S_\ell(N)$ is always 0, and $q(z)\,dz^{-2}$ is a quotient of monic polynomials. Assume the sum of the orders of vanishing of q_N on $S_\ell(N)$ is m.*

Let $z_{j,N}$ range over the noninfinite singularities of q_N, with q_N vanishing at $z_{j,N}$ to order $e_{j,N}$, and let

$$t_N = \prod_{z_{j,N} \notin S_\ell(N)} (-z_{j,N})^{e_{j,N}} \quad \text{and} \quad \alpha_N = \prod_{z_{j,N} \in S_\ell(N)} (z - t_N^{1/(m+2)} z_{j,N})^{e_j}\, dz^2.$$

Let the map F_N be the locally defined map F for the cluster $S_\ell(N)$ to the space of cluster differentials associated to $S_\ell(N)$, as defined in Proposition 4.42. Then, for an appropriate choice of the $(m+2)^{nd}$ root of t_N, $d_{\mathrm{Euclidean}}(\alpha_N, F_N(q_N)) = o(\mathrm{diam}_{q_N}(S_\ell(N)))$.

The following proposition says that if we perturb a quadratic differential but preserve the isometry type of all δ–clusters for some sufficiently small δ, then for a reasonable period coordinate chart the change in the periods is comparable to the Euclidean distance.

Proposition 4.44 *Let K be a compact subset of $\mathrm{QD}(\mathcal{T}_{g,n})$ or a space of cluster differentials, let $L > 0$, and assume that on K the systems of saddle connections associated to the Euclidean metric on K consist of saddle connections of length less than L. Then there is some $\delta_0 > 0$ such that for all $\delta < \delta_0$, the following holds: As X varies in K with singularity set Σ and orienting double cover $(\widetilde{X}, \widetilde{\Sigma})$, let $H^1_\delta(\widetilde{X}, \widetilde{\Sigma}; \mathbb{C})$ denote the subspace of $H^1(\widetilde{X}, \Sigma, \mathbb{C})$ that vanishes on all saddle connections internal to δ–clusters of singularities.*

If $\{X_m\}_{m=1}^\infty$ is a convergent sequence in K with singularities $\{\Sigma_m\}$, Y_m is another such sequence lying on a common convex T–convex period coordinate chart $U \subset \mathbb{C}^n$ with X_m for some T, U is generated by periods of saddle connections of length less than L, X_m is connected to Y_m by a line segment of length $\epsilon_m \to 0$ in U, and X_m and Y_m are such that, in local period coordinates, $X_m - Y_m \in H^1_\delta(\widetilde{X}_m, \widetilde{\Sigma}_m; \mathbb{C}) \cap H^1_{\mathrm{odd}}(\widetilde{X}_m, \Sigma_m, \mathbb{C})$, then $d_{\mathrm{Euclidean}}(X_m, Y_m) \stackrel{\cdot}{\asymp} \epsilon_m$.

5 Perturbing quadratic differentials on the sphere

We now establish a setup to discuss saddle connections in convergent sequences of quadratic differentials on $\widehat{\mathbb{C}}$. We will consider all strata of quadratic differentials with a pole fixed at ∞, and such that all poles in $\mathbb{C}$ are simple. We simply want to know that our quadratic differentials have enough saddle connections for us to apply Corollary 4.9. The construction in Definition 4.35 also applies to such differentials.

5.1 Derivatives of period coordinates

Notation 5.1 Fix integers $e_1, \dots, e_r \geq -1$. Assume $\sum_{m=0}^{r} e_m \geq -3$. For any r–tuple of distinct complex numbers $(z_1, \dots, z_r)$, let $q(z_1, \dots, z_r) = \prod_{j=1}^{r} (z - z_j)^{e_j} \, dz^2$.

Given a finite system of saddle connections γ_i for $q(w_1, \dots, w_r)$ with γ_i having endpoints w_{i_1} and w_{i_2}, there is a neighborhood U of $(w_1, \dots, w_r)$ such that for all $(z_1, \dots, z_r) \in U$ the saddle connection in $(\mathbb{C}, q(z_1, \dots, z_r))$ with endpoints z_{i_1} and z_{i_2} can be chosen Hausdorff continuously for all $\ell \in I$, such that γ_i is the collection of saddle connections at $(w_1, \dots, w_r)$. The periods vary holomorphically with respect to $(z_1, \dots, z_r)$; call these periods P_i, and their partial derivatives $\partial P_i / \partial z_j$.

Differentiation under the integral gives us the following basic formula:

Proposition 5.2 *If γ_i is a saddle connection of the quadratic differential $q(z) = f(z) \, dz^2$ with period P_i, and either $e_j \geq 1$ or v_j is not an endpoint of γ_i, then*

$$\frac{\partial P_i}{\partial z_j} = \int_{\gamma_i} \frac{e_j \sqrt{q(z)}}{2(z_j - z)}.$$

In all other cases we can recover the partial derivative $\partial P_i / \partial z_j$ by performing a 1–parameter family of Möbius transformations so that the endpoints of γ_i remain fixed, pushing forward the family of differentials, and differentiating under the integral.

Proof The only cases in which the formula for $\partial P_i / \partial z_j$ is not the immediate result of differentiation under the integral are those in which one endpoint of γ_i is z_j and $e_j \geq 1$. Suppose a and b are the endpoints of γ_i and $b = z_j$. Assume $q_h(z) = q(z)((z - b - h)/(z - b))^{e_j}$. Fix a contour of integration from a to b to be the saddle connection γ_i for the q–metric. Then

$$\frac{1}{h}\left[\int_a^{b+h} \sqrt{q_h(z)} - \int_a^b \sqrt{q(z)}\right] = \int_a^b \frac{1}{h}[\sqrt{q_h(z)} - \sqrt{q}] + \int_b^{b+h} \frac{1}{h} q_h(z).$$

We claim that the last integral on the right-hand side is $O(h^{e_j/2})$ because we can take our contour of integration from b to $b + h$ to have length h, and the integrand is $O(h^{e_j})$ along the entire contour. The remaining term limits to the desired formula.

In other cases, the integrals need not converge. To prove our remaining claim, we simply note that any holomorphic 1–parameter family of quadratic differentials $\prod (z - z_j(t))^{e_i} \, dz^2$ admits a holomorphic 1–parameter family of Möbius transformations sending $z_{j_1}(t)$, $z_{j_2}(t)$ and ∞ to $z_{j_1}(0)$ and $z_{j_2}(0)$. The pushforwards of $q(t)$ will be of the form $(z - z_{j_1})$. □

5.2 The limit of the matrix $\partial P_i / \partial z_j$

Notation 5.3 Fix a sequence of quadratic differentials $q_N = q(z_{1,N}, \dots, z_{r,N})$ that converge to an element of the partial compactification of Definition 4.35. Assume that the length-minimizing spanning trees of saddle connections are isomorphic to some fixed Γ as ribbon graphs, where each $z_{j,N}$ corresponds to a fixed vertex v_j of Γ for all N. Also assume that for each vanishing cluster S_ℓ with $z_{j,N} \in S_\ell(N)$, there is some sequence T_n as in Proposition 4.38 such that $T_N(z_{j,N})$ converges for all j, where $z_{j,N} \in S_\ell(N)$, to the corresponding singularity of the cluster differential q_ℓ corresponding to S_ℓ.

For this sequence, let γ_i vary over the edges of Γ and have period P_i. Let $\Gamma_{i,N}$ be the saddle connection corresponding to γ in $(\mathbb{C}, q_N)$. Assume we have picked $\sqrt{q_N}$ so that, for some sequence of positive real numbers t_N, the sequence $\int_{\gamma_{i,N}} t_N \sqrt{q_N}$ converges to a nonzero value, and let P_i be the holomorphic function associated to γ_i.

Finally, let M_N be the matrix $(M_N)_{ij} = \partial P_i/\partial z_j|_{(z_1,\dots,z_r)=(z_{1,N},\dots,z_{r,N})}$. Finally, let M'_N be obtained from M_N by the following operation: for each maximal vanishing cluster S_ℓ, delete the rows corresponding to the saddle connections internal to vanishing clusters, and replace the columns corresponding to singularities in vanishing clusters by a single column that is the sum of the columns corresponding to singularities in $S_\ell(N)$.

Proposition 5.4 *Let $\{q_N\}_{N=1}^\infty = \{q(z_{1,N},\dots,z_{r,N})\}_{N=1}^\infty$ be a sequence satisfying the hypotheses of Notation 5.3, and suppose that the root of the associated tree is q_0. Let m be such that q has a pole of order $m+4$ at ∞. Assume no vanishing cluster contains a pole, and there is at most one singularity of cone angle 2π in each vanishing cluster. Then we have the following:*

1. *For each i, $P_i(z_{1,N},\dots,z_{r,N})$ converges as $N \to \infty$.*
2. *M_N converges to a matrix M_∞.*
3. *If $(z_j)_\infty = (z_k)_\infty$, then $e_j(M_\infty)_{ik} = e_k(M_\infty)ij$.*
4. *M'_N converges to a matrix M'_∞*
5. *The kernel of M'_∞ has dimension 1 and it is spanned by the all-1 vector.*
6. *M'_∞ is the matrix of partial derivatives for the periods of the saddle connections and vertices of the degeneration of T that occurs as $q_n \to q_\infty$.*
7. *If γ_i corresponds to a sequence of saddle connections internal to a vanishing cluster, then $\partial P_i/\partial z_j \to 0$ for all j.*

More concisely, if $\cdot'$ denotes the object that $\cdot$ degenerates to as $N \to \infty$, then

$$\frac{\partial P_i}{\partial z_j} \to \frac{e_j}{e'_j}\frac{\partial P'_i}{\partial z'_j}.$$

If $e_j = e'_j = 0$ then $e_j/e'_j = 1$. (In particular, the period of a saddle connection that degenerates to a point has partial derivatives tending to 0.)

Proof The first claim follows easily from Corollaries 4.27, 4.28, and 4.18.

Items (2)–(7) determine the value of M, up to the value of the derivative of the saddle connection of a nonshrunk vanishing cluster.

Suppose there are no vanishing clusters. Then the period coordinates P_i are a holomorphic coordinate system for our stratum of cluster differentials in a neighborhood of q_0. Indeed, the periods of the $r-1$ edges of Γ form a local coordinate system for the stratum of the differentials $q(z_{1,N},\dots,z_{r,N})$. It is also true that the values of $r-1$ elements of $\{z_1,\dots,z_r\}$ determine a local coordinate system if the remaining z_j are fixed, since $q(z_1,\dots,z_r)$ is biholomorphically equivalent to $q(z_1+\alpha,\dots,z_r+\alpha)$ for any $\alpha\in\mathbb{C}$ (each is the pullback of the other via translation). Since any injective holomorphic map from a domain in $\mathbb{C}^n$ to a domain in $\mathbb{C}^n$ is a biholomorphism onto its image, it follows that the kernel of M consists only of perturbations of $(z_0,\dots,z_r)$ that preserve the isomorphism type of the quadratic differential q_0. This subspace is precisely the span of the all-1 vector. In this case, the dimension of the stratum of the limiting dimension equals the claimed rank of M. It follows that all claims hold in this case.

Now, recall that the sequences of saddle connections $\gamma_{i,N}\colon[0,1]\to\mathbb{C}$ each have the property that for some sequence of Möbius transformations $T_{i,N}(z)=t_{i,N}(z)+u_{i,N}$ fixing ∞ as in Proposition 4.38, $T_{i,N}\circ\gamma_{i,N}$ converges uniformly and c_1 on compact subsets of $(0,1)$ to a saddle connection in some cluster differential q_ℓ.

Assume that $T_{i,N}(z)=t_{i,N}z+u_{i,N}$ is chosen so that $T_{i,N}(z)\gamma_i(q_N)$ converges to the saddle connection corresponding to γ_i on one of the quadratic differentials q_ℓ associated to the limit of the sequence $\{q_N\}$.

Now, we would like to apply Proposition 5.2 with the change of coordinates $T_{i,N}(z)=\zeta$, so let

$$\alpha_{i,N}(\zeta)=t_{i,N}^{(m+2)}\prod_{j=1}^{r}(\zeta-\zeta_j)^{e_j}\,d\zeta^2=(T_{i,N})_*(q_N).$$

If $\gamma_{i,N}$ has endpoints $a_{i,N}$ and $b_{i,N}$ for the differential q_N, we then have

$$\frac{\partial Q_i}{\partial z_j}\bigg|_{(z_1,\dots,z_r)=(z_{1,N},\dots,z_{r,N})}=\int_{a_{i,N}}^{b_{i,N}}\tfrac{1}{2}e_j\,\frac{\sqrt{q_N(z)}}{z_{j,N}-z}=\int_{T_{i,N}(a_{i,N})}^{T_{i,N}(b_{i,N})}t_{i,N}^{-m/2}\cdot\tfrac{1}{2}e_j\,\frac{\sqrt{\alpha_{i,N}(\zeta)}}{\zeta_{j,N}-\zeta}.$$

We would like to apply the Lebesgue dominated convergence theorem to the right-hand side, parametrizing the integrals to be constant-speed $d_{(T_{i,N})_*(q_N)}$–geodesic maps from $[0,1]$ to $\mathbb{C}$. To do this, we need the following:

Claim 5.5 *A dominating function for the sequence* $\int_{T_{i,N}(a_{i,N})}^{T_{i,N}(b_{i,N})}t_{i,N}^{-m/2}\cdot\frac{1}{2}e_j\sqrt{\alpha_{i,N}(\zeta)}/(\zeta_{j,N}-\zeta)$, *where the contours are parametrized as constant-speed geodesics from* $[0,1]$ *to* $\mathbb{C}$ *with respect to* $d_{T_*q_N}$, *exists whenever* $z_{j,N}$ *is a zero of* q_N *or not an endpoint of* $\gamma_{i,N}$.

Proof Suppose that $\{\gamma_{i,N}\}$ is not internal to a vanishing cluster.

Since we are assuming that our contours of integration belong to length-minimizing spanning trees, for each z on our contour of integration, we have

$$d_{q_N}(z,(z_{j,N})\geq\min(d_{q_n}(z,(a_{i,N})),d_{q_N}(z,b_{i,N})).$$

If this were not the case, we would have

$$d_{q_N}(a_{i,N}, b_{i,N}) > \min(d_{q_N}(a_{i,N}, z_{j,N}), d_{q_N}(b_{i,N}, z_{j,N}),$$

and a saddle connection from $a_{i,N}$ to $b_{i,N}$ could not belong to a length-minimizing spanning tree.

All singularities of q_N remain bounded, the poles are all bounded away from $z_{j,N}$, and $z_{j,N}$ is a zero of q_N, so

$$d_{q_N}(z, z_{j,N}) \stackrel{.}{\prec} \int_0^{|z-z_{j,N}|} t^{1/2}\,dt = \tfrac{2}{3}|z-z_{j,N}|^{3/2}.$$

We thus have, along our contour of integration,

$$|z-z_{j,N}|^{-1} \stackrel{.}{\prec} d_{q_N}(z, z_{j,N})^{-2/3} \le \min(d_{q_N}(z, T_{i,N}(a_{i,N})), d_{q_N}(z, T_{i,N}(b_{i,N})))^{-2/3}.$$

We have parametrized our arcs to be d_{q_N} geodesics instead of d_{α_n} geodesics, but applying Lemma 4.20 to the proofs of Corollaries 4.27 and 4.28 tells us that the contours of integration converge (Hausdorff, in total length, and C^1 on compact subsets of $(0,1)$) to geodesics with respect to d_{α_∞}, where $\alpha_\infty = \lim_{n\to\infty}\alpha_n$. Therefore, for some $C > 0$, our sequence of integrals is dominated by $\int_0^1 C[x(1-x)]^{-2/3}\,dx$. This completes the proof of the claim in this case. □

5.3 The derivative of period coordinates, rescaled

We may do the same for the case of a saddle connection internal to a vanishing cluster. By translating our original sequence q_N so that one endpoint of $\gamma_{i,N}$ is 0 and not equal to z_j, we can assume that $\zeta = T_{i,N}(z) = t_{i,N}z$ with $t_{i,N} \to \infty$. If $\{S_\ell(N)\}_{N=1}^\infty$ is the minimal vanishing cluster to which $\gamma_{i,N}$ is internal, write $q_N = f_{i,N}(z)g_{i,N}(z)\,dz^2$ with $g(z) = \prod_{j\in S_\ell(N)}(z - z_{j,N}^{e_j})$. Then the sequence of integrals becomes

$$\begin{aligned}\frac{\partial P_i}{\partial z_j}\bigg|_{(z_1,\dots,z_r)=(z_{1,N},\dots,z_{r,N})} &= \int_{a_{i,N}}^{b_{i,N}} \sqrt{f(z)g(z)}\frac{e_j\,dz}{2(z-z_{j,N})}\\ &= \int_{\alpha_{i,N}}^{\beta_{i,N}} [f_N(\zeta/t_{i,N})]^{1/2}[g_N(\zeta/t_{i,N})]^{1/2}\frac{e_j\,d\zeta}{2(\zeta-\zeta_{j,N})}.\end{aligned}$$

Now, the contour of integration converges, and $f_N(\zeta/t_{i,N}) = f_N(z)$ converges to a constant, possibly zero. If m_ℓ is the degree of $g(z)$ then $[t_{i,N}^{-m}h_N(\zeta)]$ converges as a function of ζ. By Lemma 4.20 and the argument for the saddle connections not internal to vanishing clusters, for $x \in [0,1]$ we have

$$\left|\frac{h_N(\zeta)}{\zeta-\zeta_{j,N}} \circ T_{i,N}\gamma_{i,N}(x)\right| \le C[x(1-x)]^{-2/3}.$$

We therefore see that not only does a dominating function exist, but we can take the dominating function to be $t_{i,N}^{-m/2}C[x(1-x)]^{-2/3}$ for $m > 0$, if we are willing to start the sequence at some large N. This proves item (7). The rest of the claims follow from Proposition 5.2 and the dominated convergence theorem. □

Corollary 5.6 *Let $q_N = \prod_{j=1}^{s}(z - z_{j,N})^{e_j}\, dz^2$ be a sequence of cluster differentials with no poles satisfying the hypotheses of Proposition 5.4, and assume that at least one of the sequences $\{z_{j,N}\}$ does not converge to 0. Fix integers $e_{s+1}, \dots, e_r$ and let $(z_{s+1,N}, \dots, z_{r,N})$ be a sequence of $(s-r)$–tuples of complex numbers converging to $(\infty, \infty, \dots, \infty)$ in $\widehat{\mathbb{C}}^{r-s}$. Assume $\{\gamma_{i,N}\}$ is a sequence of saddle connections of the differential q_N that converges (Hausdorff), and its period P_i is locally a holomorphic function of $(z_1, \dots, z_s)$. Let $Q_{i,N}$ be the periods of a sequence saddle connections in the metric associated to the quadratic differential $\prod_{j=1}^{r}(z - z_{j,N})^{e_j}$ which converge (Hausdorff) to same limit as $\gamma_{i,N}$, which are locally given by the holomorphic function $Q_i(z_1, \dots, z_r)$.*

If $\gamma_{i,N}$ does not converge to a point, for $1 \le j \le s$ the following limits exist and are equal:

$$\lim_{N\to\infty} \frac{\partial Q_{i,N}}{\partial z_j} = \frac{\partial \log Q_{i,N}}{\partial z_j} = \lim_{N\to\infty} \frac{\partial \log P_i}{\partial z_j}.$$

If $\gamma_{i,N}$ converges to a point, then let $\gamma_{i',N}$ be another convergent sequence of saddle connections that does not converge to a point, and let them have periods $P_{i',N}$ and $Q_{i',N}$. Then

$$0 = \lim_{N\to\infty} \frac{\partial Q_{i,N}}{\partial z_j} = \frac{\partial Q_{i,N}}{\partial z_j}\frac{1}{Q_{i',N}} = \lim_{N\to\infty} \frac{\partial P_i}{\partial z_j}\frac{1}{P_{i',N}}.$$

For $s+1 \le j \le r$,

$$\lim_{N\to\infty} \frac{\partial Q_{i,N}}{\partial z_j} = \frac{\partial \log Q_{i,N}}{\partial z_j} = 0.$$

Proof We use an identical dominated convergence argument. We can treat the faraway singularities (those outside $S_\ell N$) as scalars in the limit, and by Lemma 4.20 we can use the same system of saddle connections γ_i we would if there were no finite singularities outside of $S_\ell(N)$. □

Corollary 5.7 *Given a sequence $\{q_N\}$ converging as in Proposition 5.4, let $M_{\ell,N}$ be the submatrix of M_N corresponding to saddle connections internal to the vanishing cluster $\{S_\ell(N)\}$. Then we have the following as $N \to \infty$:*

- $\partial P_i/\partial z_j|_{(z_1,\dots,z_r)=(z_{1,N},\dots,z_{r,N})} = o(|M_{\ell,N}|)$ *if* $j \notin S_\ell$.
- *If $\gamma_{i,N}$ corresponds to a row of $M_{\ell,N}$ then* $\sum_{j\in S_\ell} \partial P_i/\partial z_j|_{(z_1,\dots,z_r)=(z_{1,N},\dots,z_{r,N})} = o(|M_{\ell,N}|)$.
- *If $S_l \subsetneq S_\ell$ is a vanishing cluster and $\sum_{j\in S_l} a_j e_j = 0$, then*

$$\sum_{j\in S_l} a_j \frac{\partial P_i}{\partial z_j}\bigg|_{(z_1,\dots,z_r)=(z_{1,N},\dots,z_{r,N})} = o(|M_{\ell,N}|).$$

Proof This is immediate from Corollary 5.6 and Proposition 4.14. □

Corollary 5.8 *Let $\{q_N\}$ and $M_{\ell,N}$ be as above, and assume the center of $S_{\ell,N}$ is 0 for all N. Letting $t_N = \prod_{z_{j,N}\notin S_\ell(N)} z_j(-z_{j,N})^{e_{j,N}}$, we have that $|M_{\ell,N}| \doteqdot |t_N|^{1/2}\operatorname{diam}_{\mathbb{C}}(S_\ell(N))^{m/2}$, where m is the number of zeros in $S_\ell(N)$, counted with multiplicity. In fact, the sequence of matrices $t_N^{-1/2}\operatorname{diam}_{\mathbb{C}}(S_\ell(N))^{m/2} M_{\ell,N}$ converges to a matrix whose cokernel corresponds to the space spanned by periods of saddle connections internal to proper vanishing subclusters of $\{S_\ell(N)\}$.*

Proof We are free to consider the same system of saddle connections we would use in the absence of faraway singularities by Lemma 4.20.

Assume z_j is a zero of order e_j that degenerates to $z_{j'}$, a zero of order $e_{j'}$ times the corresponding column of the matrix of partial derivatives for the limiting differential q_ℓ. If we change coordinates on $\mathbb{C}$ by the positive real scalar $\operatorname{diam}_{\mathbb{C}}(S_\ell(N))$ so that $S_\ell(N)$ has diameter 1, then Propositions 4.16, 4.14 and 5.4 imply that column j of the matrix $|t_N|^{-1/2}\operatorname{diam}_{\mathbb{C}}(S_\ell(N))^{-m/2}M_{\ell,N}$ converges to $e_j/e_{j'}$, together with additional zeros for the saddle connections internal to proper subclusters of q_ℓ. Since the matrix of partial derivatives for q_ℓ has a nonzero limit, so does $|t_N|^{-1/2}\operatorname{diam}_{\mathbb{C}}(S_\ell(N))^{-m/2}M_{\ell,N}$, and the result follows. □

We would have liked to extend Proposition 5.4 and its corollaries to clusters that include a single pole. Unfortunately, the partial derivatives of a period need not remain bounded when a zero collides with a pole. Consider the family of quadratic differentials $(z(z-1)\,dz^2)/(z+t)$ where t varies. If t is a negative real number, then the segment [0, 1] of the real line is a vertical saddle connection; call its period $P(t)$. Then, by Proposition 5.2, we have

$$P'(t) = \frac{1}{2}\int_0^1 \sqrt{\frac{x(x-1)}{(x+t)^3}}\,dx.$$

The integrand is purely imaginary and does not change sign. Thus, by the monotone convergence theorem,

$$\lim_{t\to 0^-} |P'(t)| = \int_0^1 \frac{(1-x)^{1/2}}{x}\,dx \geq \int_0^{1/2} \frac{dx}{2x} = +\infty.$$

However, for the estimates we need, it is possible to take a double cover branched over the pole and pull back the quadratic differential; this converts the pole into a singularity with cone angle 2π at the cost of duplicating other singularities.

5.4 The main estimate

Notation 5.9 Fix a stratum of quadratic differentials in $\mathrm{QD}(\mathcal{M}_{0,n})$ with no zeros of order 2 or higher. Assume all quadratic differentials are normalized to have simple poles at ∞ and two other fixed points A and B. Then each has the form $\lambda_q[(z-A)(z-B)]^{-1}\prod_{j=1}^{s}(z-z_j)^{e_j}\,dz^2$. Fix integers $\{e_j\}_{j=1}^{s}$ with $-1 \leq e_j \leq 1$ for all j and $e_1+\cdots+e_s=-1$, and denote such a differential by $q(\lambda, z_1, \dots, z_r)$. Let

$$d_{\mathrm{Sym}}(q_1,q_2) = \inf_{\phi}\sum_{w} |\phi(w)-w| + |\lambda_{q_1}-\lambda_{q_2}|,$$

where w ranges over all noninfinite singularities of q_1, and ϕ ranges over all bijections between the singularities of q_1 and q_2 that preserve the type of singularity.

Proposition 5.10 *Suppose that $\{q_N\}$ and $\{r_N\}$ are sequences converging to possibly distinct elements of the partial compactification $\overline{\mathcal{Q}}$, but such that $d_{\mathrm{Sym}}(q_N, r_N) \to 0$ with respect to a common Möbius normalization. Let ϕ_N be the type-preserving bijection between the singularities of q_N and the singularities of*

r_N that minimizes $\sum|\phi(z_{j,N})-z_{j,N}|$ for this Möbius normalization. Let $\{S_\ell(N)\}$ be a vanishing cluster for the sequence $\{q_N\}$. Suppose that as $N\to\infty$, $d_{\mathrm{Sym}}(q_N,r_N)/\mathrm{diam}_{\mathbb{C}}(S_m(N))\to 0$ for every vanishing cluster $\{S_m(N)\}$ in q_N with $S_m(N)\supsetneq S_\ell(N)$ for all N. Then $\phi_N(S_\ell(N))$ is a vanishing cluster for r_N.

Proof It suffices to consider the minimal S_m that properly contains S_ℓ. It is clear from the definition of vanishing cluster that for any sequence of sets of singularities $S(N)$ with $S_m(N)\supseteq S(N)\supsetneq S_\ell(N)$, we must have $\mathrm{diam}_{\mathbb{C}}(T(N))\doteq\mathrm{diam}_{\mathbb{C}}(S_m(N))$, since otherwise we could pass to a subsequence along which $T(N)$ belonged to a vanishing cluster with $S_m(N)\supsetneq S(N)\supsetneq S_\ell(N)$. By minimality of S_m this can't happen, since the nesting of the vanishing clusters is determined by the limit in $\overline{\mathcal{Q}}$. □

Lemma 5.11 *Suppose $\{q_N\}_{N=1}^{\infty}=\{q((\lambda_N,z_{1,N},\dots,z_{r,N})\}_{N=1}^{\infty}$ is a sequence in $\mathrm{QD}(\mathcal{T}_{0,n})$, normalized in the sense of Notation 5.9 that converges in the sense of Definition 4.35, such that no vanishing cluster contains a pole and every vanishing cluster contains a zero. Suppose that in addition, no vanishing cluster for the sequence $\{q_N\}$ has more than k zeros (counted with multiplicity) for $k\ge 0$. Let $\{r_N\}$ be another sequence with the same Möbius normalization, converging to a point in the partial compactification of Definition 4.35, and assume that for each singularity type m and each $p\in\mathbb{C}$, the number of singularities of type m that limit to p along the sequences $\{q_N\}$ and $\{r_N\}$ is the same. (In particular, $\{q_N\}$ and $\{r_N\}$ have the same limit in $\mathrm{QD}(\mathcal{T}_{0,n})$.) Then*

$$d_{\mathrm{Euclidean}}(q_N,r_N)\dot{\succ} d_{\mathrm{Sym}}(q_N,r_N)^{2/(2+k)}\quad \text{as } N\to\infty.$$

Proof The proof is divided into two steps. In the first we fix a subsequence of potential counterexamples with specified combinatorics of the singularities belonging to vanishing clusters, and isolate some piece of the difference between q_N and r_N. In the second we show that this piece causes a nontrivial change in period coordinates that is not canceled by any other components.

Step 1 Each maximal vanishing cluster of $\{q_N\}$ is also a maximal vanishing cluster for $\{r_N\}$, and we will only consider bijections that preserve maximal vanishing clusters since d_{Sym} is only realized by such a bijection for large N. In fact, we can take this further. For each vanishing cluster $\{S_\ell(N)\}$ of $\{q_N\}$ we can pass to a subsequence whereby for each vanishing cluster $S_\ell(N)$,

$$d_{\mathrm{Sym}}(q_N,r_N)\;\dot{\star}_\ell\;\mathrm{diam}_{\mathbb{C}}(S_\ell(N)),$$

where $\dot{\star}_\ell$ depends the choice of ℓ but is one of $\dot{\prec}$, $\dot{\succ}$, or $\doteq$.

Assume

$$q_N=\frac{\lambda_N}{(z-A)(z-B)}\prod_{j=1}^{r}(z-z_{j,N})^{e_j}\,dz^2\quad\text{and}\quad r_N=\frac{\mu_N}{(z-A)(z-B)}\prod_{j=1}^{r}(z-w_{j,N})^{e_j}\,dz^2,$$

and let ϕ_N be the bijection realizing $d_{\mathrm{Sym}}(q_N,r_N)$ that takes z_j to w_j for all j. We will describe how to break the perturbation realized by ϕ_N into components, which we will call $v_{\ell,N}$ and define below.

Let $\{S_\ell(N)\}$ be either a vanishing cluster of the sequence $\{q_N\}$ or the entire set of singularities (besides ∞). Let the sequences $\{v_N\}$ and $\{v_{\ell,N}\}$ satisfy the following:

- $v_{\ell,N}$ is an $(r+1)$–tuple of complex numbers, corresponding to the coordinates $(\lambda, z_1, \dots, z_r)$ and $(\mu, w_1, \dots, w_r)$.
- If $\{S_\ell(N)\}$ is a vanishing cluster of $\{q_N\}$, then the support of $v_{\ell,N}$ is in the entries corresponding to singularities in S_ℓ.
- If $S_k \subsetneq S_\ell$ is a vanishing cluster of $\{q_N\}$, then $v_{\ell,N}$ is constant on the entries corresponding to singularities in D'. (All points in each proper subcluster move the same distance and in the same direction.)
- For each N, $\sum_\ell v_{\ell,N} = (\mu-\lambda, w_1-z_1, \dots, w_s-z_s) =: v_N$.
- If S_ℓ is a vanishing cluster and $v_{\ell,N} = (0, a_1, \dots, a_s)$ then $\sum_{j=1}^s a_j e_j = 0$. (Together with other conditions, this means that $v_{\ell,N}$ does not move the center of $S_{\ell,N}$.)

The purpose of this notation, and the idea of the rest of the proof, is that for some ℓ, the component $v_{\ell,N}$ can be detected either by the periods of saddle connections internal to $S_\ell(N)$, or by periods internal to the largest $S_k(N) \supset S_\ell(N)$ with $\operatorname{diam}_{\mathbb{C}}(S_k(N)) \dot{\prec} \|v_{\ell,N}\|$. Then we show that this does not get canceled by other components of the perturbation.

We may pass to a subsequence with the following property: for any S_k and S_ℓ, not necessarily distinct, the following converge (possibly to 0 or ∞):

$$\frac{\operatorname{diam}_{\mathbb{C}}(S_\ell(N))}{\operatorname{diam}_{\mathbb{C}}(S_k(N)))}, \quad \frac{\|v_{\ell,N}\|}{\operatorname{diam}_{\mathbb{C}} S_{k,N}}, \quad \frac{v_{\ell,N}}{\|v_N\|}.$$

In other words, the relative sizes of the various clusters and components of v_N converge, and if $v_{\ell,N}$ is not vanishingly small relative to any $v_{k,N}$, the direction of $v_{\ell,N}$ converges. Moreover, the sizes of perturbations of vanishing clusters converge relative to the sizes of all vanishing clusters.

Now pick maximal S_ℓ, subject to the following condition: either $|v_{\ell,N}| \dot{\asymp} |v_N|$, or $\sum_{S_l \subseteq S_\ell} |v_{l,N}| \dot{\asymp} |v_N|$ and $\sum_{S_l \subseteq S_\ell} |v_{\ell,N}| \dot{\succ} \operatorname{diam}_{\mathbb{C}}(S_\ell(N))$.

In particular, this means that either $S_\ell(N)$ is always the entire set of singularities, or $S_\ell(N)$ and $\phi_N(S_\ell(N))$ are both vanishing clusters of the sequence $\{r_N\}$, by Proposition 5.10.

In both cases, let $T_N(z) = a_N z + b_N$ be a sequence of Möbius transformations fixing ∞ with the property that $T_N(S_\ell(N))$ converges (Hausdorff) to the limiting differential associated to S_ℓ; we may assume $a_N \in \mathbb{R}$. Then $T_N(\phi_N(S_\ell(N)))$ also converges to the singularity set of a cluster differential along a subsequence.

Step 2 (estimating the perturbation in period coordinates) Now we have two cases. In both, we can detect a change in Euclidean distance from the periods internal to $S_\ell(N)$ by Proposition 4.42.

Case 1 ($|v_N|/\operatorname{diam}_{\mathbb{C}}(q_N) \dot{\succ} 1$) Here, for all large N we can define the projection map F associated to the clusters $S_\ell(N)$ and $\phi_N(S_\ell(N))$ in Proposition 4.42 on a connected open set containing both, by Proposition 5.10. By Lemma 4.43 and Corollary 4.15 it follows that

$$d_{\text{Euclidean}}(q_N, r_N) \dot{\succ} d_{\text{Euclidean}}(F_N(q_n), F_N(r_N)) \dot{\asymp} \max\{\operatorname{diam}_{q_N}(S_\ell(N)), \operatorname{diam}_{r_N} \phi_N(S_\ell(N))\}.$$

The last inequality can be justified as follows: First, the ratios between the factors t_N in Lemma 4.43 for q_N and r_N differ by a ratio of $1+o(1)$. Now, if two cluster differentials q and r have $d_{\mathbb{C}}$–diameter at most 1, one has diameter 1, and their distance with respect to $d_{\mathbb{C}}$ is bounded below, then $d_{\text{Euclidean}}(q,r)$ is also bounded below by compactness. Our last inequality then comes from Lemma 4.43 and the fact that scalars induce similarities on spaces of cluster differentials.

Now by Corollary 4.15 and Lemma 4.43, if $\max\{\operatorname{diam}_{\mathbb{C}}(q_N), \operatorname{diam}_{\mathbb{C}}(r_N)\} = \epsilon_N$ then

$$\max\{\operatorname{diam}_{q_N}(S_\ell(N)), \operatorname{diam}_{r_N}(\phi_N(S_\ell(N)))\} \dot{\succ} \epsilon_N^{(m+2)/2},$$
$$\max\{\operatorname{diam}_{q_N}(S_\ell(N)), \operatorname{diam}_{r_N}(\phi_N(S_\ell(N)))\} \dot{\asymp} \epsilon_N^{(m+2)/2},$$

only if q_N contains all m singularities.

Case 2 $(|v_N|/\operatorname{diam}_{\mathbb{C}}(q_N) \to 0)$ Here $|v_{\ell,N}| \dot{\asymp} |v_N|$. Again we choose F_N to be the Lipschitz projection associated to $S_\ell(N)$. Assume the center of $|v_{\ell,N}|$ is always zero and let $t_N = \prod_{z_{j,N} \notin S_\ell(M)} (-z_{j,N})^{e_{j,N}}$. Assume m is the number of zeros in $S_\ell N$ (each zero has multiplicity 1, so we don't need to specify that we are counting with multiplicity).

If $M_{\ell,N}$ is as in Corollary 5.6, then the sequence of matrices

$$\{t_N^{-1/2} \operatorname{diam}_{\mathbb{C}}(S_{\ell,N})^{-m/2} M_{\ell,N}\}_{N=1}^{\infty}$$

converges (by Corollary 5.7 and Proposition 4.14), and its kernel is the span of those vectors that are allowed to be $v_{k,N}$ with $k \neq \ell$. What we mean by this is that when we defined $v_{k,N}$, we required that it belong to a certain subspace of $\mathbb{C}^{r+1}$, and all vectors in this subspace are in the kernel of the limiting matrix. The periods of saddle connections internal to proper subclusters of $S_\ell(N)$ span the cokernel of the limiting matrix. The limiting matrix

$$\lim_{N\to\infty} t_N^{-1/2} \operatorname{diam}_{\mathbb{C}}(S_{\ell,N})^{-m/2} M_{\ell,N}$$

has full rank when restricted to the space of vectors that are allowed to be $v_{\ell,N}$.

We may move from q_N to r_N along a piecewise-smooth path β_N, so that the motion of each of the singularities $z_{j,N}$ to $\phi(z_{j,N})$ is a straight line segment with constant speed along each of two segments of β_N. On the first piece of β_N the total displacement of the vectors $(\lambda_N, z_{1,N}, \dots, z_{r,N})$ will be $v_{\ell,N}$, and on the second piece the total displacement will be the remainder.

The singularities starting at $S_\ell(N)$ will be in the stratum of $S_\ell(N)$ for all but finitely many points along each segment. For any such sequence of paths, if we pick s_N to be one point in this stratum along each β_N and construct the matrix $\widehat{M}_{\ell,N}$ analogously to the matrix $M_{\ell,N}$ at s_N, we can pass to a subsequence that converges in the sense of Notation 5.3. We also note that $\partial P_i/\partial\lambda = o(M_{\ell,N})$, since $\partial \log P_i/\partial z_j$ is not bounded for all j but $\partial P_i/\partial\lambda$ is. The limit of the sequence of matrices must be the same, since it only depended on the limiting differential for S_ℓ, which is the same for any subsequence of the sequence $\{s_N\}$. We therefore conclude that the matrices $M_{\ell,N}$ converge *uniformly after rescaling by constants c_N along β_N* in the sense that the matrix $\widehat{M}_{\ell,N}$ differs from $M_{\ell,N}$ at q_N by

$o(|M_\ell, N|)$. In particular, this does not depend on which choices of saddle connections that degenerate to the proper edges of the limiting differential for S_ℓ we picked at each s_ℓ. Moreover, the displacement of periods of S_ℓ is $M_{\ell,N}(v_{\ell,N}) + o(|M_{\ell,N}(v_N)|)$ along the first segment of β_N and the displacement is $o(|M_{\ell,N}(v_N - v_{\ell,N})|)$ along the second segment of β_N.

This, combined with Proposition 4.44 tells us that $d_{\mathrm{Euclidean}}(F_N(q_N), F_N(r_N))$ is comparable to the size of the leading-order approximation $M_{\ell,N} v_{\ell,N}$. By Corollary 5.8 we have

$$|M_{\ell,N} v_N| \stackrel{\cdot}{\asymp} |t_N|^{1/2} |s_N|^{m/2} |v_{\ell,N}| \stackrel{\cdot}{\succ} \mathrm{diam}_{\mathbb{C}}(S_{\ell,N})^{(k-m)/2} [\mathrm{diam}_{\mathbb{C}}(S_\ell(N))]^{m/2} |v_N| \cdot |v_N|^{(2+k)/2}.$$

We explain the sources of the inequalities: The size of t_N is at least comparable to the product of the zeros outside of $S_\ell(N)$, since all poles remain bounded away from 0. Since we only care about a multiplicative constant, we may ignore all zeros and poles that do not belong to a vanishing cluster containing $\{S_\ell(N)\}$, and there are at most $k-m$ additional zeros inside any such vanishing cluster, each of which is larger than $\mathrm{diam}_{\mathbb{C}}(S_\ell)$. Hence $|t_N| \stackrel{\cdot}{\succ} |v_N|^{(k-m)/2}$. The remaining inequalities are clear from $\mathrm{diam}_{\mathbb{C}}(S_\ell(N)) \stackrel{\cdot}{\succ} |v_\ell, N| \stackrel{\cdot}{\asymp} |v_N|$.

Then by Proposition 4.42 we get

$$d_{\mathrm{Euclidean}}(q_N, r_N) \stackrel{\cdot}{\succ} d_{\mathrm{Euclidean}}(F_N(q_N), F_N(r_N)) \stackrel{\cdot}{\succ} |v_N|^{(k+2)/2}. \qquad \square$$

Lemma 5.12 *Fix a compact subset H in the open upper half-plane $\mathbb{H}$. Fix a real number $C > 1$ and an integer $m \geq 0$. Consider the space K of quadratic differentials in $\mathcal{T}_{0,2m}$ with the following restrictions:*

- *They are of the form $(F(z)/G(z))\, dz^2$ with $F, G \in \mathbb{R}[z]$, $\deg(F) = 2m-4$, and $\deg(G) = 2m-1$.*
- *G may have simple roots at i and $-i$, but all other roots of G are real. (Either way, G has real coefficients.) The roots of G and ∞ are the marked points of the underlying Riemann surface.*
- *G has no repeated roots, and the distance between any two consecutive real roots of G is between $1/C$ and C.*
- *G has a root in $[-C, C]$ if i and $-i$ are roots of G; otherwise, 0 and 1 are roots of G.*
- *No root of $F(z)$ is real, and one root from each conjugate pair belongs to H.*
- *The ratio of the leading coefficients of F and G is between $1/C$ and C.*

Assume $q_1 = (F_1(z)/G_1(z))\, dz^2$ and $q_2 = (F_2(z)/G_2(z))\, dz^2$ are elements of K. Let ϕ range over all conjugation-invariant bijections of the roots of F_1 with the roots of F_2 that fix i if i is a root of G, and let ψ be the bijection of the roots of G_1 with the roots of G_2 that preserves the order of the real roots and fixes i and $-i$ if they are roots of G_1 and G_2. Let

$$\epsilon = \min_{\phi} \max\{\phi(w) - w : F_1(w) = 0\} \cup \{\psi(w) - w : G_1(w) = 0\}.$$

Then there are constants $M, \epsilon_0 > 0$ depending only on K such that the following hold:

- *If $\pm i$ are not roots of G, then $d_{\mathrm{Euclidean}}(q_1, q_2) \stackrel{\cdot}{\succ} \epsilon^{(m+2)/2}$.*
- *If $\pm i$ are roots of G, then $d_{\mathrm{Euclidean}}(q_1, q_2) \stackrel{\cdot}{\succ} \epsilon^{m+1}$.*

Proof First we will explain how both conclusions follow from Lemma 5.11. If $\pm i$ are roots of G_1 and G_2, then we can take a double cover of each that preserves the real line and is branched at $\pm i$ via $z = h(w) = (w^2 - 1)/(2w)$. Then $\pm i$ are marked points, but not poles on the pullbacks $h^*(q_1)$ and $h^*(q_2)$. Then $\iota^*(q)$ is a conjugation-invariant quadratic differential on $\widehat{\mathbb{C}}$ whose poles are all simple and occur on the projective real line. There are uniform upper and lower bounds on the distance between real poles, and the zeros with positive real part are in a fixed compact part of the upper half-plane.

Now, suppose that z_1 is a root of F_1, z_2 is a root of F_2, and $|z_1 - z_2| = \epsilon$. Then if $h(w_1) = z_1$ and $h(w_2) = z_2$, we have

$$|w_1 - w_2| = |h(w_1) - h(w_2)| \cdot \left| \frac{2w_1 w_2}{1 + w_1 w_2} \right| \dot{\succ} \epsilon.$$

We have a similar estimate for the inverse images of the roots of G_1.

We can apply Lemma 5.11 to both. □

6 Building the quasiconformal map

In this section, we build a quasiconformal map with the desired properties to prove Theorem 1.2. Note that for us it will suffice to consider the case when our two quadratic differentials lie in the principal stratum, since the principal stratum is dense in $\mathcal{QD}^1(\mathcal{T}_{g,n})$.

6.1 Cutting into triangles and nearly regular right polygons

Definition 6.1 A *nearly regular right polygon*, or *NRRP*, is a subset P of a half-translation surface (X, q) with the following properties:

- P is homeomorphic to a closed disk.
- P has at least one singularity in its interior, and at most one pole, but has no singularities on its boundary.
- ∂P is piecewise geodesic, and all pieces are vertical or horizontal segments.
- Every interior angle of ∂P is $\frac{\pi}{2}$.
- If P is doubled along its boundary to form a sphere $P \cup \overline{P}$, where $\overline{P}$ is the Schwarz reflection of P, then there is a choice of holomorphic coordinate z on $P \cup \overline{P}$ for which ∂P is the real projective line, and the quadratic differential on $P \cup \partial P$ is of the form $(F(z)/G(z))\, dz^2$ where F and G satisfy the conditions of Lemma 5.12. (That is, they have real coefficients, and G is normalized to have two roots at i and $-i$, or to have roots at 0 and 1, etc.)

Let R be the minimum q–distance from ∂P to a singularity in $P \setminus \partial P$. We call R the *radius* of the NRRP.

On a quadratic differential it is always possible to find an NRRP around each singularity with all sides of the same length. If we change the differential by a small enough amount, and in so doing break up such a singularity into multiple singularities, we can still find an NRRP that stays Hausdorff close to our original NRRP. The motivation for this definition is that if one wants to build a quasiconformal map between two half-translation surfaces, piecewise-affine maps which are affine on some triangulation will tend to have large dilatation even when two Riemann surfaces are nearly conformal, if the triangulation is nearly degenerate. Given an edgewise-linear map between boundaries of similarly shaped NRRPs in two quadratic differentials that are close in $\mathrm{QD}(\mathcal{T}_{g,n})$, we can extend to a quasiconformal map of NRRPs, and explicitly estimate its quasiconformal dilatation.

Proposition 6.2 *On any compact subset K of $\mathrm{QD}(\mathcal{T}_{g,n})$ there are constants $C = C(K) > 0$ and $\delta = \delta(K)$ such that for every surface X in K we can partition the set of singularities of X into δ–clusters and singletons* (*for some $\delta \in (0,1)$*), *and associate to each δ–cluster or singleton an NRRP containing it such that the following hold*:

- *Each NRRP contains only the singularities associated to it.*
- *Each NRRP has radius and all boundary side lengths at least $C(K)$.*
- *If we consider all lifts of the NRRPs in the universal cover of the underlying compact Riemann surface*, *the distance between any two NRRPs is more than twice as large as the side length of any NRRP.*

Proof It is clearly possible to choose a constant $C(X)$ at each $X \in K$ such that the properties hold for $C(X)$ by picking a small NRRP around each singularity. (NRRPs of a given radius persist under sufficiently small perturbation.) For each X, we can find a neighborhood of X where the conditions all hold with the constant $\frac{1}{2}C(X)$ instead of $C(X)$, and K is finitely covered by such neighborhoods. It follows that we can take $C(K)$ to be half the minimum value of $C(X)$ used in a finite subcover. □

Let Y be the closure of the complement of the union of a system of NRRPs satisfying the hypotheses of Proposition 6.2 in a half-translation surface X, and assume every singularity of X is contained in some NRRP. If we double Y along ∂Y, the foliations by vertical and horizontal segments extend by reflection, and Y acquires the structure of a quadratic differential, with singularities of cone angle 3π at the vertices of ∂Y.

Moreover, if we take the Delaunay triangulation of the resulting surface, the midpoint of any boundary edge is strictly closer to the endpoints of that edge than to any other singularity, so all edges of ∂Y belong to the Delaunay triangulation of the double cover.

Therefore, *it makes sense simply to speak of the Delaunay triangulation of Y*. That is to say, if Σ is the set of vertices of Y (which are all on the boundary of Y), then we have proved:

Proposition 6.3 *Y has a triangulation with the property that the circumcenter of each triangle is the boundary of a standard Euclidean disk of radius equal to the distance to any vertex of the triangle. The boundary edges of Y are all edges.* □

Now, we describe how to build a quasiconformal map between two nearby quadratic differentials, provided they are within some distance $c(K)$ that depends only on K.

Let X_1 and X_2 be two such Riemann surfaces, and let Y_1 be the complement of a system of NRRPs for X_1 satisfying the hypotheses of Proposition 6.2.

The circumradii of Delaunay triangles on any half-translation surface are bounded above by twice the diameter of the surface, which is bounded on K. The side lengths of Y_1 are bounded below by $C(K)$. Thus by the law of sines, the angles of Delaunay triangles are bounded below. So we could have picked $c(K)$ small enough that when we change all of the period coordinates by at most $c(K)$, all angles of the Delaunay triangulation remain bounded away from zero. So the Delaunay triangulation for Y is uniformly bounded away from degeneration, ie all of the side lengths and angles are uniformly bounded away from 0. If $d_{\text{Euclidean}}(X_1, X_2) < c(K)$, then along a path from X_1 to X_2 we can make a choice of NRRP decomposition so that the radii, edges of NRRPs, and edges of Delaunay triangles that start as edges of Y_1 vary in a Lipschitz manner with respect to arc length, and the Lipschitz constant depends only on K. Let Y_2 be the complement of the NRRPs in X_2, as chosen along the path.

By Proposition B.2 the triangulation of Y_1 persists to a triangulation Y_2, and this determines an affine map on triangles of Y_1 to a triangulation of Y_2. Since no edge gets too short and no angle gets too close to 0, it follows that if the length of the path is ϵ, then as $\epsilon \to 0$ this part of the map is $(1+O(\epsilon))$–quasiconformal on these triangles, and the implied constants in the $O(\epsilon)$ depend only on K.

6.2 The Beurling–Ahlfors extension

The final step is to show how to build a quasiconformal map between NRRPs, and to estimate its dilatation. We will do this by means of a Beurling–Ahlfors extension. Given an orientation-preserving self-homeomorphism h of $\mathbb{R}^+$, we have the following:

Definition 6.4 The *Beurling–Ahlfors extension* of h to the upper half-plane *with parameter* $r > 0$ is the function f_r given by

$$f_r(x+iy) = \frac{1}{2}\int_0^1 [h(x+yt)+h(x-yt)]\,dt + \frac{ir}{2}\int_0^1 [h(x+yt)-h(x-yt)]\,dt.$$

Definition 6.5 Let $\rho \in [1,\infty)$. We say that an orientation-preserving homeomorphism $h\colon \mathbb{R} \to \mathbb{R}$ is *ρ–quasisymmetric* on $\mathbb{R}$ if for all $x \in \mathbb{R}, t > 0$ we have

$$\frac{1}{\rho} \le \frac{h(x+t)-h(x)}{h(x)-h(x-t)} \le \rho.$$

Theorem 6.6 [Beurling and Ahlfors 1956] *If h is ρ–quasisymmetric, then for some explicit choice of r the Beurling–Ahlfors extension f_r is ρ^2–quasiconformal. Moreover, $r = 2 + O(\rho - 1)$ as $\rho \to 1^+$.*

Note that $r = 2$ extends a Möbius transformation of $\mathbb{RP}^1$ to a Möbius transformation of $\mathbb{CP}^1$. Any choice of r yields a quasiconformal extension, but possibly with worse dilatation.

In the event that an NRRP does not contain a marked point, the Beurling–Ahlfors extension, together with the piecewise-affine maps with compatible boundary conditions, completes our construction of a quasiconformal map. In the case with marked points, we will have to compose with a quasiconformal map on each NRRP which is trivial on the boundary and moves the marked point to the correct location. For this we will uniformize our map to the upper half-plane and apply the unique $\mathbb{R}$–linear map on the complex plane fixing the real line pointwise and taking our marked point to its desired location.

We have established uniform control over the NRRPs. If the only interior singularity is a pole, then our map of NRRPs is in fact already an isometry in the singular metric, and hence conformal. In all other cases we have at least four poles on the boundary, and we can assume three of them are 0, 1, and ∞. Subject to this normalization, if we double each NRRP along its boundary, the singularities in the interior of each NRRP can be chosen to belong to some fixed compact subset of the upper half-plane.

This motivates the following:

Lemma 6.7 *Suppose that P_1 and P_2 are NRRPs. Let $h_q\colon \partial P_1 \to \partial P_2$ be a piecewise-affine identification of sides with respect to the singular metrics on P_1 and P_2, and affine on each pair of corresponding sides. Assume that when we uniformize each of P_1 and P_2 to the upper half-plane, three pairs of corresponding corners are sent to 0, 1, and ∞, the distance between any two finite singularities on the boundary is between $1/B$ and B for $B > 1$, and all interior singularities belong to some fixed compact subset K of the open upper half-plane. For all sufficiently small ϵ, if the singularities of P_1 and P_2 can be put in bijection so that corresponding singularities are distance at most ϵ apart in the plane when uniformized, then the quasisymmetry constant ρ of the boundary map $h\colon \mathbb{R} \to \mathbb{R}$ induced by h_q and our chosen uniformizations is at most $1 + C\epsilon$, for some C depending only on K, the number of sides of P_1 and P_2, and B.*

Proof The boundary map h is continuous, and it is differentiable except possibly at points corresponding to vertices of the NRRP. It is enough to show that the derivative of the boundary map satisfies $h'(x)-1 = O(\epsilon)$ uniformly on the complement of these points.

Suppose that one pair of corresponding sides is $h([a_1, b_1]) = [a_2, b_2]$, where a_1, a_2, b_1, and b_2 are all finite. (We will deal with the intervals with an endpoint at ∞ later.) Assume the quadratic differentials are $(p_i(z)/((z-a_i)(z-b_i)q_i(z)))\,dz^2$ on $[a_i, b_i]$ where p_i and q_i are polynomials. Then for $a_1 < c < b_1$, our boundary map h is defined so that

$$\frac{\int_{a_1}^{c}\bigl(\frac{p_1(z)}{(z-a_1)(z-b_1)q_1(z)}\bigr)^{1/2}\,dz}{\int_{a_1}^{b_1}\bigl(\frac{p_i(z)}{(z-a_i)(z-b_i)q_i(z)}\bigr)^{1/2}\,dz} = \frac{\int_{a_2}^{h(c)}\bigl(\frac{p_2(z)}{(z-a_2)(z-b_2)q_2(z)}\bigr)^{1/2}\,dz}{\int_{a_2}^{b_2}\bigl(\frac{p_2(z)}{(z-a_2)(z-b_2)q_2(z)}\bigr)^{1/2}\,dz}.$$

We differentiate both sides of the above equation in c and solve for $h'(c)$:

$$h'(c) = \frac{\int_{a_2}^{b_2}\bigl(\frac{p_2(z)}{(z-a_2)(z-b_2)q_2(z)}\bigr)^{1/2}\,dz}{\int_{a_1}^{b_1}\bigl(\frac{p_1(z)}{(z-a_1)(z-b_1)q_1(z)}\bigr)^{1/2}\,dz}\left(\frac{p_1(c)q_2(h(c))(h(c)-a_2)(h(c)-b_2)}{p_2(h(c))q_1(c)(c-a_1)(c-b_1)}\right)^{1/2}.$$

By applying changes of coordinates consisting of translations by real numbers to each NRRP, we can make a few simplifying assumptions: at the cost of increasing ϵ to 2ϵ we can do a translation to assume $a_1 = a_2 = 0$. After this simplifying assumption, our argument can be rephrased. Let

$$f_i(z) = \frac{\left(\frac{p_i(z)}{z(z-b_i)q_i(z)}\right)^{1/2}}{\int_0^{b_i}\left(\frac{p_i(t)}{t(t-b_i)q_i(t)}\right)^{1/2}dt}.$$

Then $h'(c) = f_1(c)/f_2(h(c))$. So we need to show $\log[f_1(c)/f_2(h(c))] = O(\epsilon)$.

We will actually do this under the assumption that $0 < c \le \frac{2}{3}b_i$. A basically identical argument will cover the cases $\frac{1}{3}b_i < c < b$, and a final argument will extend this to cover the two pairs of sides of P_1 and P_2 that have ∞ as an endpoint after we uniformize.

In what follows, C will be used to denote various positive constants that depend only on choices we have already made — its meaning may vary from one line to the next. The following will be immediate from our compactness assumptions. To simplify matters, we first prove:

Claim 6.8 $$|\log(h(c)/c)| \le C\epsilon.$$

The idea is that the f_i are equal to $z^{-1/2}$ times two very similar functions. Recall that h is defined so that $\int_0^c f_1(z)\,dz = \int_0^{h(c)} f_2(z)\,dz$.

Define $g_i(z) = z^{1/2} f_i(z)$.

We know the following:

- The g_i are either both purely imaginary or both purely real, and never change sign.
- $-C \le \log|g_i(z)| \le C$ and $|g_i'(z)| \le C$ for $z \in \left[0, \frac{3}{4}b_i\right]$.
- $|\log(f_1(z)/f_2(z))| = |\log(g_1(z)/g_2(z))| < C\epsilon$ for $z \in \left(0, \frac{3}{4}b_i\right]$.
- $-C \le \log\left[z^{1/2}\int_0^z f_i(t)\,dt\right] \le C$.
- $\frac{d}{dz}\log\left|\int_0^z t^{-1/2}g_i(t)\,dt\right| \ge z^{-1/2}/(Cz^{1/2}) = 1/(Cz)$ for $z \in \left(0, \frac{3}{4}b_i\right]$.

From the last observation we see that the amount that we need to perturb z to change the value of $\log\int_0^z t^{-1/2}g_i(t)\,dt$ by ϵ is at most $Cz\epsilon$, for all sufficiently small ϵ. In particular, for $i = 2$ we can start with $\int_0^c f_2(t)\,dt$, and by moving the upper limit of integration most $C\epsilon c$ away from c we can adjust the value of the integral of f_2 by a factor of at least $1 + C\epsilon$. The intermediate value theorem implies that the adjustment of the upper limit of integration needed to get from c to $h(c)$ is at most $C\epsilon c$. This establishes the claim.

Now that we have the claim, the main case of the lemma follows quite easily: the logarithmic derivatives of all factors of f_i are bounded by some constant C on $[c, h(c)]$ or $[h(c), c]$, and $\log(f_1/f_2)$ is uniformly close to $\left(0, \frac{3}{4}b\right)$. So the claim, plus boundedness of logarithmic derivatives of the remaining factors, gives us $g_1(c)/g_2(h(c)) = 1 + O(\epsilon)$, and by the claim, $c/h(c) = 1 + O(\epsilon)$. This proves the main case of the lemma. Interchanging the roles of a_1 and b_1 handles the case in which c is in the right half of the interval $[a_1, b_1]$.

Now we must handle the infinite intervals. For $1 < c < \infty$, we can conjugate $h(z)$ by the map $z \to 1/z$ to a function $H(c)$ satisfying an estimate of the type we originally had. The lemma and its proof hold for H; in particular $H(1/c)/(1/c) = 1 + O(\epsilon)$. Finally, we have

$$h'(c) = \frac{d}{dc}\left(\frac{1}{H(1/c)}\right) = H'(1/c)\frac{1}{c^2 H(1/c)^2} = 1 + O(\epsilon). \qquad \square$$

Proof of Theorem 1.2 We describe how to build a quasiconformal map between two nearby quadratic differentials, provided they are within some distance $c(K)$ depending only on a compact set $K \subset \mathrm{QD}(\mathcal{T}_{g,n})$.

Let X_1 and X_2 be two such Riemann surfaces, and let Y_1 be the complement of a system of NRRPs for X_1 satisfying the hypotheses of Proposition 6.2.

The circumradii of Delaunay triangles on any half-translation surface are bounded above by twice the diameter of the surface, which is bounded on K. The side lengths of Y_1 are bounded below by $C(K)$. Thus by the law of sines, the angles of Delaunay triangles are bounded below. So we could have picked $c(K)$ small enough that when we change all of the period coordinates by at most $c(K)$, all angles of the Delaunay triangulation remain bounded away from zero. So the Delaunay triangulation for Y is uniformly bounded away from degeneration, ie all of the side lengths and angles are uniformly bounded away from 0. If $d_{\text{Euclidean}}(X_1, X_2) < c(K)$, then along a path from X_1 to X_2 we can make a choice of NRRP decomposition so that the radii, edges of NRRPs, and edges of Delaunay triangles that start as edges of Y_1 vary in a Lipschitz manner with respect to arc length, and the Lipschitz constant depends only on K. Let Y_2 be the complement of the NRRPs in X_2, as chosen along the path.

Since the triangulation persists, this determines an affine map on triangles of Y_1 to a triangulation of Y_2. Since no edge gets too short, and no angle gets too close to 0, it follows that if the length of the path is ϵ, then as $\epsilon \to 0$ this part of the map is $(1+O(\epsilon))$–quasiconformal on these triangles, and the implied constants in the $O(\epsilon)$ depend only on K. We need only extend this map to the interiors of the disks, which are pairs of NRRPs in Y_1 and Y_2.

The NRRPs themselves are represented by quadratic differentials on disks, which can be uniformized to the closure of the upper half-plane in $\widehat{\mathbb{C}}$, and the metrics induce quadratic differentials on these close disks.

We may assume three pairs of corresponding vertices map to 0, 1, and ∞ under this uniformization if the NRRPs do not contain marked points; we may assume that one vertex is ∞ and the marked point is sent to i if there is a marked point. The quadratic differentials on corresponding pairs of NRRPs are Euclidean distance $O(\epsilon)$ in the moduli spaces of quadratic differentials in Lemma 5.12.

Thus, the locations of the zeros and poles in uniformized coordinates differ by $O(\epsilon^{\alpha_{g,n}})$, where $\alpha_{g,n} = 2/[2 + a_n(4g - 4 + n)]$, by Lemma 5.11. This allows us to use Lemma 6.7 to estimate the dilatation of the map on the NRRPs as $1 + O(\epsilon^{\alpha_{g,n}})$.

We have now produced a quasiconformal map f between the compact Riemann surfaces that are the completions of X_1 and X_2 with dilatation $1 + O(\epsilon^{\alpha_{g,n}})$. Unfortunately, it might not send marked points on X_1 to marked points on X_2, and this problem occurs precisely when the marked points are in NRRPs.

However, if we can find another quasiconformal map $\tau\colon X_2 \to X_2$ with dilatation $1 + O(\epsilon^{\alpha_{g,n}})$ that fixes the complement of each NRRP in X_2, but such that $\tau(f(p_1)) = p_2$ whenever p_1 and p_2 are corresponding pairs of marked points in X_1 and X_2, respectively, then the log of the dilatation of $\tau \circ f$ is $O(\epsilon^{\alpha_{g,n}})$, ie the Teichmüller distance between X_1 and X_2 is $O(\epsilon^{\alpha_{g,n}})$.

In uniformized coordinates, the map from the boundary of each NRRP in X_1 to the corresponding NRRP of X_2 is nearly the identity on any compact subset of the real line. More precisely, if S is a compact subset of the real line, then for all $x \in S$ we have

$$f(x) - x = O(\epsilon^{\alpha_{g,n}}).$$

It follows that for any compact $K \subset \mathbb{H}$ the Beurling–Ahlfors extension moves points in K by $O(\epsilon^{\alpha_{g,n}})$ as well.

We therefore pick τ to be the $\mathbb{R}$–linear map that is the identity on $\partial\mathbb{H}$ and moves the marked point in K from $f(i)$ to i, which is a distance of $O(\epsilon^{\alpha_{g,n}})$. Then on the interior of each NRRP, τ also has constant dilatation that is $1 + O(\epsilon^{\alpha_{g,n}})$.

We have constructed a homeomorphism that satisfies the dilatation bound at almost every point, as its first partial derivatives are defined off of a measure 0 set, and these partial derivatives are clearly bounded away from every point except possibly the singularities and vertices of the NRRPs since there is a conformal metric in which the partial derivatives converge piecewise. This shows that if we delete a finite set, the first partial derivatives are in L^2_{loc}. To finish, we simply recall that a homeomorphism which is M–quasiconformal on the complement of a finite set is M–quasiconformal; see eg [Hubbard 2006]. □

Appendix A Local finiteness of period coordinate systems

In this appendix we show the following, leaving the proof to the end of the section:

Proposition A.1 *There are only finitely many systems of period coordinates represented by saddle connections of length at most D in any compact subset K of $\mathrm{QD}^1(\mathcal{T}_{g,n})$.*

To this end, we will show that every such period coordinate system is related to a basis of the edges of the Delaunay triangulation by one of finitely many transition matrices, and there are only finitely many choices of Delaunay triangulation on K. We will always assume that our triangulations are *labeled* and *marked*, that is, the vertices and edges are distinguishable, we know which marked points (if any) on the base surface correspond to which singularities, and we know which homotopy classes of curves on the surface are represented by which homotopy classes of curves on the graph. Finally, we will also include in the data of the Delaunay triangulation the sign of the slope of each saddle connection.

Proposition A.2 *Up to the action of the mapping class group, there are only finitely many triangulations that can occur as Delaunay triangulations in $\mathrm{QD}^1(\mathcal{T}_{g,n})$.*

Proof The number of triangles is a function only of the stratum, since the sum of the angles of a triangle is π and the sum of the angles of all triangles is the sum of the cone angles of the singularities. (There is one triangle for each vertical separatrix emanating from a singularity.) □

Proposition A.3 *Only finitely many Delaunay triangulations occur in any compact* $K \subset \mathrm{QD}^1(\mathcal{T}_{g,n})$.

Proof If not, then infinitely many such triangulations occur in a single orbit of the $\mathrm{Mod}(S_{g,n})$. Given a fixed Delaunay triangulation T, the space of complex lengths we can assign to the edges and still have a unit-area quadratic differential with systole $> \epsilon$ and such that the triangulation remains Delaunay (but possibly becomes degenerate) is compact. (If there are no short simple closed curves, Mumford's compactness criterion implies the diameter of the surface, and hence the diameter of each Delaunay triangle, is bounded. The inequalities that guarantee that the triangulation is a Delaunay triangulation are closed conditions.) This contradicts the proper discontinuity of the action of $\mathrm{Mod}(S_g, n)$ on $\mathrm{QD}^1(\mathcal{T}_{g,n})$. □

Proposition A.4 *There is a uniform bound on the number of geodesic concatenations of saddle connections of length at most* D *on* K. *In particular, there is a uniform bound on the number of saddle connections of length at most* D *on* K.

Proof We first reduce to the case in which $(X, q) \in K$ has no poles. If q has marked points, we simply find q' belonging to a compact K' in a moduli space of higher-genus surfaces by taking a pole-free degree-2 branched cover that is branched over all poles such that any geodesic concatenation of saddle connections on (X, q) can be lifted to a saddle connection or geodesic concatenation of saddle connections on (X', q'); we then bound the number of geodesic segments joining cone points and marked points on (X', q').

To form (X', q') in a way that guarantees that X' stays in a compact K', the branching will be over all marked points, and, if the number of marked points is odd, one additional point. We pick this extra branch point to be as far away as possible from the marked points of X. To this double cover we give the metric associated to the quadratic differential that is $1/\sqrt{2}$ times the pullback of q (this is so that the area is 1). Because we have a lower bound on the distance between branch points in (X, q) in K, the space K' of double covers has a lower bound on the injectivity radius, ie it satisfies Mumford's compactness criterion. Thus we have reduced to the case of dealing with a compact subset of a moduli space without poles.

To deal with surfaces without poles, we fix a word metric on $\pi_1(S_g)$. We first claim that there exist real numbers $k, c > 0$ such that whenever $(X, q) \in K$, the fundamental group of S_g in this word metric is equivariantly (k, c)–quasi-isometric to the universal cover $(\widetilde{X}, \tilde{q})$ of $(X, q) \in K$ via a (k, c) quasi-isometry $\phi_{(X,q)}$ that sends $h \in \pi_1(K)$ to $h(x, q)$, where $x \in X$ is chosen in a continuous section of the universal disk bundle over $\mathrm{QD}^1(\mathcal{T}_g)$, ie the universal cover of the surface bundle over $\mathrm{QD}(\mathcal{T}_{g,n})$. The uniformity of quasi-isometries then follows from finiteness of $\{h \subset \pi_1(S_g) : d_{\tilde{q}}(x, hx) < R \text{ for some } x \in K\}$.

Now, for each Delaunay triangulation T that appears in K, we make the quasi-isometries between the universal covers of surfaces with triangulation T more explicit by picking the basepoint to be a vertex p. By virtue of the quasi-isometry constants between the fundamental group of X and the universal cover of X, there is a constant M depending on K and D such a ball of radius D in the universal cover of (X,q) can contain at most M lifts of any point in X distance less than D from any fixed lift of the basepoint.

If this were not the case, then there would be no upper bound on the number of closed geodesics in X with basepoint p of length at most $D+\max_{(X,q)\in K}\operatorname{diam}_q(X)$ as X varies over K; this would contradict the existence of the uniform quasi-isometry constants between the fundamental group and the universal cover of (X,q).

In particular, we conclude that for any Delaunay vertices v_1 and v_2, and a fixed lift w of v_1, there are at most M lifts of geodesic paths of length at most D that start at w and end at lifts of v_2, and therefore at most M geodesics from v_1 to v_2 on X of length at most D.

Since we may have chosen T to be any Delaunay triangulation and p to be any vertex, and the number of vertices is obviously bounded in K, this means that there is an upper bound on the number of geodesics of length at most D whose endpoints are singularities of q–metrics for all (X,q). □

Proposition A.5 *If $q_n \in K$ all have the same Delaunay triangulation and $q_n \to q$, then the common Delaunay triangulation of the q_n is a possibly degenerate Delaunay triangulation of q.*

Proof The conditions for a triangulation to be Delaunay are nonstrict inequalities in the absolute values of period coordinates, which persist under taking limits. □

Corollary A.6 *For any $D>0$ and any compact $K \subset \mathrm{QD}^1(\mathcal{T}_{g,n})$ there are only finitely many period coordinate systems consisting of saddle connections of length less than D.*

Proof There are only finitely many period coordinate systems that can be Delaunay triangulations in a neighborhood of each point in K because there are only finitely many Delaunay triangulations at each point, and only finitely many Delaunay triangulations that can degenerate to a triangulation T without changing the topology. Thus each point in K has a neighborhood with a uniformly bounded number of possible Delaunay triangulations, and a simple compactness argument tells us that there are only finitely many Delaunay triangulations of surfaces in K. For each subset of K sharing a common Delaunay triangulation we can fix a basis of period coordinates coming from edges of the triangulation, and we can write the period of every homotopy class of geodesic arc whose endpoints are singularities in terms of this basis.

Finally, we claim that the saddle connections in a fixed homotopy class can only have finitely many representations in terms of any fixed basis for $H_1^{\mathrm{odd}}(\widetilde{X},\widetilde{\Sigma})$ consisting of saddle connections that are edges of the Delaunay triangulation, if we force (X,q) to remain in K. Indeed, by uniform quasi-isometry of the fundamental group with $\widetilde{X}$ (endowed with the pulled back q metric), such a saddle connection can only pass through a fixed finite list of Delaunay triangles, and it cannot leave and then enter a triangle since triangles are geodesically convex in the q–metrics. Thus there is a finite list of sequences of triangles a saddle connection can pass through, and by developing this chain of triangles in the plane, we can write its period as one of finitely many sums of sides of periods of the edges of the triangles. □

Corollary A.7 *For each compact $K \subset \mathcal{T}_{g,n}$ there is a finite set of triangulations $\{T_i\}_i \in I$ such that all quadratic differentials lying over K belong to a T_i–convex set for some $i \in I$.*

Proof the q–metric on (X, q) is obtained by pulling back the Euclidean metric on $\mathbb{R}^2$ from a collection of charts, but one could just as easily have pulled back the L^∞–metric $d(a+bi, c+di) = \max(|a-c|, |b-d|)$ to form an L^∞ q–metric on X. We may consider Delaunay triangulations with respect to the L^∞ flat metric, following [Guéritaud 2016]. Every triangle in the L^∞ Delaunay triangulation has its vertices on the boundary of an open square that is maximal with respect to the conditions that its lifts to the universal cover are embedded and that its edges be vertical and horizontal. If a triangulation is the L^∞ triangulation for a set of surfaces in $\mathrm{QD}(\mathcal{T}_{g,n})$ that has nonzero Masur–Veech measure, then no edge of the triangulation is always horizontal or always vertical. So we may further restrict each saddle connection in the triangulation to have nonpositive slope or nonnegative slope, and include that in the combinatorial data of the triangulation. We may also disregard any triangulations that arise only on sets of measure zero.

We may now claim the conditions that a triangulation be L^∞ Delaunay are given by (nonstrict) linear inequalities in the period coordinates. Moreover, no triangle can have three edges of positive slope or three edges of negative slope. Hence, if we orient all edges so that the imaginary parts of periods are nonnegative, then for each L^∞ triangulation T the space of surfaces for which T is an L^∞ triangulation is T–convex.

Moreover, for each triangle there is a point p that is equidistant from the three vertices in the L^∞–metric, and those three vertices are the nearest three vertices to p in the L^∞ metric (possibly along with some others, if the L^∞ triangulation is not unique). If this distance is d then they do not all have positive slope or all negative slope, since they touch at least three sides of a $2d \times 2d$ square with vertical and horizontal sides centered at p. This imposes a uniform bound on the lengths of saddle connections in the L^∞ Delaunay triangulation of $2\sqrt{2}$ times the diameter of the surface. Since it is obviously enough to consider unit-area surfaces lying over K, which have bounded diameter, this imposes a bound on the number of possible L^∞ Delaunay triangulations. □

Proof of Proposition A.1 There are locally finitely many L^∞–Delaunay triangulations; for each such triangulation we can locally fix names of the singularities. The collection of all singularities in the universal cover of (X, q) is quasi-isometric to the fundamental group of the compact surface $\overline{X}$, uniformly for all $(X, q) \in K$. In particular, there are finitely many ways to choose the names of the endpoints of a saddle connection of length at most L in the metric universal cover of $(\overline{X}, q)$, up to the action of $\pi_1(\overline{X})$, and therefore only finitely many systems of saddle connections of length at most L for each L^∞ Delaunay triangulation that occurs in K. □

Appendix B δ–clusters and the Euclidean metric

First, we would like to establish a basic fact about systems of period coordinates persisting under perturbation. This will be useful throughout:

Proposition B.1 *Let $\mathcal{Q}(\mu)$ be a stratum of cluster differentials. Then there is a constant $\epsilon > 0$, depending only on μ, such that whenever Γ is a length-minimizing system of saddle connections on the vertex set of singularities of $q \in \mathcal{Q}(\mu)$ there is a holomorphic coordinate system $U \subset \mathbb{C}^r$ given by the logs of the periods $(\log(P_1), \dots, \log(P_r))$ of saddle connections in Γ, and moreover U contains an ϵ–ball about q in which the coordinates are logarithms of periods of saddle connections.*

Proof First we reduce to the case in which the stratum has no poles, except for a higher-order pole at ∞. In the case of a cluster differential there is a length-minimizing tree of saddle connections Γ, and we claim that the length-minimizing tree $\widetilde{\Gamma}$ for the double cover $(\widetilde{X}, \tilde{q})$ of (X, q) branched only over the pole contains the lifts of all the edges of the base graph Γ. Therefore a small perturbation of the logs of the periods of Γ arises as the quotient of a perturbation of the logs of the periods of $\widetilde{\Gamma}$ which descends. The proof is by following the greedy algorithm for length-minimizing spanning trees: repeatedly pick the shortest edge that does not form a cycle with previous edges. Every saddle connection of $(\widetilde{X}, \tilde{q})$ is a lift of a saddle connection of (X, q). Following the greedy algorithm to build $\widetilde{\Gamma}$ will thus start with a lift of the shortest edge of Γ and then the other lift; this does not form a cycle because the double cover of the graph Γ branched over the pole is still a tree. Now, continuing inductively, assume that the first $2k$ edges of $\widetilde{\Gamma}$ are the lifts of the first k edges of Γ. The shortest saddle connection of Γ that has not been picked has lifts which do not form cycles with the previously picked edges of $\widetilde{\Gamma}$, because the double cover of Γ is a tree. However, any shorter edge would project down to an edge e, which, when added to the first k edges of Γ, creates a cycle. Adding the two lifts of e to Γ and Γ' would therefore produce a graph with first homology group of rank 2, since it would be obtained by taking two disjoint copies of a graph with first integral homology group $\mathbb{Z}$ and identifying a pair of vertices. Since deleting one edge of a graph can only decrease the rank of the first homology group by 1, it follows that adding just one lift of e would create a cycle with the first $2k$ edges of Γ. Therefore the lift of the saddle connection e is not available to pick as the next edge of $\widetilde{\Gamma}$. Hence the two lifts of the $(k+1)^{\text{st}}$ edge of Γ can be chosen by the greedy algorithm, and it is possible to choose one and then the other as the next two edges of $\widetilde{\Gamma}$ while following the greedy algorithm. By induction, our claim follows, as does the reduction to the case with no poles.

Now suppose the proposition is false. We can find a sequence of counterexamples $\{q_m\}$ that would converge in the sense of Definition 4.35, and the graph Γ, as well as which collections of vertices correspond to vanishing clusters, are the same for all m. There would be perturbation vectors h_m of the logs of periods with $\|h_m\| \to 0$ such that $\{(\log P_1(q_m), \dots, \log P_r(q_m)) + t h_m : t \in [0, 1)\}$ represents log-period coordinates of saddle connections with the graph Γ, but $(\log P_1(q_m), \dots, \log P_r(q_m)) + h_m$ does not. If the sequence $\{q_m\}$ has no vanishing clusters, this follows easily from the fact that $\log P_i$ is a system of period coordinates. So we induct on the number of nested clusters.

First, we note that the edges of Γ contain a spanning tree for each vanishing cluster of $\{q_m\}$, ie there are $n-1$ edges that are internal to each vanishing cluster with n singularities. Otherwise there would be a trivial improvement of Γ by adding an edge joining two singularities in the same vanishing cluster that were not connected by a path in Γ that does not go outside the cluster, and deleting some other edge. In fact, the edge set of Γ is naturally in bijection with the disjoint union of the edge sets of length-minimizing trees of saddle connections on the quadratic differentials corresponding the various cluster differentials (X_ℓ, q_ℓ) associated to the limit of the sequence $\{q_m\}$ in the compactification of Definition 4.35. To obtain the saddle connections corresponding to a length-minimizing tree in (X_ℓ, q_ℓ), simply delete all singularities outside S_ℓ and contract all edges belonging to each $S_j \subsetneq S_\ell$; the remaining graph will have edges corresponding to S_ℓ.

We will show that for large enough m, each singularity is bounded away from each edge of Γ of which it is not a vertex along $\gamma_m([0,1)) := \{(\log P_1(q_m), \dots, \log P_r(q_m)) + t h_m : t \in [0,1)\}$, contradicting the fact that the coordinate system does not extend the point where $t = 1$.

Step 1 For any two singularities a and b, $d_{q_m}(a,b)$ and the distance between a and b along the metric graph Γ is uniformly comparable as $m \to \infty$, ie $d_{q_m}(a,b) \stackrel{.}{\asymp} d_{q_m}|_\Gamma(a,b)$. The implied constants in $\stackrel{.}{\asymp}$ depend on the sequence but not on m. Moreover, both of these distances are comparable to ($\stackrel{.}{\asymp}$) the q_m–diameter of the smallest vanishing cluster $S_\ell(a,b)$ containing a and b, or the entire singularity set if no such vanishing cluster exists, and comparable to the length of any saddle connection that is internal to $S_\ell(a,b)$ but not to a proper vanishing subcluster.

Proof This is clear from the definitions and Proposition 4.29, since any saddle connection whose limit is less than every positive multiple of $\operatorname{diam}_q(S_\ell(a,b))$ would belong to a proper subcluster. □

Step 2 Let $d_{q_m + t h_m}$ denote the metric at time t along γ_m. For any two singularities a and b, $(d/dt) \log d_{q_m + t h_m}(a,b)$ is defined for almost every t, and converges to 0 uniformly along points of definition as $m \to \infty$. Moreover, $\log d_{q_m + t h_m}(a,b)$ is locally absolutely continuous, so the integral of this derivative represents the change in length.

Proof Since the angle formed by two saddle connections varies real-analytically, the set of times at which the configuration of saddle connections on the geodesic from a to b changes is discrete, consisting only of isolated times at which two of the saddle connections form an angle of π. This establishes the final claim, so we must now bound the logarithmic derivatives of distances. This is done by induction on the nesting of the cluster containing a and b.

First, we note that the distance from a to b is the sum of the lengths of finitely many saddle connections, and a saddle connection from c to d has period equal to the sum of the periods of saddle connections of the path from c to d in Γ with appropriate signs. There are therefore only finitely many ways to express the periods of saddle connections comprising the geodesic from a to b as linear combinations of periods from Γ, and since the geodesic does not revisit an edge, there is a bound on how many saddle connections are in the geodesic. So the derivatives (but not necessarily logarithmic derivatives) of distances are all uniformly bounded and tend to 0. In particular, this means that for each $\delta > 0$, every maximal vanishing cluster remains a δ–cluster along γ_m for almost every m.

Now, by induction, assume that some vanishing cluster S_ℓ is, for each $\delta > 0$, a δ–cluster in the singularity set of each point in γ_m for all sufficiently large m. Moreover assume that for all proper subclusters

$S_k \supsetneq S_\ell$, the conclusion holds for any singularities that belong to S_k but not to a common subcluster of S_k. Then along all but finitely many γ_m, the unique geodesic joining each pair of singularities in S_k consists entirely of saddle connections that are internal to S_k. We can then rescale each q_m to assume S_k has q_m–diameter 1, and ignore all singularities outside of S_k. By the same arguments, the conclusion holds for all pairs of singularities in S_k that are not in any proper vanishing subcluster. By induction, the conclusion holds for all vertices a and b of Γ. □

Step 3 By Step 2, no singularities collide along γ_m as $t \to 1$ and γ_m may be extended to all $[0, 1]$, but with the possibility that some edges of Γ may degenerate to concatenations of saddle connections rather than saddle connections at $\gamma(1)$. Moreover, for each pair of singularities a and b, the direction of the geodesic from a to b varies almost everywhere differentiably, locally absolutely continuously, and with derivative uniformly tending to 0 on points of definition as $m \to 0$. In particular, for each saddle connection that exists along γ, the angle changes absolutely continuously, and differentiably with uniformly bounded derivative.

Proof It is sufficient to show that the imaginary part of the logarithmic derivative of the period of every saddle connection goes to 0 as $\gamma_m \to \infty$. If a and b are joined by a saddle connection at some point along γ_m and m is sufficiently large, let $S_\ell(a,b)$ be the minimal vanishing cluster containing a and b. For all points $q_m + th_m$ on γ we have $d_{q_m+th_m}(a,b) \stackrel{.}{\asymp} \operatorname{diam}_{q_m}(S_\ell(a,b))$. The period of the saddle connection joining a and b is a bounded linear combination of periods of saddle connections of Γ internal to $S_\ell(a,b)$, all of which are at most a bounded multiple of $d_{q_m+th_m}(a,b)$, and each of these periods has logarithmic derivative going to 0 as $m \to \infty$. It follows that the logarithmic derivative of the saddle connection joining a to b goes to 0. □

Step 4 By Step 2 no singularities collide along γ_m as $t \to 1$, so it must be the case that the distance from some singularity a to an edge of Γ joining two vertices $b, c \neq a$ goes to 0 along γ_m as $t \to 1$. The only way this can happen is if $d_{q_m}(a,b)/d_{q_m}(b,c) \to 0$ or $d_{q_m}(a,c)/d_{q_m}(b,c) \to 0$.

Proof By the length-minimizing property, each saddle connection of Γ is a side of two equilateral singularity-free triangles in q_m. Then $d(a,b) + d(a,c) \to d(b,c)$ along γ_m as $t \to 1$, but since distances are preserved up to a small multiplicative error, $\lim_{m\to\infty}(d_{q_m}(a,b) + d_{q_m}(a,c))/d_{q_m}(b,c) \to 1$. Now, if b and c belong to a vanishing cluster not containing a this is impossible, and if the only vanishing clusters containing a and one of b or c contains all three, then a is not on the geodesic joining b and c in the cluster differential associated to this vanishing cluster. □

Step 5 Assume that as $t \to 1$ along infinitely many γ_m, the singularities a and b belong to a vanishing cluster S_ℓ, and b is joined to a vertex $c \notin S_\ell$ by an edge of Γ. Then the angle formed at b by the geodesic from b to c and the geodesic from b to a remains bounded away from 0 along γ_m.

Proof This is obvious from Step 3, since the geodesic joining a to b is a concatenation of saddle connections. This completes the proof, since as $t \to 1$, the distance between a and any saddle connection in Γ not containing a does not tend to 0 along γ_m as $t \to 1$. □

Proposition B.2 *Let K be a compact subset of $\mathcal{M}_{g,n}$ or a moduli space of cluster differentials. Then there is some $\epsilon(K) > 0$ such that for all half-translation surfaces (X, q, Σ), length-minimizing systems of saddle connections whose pairs of lifts added with opposite signs form a basis for $H^1_{\mathrm{odd}}(\widetilde{X}, \widetilde{\Sigma})$ persist when the logs of their periods all change by at most $\epsilon(K)$.*

Proof The first thing we observe is that given a sequence that converges in the sense of Definition 4.35, the collection of saddle connections that belong to saddle connections internal to vanishing clusters is either linearly independent or has exactly one dependence relation, which we will describe. Cutting $\widetilde{X}$ along this system of saddle connections produces either one connected component or two, and if produces two components the relation is that the sum of the lifts of the edges that divide the length-minimizing tree into two components, each of which contains an odd number of cone points of cone angle an odd multiple of π, taken with appropriate signs, is the shared boundary of two surfaces. Therefore all but the longest such edge will be included in any length-minimizing system of saddle connections. By Proposition B.2 all of the saddle connections in the length-minimizing system internal to vanishing clusters will persist, since a perturbation that changes the log-periods of all k of the remaining saddle connections by at most ϵ changes the log-period of the longest one by $O(k\epsilon)$. The proof that saddle connections that are not internal to vanishing clusters persist is essentially the same as in the cluster differential case, since $(\widetilde{X}, \tilde{q})$ admits a cover branched only over the poles, and the universal cover of this is complete and nonpositively curved, so there is a unique geodesic joining each pair of points. □

Proof of Lemma 4.43 This is immediate from Lemma 4.20, and Propositions 4.13 and B.2. □

Definition B.3 A length-minimizing period-coordinate system for (X, q) is *persistent* for (X', q') if for each coordinate, the change in the log-period is less than the quantity $\epsilon(K)$ for some compact set K containing (X, q) in its interior.

Proof of Proposition 4.41 First, we note that there is an upper bound on the systole of (X, q) and a lower bound on the diameter of (X, q) coming from K. We can fix a compact set $K' \subset \mathrm{QD}^1(\mathcal{M}_{g,n})$ which contains the projection of K to the moduli space $\mathcal{M}_{g,n}$, defined to be all surfaces (X, q) with some fixed lower bound on the d_q–systole and some fixed upper bound on the d_q–diameter. For all period coordinate systems on K' with upper bound L on the length of the saddle connections used, as (X, q) varies, the q–distances between pairs of singularities, the systole, and the diameter of (X, q) are uniformly Lipschitz with respect to the path metric $d_{\mathrm{Euclidean}}$, since there are locally finitely many period coordinate systems by Proposition A.1. Since there is a neighborhood of K which projects to K', and the path γ can be assumed to lie entirely in K', the proposition follows easily from the Lipschitz property. □

Proof of Proposition 4.44 Clearly the only way a sequence of counterexamples is possible is if

$$\frac{d_{\mathrm{Euclidean}}(X_m, Y_m)}{\epsilon_m} \to 0.$$

Obviously a sequence of counterexamples contains a subsequence in which $\{X_m\}$ converges in the sense of Notation 5.3 by passing to a subsequence. Moreover, we can assume that for every vanishing cluster D of $\{X_m\}$, the ratio $\operatorname{diam}_{q_m}(D)/\epsilon_m$ converges in $[0,\infty]$.

Moreover, we can pick a system of period coordinates that consists of a maximal set of saddle connections internal to vanishing clusters whose periods are linearly independent in $H^1(\widetilde{X},\widetilde{\Sigma};\mathbb{C})$, and a length-minimizing set of complementary geodesic arcs in X_m whose endpoints are singularities, that complete the system of period coordinates. Note that the saddle connections defining coordinate system are internal to vanishing clusters if and only if their lengths go to 0. Then, in these systems of period coordinates, we make the following claim:

Claim B.4 *The complementary geodesic arcs can be chosen* (*Hausdorff*) *continuously on the straight line-segment path from X_m to Y_m, and the change in their periods is $\dot{\asymp}\ \epsilon_m$ as $m\to\infty$.*

Proof The only periods that change are those whose lengths are bounded away from zero, and the perturbations are close to zero. Thus the length-minimizing coordinate system is persistent. The claim is thus immediate from Proposition B.2 and the fact that distance in each of the finite permissible coordinate systems is uniformly comparable to distance in the length-minimizing system.

If a sequence of counterexamples exists then we can of course choose them to lie in the principal stratum, since the principal stratum is dense on each coordinate chart.

Now, we assume that we have a counterexample sequence $\{(X_m,Y_m)\}_{m=1}^{\infty}$ with

$$d_{\text{Euclidean}}(X_m,Y_m)<\epsilon'_m=o(\epsilon_m),$$

ie $\lim_{m\to\infty}\epsilon'_m/\epsilon_m=0$. Then we will show that pairs of corresponding complementary saddle connections of X_m and Y_m in the persistent coordinate chart have periods differing by $o(\epsilon_m)$.

Consider a path $\gamma_m(t)$ from X_m to Y_m that is rectifiable, parametrized by arc length, and of length ϵ'_m with respect to $d_{\text{Euclidean}}$. For $1\le j\le 4g-4+2n$, make Hausdorff continuous choices of singularities $a_m^j(t)$ with the property that for each t, the collection of cone points $a_m^j(t)$, counted with multiplicity, are the singularities of $\gamma_m(t)$, where $a_m^j(t)$ for $1\le j\le n$ correspond to the marked points and $a_m^j(t)$ for $n+1\le j\le 4g-4+n$ correspond to the zeros with multiplicity (or multiplicity -1 if they collide with marked points). These choices are not unique, but they are unique up to a permutation of the singularities that preserves maximal vanishing clusters, since all nonuniqueness is caused by collisions of singularities, and this can only happen to two singularities in the same vanishing cluster. For a proof that it is possible to continuously choose a root along the path γ, see eg [Armstrong 1968, Lemma 1].

Then t runs from 0 to ϵ'_m, and let $\ell_m(t)$ be a continuous choice of a simple (non-self-intersecting) geodesic joining $a_m^j(t)$ and $a_m^k(t)$ such that $\ell_m(0)$ is one of the complementary saddle connections that is part of the length-minimizing system for X_m. We note that this simple geodesic is not quite determined by its endpoints and homotopy class because there may be poles, but there are at most two choices for each pair of endpoints and they can be chosen continuously by passing to an appropriate finite branched cover branched only over marked points and branched over all marked points, if such exist. (In particular, this operation is defined over the base surface, so this finite cover varies in a Lipschitz manner over $\mathrm{QD}(\mathcal{T}_{g,n})$.) In fact, we may assume that we picked the functions $a_m^r(t)$ to be projections of continuous choices of singularities in this branched double cover; each such choice determines a unique choice of $\ell_m(t)$. For the remainder of the proof we may assume $\ell_m(t)$ is chosen continuously in the universal cover of some branched cover.

In particular, the length of $\ell_m(t)$ is bounded above and bounded away from 0. Then, by the definition of the Euclidean path metric, the following are uniformly Lipschitz in t, ie Lipschitz with a constant that holds for all sufficiently large m:

(1) the length of $\ell_m(t)$,

(2) the distance between any two choices of $\alpha_m^r(t)$ and $\beta_m^r(t)$, if these are two possible ways of choosing $a_m^r(t)$, for each r,

(3) the diameter of the maximal vanishing cluster containing a_m^r along $\gamma_m(t)$, for each r,

(4) the distance from any singularity $a_m^r(t)$ to $\ell_m(t)$, and

(5) the slope of the longest segment of $\ell_m(t)$, measured as an angle in $\mathbb{R}/\pi\mathbb{Z}$.

Let b_m^j be the singularity that corresponds to $a_m^j(0)$ in the persistent coordinate system for X_m that extends to Y_m. In order for us to get a contradiction it must be the case that for some choice of a_j and a_k the singularity $a_m^j(\epsilon_m)$ in Y_m is not equal to b_m^j, since otherwise the length and angle of $\ell_m(\epsilon_m)$, and hence also the period of the corresponding saddle connection, would each differ by $O(\epsilon_m)$ from $\ell_m(0)$. In fact, it must be the case that for some j, $\epsilon'_m/d_Y(b_m, a_m^j(\epsilon'_m)) \to 0$. Otherwise, we could move the endpoints of the geodesic joining $a_m^j(t)$ and $a_m^k(t)$ to b_m^j and b_m^k, and so the length of the geodesic, and the slope of the longest segment, would vary in a manner that is Lipschitz in the change of the endpoints. The resulting geodesic joining b_m^j and b_m^k would be a single saddle connection, whose length and angle are $o(\epsilon_m)$ away from the period of the saddle connection joining $a_m(0)$ and $b_m(0)$.

Finally, we show that $a_m^j(\epsilon'_m)$ and $a_m^k(\epsilon'_m)$ are $O(\epsilon'_m)$ away from the singularities b_m^j and b_m^k in Y_m, which makes this contradiction impossible. If we vary the endpoints in Y_m along rectifiable curves, the distance and angle of the longest segment again vary in a uniformly Lipschitz manner with respect to the endpoints. So we can therefore move the endpoints to the singularities corresponding to a_m and b_m, and conclude that their lengths and angles have changed by $o(\epsilon_m)$. However, this is also impossible, since a singularity that is not an endpoint of a saddle connection of length M in a length-minimizing system and is distance d away from both ends of the saddle connection is distance at least $\min\bigl(\frac{\sqrt{3}}{2}d, M\bigr)$ away from the saddle connection. (This is due to the fact that the original saddle connection is a side of two singularity-free equilateral triangles). This gives a contradiction to item (4) above. □

References

[Armstrong 1968] **M A Armstrong**, *The fundamental group of the orbit space of a discontinuous group*, Proc. Cambridge Philos. Soc. 64 (1968) 299–301 MR Zbl

[Bainbridge et al. 2019] **M Bainbridge**, **D Chen**, **Q Gendron**, **S Grushevsky**, **M Möller**, *Strata of k–differentials*, Algebr. Geom. 6 (2019) 196–233 MR Zbl

[Beurling and Ahlfors 1956] **A Beurling**, **L Ahlfors**, *The boundary correspondence under quasiconformal mappings*, Acta Math. 96 (1956) 125–142 MR Zbl

[Farb and Margalit 2012] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[Fathi et al. 2012] **A Fathi**, **F Laudenbach**, **V Poénaru**, *Thurston's work on surfaces*, Math. Notes 48, Princeton Univ. Press (2012) MR Zbl

[Guéritaud 2016] **F Guéritaud**, *Veering triangulations and Cannon–Thurston maps*, J. Topol. 9 (2016) 957–983 MR Zbl

[Hubbard 2006] **J H Hubbard**, *Teichmüller theory and applications to geometry, topology, and dynamics, I: Teichmüller theory*, Matrix Editions, Ithaca, NY (2006) MR Zbl

[Hubbard and Masur 1979] **J Hubbard**, **H Masur**, *Quadratic differentials and foliations*, Acta Math. 142 (1979) 221–274 MR Zbl

[Maskit 1985] **B Maskit**, *Comparison of hyperbolic and extremal lengths*, Ann. Acad. Sci. Fenn. Math. 10 (1985) 381–386 MR Zbl

[Masur 1982] **H Masur**, *Interval exchange transformations and measured foliations*, Ann. of Math. 115 (1982) 169–200 MR Zbl

[Masur and Smillie 1991] **H Masur**, **J Smillie**, *Hausdorff dimension of sets of nonergodic measured foliations*, Ann. of Math. 134 (1991) 455–543 MR Zbl

[Mumford 1971] **D Mumford**, *A remark on Mahler's compactness theorem*, Proc. Amer. Math. Soc. 28 (1971) 289–294 MR Zbl

[Rafi 2007] **K Rafi**, *Thick-thin decomposition for quadratic differentials*, Math. Res. Lett. 14 (2007) 333–341 MR Zbl

[Royden 1971] **H L Royden**, *Automorphisms and isometries of Teichmüller space*, from "Advances in the theory of Riemann surfaces", Ann. of Math. Stud. 66, Princeton Univ. Press (1971) 369–383 MR Zbl

[Veech 1986] **W A Veech**, *The Teichmüller geodesic flow*, Ann. of Math. 124 (1986) 441–530 MR Zbl

Monroe Township, NJ, United States

ian.frankel@gmail.com

Received: 16 April 2021 Revised: 15 April 2023

Algebraic & Geometric Topology 24:5 (2024) 2509–2536
DOI: 10.2140/agt.2024.24.2509
Published: 19 August 2024

Topological Hochschild homology of truncated Brown–Peterson spectra, I

Gabriel Angelini-Knoll
Dominic Leon Culver
Eva Höning

We compute the topological Hochschild homology of sufficiently structured forms of truncated Brown–Peterson spectra with coefficients. In particular, we compute $\mathrm{THH}_*(B\langle n\rangle; H\mathbb{Z}_{(p)})$ for all n, where $B\langle n\rangle$ is an E_3 form of $\mathrm{BP}\langle n\rangle$ for certain primes p, and $\mathrm{THH}_*(B\langle 2\rangle; M)$ for $M \in \{k(1), k(2)\}$. For example, this gives a computation of $\mathrm{THH}(\mathrm{taf}^D; M)$ for $M \in \{H\mathbb{Z}_{(3)}, k(1), k(2)\}$ where taf^D is the E_∞ form of $\mathrm{BP}\langle 2\rangle$ constructed by Hill and Lawson.

16E40, 19D55, 55N22, 55P43, 55Q51; 55P42, 55Q10, 55T99

1 Introduction

Topological Hochschild homology and cohomology are rich invariants of rings, or more generally ring spectra, with applications to such fields as string topology [Cohen and Jones 2002], deformation theory of A_∞ algebras [Angeltveit 2008], and integral p–adic Hodge theory [Bhatt et al. 2019]. Topological Hochschild homology is also a first order approximation to algebraic K–theory in a sense made precise using Goodwillie calculus by [Dundas and McCarthy 1994].

Algebraic K–theory of ring spectra that arise in chromatic stable homotopy theory are of particular interest because of the program of Ausoni and Rognes [2002], which suggests that algebraic K–theory shifts chromatic complexity up by one, a higher chromatic height analogue of conjectures of Lichtenbaum [1973] and Quillen [1975]. A higher chromatic height analogue of one of the Lichtenbaum–Quillen conjectures was recently proven for truncated Brown–Peterson spectra $\mathrm{BP}\langle n\rangle$ by [Hahn and Wilson 2018]. However, it is still desirable to have a more explicit computational understanding of algebraic K–theory of $\mathrm{BP}\langle n\rangle$ in order to understand the étale cohomology of $\mathrm{BP}\langle n\rangle$ as suggested by Rognes [2014, Sections 5 and 6].

One of the most fundamental objects in chromatic stable homotopy theory is the Brown–Peterson spectrum BP, which is a complex oriented cohomology theory that carries the universal p–typical formal group. The coefficients of BP are the symmetric algebra over $\mathbb{Z}_{(p)}$ on generators v_i for $i \geq 1$, and we may form truncated versions of BP, denoted by $\mathrm{BP}\langle n\rangle$, by coning off the regular sequence $(v_{n+1}, v_{n+2}, \dots)$. More generally, we consider forms of $\mathrm{BP}\langle n\rangle$, in the spirit of [Morava 1989], which are constructed by coning off some sequence $(v'_{n+1}, v'_{n+2}, \dots)$ of indecomposable algebra generators in BP_* where $|v'_k| = |v_k|$ (see Definition 2.1 for a precise definition). We will be most interested in working with forms of $\mathrm{BP}\langle n\rangle$ that are E_m–ring spectra for sufficiently large m. We will refer to such spectra as E_m forms of $\mathrm{BP}\langle n\rangle$. For example, the spectrum $H\mathbb{Z}_{(p)}$ is an E_∞ form of $\mathrm{BP}\langle 0\rangle$, and ℓ is an E_∞ form of $\mathrm{BP}\langle 1\rangle$ at all primes by [Baker and Richter 2008].

In the last decade E_∞ forms of $\mathrm{BP}\langle 2\rangle$ were constructed at the prime $p=2$ by [Lawson and Naumann 2012] and $p=3$ by [Hill and Lawson 2010]. Lawson and Naumann [2012] used the moduli stack of formal groups with a $\Gamma_1(3)$–structure to construct an E_∞ form of $\mathrm{BP}\langle 2\rangle$ at the prime 2 denoted by $\mathrm{tmf}_1(3)$. Hill and Lawson [2010] used a quaternion algebra D of discriminant 14 and its associated Shimura curve $\mathscr{X}^D$ to construct an E_∞ form of $\mathrm{BP}\langle 2\rangle$ at the prime $p=3$, denoted by taf^D. Even more recently, Hahn and Wilson [2022] constructed an E_3 form of $\mathrm{BP}\langle n\rangle$ at all primes and for all n, which we denote by $\mathrm{BP}\langle n\rangle'$. This is especially interesting since no $E_{2(p^2+2)}$ form of $\mathrm{BP}\langle n\rangle$ exists for $n\geq 4$ by Lawson [2018] at the prime $p=2$ and Senger [2017] at primes $p>2$. Highly structured models for truncated Brown–Peterson spectra make computations of invariants of these truncated Brown–Peterson spectra more tractable, and therefore they will be important for our calculations.

For small values of n, the calculations of $\mathrm{THH}_*(\mathrm{BP}\langle n\rangle)$ are known and of fundamental importance. The first known computations of topological Hochschild homology are those of Bökstedt [1985] for $\mathrm{THH}_*(H\mathbb{F}_p)$ and $\mathrm{THH}_*(H\mathbb{Z}_{(p)})$. To illustrate how fundamental these computations are, we point out that the computation $\mathrm{THH}_*(H\mathbb{F}_p)\cong P(\mu_0)$ where $|\mu_0|=2$ is the linchpin for a new proof of Bott periodicity [Hesselholt and Nikolaus 2020]. McClure and Staffeldt [1993] computed the Bockstein spectral sequence

$$\mathrm{THH}_*(\ell;H\mathbb{F}_p)[v_1]\Rightarrow \mathrm{THH}_*(\ell;k(1)),$$

which was extended by Angeltveit, Hill, and Lawson [Angeltveit et al. 2010] to compute the square of spectral sequences

$$\begin{array}{ccc}\mathrm{THH}_*(\mathrm{BP}\langle 1\rangle;H\mathbb{F}_p)[v_0,v_1] & \Longrightarrow & \mathrm{THH}_*(\mathrm{BP}\langle 1\rangle;H\mathbb{Z}_{(p)})_p[v_1]\\ \Downarrow & & \Downarrow\\ \mathrm{THH}_*(\mathrm{BP}\langle 1\rangle;k(1))[v_0] & \Longrightarrow & \mathrm{THH}_*(\mathrm{BP}\langle 1\rangle;\mathrm{BP}\langle 1\rangle)_p\end{array}$$

This gives a complete computation of $\mathrm{THH}_*(\mathrm{BP}\langle 1\rangle)$.

Let $B\langle n\rangle$ denote an E_3 form of $\mathrm{BP}\langle n\rangle$ (see Definition 2.1).[1] In Proposition 2.7 we compute

$$\mathrm{THH}_*(B\langle n\rangle;H\mathbb{F}_p)\cong E(\lambda_1,\dots,\lambda_{n+1})\otimes P(\mu_{n+1}),$$

where $|\lambda_i|=2p^i-1$ and $|\mu_{n+1}|=2p^{n+1}$, as a consequence of work of [Angeltveit and Rognes 2005]. Hahn and Wilson [2018] calculated the groups $\mathrm{THH}_*(B\langle n\rangle/\mathrm{MU})$, but working over MU significantly simplifies the calculation. Ausoni and Richter [2020] computed $\mathrm{THH}_*(E(2))$ under the assumption that $E(2)=\mathrm{BP}\langle 2\rangle[v_2^{-1}]$ has an E_∞–ring structure and gave a conjectural answer for $\mathrm{THH}_*(E(n))$, which is consistent with our calculations. These are currently the only known results for $n\geq 2$.

[1]Note that there is a spectrum commonly denoted by $B(n)=v_n^{-1}P(n)$ in other references (eg [Ravenel 1986]) and our notation and meaning is distinct.

The main three results of this paper are computations of the Bockstein spectral sequences

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p)[v_0] \Rightarrow \mathrm{THH}_*(B\langle n\rangle; H\mathbb{Z}_{(p)})_p, \tag{1-1}$$

$$\mathrm{THH}_*(B\langle 2\rangle; H\mathbb{F}_p)[v_1] \Rightarrow \mathrm{THH}_*(B\langle 2\rangle; k(1)), \tag{1-2}$$

$$\mathrm{THH}_*(B\langle 2\rangle; H\mathbb{F}_p)[v_2] \Rightarrow \mathrm{THH}_*(B\langle 2\rangle; k(2)), \tag{1-3}$$

where $B\langle n\rangle$ is an E_3 form of $BP\langle n\rangle$ and we assume $p \geq 3$ for our computation of the spectral sequence (1-2). The Bockstein spectral sequences (1-2) and (1-3) are of similar computational complexity to the main result of McClure and Staffeldt [1993] and we were inspired by their work.

We summarize our three main results as follows: First, we compute the topological Hochschild homology of an E_3 form of $\mathrm{BP}\langle n\rangle$ with $H\mathbb{Z}_{(p)}$ coefficients.

Theorem A (Theorem 3.8) *Let $B\langle n\rangle$ be an E_3 form of $\mathrm{BP}\langle n\rangle$ and at $p > 2$ assume the error term* (3-7) *vanishes. Then there is an isomorphism of graded $\mathbb{Z}_{(p)}$–modules*

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{Z}_{(p)}) \cong E_{\mathbb{Z}_{(p)}}(\lambda_1, \dots, \lambda_n) \otimes (\mathbb{Z}_{(p)} \oplus T_0^n)$$

where T_0^n is an explicit torsion $\mathbb{Z}_{(p)}$–module defined in (3-11).

In particular, the error term (3-7) vanishes for any E_4 form of $\mathrm{BP}\langle n\rangle$ such as $B\langle 2\rangle = \mathrm{taf}^D$. It is possible that the error term (3-7) also vanishes for $B\langle n\rangle = \mathrm{BP}\langle n\rangle'$ where $\mathrm{BP}\langle n\rangle'$ is the E_3 form of $\mathrm{BP}\langle n\rangle$ constructed by Hahn and Wilson [2022] at odd primes, but it is not known to the authors. Theorem 3.8 also holds for $B\langle 2\rangle = \mathrm{tmf}_1(3)$ and $B\langle n\rangle = \mathrm{BP}\langle n\rangle'$, where $\mathrm{BP}\langle n\rangle'$ is the E_3 form of $\mathrm{BP}\langle n\rangle$ at the prime 2 constructed by Hahn and Wilson [2022].

Second, we compute the topological Hochschild homology of an E_3 form $B\langle 2\rangle$ of $\mathrm{BP}\langle 2\rangle$ at $p \geq 3$ with $k(1)$ coefficients.

Theorem B (Theorem 4.6) *Let $B\langle 2\rangle$ denote an E_3 form of $\mathrm{BP}\langle 2\rangle$ at an odd prime p. There is an isomorphism of $P(v_1)$–modules*

$$\mathrm{THH}_*(B\langle 2\rangle; k(1)) \cong E(\lambda_1) \otimes (P(v_1) \oplus T_1^2)$$

where T_1^2 is an explicit v_1–torsion $P(v_1)$–module defined in (4-3).

In particular, this result holds for $B\langle 2\rangle = \mathrm{taf}^D$ and $\mathrm{BP}\langle 2\rangle'$ at odd primes.

Finally, we compute topological Hochschild homology of any E_3 form of $\mathrm{BP}\langle 2\rangle$ with $k(2)$ coefficients.

Theorem C (Theorem 5.5) *Let $B\langle 2\rangle$ be an E_3 form of $\mathrm{BP}\langle 2\rangle$. There is an isomorphism of $P(v_2)$–modules*

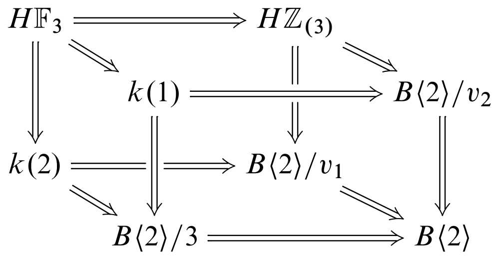
$$\mathrm{THH}_*(B\langle 2\rangle; k(2)) \cong P(v_2) \oplus T_2^2,$$

where T_2^2 is an explicit v_2–torsion $P(v_2)$–module defined in (5-2).

In particular, this result holds for $B\langle 2\rangle = \mathrm{taf}^D$, $B\langle 2\rangle = \mathrm{tmf}_1(3)$, and $BP\langle 2\rangle'$ at any prime. We end with a conjectural answer (see Conjecture 5.6) for $\mathrm{THH}_*(B\langle n\rangle; k(m))$ for all integers $1 \le m \le n$ and any E_3 form of $B\langle n\rangle$ at a prime p.

We now outline our approach to computing $\mathrm{THH}_*(\mathrm{taf}^D)$ in the sequels to this paper. There is a cube of Bockstein spectral sequences

(1-4)



where we use the abbreviation $M/x \Rightarrow M$ for the Bockstein spectral sequence with signature

$$\mathrm{THH}_*(\mathrm{taf}^D; M/x)[x] \Rightarrow \mathrm{THH}_*(\mathrm{taf}^D; M),$$

where $M \in \{H\mathbb{Z}_{(3)}, k(1), k(2), \mathrm{taf}^D/3, \mathrm{taf}^D/v_1, \mathrm{taf}^D/v_2, \mathrm{taf}^D\}$. Here we write taf^D/x for the cofiber of a representative of an element $x \in \pi_{2k}\mathrm{taf}^D$ regarded as a taf^D–module map $\Sigma^{2k}\mathrm{taf}^D \to \mathrm{taf}^D$. In the sequels to this paper, we plan to compute $\mathrm{THH}_*(\mathrm{taf}^D; M)$ for $M = \mathrm{taf}^D/3$ and $M = \mathrm{taf}^D/v_1$ by comparing the edges of the cube of Bockstein spectral sequences to the Hochschild–May spectral sequence [Angelini-Knoll and Salch 2018] and the Brun spectral sequence [Höning 2020], which compute the diagonals of the faces of the cube directly. Finally, we plan to compute $\mathrm{THH}_*(\mathrm{taf}^D)$ by again comparing the Hochschild–May spectral sequence to the relevant Bockstein spectral sequences in addition to cosimplicial descent techniques.

Conventions We write F_*X for $\pi_*(F \wedge X)$ for any spectra F and X. We also use the shorthand $H_*(X)$ for $(H\mathbb{F}_p)_*X$ for any spectrum X. We write $\doteq$ to mean that an equality holds up to multiplication by a unit. The dual Steenrod algebra $H_*(H\mathbb{F}_p)$ will be denoted by $\mathscr{A}_*$ with coproduct $\Delta\colon \mathscr{A}_* \to \mathscr{A}_* \otimes \mathscr{A}_*$. Given a left $\mathscr{A}_*$–comodule M, its left coaction will be denoted by $\nu\colon \mathscr{A}_* \to \mathscr{A}_* \otimes M$, where the comodule M is understood from the context. The antipode $\chi\colon \mathscr{A}_* \to \mathscr{A}_*$ will not play a role except that we will write $\bar{\xi}_i := \chi(\xi_i)$ and $\bar{\tau}_i := \chi(\tau_i)$.

When not otherwise specified, tensor products will be taken over $\mathbb{F}_p$ and $\mathrm{HH}_*(A)$ denotes the Hochschild homology of a graded $\mathbb{F}_p$–algebra relative to $\mathbb{F}_p$. We will let $P_R(x)$, $E_R(x)$ and $\Gamma_R(x)$ denote a polynomial algebra, exterior algebra, and divided power algebra over R on a generator x. When $R = \mathbb{F}_p$, we omit it from the notation. Let $P_i(x)$ denote the truncated polynomial algebra $P(x)/(x^i)$.

Acknowledgements The authors would like to thank Tyler Lawson for helpful conversations. The authors would also like to thank John Rognes and the referee for their careful reading of the paper and useful suggestions for improvement. Parts of this paper were written while Culver was in residence at the

Max Planck Institute for Mathematics. He would like to thank the institute for their support. Höning was funded by the DFG priority program SPP 1786 *Homotopy theory and algebraic geometry* and the Radboud Excellence Initiative. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 1010342555.

2 Topological Hochschild homology mod $(p, \dots, v_n)$

We begin by giving a precise definition of an E_m form $B\langle n\rangle$ of $\mathrm{BP}\langle n\rangle$. We then compute topological Hochschild homology of an E_3 form $B\langle n\rangle$ of $\mathrm{BP}\langle n\rangle$ at an arbitrary prime p with coefficients in $H\mathbb{F}_p$. First, recall that there is an isomorphism $\mathrm{BP}_* \cong \mathbb{Z}_{(p)}[v_i \mid i \geq 1]$ and an isomorphism

$$\mathrm{BP}_*\mathrm{BP} \cong \mathbb{Z}_{(p)}[v_i \mid i \geq 1][t_i \mid i \geq 1]$$

where the degrees of the generators are $|v_i| = |t_i| = 2p^i - 2$ for $i \geq 1$. The generators t_i are determined by the canonical strict isomorphism f from the universal p–typical formal group law to itself given by the power series

$$f^{-1}(x) = \sum_{i\geq 0}^{F} t_i x^{p^i}$$

where F is the universal p–typical formal group law [Ravenel 1986, Lemma A2.1.26]. We let v_i be the Araki generators. Note that the Araki generators agree with Hazewinkel generators mod p [Ravenel 1986, Theorem A2.2.3].

2.1 Forms of $\mathrm{BP}\langle n\rangle$

We fix a precise notion of a form of the truncated Brown–Peterson spectrum in the spirit of [Morava 1989] below.

Definition 2.1 (cf [Lawson and Naumann 2014, Definition 4.1]) Fix integers $m \geq 1$ and $n \geq 0$. By an E_m form of $\mathrm{BP}\langle n\rangle$ (at the prime p), we mean a p–local E_m–ring spectrum R equipped with a complex orientation $\mathrm{MU}_{(p)} \to R$ such that the composite

$$\mathbb{Z}_{(p)}[v_1, \dots, v_n] \to \mathrm{BP}_* \to \pi_*\mathrm{MU}_{(p)} \to \pi_* R$$

is an isomorphism.

Remark 2.2 Note that we do not assume that an E_m form of $\mathrm{BP}\langle n\rangle$ at the prime p is an E_m MU–algebra, and therefore Definition 2.1 differs slightly from the definition of an E_m MU–algebra form of $\mathrm{BP}\langle n\rangle$ appearing in work of Hahn and Wilson [2022, Definition 2.0.1]. An E_m MU–algebra form of $\mathrm{BP}\langle n\rangle$ in the sense of [Hahn and Wilson 2022, Definition 2.0.1] is an E_m form of $\mathrm{BP}\langle n\rangle$ at the prime p in the sense of Definition 2.1. The distinction arises because, for example, taf^D is an E_∞ form of $BP\langle 2\rangle$; however it is not known, at least to the authors, whether the complex orientation $\mathrm{MU} \to \mathrm{taf}^D$ can be elevated to an E_∞–ring spectrum map. Nonetheless, we know that the map $\mathrm{MU} \to \mathrm{taf}^D$ is an E_2–ring spectrum map by [Chadwick and Mandell 2015, Theorem 1.2], which is sufficient for our purposes.

Notation 2.3 Throughout, we let $B\langle n\rangle$ denote an E_3 form of $\mathrm{BP}\langle n\rangle$ at the prime p in the sense of Definition 2.1 for $n \geq 0$.

We collect some consequences of Definition 2.1.

Proposition 2.4 *Since $B\langle n\rangle$ is an E_3 form of $\mathrm{BP}\langle n\rangle$ at the prime p for $m \geq 3$, the following hold*:

(1) *There are indecomposable algebra generators v_i' with $v_i' = v_i$ for $1 \leq i \leq n$ such that*

$$\mathrm{BP}_*/(v_k' \mid k \geq n+1) \cong \pi_* B\langle n\rangle.$$

(2) *The orientation $\mathrm{MU}_{(p)} \to B\langle n\rangle$ lifts to an E_2–ring spectrum map and consequently there is an E_2–ring spectrum map $\mathrm{BP} \to B\langle n\rangle$ realizing the canonical quotient map $\mathrm{BP}_* \to \mathrm{BP}_*/(v_k' \mid k \geq n+1)$ on homotopy groups.*

(3) *There is an E_3–ring spectrum map $B\langle n\rangle \to H\mathbb{Z}_{(p)}$ and the map induced by the composite*

$$B\langle n\rangle \to H\mathbb{Z}_{(p)} \to H\mathbb{F}_p \tag{2-1}$$

in mod p homology provides an isomorphism

$$H_*(B\langle n\rangle) \cong \mathscr{A}//E(n)_* \subset \mathscr{A}_*$$

of $\mathscr{A}_*$–comodule $\mathbb{F}_p$–algebras onto its image in the dual Steenrod algebra.

(4) *If $B\langle n\rangle$ is E_3 and $x_1, \dots, x_n$ is a regular sequence of elements in $B\langle n\rangle_*$, then one can construct the spectrum $B\langle n\rangle/(x_1, x_2, \dots, x_n)$ as an E_1 $B\langle n\rangle$–algebra.*

(5) *The p–completion of $B\langle n\rangle$ is weakly equivalent to the p–completion of any other E_m form of $\mathrm{BP}\langle n\rangle$ at the prime p in the category of spectra.*

Proof For part (1) set $v_i' := v_i - f_i(v_1, \dots, v_n)$ for $i \geq n+1$, where $f_i(v_1, \dots, v_n)$ is the image of v_i under $\mathrm{BP}_* \to \mathrm{BP}\langle n\rangle_* \cong \mathbb{Z}_{(p)}[v_1, \dots, v_n]$. Part (2) follows by applying [Chadwick and Mandell 2015, Theorem 1.2]. Part (3) is [Lawson and Naumann 2014, Theorem 4.4]. Part (4) follows from [Angeltveit 2008, Section 3] (cf [Hahn and Wilson 2018, Theorem A]). Part (5) is [Angeltveit and Lind 2017, Theorem A]. □

Example 2.5 The Eilenberg–Mac Lane spectrum $H\mathbb{Z}_{(p)}$ is an E_∞ form of $\mathrm{BP}\langle 0\rangle$. The Adams summand ℓ is an E_∞ form of $\mathrm{BP}\langle 1\rangle$ by [Baker and Richter 2008, Corollary 1.4].

Notation 2.6 Let $\mathrm{tmf}_1(3)$ denote the E_∞ form of $\mathrm{BP}\langle 2\rangle$ constructed by Lawson and Naumann [2012] at $p = 2$. Let taf^D denote the E_∞ form of $\mathrm{BP}\langle 2\rangle$ constructed by Hill and Lawson [2010] at $p = 3$. Let $\mathrm{BP}\langle n\rangle'$ denote the E_3 form of $\mathrm{BP}\langle n\rangle$ constructed by Hahn and Wilson [2022] at all primes.

2.2 Topological Hochschild homology mod $(p, \dots, v_n)$

The mod p homology of $\mathrm{THH}(\mathrm{BP}\langle n\rangle)$ has been calculated by Angeltveit and Rognes [2005, Theorem 5.12] assuming that $\mathrm{BP}\langle n\rangle$ is an E_3–ring spectrum. Their argument also applies to topological Hochschild homology of any E_3 form $B\langle n\rangle$ of $\mathrm{BP}\langle n\rangle$ at a prime p, as we now explain. By Proposition 2.4, the linearization map (2-1) induces an isomorphism

$$H_*(B\langle n\rangle) \cong \begin{cases} P(\bar\xi_1, \bar\xi_2, \dots) \otimes E(\bar\tau_{n+1}, \bar\tau_{n+2}, \dots) & \text{if } p \geq 3, \\ P(\bar\xi_1^2, \dots, \bar\xi_{n+1}^2, \bar\xi_{n+2}, \dots) & \text{if } p = 2, \end{cases}$$

with its image in $\mathscr{A}_*$ as an $\mathscr{A}_*$–subcomodule algebra of $\mathscr{A}_*$. By [Brun et al. 2007, Theorem 3.4], the spectrum $\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p)$ is an E_2–ring spectrum and the unit map

$$H\mathbb{F}_p \to \mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p)$$

is a map of E_2–ring spectra. Using [Brun et al. 2007, Section 3.3], the proof of [Angeltveit and Rognes 2005, Proposition 4.3] carries over mutatis mutandis and implies that the Bökstedt spectral sequence with signature

$$E^2_{*,*} = \mathrm{HH}_{*,*}(H_*(B\langle n\rangle); \mathscr{A}_*) \Rightarrow H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$$

is a spectral sequence of $\mathscr{A}_*$–comodule algebras. As in [Angeltveit and Rognes 2005, Section 5.2], the spectral sequence collapses at the E^2–page if $p = 2$. If $p \geq 3$, one can use the map to the Bökstedt spectral sequence with signature

$$E^2_{*,*} = \mathrm{HH}_{*,*}(\mathscr{A}_*) \Rightarrow H_*(\mathrm{THH}(H\mathbb{F}_p))$$

to determine the differentials (cf [Angeltveit and Rognes 2005, Section 5.4]). Since $B\langle n\rangle$ is an E_3–ring spectrum, Dyer–Lashof operations are defined on $H_*(B\langle n\rangle)$ and $H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$ in a range that is sufficient to resolve the multiplicative extensions (see [Angeltveit and Rognes 2005, Proof of Theorem 5.12]). We get an isomorphism of $\mathscr{A}_*$–comodule $\mathscr{A}_*$–algebras

$$H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p)) \cong \begin{cases} \mathscr{A}_* \otimes E(\sigma\bar\xi_1, \dots, \sigma\bar\xi_{n+1}) \otimes P(\sigma\bar\tau_{n+1}) & \text{if } p \geq 3, \\ \mathscr{A}_* \otimes E(\sigma\bar\xi_1^2, \dots, \sigma\bar\xi_{n+1}^2) \otimes P(\sigma\bar\xi_{n+2}) & \text{if } p = 2. \end{cases} \tag{2-2}$$

Since $\sigma\colon H_*(B\langle n\rangle) \to H_{*+1}(\mathrm{THH}(B\langle n\rangle)) \to H_{*+1}(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$ is a comodule map and a derivation, the $\mathscr{A}_*$–coaction of

$$H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$$

can be deduced from that of $H_*(B\langle n\rangle) \subseteq \mathscr{A}_*$ (cf [Angeltveit and Rognes 2005, Proof of Theorem 5.12]): for $p \geq 3$ the classes $\sigma\bar\xi_i$ for $1 \leq i \leq n+1$ are $\mathscr{A}_*$–comodule primitives and we have

$$\nu(\sigma\bar\tau_{n+1}) = 1 \otimes \sigma\bar\tau_{n+1} + \bar\tau_0 \otimes \sigma\bar\xi_{n+1}. \tag{2-3}$$

For $p = 2$ the classes $\sigma\bar\xi_i^2$ for $1 \leq i \leq n+1$ are $\mathscr{A}_*$–comodule primitives and we have

$$\nu(\sigma\bar\xi_{n+2}) = 1 \otimes \sigma\bar\xi_{n+2} + \bar\xi_1 \otimes \sigma\bar\xi_{n+1}^2. \tag{2-4}$$

Proposition 2.7 *Let $B\langle n\rangle$ be an E_3 form of $\mathrm{BP}\langle n\rangle$. There is an isomorphism of graded $\mathbb{F}_p$–algebras*

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p) \cong E(\lambda_1, \dots, \lambda_{n+1}) \otimes P(\mu_{n+1}), \tag{2-5}$$

where the degrees of the algebra generators are $|\lambda_i| = 2p^i - 1$ for $1 \le i \le n+1$ and $|\mu_{n+1}| = 2p^{n+1}$.

Proof Since $\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p)$ is an $H\mathbb{F}_p$–module, the Hurewicz homomorphism induces an isomorphism between $\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p)$ and the subalgebra of comodule primitives in $H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$. For $1 \le i \le n+1$ we write $\lambda_i := \sigma\bar{\xi}_i$ if $p \ge 3$ and $\lambda_i := \sigma\bar{\xi}_i^2$ if $p = 2$. We also define

$$\mu_{n+1} := \begin{cases} \sigma\bar{\tau}_{n+1} - \bar{\tau}_0\sigma\bar{\xi}_{n+1} & \text{if } p \ge 3, \\ \sigma\bar{\xi}_{n+2} - \bar{\xi}_1\sigma\bar{\xi}_{n+1}^2 & \text{if } p = 2. \end{cases}$$

Then it is clear that the subalgebra of $H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$ consisting of comodule primitives is as claimed. □

3 Topological Hochschild homology mod $(v_1, \dots, v_n)$

We begin by setting up the Bockstein spectral sequence. In order to ensure that this spectral sequence is multiplicative, we compare it with the Adams spectral sequence.

3.1 Bockstein and Adams spectral sequences

Let $B\langle n\rangle$ be an E_3 form of $\mathrm{BP}\langle n\rangle$ at the prime p which is equipped with a choice of generators v_i in degrees $|v_i| = 2p^i - 2$ for $0 < i \le n$ such that $B\langle n\rangle_* = \mathbb{Z}_{(p)}[v_1, \dots, v_n]$. Let $v_0 = p$ by convention. Let

$$k(i) = B\langle n\rangle/(p, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$$

be the E_1 $B\langle n\rangle$–algebra constructed in Proposition 2.4 (4) where $k(0) = H\mathbb{Z}_{(p)}$. We regard $k(i)$ as a right $B\langle n\rangle \wedge B\langle n\rangle^{\mathrm{op}}$–module by restriction along the map

$$B\langle n\rangle \wedge B\langle n\rangle^{\mathrm{op}} \to B\langle n\rangle \to k(i).$$

For $0 \le i \le n$ we have cofiber sequences of right $B\langle n\rangle \wedge B\langle n\rangle^{\mathrm{op}}$–modules

$$\Sigma^{|v_i|} k(i) \xrightarrow{\cdot v_i} k(i) \to H\mathbb{F}_p.$$

Applying the functor $- \wedge_{B\langle n\rangle \wedge B\langle n\rangle^{\mathrm{op}}} B\langle n\rangle$ produces the cofiber sequence

$$\Sigma^{|v_i|}\,\mathrm{THH}(B\langle n\rangle; k(i)) \to \mathrm{THH}(B\langle n\rangle; k(i)) \to \mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p).$$

Iterating this, we produce the tower

$$\begin{array}{ccccccc} \cdots \longrightarrow & \Sigma^{2|v_i|} T(k(i)) & \xrightarrow{\cdot v_i} & \Sigma^{|v_i|} T(k(i)) & \xrightarrow{\cdot v_i} & T(k(i)) \\ & \downarrow & & \downarrow & & \downarrow \\ & \Sigma^{2|v_i|} T(H\mathbb{F}_p) & & \Sigma^{|v_i|} T(H\mathbb{F}_p) & & T(H\mathbb{F}_p) \end{array} \tag{3-1}$$

where $T(k(i)) := \mathrm{THH}(B\langle n\rangle; k(i))$ and $T(H\mathbb{F}_p) := \mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p)$.

This yields an exact couple after applying homotopy groups and it produces the v_i–Bockstein spectral sequence with E_1–page

$$E_1^{*,*} = \mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p)[v_i]. \tag{3-2}$$

Note that the fact that $B\langle n\rangle$ and $k(i)$ are connective and have homotopy groups that are degreewise finitely generated $\mathbb{Z}_{(p)}$–modules implies that the homotopy groups of $\mathrm{THH}(B\langle n\rangle; k(i))$ are degreewise finitely generated $\mathbb{Z}_{(p)}$–modules, too. It follows that $\mathrm{THH}_*(B\langle n\rangle; k(i))$ has the form

$$\mathrm{THH}_*(B\langle n\rangle; k(i)) \cong \bigoplus_l P(v_i)\{\alpha_l\} \oplus \bigoplus_k P_{r_k}(v_i)\{\beta_k\}$$

for some classes α_l and β_k. Here, for $i = 0$, $P(v_i)$ is defined to be $\mathbb{Z}_{(p)}$ and $P_r(v_i)$ is $\mathbb{Z}/p^i$. We get that

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p) \cong \bigoplus_l \mathbb{F}_p\{a_l\} \oplus \bigoplus_k \mathbb{F}_p\{b_k\} \oplus \bigoplus_k \mathbb{F}_p\{c_k\},$$

where a_l and b_k are the images of α_l and β_k under the map $\mathrm{THH}_*(B\langle n\rangle; k(i)) \to \mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p)$, and c_k is a preimage of $v_i^{r_k-1}\beta_k$ under the map $\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p) \to \Sigma^{|v_i|+1}\,\mathrm{THH}_*(B\langle 2\rangle; k(i))$. The differentials in the spectral sequence are given as follows: The classes a_l and b_k are infinite cycles. The class c_k survives to the E_{r_k}–page and we have

$$d_{r_k}(c_k) = v_i^{r_k} b_k.$$

The spectral sequence converges strongly to $\mathrm{THH}_*(B\langle n\rangle; k(i))$ for $0 < i \le n$ and $\pi_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})_p)$ for $i = 0$. The cofibers in the tower (3-1) are $H\mathbb{F}_p$–module spectra.

We now relate the Bockstein spectral sequence to the Adams spectral sequence. In order to do this, we show that the tower (3-1) is also an Adams resolution. For the definition of an Adams resolution, the reader is referred to [Ravenel 1986, Definition 2.1.3]. In order to show that this tower is an Adams resolution, it must be shown that the vertical morphisms

$$\Sigma^{m|v_i|}\,\mathrm{THH}(B\langle n\rangle; k(i)) \to \Sigma^{m|v_i|}\,\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p) \tag{3-3}$$

induce monomorphisms in mod p homology. We have equivalences of spectra

$$\mathrm{THH}(B\langle n\rangle; M) \simeq M \wedge_{B\langle n\rangle} \mathrm{THH}(B\langle n\rangle)$$

for $M \in \{H\mathbb{F}_p, k(i) \mid 0 \le i \le n\}$ by [Hahn and Wilson 2022, Remark 6.1.4] and consequently there is an Eilenberg–Moore spectral sequence

$$\mathrm{Tor}_{*,*}^{H_*B\langle n\rangle}(H_*(M), H_*(\mathrm{THH}(B\langle n\rangle))) \Rightarrow H_*(\mathrm{THH}(B\langle n\rangle; M))$$

for each $M \in \{H\mathbb{F}_p, k(i) \mid 0 \le i \le n\}$. Since $H_*(\mathrm{THH}(B\langle n\rangle))$ is a free $H_*(B\langle n\rangle)$–module by [Angeltveit and Rognes 2005, Theorem 5.12], the Eilenberg–Moore spectral sequence collapses at the E^2–page without room for differentials. Furthermore, the morphism (3-3) induces a morphism of Eilenberg–Moore spectral sequences. Thus, we observe that the morphism (3-3) induces the map

$$H_*(k(i)) \otimes_{H_*(B\langle n\rangle)} H_*(\mathrm{THH}(B\langle n\rangle)) \to \mathscr{A}_* \otimes_{H_*(B\langle n\rangle)} H_*(\mathrm{THH}(B\langle n\rangle)) \tag{3-4}$$

in mod p homology where the map on the first factor is induced by the linearization map $k(i) \to H\mathbb{F}_p$. The map (3-4) is an injection. Since $H_*(\mathrm{THH}(B\langle n\rangle))$ is a free $H_*(B\langle n\rangle)$–module, the map (3-3) induces an injection on mod p homology. Thus, we have shown the following proposition.

Proposition 3.1 *The tower* (3-1) *is an Adams resolution.*

Thus, the Adams spectral sequence for $\mathrm{THH}(B\langle n\rangle; k(i))$ agrees with the Bockstein spectral sequence for $0 \le i \le n$. By [Ravenel 1986, Theorem 2.3.3], we know that the Adams spectral sequence for $\mathrm{THH}(B\langle n\rangle; k(i))$, and consequently the Bockstein spectral sequence, is multiplicative for $0 \le i \le n$ from the E_2–page onwards. To see that the Adams spectral sequence is in fact multiplicative from the E_1–page onwards, we prove explicitly in the case $i = 0$ that the d_1 differential satisfies the Leibniz rule in Lemma 3.4. In the case $i > 0$, we can apply a change of rings isomorphism and compute explicitly that the E_2–page is

$$\mathrm{Ext}^{*,*}_{E(Q_i)_*}(\mathbb{F}_p, E(\lambda_1, \dots, \lambda_{n+1}) \otimes P(\mu_{n+1})) = P(v_i) \otimes E(\lambda_1, \dots, \lambda_{n+1}) \otimes P(\mu_{n+1})$$

using the coactions discussed previously on λ_i and μ_{n+1}. Consequently, when $i > 0$ there are no nontrivial d_1 differentials. Altogether, this proves the following corollary.

Corollary 3.2 *The v_i–Bockstein spectral sequence computing $\mathrm{THH}_*(B\langle n\rangle; k(i))$ in the case $i \ge 1$ and $\pi_* \mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})_p$ in the case $i = 0$ is multiplicative from the E_1–page onwards.*

3.2 Rational topological Hochschild homology

We use the $H\mathbb{Q}$–based Bökstedt spectral sequence to compute

$$\pi_*(L_0 \,\mathrm{THH}(B\langle n\rangle)) = H\mathbb{Q}_* \mathrm{THH}(B\langle n\rangle) = \pi_* \mathrm{THH}(B\langle n\rangle) \otimes \mathbb{Q}$$

for $0 \le n \le \infty$ where $B\langle \infty\rangle = \mathrm{BP}$ and $L_0 = L_{H\mathbb{Q}}$ is the Bousfield localization at $H\mathbb{Q}$. Since BP and $B\langle n\rangle$ are E_3–ring spectra, the $H\mathbb{Q}$–based Bökstedt spectral sequences are strongly convergent multiplicative spectral sequence with signature

$$E^2_{**} = \mathrm{HH}^{\mathbb{Q}}_{*,*}(H\mathbb{Q}_* B\langle n\rangle) \Rightarrow H\mathbb{Q}_* \mathrm{THH}(B\langle n\rangle)$$

for $0 \le n \le \infty$. Recall that the rational homology of $B\langle n\rangle$ is

$$H\mathbb{Q}_* B\langle n\rangle \cong P_{\mathbb{Q}}(v_1, \dots, v_n)$$

with $|v_i| = 2p^i - 2$ for $1 \le i \le n \le \infty$. Thus, the E^2–term of the Bökstedt spectral sequence is

$$E^2_{*,*} = P_{\mathbb{Q}}(v_1, \dots, v_n) \otimes_{\mathbb{Q}} E_{\mathbb{Q}}(\sigma v_1, \dots, \sigma v_n)$$

where the bidegree of σv_i is $(1, 2(p^i-1))$ for $1 \le i \le n \le \infty$. Since the E^2–page is generated as a $\mathbb{Q}$–algebra by classes in Bökstedt filtration degree 0 and 1, the first quadrant spectral sequence collapses at the E^2–page and $E^2_{*,*} = E^{\infty}_{*,*}$. There are no multiplicative extensions, because the E^{∞}–pages are free graded-commutative $\mathbb{Q}$–algebras. Therefore, we produce isomorphisms of graded $\mathbb{Q}$–algebras

$$\mathrm{THH}_*(B\langle n\rangle) \otimes \mathbb{Q} \cong P_{\mathbb{Q}}(v_1, \dots, v_n) \otimes_{\mathbb{Q}} E_{\mathbb{Q}}(\sigma v_1, \dots, \sigma v_n)$$

with $|\sigma v_i| = 2p^i - 1$ for $1 \le i \le n \le \infty$. It follows that there is an equivalence

$$L_0\,\mathrm{THH}(B\langle n\rangle) \simeq \bigvee_{x \in B_n} \Sigma^{|x|} L_0 B\langle n\rangle,$$

where B_n is a graded basis for $E_{\mathbb{Q}}(\sigma v_1, \dots, \sigma v_n)$ as a graded $\mathbb{Q}$–vector space, since L_0 is a smashing localization. We may also let $n = \infty$ and in this case $B\langle\infty\rangle = \mathrm{BP}$ and B_∞ is a graded basis for $E_{\mathbb{Q}}(\sigma v_1, \sigma v_2, \dots)$ as a graded $\mathbb{Q}$–vector space.

By Proposition 2.4, the linearization map $\mathrm{BP}\langle n\rangle \to H\mathbb{Z}_{(p)}$ is an E_3–ring spectrum map. Since the localization map $H\mathbb{Z}_{(p)} \to H\mathbb{Q}$ is an E_∞–ring spectrum map, we may infer that the Bökstedt spectral sequence

$$E^2_{**} = \mathrm{HH}^{\mathbb{Q}}_{*,*}(H\mathbb{Q}_* B\langle n\rangle; \mathbb{Q}) \Rightarrow H\mathbb{Q}_*\,\mathrm{THH}(B\langle n\rangle; H\mathbb{Q})$$

is a spectral sequence of $\mathbb{Q}$–algebras by using [Brun et al. 2007, Section 3.3] to adapt the proof of [Angeltveit and Rognes 2005, Proposition 4.3]. This spectral sequence collapses without extensions by the same argument as before. All of these computations are functorial with respect to the map of E_2–ring spectra $\mathrm{BP} \to B\langle n\rangle$ from Proposition 2.4. This proves the following result.

Proposition 3.3 *There is an isomorphism of graded $\mathbb{Q}$–algebras*

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{Q}) \cong E_{\mathbb{Q}}(\sigma v_1, \dots, \sigma v_n) \tag{3-5}$$

for all $0 \le n \le \infty$. The map

$$\mathrm{THH}_*(\mathrm{BP}; H\mathbb{Q}) \to \mathrm{THH}_*(B\langle n\rangle; H\mathbb{Q})$$

sends σv_i to σv_i for $0 \le i \le n$.

3.3 The v_0–Bockstein spectral sequence

In this section, we compute the v_0–Bockstein spectral sequence with signature

$$E_1^{*,*} = \mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p)[v_0] \Rightarrow \mathrm{THH}_*(B\langle n\rangle; H\mathbb{Z}_{(p)})_p \tag{3-6}$$

where $B\langle n\rangle$ is an E_3 form of $\mathrm{BP}\langle n\rangle$. At odd primes, we must assume that a certain error term (3-7) vanishes. This error term vanishes for any E_4 form of $\mathrm{BP}\langle n\rangle$ at odd primes, for example taf^D.

Lemma 3.4 *There is a differential*

$$d_1(\mu_{n+1}) \doteq v_0 \lambda_{n+1}$$

in the v_0–Bockstein spectral sequence (3-6) *and the d_1 differential satisfies the Leibniz rule.*

Proof We just give the argument for $p \geq 3$ to simplify the discussion since the argument for $p = 2$ is the same up to a change of symbols. Recall that the classes μ_{n+1} and λ_{n+1} in $\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p)$ correspond to the comodule primitives $\sigma\bar{\tau}_{n+1} - \bar{\tau}_0\sigma\bar{\xi}_{n+1}$ and $\sigma\bar{\xi}_{n+1}$ in $H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$. We therefore have to show that $\sigma\bar{\tau}_{n+1} - \bar{\tau}_0\sigma\bar{\xi}_{n+1}$ maps to $\sigma\bar{\xi}_{n+1}$ under the map β_1 that is given by applying $H_*(-)$ to

$$\Sigma^{-1}\,\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p) \to \mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)}) \to \mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p).$$

As above, one sees that

$$H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})) \cong H_*(H\mathbb{Z}_{(p)}) \otimes E(\sigma\bar{\xi}_1, \dots, \sigma\bar{\xi}_{n+1}) \otimes P(\sigma\bar{\tau}_{n+1}).$$

The map

$$H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})) \to H_*(\mathrm{THH}(B\langle n\rangle; H\mathbb{F}_p))$$

is induced by the inclusion $H_*(H\mathbb{Z}_{(p)}) \to H_*(H\mathbb{F}_p)$. Since the elements $\sigma\bar{\xi}_{n+1}$ and $\sigma\bar{\tau}_{n+1}$ are in the image of this map, they map to zero under β_1. Since $\bar{\tau}_0$ is not in the image, it maps to 1 under β_1 (up to a unit). Since β_1 is a derivation, we get $\beta_1(\sigma\bar{\tau}_{n+1} - \bar{\tau}_0\sigma\bar{\xi}_{n+1}) \doteq \sigma\bar{\xi}_{n+1}$.[2] Finally, we observe that the d_1–differential satisfies the Leibniz rule because the Hurewicz map is a ring map and the Bockstein operator β_1 is a derivation. □

To compute the differentials d_r for $r > 1$ we use [May 1970, Proposition 6.8].

Lemma 3.5 [May 1970, Proposition 6.8] *If $d_{r-1}(x) \neq 0$ in the v_0–Bockstein spectral sequence* (3-6) *and $|x| = 2q$, then*

$$d_r(x^p) \doteq v_0 x^{p-1} d_{r-1}(x)$$

if $r > 2$. If $r = 2$ and $p = 2$, then

$$d_r(x^p) \doteq v_0 x^{p-1} d_{r-1}(x) + Q^{|x|}(d_1(x)).$$

If $r = 2$ and $p > 2$, then

$$d_r(x^p) \doteq v_0 x^{p-1} d_{r-1}(x) + \mathsf{E},$$

where

$$\mathsf{E} = \sum_{j=1}^{(p-1)/2} j[d_1(x)x^{j-1}, d_1(x)x^{p-j-1}]_1 \tag{3-7}$$

and $[-,-]_1$ denotes the Browder bracket.

Remark 3.6 The result above also appears in [Bruner 1977] in the context of the Adams spectral sequence for an H_∞–ring spectrum (cf [Bruner et al. 1986, Chapter VI Theorems 1.1 and 1.2]).

We note that in order to apply [May 1970, Proposition 6.8], we need the $\cup_1$–product on $\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})$ to satisfy the Hirsch formula, which states that $-\cup_1 c$ is a derivation. We observe that the $\cup_1$–product is

[2] Note that the Bockstein operator β_1 is defined for any $H\mathbb{Z}$–algebra R and it is a derivation at this level of generality by [Browder 1961; Shipley 2007].

a chain homotopy from $x \cdot y$ to $y \operatorname{cof} x$, which corresponds to a braiding in a braided monoidal category. From this perspective, the Hirsch formula corresponds to the first Hexagon axiom in the definition of a braided monoidal category [Joyal and Street 1985, Section 1, B1]. It is well documented that there is an E_2–operad in small categories with the property that algebras over this operad are braided monoidal categories [Dunn 1997]. The n^{th} category in this operad is the translation groupoid $\mathrm{Br}_n \int \Sigma_n$ of the action of the pure Artin braid group Br_n on Σ_n via the canonical inclusion $\mathrm{Br}_n \to \Sigma_n$. We consider the corresponding operad $\mathcal{B}_2$ in $H\mathbb{Z}$–modules by applying the nerve of the category $\mathrm{Br}_n \int \Sigma_n$ and then applying the functor $H\mathbb{Z}_{(p)} \wedge -$. In other words, the n^{th} chain complex in the operad in chain complexes is $\mathcal{B}_2(n) = H\mathbb{Z}_{(p)} \wedge N(\mathrm{Br}_n \int \Sigma_n)_+$. The fact that $\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})$ satisfies the Hirsch formula now follows from two facts:

1. algebras over the operad $\mathcal{B}_2$ in chain complexes satisfy the Hirsch formula (cf [Dunn 1997, Theorem 1.6]), and
2. using [May 1972, Construction 9.6], we replace the E_2 $H\mathbb{Z}_{(p)}$–algebra $\mathrm{THH}(B\langle n\rangle; \mathbb{Z}_{(p)})$ with an $\mathcal{B}_2$ algebra without changing the underlying spectrum.

We therefore tacitly replace our E_2–ring spectrum $\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})$ in $H\mathbb{Z}_{(p)}$–modules with an algebra over the operad $\mathcal{B}_2$ throughout the remainder of the section. The authors thank T Lawson for suggesting this argument.

We can consequently prove the following differential pattern.

Corollary 3.7 *In the spectral sequence* (3-6), *there are differentials*

$$d_{r+1}(\mu_{n+1}^{p^r}) \doteq v_0^{r+1} \mu_{n+1}^{p^r-1} \lambda_{n+1} \tag{3-8}$$

when $p = 2$ under the assumption that $B\langle n\rangle$ is an E_3 form. Consequently, there are differentials

$$d_{\nu_p(k)+1}(\mu_{n+1}^k) \doteq v_0^{\nu_p(k)+1} \mu_{n+1}^{k-1} \lambda_{n+1}$$

where $\nu_p(k)$ denotes the p–adic valuation of k. The same formulas hold for $p \geq 3$ when the error term (3-7) *vanishes, for example when $B\langle n\rangle$ is an E_4 form of* $\mathrm{BP}\langle n\rangle$.

Proof There is a differential

$$d_1(\mu_{n+1}) \doteq v_0 \lambda_{n+1}$$

by Lemma 3.4 for any prime p. We will argue that this differential implies the differentials (3-8) for $r \geq 1$ by applying Lemma 3.5 and observing that the obstructions vanish.

When $r = 1$ and $p > 2$, the formula (3-8) holds whenever the error term (3-7) vanishes by Lemma 3.5. The Browder bracket $[-,-]_1$ vanishes by [May 1970, Proposition 6.3(iii)] when $B\langle n\rangle$ is an E_4 form of $\mathrm{BP}\langle n\rangle$ since in that case $\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})$ is an E_3–ring spectrum. This completes the base step in the induction for $p > 2$.

If $p=2$ and $r=1$, Lemma 3.4 implies that the error term for $d_2(\mu_{n+1}^2)$ is $Q^{2^{n+2}}\lambda_{n+1}$. At $p=2$,

$$Q^{2^{n+2}}\lambda_{n+1} = Q^{2^{n+2}}(\sigma\bar{\xi}_{n+1}^2) = \sigma(Q^{2^{n+2}}(\bar{\xi}_{n+1}^2)) = \sigma((Q^{2^{n+1}}\bar{\xi}_{n+1})^2) = \sigma(\bar{\xi}_{n+2}^2) = 0 \tag{3-9}$$

as we now explain. First, the operation $Q^{2^{n+2}}$ is defined on λ_{n+1} because $2^{n+2} = |\lambda_{n+1}|+1$ and $B\langle n\rangle$ is an E_3 form of $\mathrm{BP}\langle n\rangle$ by assumption. The first equality in (3-9) holds by definition of λ_3, the second equality holds because σ commutes with Dyer–Lashof operations by [Angeltveit and Rognes 2005, Proposition 5.9], the third equality holds by [Bruner et al. 1986, Chapter III, Theorem 2.2], and the last equality holds because σ is a derivation in mod p homology, by [Angeltveit and Rognes 2005, Proposition 5.10]. This completes the base step in the induction at $p=2$.

Now let $\alpha = \nu_p(k)$ and let p be any prime. We have that $k = p^\alpha j$ where p does not divide j. So, by the Leibniz rule,

$$\begin{aligned} d_{\alpha+1}(\mu_{n+1}^k) = d_{\alpha+1}((\mu_{n+1}^{p^\alpha})^j) &= j\mu_{n+1}^{p^\alpha(j-1)} d_{\alpha+1}(\mu_{n+1}^{p^\alpha}) \\ &= j v_0^{\alpha+1}\mu_{n+1}^{p^\alpha(j-1)}\mu_{n+1}^{p^\alpha-1}\lambda_{n+1} = v_0^{\alpha+1}\mu_{n+1}^{k-1}\lambda_{n+1} \end{aligned}$$

since j is not divisible by p and therefore is a unit in $\mathbb{F}_p$. □

We now argue that the classes λ_i for $1\le i\le n$ are not p–torsion in $\mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)})$. Recall from Proposition 3.3 that there is an isomorphism

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{Q}) \cong E_{\mathbb{Q}}(\sigma v_1,\dots,\sigma v_n).$$

We claim that the map

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{Z}_{(p)}) \to \mathrm{THH}_*(B\langle n\rangle; H\mathbb{Q})$$

sends λ_i to $p^{-1}\sigma v_i$ $1\le i\le n$. To see this, we note that there is a map of E_2–ring spectra $\mathrm{BP}\to B\langle n\rangle$ by Proposition 2.4 and this produces a commutative diagram

$$\begin{array}{ccc} \mathrm{THH}(\mathrm{BP}) & \longrightarrow & \mathrm{THH}(\mathrm{BP}; H\mathbb{Q}) \\ \downarrow & & \downarrow \\ \mathrm{THH}(B\langle n\rangle; H\mathbb{Z}_{(p)}) & \longrightarrow & \mathrm{THH}(B\langle n\rangle; H\mathbb{Q}) \end{array}$$

of E_1–ring spectra by [Brun et al. 2007]. By Proposition 3.3, we know σv_i maps to σv_i for $1\le i\le n$ under the left vertical map. By [Rognes 2020, Theorem 1.1], we know that

$$\sigma v_i \equiv p\tilde{\lambda}_i \mod (v_i \mid i\ge 1)$$

up to a unit for some classes $\tilde{\lambda}_i = \sigma t_i$. Note that the choice of generators v_i in [Rognes 2020, Theorem 1.1] differ from ours, but they are the same up to a unit and modulo decomposables. Therefore there isn't a difference up to a unit modulo $(v_i \mid i\ge 1)$ after applying the derivation σ. There is an isomorphism

$$\mathrm{THH}_*(\mathrm{BP}) \cong E_{\mathrm{BP}_*}(\tilde{\lambda}_k \mid k\ge 1)$$

and we know that $\tilde{\lambda}_i$ maps to λ_i under the map

$$\mathrm{THH}_*(\mathrm{BP}) \to \mathrm{THH}_*(B\langle n\rangle; \mathbb{Z}_{(p)})$$

for $1 \le i \le n$ by Zahler [1971] and this does not depend on our choice of E_3 form of $\mathrm{BP}\langle n\rangle$. Therefore, the elements $\lambda_1, \dots, \lambda_n$ are not p–torsion and there are no further differentials in the v_0–Bockstein spectral sequence (3-6). We define

$$\lambda_s := \begin{cases} \lambda_s & \text{if } 1 \le s \le n+1, \\ \lambda_{s-1}\mu_{n+1}^{p^{s-(n+2)}(p-1)} & \text{if } s > n+1. \end{cases} \tag{3-10}$$

Note that $\mathrm{THH}_*(B\langle n\rangle; \mathbb{Z}_{(p)})$ is finite type so we can compute $\mathrm{THH}_*(B\langle n\rangle; \mathbb{Z}_{(p)})$ from $\mathrm{THH}_*(B\langle n\rangle; \mathbb{Q})$ and $\mathrm{THH}_*(B\langle n\rangle; \mathbb{Q}_p)$ using the arithmetic fracture square

$$\begin{array}{ccc} \mathrm{THH}_*(B\langle n\rangle; \mathbb{Z}_{(p)}) & \longrightarrow & \prod_p \mathrm{THH}_*(B\langle n\rangle; \mathbb{Z}_p) \\ \downarrow & & \downarrow \\ \mathrm{THH}_*(B\langle n\rangle; \mathbb{Q}) & \longrightarrow & \prod_p \mathrm{THH}_*(B\langle n\rangle; \mathbb{Q}_p) \end{array}$$

This proves the following theorem.

Theorem 3.8 *Let $B\langle n\rangle$ be an arbitrary E_3 form of $\mathrm{BP}\langle n\rangle$ and at $p > 2$ assume the error term* (3-7) *vanishes. Then there is an isomorphism of graded $\mathbb{Z}_{(p)}$–modules*

$$\mathrm{THH}_*(B\langle n\rangle; H\mathbb{Z}_{(p)}) \cong E_{\mathbb{Z}_{(p)}}(\lambda_1, \dots, \lambda_n) \otimes (\mathbb{Z}_{(p)} \oplus T_0^n),$$

where T_0^n is a torsion $\mathbb{Z}_{(p)}$–module defined by

$$T_0^n = \bigoplus_{s\ge 1} \mathbb{Z}/p^s \otimes P_{\mathbb{Z}_{(p)}}(\mu_{n+1}^{p^s}) \otimes \mathbb{Z}_{(p)}\{\lambda_{n+s}\mu_{n+1}^{jp^{s-1}} \mid 0 \le j \le p-2\}. \tag{3-11}$$

4 Topological Hochschild homology mod (p, v_2)

In this section, we compute topological Hochschild homology of $B\langle 2\rangle$ with coefficients in $k(1)$. First we compute topological Hochschild homology with coefficients in $K(1)$.

4.1 $K(1)$–local topological Hochschild homology

In this section we assume that $p \ge 3$ and write $B\langle 2\rangle$ for an E_3 form of $\mathrm{BP}\langle 2\rangle$. Write $k(1) = B\langle 2\rangle/(p, v_2)$ for the E_1 $B\langle 2\rangle$–algebra constructed as in Proposition 2.4 and let $K(1) = k(1)[v_1^{-1}]$. In order to determine the topological Hochschild homology of $B\langle 2\rangle$ with coefficients in $k(1)$, we first determine

$$\mathrm{THH}(B\langle 2\rangle; K(1)) = \mathrm{THH}(B\langle 2\rangle; K(1)).$$

To compute the multiplicative Bökstedt spectral sequence

$$E^2_{*,*} = \mathrm{HH}_{*,*}^{K(1)_*}(K(1)_*B\langle 2\rangle) \Rightarrow K(1)_*\,\mathrm{THH}(B\langle 2\rangle),$$

we first need to compute $K(1)_*B\langle 2\rangle$. To compute $K(1)_*B\langle 2\rangle$ we first relate it to $\mathrm{BP}_*\mathrm{BP}$. Recall that

$$\mathrm{BP}_*\mathrm{BP} = \mathrm{BP}_*[t_1, t_2, \dots]$$

with $|t_i| = 2p^i - 2$. By [Ravenel 1986, Theorem A2.2.6], the right unit η_R is determined by

$$\sum_{i,j\geq 0}^{F} t_i \eta_R(v_j)^{p^i} = \sum_{i,j\geq 0}^{F} v_i t_j^{p^i} \tag{4-1}$$

where $t_0 = 1$ and $v_0 = p$.

Lemma 4.1 *The composite map*

$$K(1)_* \otimes_{\mathrm{BP}_*} \mathrm{BP}_*\mathrm{BP} \otimes_{\mathrm{BP}_*} B\langle 2\rangle_* \to \pi_*(K(1) \wedge_{\mathrm{BP}} (\mathrm{BP}\wedge\mathrm{BP}) \wedge_{\mathrm{BP}} B\langle 2\rangle) \cong K(1)_* B\langle 2\rangle$$

is an isomorphism.

Proof Consider the commutative diagram

$$\begin{array}{ccc} \pi_*(K(1)\wedge B\langle 2\rangle) & \longrightarrow & \pi_*(K(1)\wedge B\langle 2\rangle[v_1^{-1}]) \\ \cong\uparrow & & \cong\uparrow \\ \pi_*(K(1)\wedge_{\mathrm{BP}}(\mathrm{BP}\wedge\mathrm{BP})\wedge_{\mathrm{BP}} B\langle 2\rangle) & \longrightarrow & \pi_*(K(1)\wedge_{\mathrm{BP}}(\mathrm{BP}\wedge\mathrm{BP})\wedge_{\mathrm{BP}} B\langle 2\rangle[v_1^{-1}]) \\ \uparrow & & \uparrow \\ K(1)_*\otimes_{\mathrm{BP}_*}\mathrm{BP}_*\mathrm{BP}\otimes_{\mathrm{BP}_*} B\langle 2\rangle_* & \longrightarrow & \pi_*(K(1))\otimes_{\mathrm{BP}_*}\mathrm{BP}_*\mathrm{BP}\otimes_{\mathrm{BP}_*} B\langle 2\rangle_*[v_1^{-1}] \end{array} \tag{4-2}$$

Since $B\langle 2\rangle[v_1^{-1}]$ is Landweber exact, the right-hand vertical map is an isomorphism. In (4-1) the F–summands in degree $\leq 2p-2$ are $\eta_R(v_0)$, $t_1\eta_R(v_0)^p$, $\eta_R(v_1)$, v_0, v_1 and v_0t_1. Thus, $\eta_R(v_1) = v_1$ in $K(1)_*\otimes_{\mathrm{BP}_*}\mathrm{BP}_*\mathrm{BP} = K(1)_*[t_i \mid i\geq 1]$, because $p=0$ in this ring. In $K(1)_*\otimes_{\mathrm{BP}_*}\mathrm{BP}_*\mathrm{BP}\otimes_{\mathrm{BP}_*} B\langle 2\rangle_*$,

$$v_1\otimes 1\otimes 1 = 1\otimes v_1\otimes 1 = 1\otimes\eta_R(v_1)\otimes 1 = 1\otimes 1\otimes v_1$$

holds. This implies that the upper and lower horizontal map in the diagram are isomorphisms. It follows that the left vertical map is an isomorphism too. □

Notation 4.2 Let $f_i(v_1, v_2) \in B\langle 2\rangle_* = \mathbb{Z}_{(p)}[v_1, v_2]$ be the image of v_i under $\mathrm{BP}_* \to B\langle 2\rangle_*$. Define

$$v_i' := v_i - f_i(v_1, v_2) \in \mathrm{BP}_*.$$

Then v_i' is in the kernel of $\mathrm{BP}_* \to B\langle 2\rangle_*$ and $\mathrm{BP}_* = \mathbb{Z}_{(p)}[v_1, v_2, v_3', \dots]$.

By Lemma 4.1,

$$\begin{aligned} K(1)_*B\langle 2\rangle &= (K(1)_*\otimes_{\mathrm{BP}_*}\mathrm{BP}_*[t_1,\dots])\otimes_{\mathbb{Z}_{(p)}[v_1,v_2,v_3',\dots]}\mathbb{Z}_{(p)}[v_1,v_2] \\ &= K(1)_*[t_i \mid i\geq 1]/(\eta_R(v_3'),\dots). \end{aligned}$$

Lemma 4.3 *For $i\geq 0$ the element $\eta_R(v_{i+1}) \in K(1)_*[t_i \mid i\geq 1]$ actually lies in $K(1)_*[t_1,\dots,t_i]$. In fact,*

$$\eta_R(v_{i+1}) = v_{i+1} + v_1 t_i^p - v_1^{p^i} t_i + g_i,$$

where $g_i \in K(1)_*[t_1,\dots,t_{i-1}]$.

Proof We will prove the claim in $k(1)_*[t_i \mid i\geq 1]$; from this the result will follow. The reason we do this is because we will want to make degree arguments, and hence will want to avoid negative gradings.

In $\mathrm{BP}_*\mathrm{BP}/(p)$, we have $\eta_R(v_1) = v_1$. It also follows from (4-1) that, for $i \geq 0$,

$$\eta_R(v_{i+1}) \equiv v_{i+1} + v_1 t_i^p - v_1^{p^i} t_i \mod (t_1, t_2, \dots, t_{i-1})$$

in $\mathrm{BP}_*\mathrm{BP}/(p)$. Thus, this congruence also holds in $k(1)_*[t_i \mid i \geq 1]$. Since $\eta_R(v_{i+1})$ lifts to $\mathrm{BP}_*\mathrm{BP}/(p)$ we may make our degree arguments in $k(1)_*[t_i \mid i \geq 1]$. In the ring $k(1)_*[t_i \mid i \geq 1]$, we therefore have

$$\eta_R(v_{i+1}) = v_{i+1} + v_1 t_i^p - v_1^{p^i} t_i + g_i,$$

where g_i is a polynomial in the ideal generated by $t_1, t_2, \dots, t_{i-1}$. Thus far we have not excluded the possibility that a monomial divisible by t_j with $j \geq i$ occurs as a summand of g_i.

For $j > i+1$, we can exclude this possibility for degree reasons. Indeed, $\eta_R(v_{i+1})$ is homogenous of degree $2(p^{i+1}-1)$, and when $j > i+1$ the element t_j has degree greater than $2(p^{i+1}-1)$. Consider the case when $j = i+1$. To exclude this case, suppose there exists a monomial m in $k(1)_*[t_1, \dots, t_i]$ which is a summand of g_i and is divisible by t_{i+1}. Then as the degrees of t_{i+1} and $\eta_R(v_{i+1})$ are the same, it follows that $m = a t_{i+1}$ for some $a \in \mathbb{F}_p$. If $a \neq 0$, then this contradicts the assumption that g_i is in the ideal $(t_1, \dots, t_{i-1})$. This shows that $g_i \in k(1)_*[t_1, \dots, t_i]$.

We now exclude the possibility that a monomial divisible by t_i occurs as a summand of g_i. Note that the summands $t_k \eta_R(v_j)^{p^k}$ and $v_k t_j^{p^k}$ in (4-1) both have degree $2(p^{k+j}-1)$. Cross terms in (4-1) from those summands with degree less than or equal than $2(p^{i+1}-1)$ could potentially produce a t_i divisible monomial as a summand of g_i. On the right-hand side of (4-1), the possible summands are those of the form $v_j t_i^{p^j}$. As this must have degree at most $2(p^{i+1}-1)$, we must have $j = 0, 1$. These correspond, respectively, to $v_0 t_i = p t_i$ and $v_1 t_i^p$. But $p = 0$ in $k(1)_*$, so the only one to consider is $v_1 t_i^p$. This has degree exactly $2(p^{i+1}-1)$, and so a monomial divisible by this element does not occur in g_i. In fact, it has already been accounted for.

For the left-hand side, we similarly find that the only summand which could potentially produce a t_i divisible monomial as a summand of $\eta_R(v_{i+1})$ is

$$t_i \eta_R(v_1)^{p^i} = t_i v_1^{p^i}.$$

As this has exactly degree $2(p^{i+1}-1)$, it does not occur in g_i because it cannot be written as an element in the ideal $(t_1, \dots, t_{i-1})$ for degree reasons. In fact, this element has already been accounted for. Thus there are no t_i divisible monomials appearing as summands of g_i. Consequently, we have shown that $g_i \in k(1)_*[t_1, \dots, t_{i-1}]$ as desired. □

Recall from the proof of Proposition 2.4 that

$$v_i' = v_i - f_i(v_1, v_2)$$

for some $f_i \in \mathbb{Z}_{(p)}[x, y]$. In light of the previous lemma, we conclude that the class

$$\eta_R(v_i') = \eta_R(v_i) - f_i(\eta_R(v_1), \eta_R(v_2)) \in K(1)_*[t_i \mid i \geq 1]$$

also lies in $K(1)_*[t_1, \dots, t_{i-1}]$ for each $i \geq 3$.

Lemma 4.4 *The maps of commutative $K(1)_*$–algebras*

$$K(1)_*[t_1,\dots,t_{i-1}]/(\eta_R(v_3'),\dots,\eta_R(v_i')) \to K(1)_*[t_1,\dots,t_i]/(\eta_R(v_3'),\dots,\eta_R(v_{i+1}'))$$

induced by precomposing the canonical quotient map with the canonical inclusion map are étale for $i \geq 2$.

Proof For ease of notation, set

$$A_i := K(1)_*[t_1,\dots,t_{i-1}]/(\eta_R(v_3'),\dots,\eta_R(v_i'))$$

for $i \geq 2$. Note that Lemma 4.3 allows us to make this definition. Note also that $A_{i+1} = A_i[t_i]/(\eta_R(v_{i+1}'))$. We wish to show that the map

$$A_i \to A_{i+1}$$

is an étale morphism. To do this, it is enough to show that the partial derivative of $\eta_R(v_{i+1}')$ with respect to t_i is a unit in A_i. Write ∂_i for the partial derivative with respect to t_i. Since

$$v_{i+1}' = v_{i+1} - f_{i+1}(v_1, v_2)$$

for some $f_{i+1} \in \mathbb{Z}_{(p)}[x, y]$, we can infer that

$$\eta_R(v_{i+1}') = \eta_R(v_{i+1}) - f_{i+1}(\eta_R(v_1), \eta_R(v_2)).$$

In $K(1)_*[t_1, t_2, \dots]$, we know $\eta_R(v_1) = v_1$ since $p = 0$ in $K(1)_*$, and we have

$$\eta_R(v_2) = v_1 t_1^p - v_1^p t_1.$$

Thus,

$$\partial_i \eta_R(v_{i+1}') = \partial_i \eta_R(v_{i+1})$$

for $i \geq 2$ and it suffices to show that $\partial_i \eta_R(v_{i+1})$ is a unit in A_i.

By Lemma 4.3, we have the formula

$$\eta_R(v_{i+1}) = v_{i+1} + v_1 t_i^p - v_1^{p^i} t_i + g_i,$$

where $g_i \in K(1)_*[t_1,\dots,t_{i-1}]$. Thus, we conclude that

$$\partial_i \eta_R(v_{i+1}) = -v_1^{p^i} \in K(1)_*[t_1,\dots,t_{i-1}].$$

Since $v_1^{p^i}$ is a unit in $K(1)_*$, this shows that $\partial_i \eta_R(v_{i+1})$ is a unit in A_i. □

We continue to use the notation from the proof of the previous lemma. Since each map $A_i \to A_{i+1}$ is étale, we may apply [Weibel and Geller 1991, Theorem 0.1] to conclude that

$$\begin{aligned}
\mathrm{HH}_{*,*}^{K(1)_*}(K(1)_* B\langle 2\rangle) &= \operatorname{colim} \mathrm{HH}_{*,*}^{K(1)_*}(A_i) \\
&= \operatorname{colim} \mathrm{HH}_{*,*}^{K(1)_*}(A_2) \otimes_{A_2} A_i \\
&= \mathrm{HH}_{*,*}^{K(1)_*}(A_2) \otimes_{A_2} K(1)_* B\langle 2\rangle \\
&= E(\sigma t_1) \otimes K(1)_* B\langle 2\rangle.
\end{aligned}$$

Since this is concentrated in Bökstedt filtration 0 and 1, the Bökstedt spectral sequence collapses, yielding

$$E(\sigma t_1) \otimes K(1)_* B\langle 2\rangle \cong K(1)_* \operatorname{THH}(B\langle 2\rangle).$$

In the Hopf algebroid $(\mathrm{BP}_*, \mathrm{BP}_*\mathrm{BP})$, we have the formula

$$\sum_{i\geq 0}^{F} \Delta(t_i) = \sum_{i,j\geq 0}^{F} t_i \otimes t_j^{p^i}$$

by [Ravenel 1986, Theorem A2.1.27]. Since the $\mathrm{BP}_*\mathrm{BP}$–coaction on t_i agrees with the coproduct, it is determined by the formula

$$\Delta(t_1) = 1 \otimes t_1 + t_1 \otimes 1.$$

Note that $(K(1)_*, K(1)_*K(1))$ is a flat Hopf algebroid and $K(1)_*(X)$ is a left $K(1)_*K(1)$–comodule for every spectrum X. By naturality, we observe that $t_1 \in K(1)_* B\langle 2\rangle$ has the $K(1)_*K(1)$–coaction $1 \otimes t_1 + t_1 \otimes 1$. Let

$$\sigma\colon K(1)_* B\langle 2\rangle \to K(1)_{*+1} \operatorname{THH}(B\langle 2\rangle)$$

be the usual σ operator analogous to the one defined in [McClure and Staffeldt 1993]. By [Angeltveit and Rognes 2005, Proposition 5.10], which also applies to our setting because the Hopf element $\eta = 0 \in K(1)_*$, the operator σ is a derivation. It is also clear that σ is compatible with the $K(1)_*K(1)$–comodule action in the sense that

$$\psi(\sigma x) = (1 \otimes \sigma)(\psi(x)),$$

where

$$\psi\colon K(1)_* \operatorname{THH}(B\langle 2\rangle) \to K(1)_*K(1) \otimes K(1)_* \operatorname{THH}(B\langle 2\rangle).$$

It follows that $\sigma t_1 \in K(1)_* \operatorname{THH}(B\langle 2\rangle)$ is a comodule primitive. Since there is a weak equivalence $\operatorname{THH}(B\langle 2\rangle, K(1)) \simeq K(1) \wedge_{B\langle 2\rangle} \operatorname{THH}(B\langle 2\rangle)$ by [Hahn and Wilson 2022, Remark 6.1.4], we may infer from the Künneth isomorphism that there is an isomorphism of $K(1)_*$–modules

$$K(1)_* \operatorname{THH}(B\langle 2\rangle; K(1)) \cong K(1)_*K(1) \otimes E(\sigma t_1),$$

where σt_1 is a comodule primitive. Since $\operatorname{THH}(B\langle 2\rangle; K(1))$ is a $K(1)$–module spectrum and $K(1)_*$ is a graded field, we have that it splits as a sum of suspensions of $K(1)$ and that its homotopy is isomorphic to the comodule primitives in $K(1)_* \operatorname{THH}(B\langle 2\rangle; K(1))$. Thus, there is an isomorphism of $K(1)_*$–modules

$$\operatorname{THH}_*(B\langle 2\rangle; K(1)) = K(1)_* \otimes E(\sigma t_1).$$

Since σt_1 lifts to a class in $\tilde{\lambda}_1 \in \operatorname{THH}_*(B\langle 2\rangle; k(1))$ which projects onto λ_1 via the map

$$\operatorname{THH}_*(B\langle 2\rangle; k(1)) \to \operatorname{THH}_*(B\langle 2\rangle; H\mathbb{F}_p)$$

induced by the linearization map $k(1) \to H\mathbb{F}_p$ by [Zahler 1971], we simply rename this class λ_1.

In summary, we have proven the following theorem.

Theorem 4.5 *For $B\langle 2\rangle$ an E_3 form of $\mathrm{BP}\langle 2\rangle$ and $p \geq 3$, the following hold:*

(1) *There is a weak equivalence*

$$K(1) \vee \Sigma^{2p-1} K(1) \simeq \mathrm{THH}(B\langle 2\rangle; K(1)).$$

(2) *The $P(v_1)$–module $\mathrm{THH}_*(B\langle 2\rangle; k(1))$, modulo v_1–torsion, is freely generated by 1 and λ_1.*

4.2 The v_1–Bockstein spectral sequence

We compute $\mathrm{THH}_*(B\langle 2\rangle; k(1))$ using the spectral sequence (3-2) for $n = 2$ and $i = 1$. For $s \geq 4$, we recursively define

$$\lambda_s := \lambda_{s-2} \mu_3^{p^{s-4}(p-1)}.$$

For $s \geq 1$, we define

$$r(s,1) := \begin{cases} p^{s+1} + p^{s-1} + \cdots + p^2 & \text{if } s \equiv 1 \bmod 2, \\ p^{s+1} + p^{s-1} + \cdots + p^3 & \text{if } s \equiv 0 \bmod 2. \end{cases}$$

Theorem 4.6 *Let $B\langle 2\rangle$ be an E_3 form of $\mathrm{BP}\langle 2\rangle$ and let $p \geq 3$. There is an isomorphism of $P(v_1)$–modules*

$$\mathrm{THH}_*(B\langle 2\rangle; k(1)) \cong E(\lambda_1) \otimes (P(v_1) \oplus T_1^2),$$

where

$$T_1^2 = \bigoplus_{s \geq 1} P_{r(s,1)}(v_1) \otimes E(\lambda_{s+2}) \otimes P(\mu_3^{p^s}) \otimes \mathbb{F}_p\{\lambda_{s+1} \mu_3^{j p^{s-1}} \mid 0 \leq j \leq p-2\}. \tag{4-3}$$

Proof We prove by induction on $s \geq 1$ that

$$E_{r(s,1)}^{*,*} = E(\lambda_1) \otimes \big(P(v_1) \otimes E(\lambda_{s+1}, \lambda_{s+2}) \otimes P(\mu_3^{p^{s-1}}) \oplus M_s\big)$$

with

$$M_s = \bigoplus_{t=1}^{s-1} P_{r(t,1)}(v_1) \otimes E(\lambda_{t+2}) \otimes P(\mu_3^{p^t}) \otimes \mathbb{F}_p\{\lambda_{t+1} \mu_3^{j p^{t-1}} \mid 0 \leq j \leq p-2\},$$

that we have a differential $d_{r(s,1)}(\mu_3^{p^{s-1}}) \doteq v_1^{r(s,1)} \lambda_{s+1}$, and that the classes λ_{s+1} and λ_{s+2} are infinite cycles. This implies the statement.

By Theorem 4.5, the elements v_1^s are permanent cycles for every s, so the classes λ_2 and λ_3 cannot support differentials and thus are infinite cycles. Note that we use $p \geq 3$ here; for $p = 2$ we would have a possible differential $d_2(\lambda_3) \doteq v_1^2 \lambda_1 \lambda_2$. Since the classes $v_1^n \lambda_1$ survive by Theorem 4.5, the only possible differential on μ_3 is

$$d_{p^2}(\mu_3) \doteq v_1^{p^2} \lambda_2$$

for bidegree reasons. This differential must exist because otherwise the spectral sequence would collapse at the E_2–page by multiplicativity which would contradict Theorem 4.5. This proves the base step $s = 1$ of the induction. Now, assume that the statement holds for some $s \geq 1$. We then get

$$E_{r(s,1)+1}^{*,*} = E(\lambda_1) \otimes (P(v_1) \otimes E(\lambda_{s+2}, \lambda_{s+1} \mu_3^{p^{s-1}(p-1)}) \otimes P(\mu_3^{p^s}) \oplus M_{s+1},$$

and it suffices to show that $\lambda_{s+3} = \lambda_{s+1}\mu_3^{p^{s-1}(p-1)}$ is an infinite cycle and that

$$d_{r(s+1,1)}(\mu_3^{p^s}) \doteq v_1^{r(s+1,1)}\lambda_{s+2}.$$

Note that the class λ_{s+2} is an infinite cycle by the induction hypothesis. The class λ_{s+3} is an infinite cycle for bidegree reasons and because the classes v_1^s are permanent cycles. Note that we use $p \geq 3$ here; for $p = 2$ and s even we would have a possible differential $d_{r(s,1)+p}(\lambda_{s+3}) \doteq v_1^{r(s,1)+p}\lambda_1\lambda_{s+2}$. The class $\mu_3^{p^s}$ must support a differential because otherwise the spectral sequence would collapse at this stage which would contradict Theorem 4.5. Since the classes $v_1^n\lambda_1$ are permanent cycles,

$$d_{r(s+1,1)}(\mu_3^{p^s}) \doteq v_1^{r(s+1,1)}\lambda_{s+2}$$

for bidegree reasons. Here note that $v_1^{r(s,1)}\lambda_{s+3}$ has the right topological degree, but the filtration degree is too low for it to be the target of a differential on $\mu_3^{p^s}$ at the E_ℓ–page for $\ell > r(s,1)$. This completes the induction step. □

5 Topological Hochschild homology mod (p, v_1)

In this section $B\langle 2\rangle$ is again an E_3 form of $\mathrm{BP}\langle 2\rangle$, eg $\mathrm{tmf}_1(3)$ at $p = 2$, taf^D at $p = 3$, or $\mathrm{BP}\langle n\rangle'$ at an arbitrary prime p. We let $k(2) := B\langle 2\rangle/(p, v_1)$ be the E_1 $B\langle 2\rangle$–algebra constructed in Proposition 2.4 and let $K(2) = k(2)[v_2^{-1}]$. The goal of this section is to compute the homotopy groups of $\mathrm{THH}(B\langle 2\rangle; K(2))$. In Section 5.1, we first show that the unit map

$$K(2) \to \mathrm{THH}_*(B\langle 2\rangle; K(2))$$

is an equivalence. This implies that in the abutment of the v_2–Bockstein spectral sequence

$$\mathrm{THH}_*(B\langle 2\rangle; H\mathbb{F}_p)[v_2] \Rightarrow \mathrm{THH}_*(B\langle 2\rangle; k(2))$$

all classes are v_2–torsion besides the powers of v_2. This allows us to compute this spectral sequence in Section 5.2.

5.1 $K(2)$–local topological Hochschild homology

Considering a diagram analogous to (4-2), one sees that we have an isomorphism

$$K(2)_* \otimes_{\mathrm{BP}_*} \mathrm{BP}_*\mathrm{BP} \otimes_{\mathrm{BP}_*} B\langle 2\rangle_* \to \pi_*(K(2) \wedge B\langle 2\rangle).$$

For this, note that

$$\eta_R(v_1) = v_1 = 0 \in K(2)_* \otimes_{\mathrm{BP}_*} \mathrm{BP}_*\mathrm{BP} = K(2)_*[t_i \mid i \geq 1]$$

and therefore $\eta_R(v_2) = v_2$. This implies that the equality

$$v_2 \otimes 1 \otimes 1 = 1 \otimes 1 \otimes v_2$$

holds in the tensor product

$$K(2)_* \otimes_{\mathrm{BP}_*} \mathrm{BP}_*\mathrm{BP} \otimes_{\mathrm{BP}_*} B\langle 2\rangle_*.$$

From this, we determine that

$$K(2)_*B\langle 2\rangle = K(2)_*[t_i \mid i \geq 1]/(\eta_R(v_3'), \dots).$$

In particular, this is a graded commutative $K(2)_*$–algebra even at $p=2$ where $K(2)$ is not homotopy commutative (cf [Angeltveit and Rognes 2005, Lemma 8.9]).

Lemma 5.1 *In* $K(2)_*[t_1 \mid i \geq 1]$,

$$\eta_R(v_{i+2}) = v_{i+2} + v_2 t_i^{p^2} - v_2^{p^i} t_i + g_i$$

where $g_i \in K(2)_*[t_1, \dots, t_{i-1}]$.

Proof We argue similarly to Lemma 4.3 and make our arguments in the ring $k(2)_*[t_i \mid i \geq 1]$. The result will follow from this. We have that

$$\eta_R(v_{i+2}) \equiv v_{i+2} + v_2 t_i^{p^2} - v_2^{p^i} t_i \mod (t_1, t_2, \dots, t_{i-1}),$$

in $\mathrm{BP}_*\mathrm{BP}/(p, v_1)$ (see [Ravenel 1986, Proof of Theorem 4.3.2]). Consequently, this formula also holds in $k(2)_*[t_i \mid i \geq 1]$. This shows that in $k(2)_*[t_i \mid i \geq 1]$,

$$\eta_R(v_{i+2}) = v_{i+2} + v_2 t_i^{p^2} - v_2^{p^i} t_i + g_i$$

for some g_i in the ideal $(t_1, t_2, \dots, t_{i-1})$. Since $\eta_R(v_{i+2})$ lifts to the graded abelian group $\mathrm{BP}_*\mathrm{BP}/(p, v_1)$, we may also make degree arguments in $k(2)_*[t_i \mid i \geq 1]$.

Note that for degree reasons, there can be no instance of a t_j with $j > i+2$ dividing a monomial summand of g_i. We can also exclude the possibility of t_{i+2} dividing a monomial in g_i. Indeed, a monomial in g_i divisible by t_{i+2} would necessarily be just t_{i+2} itself, contradicting that g_i is in the ideal $(t_1, \dots, t_{i-1})$. This shows that

$$\eta_R(v_{i+2}) \in k(2)_*[t_1, \dots, t_{i+1}].$$

for all $i \geq 0$.

We now exclude the possibility that t_{i+1} divides a monomial in $\eta_R(v_{i+2})$. To do this, we note that a t_{i+1} divisible monomial in g_i could arise from cross terms involving the universal p–typical formal group law and the formula (4-1). Note that the only terms to consider on the right-hand side are $v_0 t_{i+1}$ and $v_1 t_{i+1}^p$, which are 0 since $p = v_1 = 0 \in k(2)_*$. On the left-hand side, we only need to consider the terms $t_k \eta_R(v_{j+2})^{p^k}$ of degree less than or equal to $2(p^{i+2}-1)$. This immediately implies that $j \leq i$. For $k = i+1$, the term of smallest degree is $t_{i+1}\eta_R(v_2)^{p^{i+1}}$. The degree of this term is $2(p^{i+3}-1)$, which is too large. Thus we can exclude the possibility that $k = i+1$. Now as $j \leq i$ and since we have shown that $\eta_R(v_{j+2}) \in k(2)_*[t_1, \dots, t_{j+1}]$, we see that none of the relevant terms on the left-hand side can contribute a t_{i+1} divisible monomial summand to $\eta_R(v_{i+2})$. Thus we have that $g_i \in K(2)_*[t_1, \dots, t_i]$.

We are left to consider whether a t_i divisible monomial could occur as a summand of g_i via the cross terms coming from the formal group law F in (4-1). On the right-hand side, we only need to consider the term $v_2 t_i^{p^2}$. Here we use the fact that $v_1 = 0 \in k(2)_*$. This term has already been accounted for and is not in g_i. On the left-hand side, since we have shown that $\eta_R(v_{j+2}) \in k(2)_*[t_1, \dots, t_j]$, the only term we need to consider is $t_i v_2^{p^i}$. Again, we have already considered this term. We can therefore conclude that $g_i \in k(2)_*[t_1, \dots, t_{i-1}]$. □

Definition 5.2 We define commutative $K(2)_*$–algebras

$$C_0 := K(2)_*,$$
$$C_i := C_{i-1}[t_i]/\eta_R(v'_{i+2}), \quad i \geq 1,$$

and write $h_i \colon C_{i-1} \to C_i$ for the map of commutative $K(2)_*$–algebras defined as the composite of the canonical inclusion map $C_{i-1} \to C_{i-1}[t_i]$ with the canonical quotient map $C_{i-1}[t_i] \to C_{i-1}[t_i]/\eta_R(v'_{i+2})$.

Thus we have

$$C_i = K(2)_*[t_1, \dots, t_i]/(\eta_R(v'_3), \dots, \eta_R(v'_{i+2}))$$

for $i \geq 1$ and

$$K(2)_* B\langle 2\rangle = \operatorname*{colim}_i C_i.$$

We proceed in the same fashion as in Section 4.1 and argue that $h_i \colon C_{i-1} \to C_i$ is étale by examining the derivative of $\eta_R(v'_{i+2})$ with respect to t_i.

Lemma 5.3 *The map of commutative rings $h_i \colon C_{i-1} \to C_i$ from Definition 5.2 is étale.*

Proof We have that

$$v'_{i+2} = v_{i+2} - f_{i+2}(v_1, v_2) = v_{i+2} - f_{i+2}(0, v_2).$$

Hence,

$$\eta_R(v'_{i+2}) = \eta_R(v_{i+2}) - f_{i+2}(0, v_2).$$

Let ∂_i denote the partial derivative with respect to t_i. Since $C_i = C_{i-1}[t_i]/(\eta_R(v'_{i+2}))$, to show the morphism $C_{i-1} \to C_i$ is étale, it is enough to show that $\partial_i \eta_R(v'_{i+2})$ is a unit. We have

$$\partial_i \eta_R(v'_{i+2}) = \partial_i \eta_R(v_{i+2}) - \partial_i f_{i+2}(0, v_2) = \partial_i \eta_R(v_{i+2}).$$

From Lemma 5.1, we find that $\partial_i g_i = 0$, and hence

$$\partial_i \eta_R(v_{i+2}) = \partial_i (v_{i+2} + v_2 t_i^{p^2} - v_2^{p^i} t_i + g_i) = -v_2^{p^i}$$

which is a unit. □

Since each map $C_i \to C_{i+1}$ is étale, we may apply [Weibel and Geller 1991, Theorem 0.1] to conclude that the unit map

$$K(2)_* B\langle 2\rangle \to \mathrm{HH}_{*,*}^{K(2)_*}(K(2)_* B\langle 2\rangle) \tag{5-1}$$

is an isomorphism of graded commutative $\mathbb{F}_p$–algebras (even at $p=2$). The unit map

$$K(2)_*B\langle 2\rangle \to K(2)_*\,\mathrm{THH}(B\langle 2\rangle)$$

is the edge homomorphism in the Bökstedt spectral sequence

$$E^2_{*,*} = \mathrm{HH}^{K(2)_*}_{*,*}(K(2)_*B\langle 2\rangle) \Rightarrow K(2)_*\,\mathrm{THH}(B\langle 2\rangle)$$

and the input is concentrated in Bökstedt filtration zero by (5-1), so the spectral sequence collapses without extensions yielding an isomorphism

$$K(2)_*B\langle 2\rangle \cong K(2)_*\,\mathrm{THH}(B\langle 2\rangle)$$

of graded commutative $\mathbb{F}_p$–algebras (even at the prime $p=2$).

By the Künneth isomorphism, the map

$$K(2)_*K(2) \to K(2)_*\,\mathrm{THH}(B\langle 2\rangle, K(2))$$

is an isomorphism as well. Since both $K(2)$ and $\mathrm{THH}(B\langle 2\rangle; K(2))$ are $K(2)$–local, we obtain the following result.

Corollary 5.4 *The unit map*

$$\eta\colon K(2) \to \mathrm{THH}(B\langle 2\rangle; K(2))$$

is an equivalence. Consequently, the $P(v_2)$–module $\mathrm{THH}_*(B\langle 2\rangle; k(2))$ modulo v_2–torsion is freely generated by 1.

5.2 The v_2–Bockstein spectral sequence

Recall from Section 3.1 that the tower of spectra used to build the Bockstein spectral sequence (3-1) can be identified as an Adams tower and therefore the Bockstein spectral sequence is multiplicative.

For $s \ge 4$, recursively define

$$\lambda_s := \lambda_{s-3}\mu_3^{p^{s-4}(p-1)}.$$

For $s \ge 1$, set

$$r(s,2) = \begin{cases} p^s + p^{s-3} + \cdots + p^4 + p & \text{if } s \equiv 1 \bmod 3,\\ p^s + p^{s-3} + \cdots + p^5 + p^2 & \text{if } s \equiv 2 \bmod 3,\\ p^s + p^{s-3} + \cdots + p^6 + p^3 & \text{if } s \equiv 0 \bmod 3.\end{cases}$$

Theorem 5.5 *Let $B\langle 2\rangle$ be an E_3 form of $\mathrm{BP}\langle 2\rangle$. There is an isomorphism of $P(v_2)$–modules*

$$\mathrm{THH}_*(B\langle 2\rangle; k(2)) \cong P(v_2) \oplus T_2^2,$$

where

$$T_2^2 \cong \bigoplus_{s\ge 1} P_{r(s,2)}(v_2) \otimes E(\lambda_{s+1}, \lambda_{s+2}) \otimes P(\mu_3^{p^s}) \otimes \mathbb{F}_p\{\lambda_s \mu_3^{jp^{s-1}} \mid 0 \le j \le p-2\}. \tag{5-2}$$

Proof We prove by induction on $s \geq 1$ that

$$E^{*,*}_{r(s,2)} = P(v_2) \otimes E(\lambda_s, \lambda_{s+1}, \lambda_{s+2}) \otimes P(\mu_3^{p^{s-1}}) \oplus M_s$$

with

$$M_s = \bigoplus_{t=1}^{s-1} P_{r(t,2)}(v_2) \otimes E(\lambda_{t+1}, \lambda_{t+2}) \otimes P(\mu_3^{p^t}) \otimes \mathbb{F}_p\{\lambda_t \mu_3^{jp^{t-1}} \mid 0 \leq j \leq p-2\},$$

that λ_s, λ_{s+1} and λ_{s+2} are infinite cycles, and that $d_{r(s,2)}(\mu_3^{p^{s-1}}) \doteq v_2^{r(s,2)} \lambda_s$. This implies the statement. Since the v_2^n survive to the E_∞–page by Corollary 5.4, the classes λ_1, λ_2 and λ_3 are infinite cycles. The class μ_3 needs to support a differential, because otherwise the spectral sequence would collapse at the E_2–page by multiplicativity, which is a contradiction to Corollary 5.4. For bidegree reasons the only possibility is

$$d_p(\mu_3) \doteq v_2^p \lambda_1.$$

This proves the base step $s = 1$ of the induction. We now assume that the statement holds for some $s \geq 1$. We then get

$$E^{*,*}_{r(s,2)+1} = P(v_2) \otimes E(\lambda_{s+1}, \lambda_{s+2}, \lambda_s \mu_3^{p^{s-1}(p-1)}) \otimes P(\mu_3^{p^s}) \oplus M_{s+1}.$$

It now suffices to show that $\lambda_{s+3} = \lambda_s \mu_3^{p^{s-1}(p-1)}$ is an infinite cycle and that we have a differential $d_{r(s+1,2)}(\mu_3^{p^s}) \doteq v_2^{r(s+1,2)} \lambda_{s+1}$. We cannot have a differential of the form

$$d_r(\lambda_{s+3}) \doteq v_2^n \lambda_{s+1} \lambda_{s+2}$$

for degree reasons, so λ_{s+3} is an infinite cycle. The class $\mu_3^{p^s}$ must support a differential, because otherwise the spectral sequence would collapse at this stage, which is a contradiction to Corollary 5.4. For bidegree reasons the only possibility is

$$d_{r(s+1,2)}(\mu_3^{p^s}) \doteq v_2^{r(s+1,2)} \lambda_{s+1}.$$

Note that $v_2^{r(s,2)} \lambda_{s+3}$ has the right topological degree, but a too small filtration degree to be the target of a differential on $\mu_3^{p^s}$. This completes the inductive step. □

We end with a conjectural answer for $\mathrm{THH}(\mathrm{BP}\langle n\rangle; k(m))$ for all $1 \leq m \leq n$.

Conjecture 5.6 Suppose $1 \leq m \leq n$. Let $B\langle n\rangle$ be an E_3 form of $\mathrm{BP}\langle n\rangle$. There is an isomorphism

$$\mathrm{THH}_*(B\langle n\rangle; k(m)) \cong E(\lambda_1, \dots, \lambda_{n-m}) \otimes (P(v_m) \oplus T^n_m),$$

where

$$T^n_m = \bigoplus_{s \geq 1} P_{r_n(s,m)}(v_m) \otimes E(\lambda_{n-m+s+1}, \dots, \lambda_{n+s}) \otimes P(\mu_{n+1}^{p^s}) \otimes \mathbb{F}_p\{\lambda_{n-m+s} \mu_{n+1}^{p^{\ell p^{s-1}}} \mid 0 \leq \ell \leq p-2\}$$

and by convention $E(\lambda_1, \dots, \lambda_{n-m}) = \mathbb{F}_p$ when $n = m$. The sequence of integers $r_n(s,m)$ is defined by

$$r_n(s,m) = p^{n-m+s} + p^{n-m+s-(m+1)} + \cdots + p^{n+j-m},$$

where j is the unique element in $\{1, \dots, m+1\}$ such that $s \equiv j \mod m+1$.

Here the class λ_s is defined recursively by the formula

$$\lambda_s := \lambda_{s-(m+1)} \mu_{n+1}^{p^{s-(n+2)}(p-1)}$$

for $s \geq n+2$ and we name the classes in the abutment that are not divisible by v_n by their projection to $\mathrm{THH}_*(B\langle n\rangle; H\mathbb{F}_p)$.

Remark 5.7 When $m = 1$ and $n = 2$, we observe that this is consistent with Theorem 4.6 where $r_2(s,1) = r(s,1)$. When $m = 2$ and $n = 2$, we observe that this is consistent with Theorem 5.5 where $r_2(s,2) = r(s,2)$.

References

[Angelini-Knoll and Salch 2018] **G Angelini-Knoll**, **A Salch**, *A May-type spectral sequence for higher topological Hochschild homology*, Algebr. Geom. Topol. 18 (2018) 2593–2660 MR Zbl

[Angeltveit 2008] **V Angeltveit**, *Topological Hochschild homology and cohomology of A_∞ ring spectra*, Geom. Topol. 12 (2008) 987–1032 MR Zbl

[Angeltveit and Lind 2017] **V Angeltveit**, **J A Lind**, *Uniqueness of* $\mathrm{BP}\langle n\rangle$, J. Homotopy Relat. Struct. 12 (2017) 17–30 MR Zbl

[Angeltveit and Rognes 2005] **V Angeltveit**, **J Rognes**, *Hopf algebra structure on topological Hochschild homology*, Algebr. Geom. Topol. 5 (2005) 1223–1290 MR Zbl

[Angeltveit et al. 2010] **V Angeltveit**, **M A Hill**, **T Lawson**, *Topological Hochschild homology of ℓ and ko*, Amer. J. Math. 132 (2010) 297–330 MR Zbl

[Ausoni and Richter 2020] **C Ausoni**, **B Richter**, *Towards topological Hochschild homology of Johnson–Wilson spectra*, Algebr. Geom. Topol. 20 (2020) 375–393 MR Zbl

[Ausoni and Rognes 2002] **C Ausoni**, **J Rognes**, *Algebraic K–theory of topological K–theory*, Acta Math. 188 (2002) 1–39 MR Zbl

[Baker and Richter 2008] **A Baker**, **B Richter**, *Uniqueness of E_∞ structures for connective covers*, Proc. Amer. Math. Soc. 136 (2008) 707–714 MR Zbl

[Bhatt et al. 2019] **B Bhatt**, **M Morrow**, **P Scholze**, *Topological Hochschild homology and integral p–adic Hodge theory*, Publ. Math. Inst. Hautes Études Sci. 129 (2019) 199–310 MR Zbl

[Bökstedt 1985] **M Bökstedt**, *The topological Hochschild homology of $\mathbb{Z}$ and $\mathbb{Z}/p$*, unpublished manuscript (1985)

[Browder 1961] **W Browder**, *Torsion in H–spaces*, Ann. of Math. 74 (1961) 24–51 MR Zbl

[Brun et al. 2007] **M Brun**, **Z Fiedorowicz**, **R M Vogt**, *On the multiplicative structure of topological Hochschild homology*, Algebr. Geom. Topol. 7 (2007) 1633–1650 MR Zbl

[Bruner 1977] **R R Bruner**, *The Adams spectral sequence of H_∞ ring spectra*, PhD thesis, The University of Chicago (1977) MR Available at `https://www.proquest.com/docview/302871719`

[Bruner et al. 1986] **R R Bruner**, **J P May**, **J E McClure**, **M Steinberger**, *H_∞ ring spectra and their applications*, Lecture Notes in Math. 1176, Springer (1986) MR Zbl

[Chadwick and Mandell 2015] **S G Chadwick**, **M A Mandell**, *E_n genera*, Geom. Topol. 19 (2015) 3193–3232 MR Zbl

[Cohen and Jones 2002] **R L Cohen**, **J D S Jones**, *A homotopy theoretic realization of string topology*, Math. Ann. 324 (2002) 773–798 MR Zbl

[Dundas and McCarthy 1994] **B I Dundas**, **R McCarthy**, *Stable K–theory and topological Hochschild homology*, Ann. of Math. 140 (1994) 685–701 MR Zbl

[Dunn 1997] **G Dunn**, *Lax operad actions and coherence for monoidal N–categories, A_∞ rings and modules*, Theory Appl. Categ. 3 (1997) 50–84 MR Zbl

[Hahn and Wilson 2018] **J Hahn**, **D Wilson**, *Quotients of even rings*, preprint (2018) arXiv 1809.04723

[Hahn and Wilson 2022] **J Hahn**, **D Wilson**, *Redshift and multiplication for truncated Brown–Peterson spectra*, Ann. of Math. 196 (2022) 1277–1351 MR Zbl

[Hesselholt and Nikolaus 2020] **L Hesselholt**, **T Nikolaus**, *Topological cyclic homology*, from "Handbook of homotopy theory" (H Miller, editor), CRC, Boca Raton, FL (2020) 619–656 MR Zbl

[Hill and Lawson 2010] **M Hill**, **T Lawson**, *Automorphic forms and cohomology theories on Shimura curves of small discriminant*, Adv. Math. 225 (2010) 1013–1045 MR Zbl

[Höning 2020] **E Höning**, *On the Brun spectral sequence for topological Hochschild homology*, Algebr. Geom. Topol. 20 (2020) 817–863 MR Zbl

[Joyal and Street 1985] **A Joyal**, **R Street**, *Braided monoidal categories*, Macquarie Math. Reports 850067 (1985) Available at `http://science.mq.edu.au/~street/BMC850067.pdf`

[Lawson 2018] **T Lawson**, *Secondary power operations and the Brown–Peterson spectrum at the prime 2*, Ann. of Math. 188 (2018) 513–576 MR Zbl

[Lawson and Naumann 2012] **T Lawson**, **N Naumann**, *Commutativity conditions for truncated Brown–Peterson spectra of height* 2, J. Topol. 5 (2012) 137–168 MR Zbl

[Lawson and Naumann 2014] **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 Zbl

[Lichtenbaum 1973] **S Lichtenbaum**, *Values of zeta-functions, étale cohomology, and algebraic K–theory*, from "Algebraic K–theory, II: 'Classical' algebraic K–theory and connections with arithmetic" (H Bass, editor), Lecture Notes in Math. 342, Springer (1973) 489–501 MR Zbl

[May 1970] **J P May**, *A general algebraic approach to Steenrod operations*, from "The Steenrod algebra and its applications" (F P Peterson, editor), Lecture Notes in Math. 168, Springer (1970) 153–231 MR Zbl

[May 1972] **J P May**, *The geometry of iterated loop spaces*, Lecture Notes in Math. Vol. 271, Springer (1972) MR Zbl

[McClure and Staffeldt 1993] **J E McClure**, **R E Staffeldt**, *On the topological Hochschild homology of bu, I*, Amer. J. Math. 115 (1993) 1–45 MR Zbl

[Morava 1989] **J Morava**, *Forms of K–theory*, Math. Z. 201 (1989) 401–428 MR Zbl

[Quillen 1975] **D Quillen**, *Higher algebraic K–theory*, from "Proceedings of the International Congress of Mathematicians, I" (R D James, editor), Canad. Math. Congr., Montreal, QC (1975) 171–176 MR Zbl

[Ravenel 1986] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Mathematics 121, Academic, Orlando, FL (1986) MR Zbl

[Rognes 2014] **J Rognes**, *Algebraic K–theory of strict ring spectra*, from "Proceedings of the International Congress of Mathematicians, II" (S Y Jang, Y R Kim, D-W Lee, I Yie, editors), Kyung Moon Sa, Seoul (2014) 1259–1283 MR Zbl

[Rognes 2020] **J Rognes**, *The circle action on topological Hochschild homology of complex cobordism and the Brown–Peterson spectrum*, J. Topol. 13 (2020) 939–968 MR Zbl

[Senger 2017] **A Senger**, *The Brown–Peterson spectrum is not* $\mathbb{E}_{2(p^2+2)}$ *at odd primes*, preprint (2017) arXiv 1710.09822

[Shipley 2007] **B Shipley**, $H\mathbb{Z}$*–algebra spectra are differential graded algebras*, Amer. J. Math. 129 (2007) 351–379 MR Zbl

[Weibel and Geller 1991] **C A Weibel**, **S C Geller**, *Étale descent for Hochschild and cyclic homology*, Comment. Math. Helv. 66 (1991) 368–388 MR Zbl

[Zahler 1971] **R Zahler**, *The Adams–Novikov spectral sequence for the spheres*, Bull. Amer. Math. Soc. 77 (1971) 169–174 MR Zbl

Universite Sorbonne Paris Nord, LAGA, CNRS
Villetaneuse, France

Max Planck Institute for Mathematics
Bonn, Germany

Department of Mathematics, Radboud University
Nijmegen, Netherlands

angelini-knoll@math.univ-paris13.fr, dominic.culver@gmail.com, eva.hoening@gmail.com

Received: 24 June 2021 Revised: 4 May 2023

Algebraic & Geometric Topology **24**:5 (2024) 2537–2570
DOI: 10.2140/agt.2024.24.2537
Published: 19 August 2024

Points of quantum SL_n coming from quantum snakes

DANIEL C DOUGLAS

We show that the quantized Fock–Goncharov monodromy matrices satisfy the relations of the quantum special linear group SL_n^q. The proof employs a quantum version of the technology of Fock and Goncharov, called snakes. This relationship between higher Teichmüller theory and quantum group theory is integral to the construction of an SL_n–quantum trace map for knots in thickened surfaces, partially developed in previous work of the author.

20G42, 32G15, 57K31

Introduction

For a finitely generated group Γ and a suitable Lie group G, a primary object of study in low-dimensional geometry and topology is the G–character variety

$$\mathcal{R}_G(\Gamma) = \{\rho\colon \Gamma \to G\} /\!\!/ G$$

consisting of group homomorphisms ρ from Γ to G, considered up to conjugation. Here the quotient is taken in the algebraic geometric sense of geometric invariant theory; see Mumford, Fogarty, and Kirwan [25]. Character varieties can be explored using a wide variety of mathematical skill sets. Some examples include the Higgs bundle approach of Hitchin [18], the dynamics approach of Labourie [23], and the representation theory approach of Fock and Goncharov [9].

In the case where the group $\Gamma = \pi_1(\mathfrak{S})$ is the fundamental group of a punctured surface $\mathfrak{S}$ of finite topological type, and where the Lie group $G = \mathrm{SL}_n(\mathbb{C})$ is the special linear group, we are interested in studying a relationship between two competing deformation quantizations of the character variety $\mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S}) := \mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\pi_1(\mathfrak{S}))$. Here a deformation quantization $\{\mathcal{R}^q\}_q$ of a Poisson space $\mathcal{R}$ is a family of noncommutative algebras $\mathcal{R}^q$ parametrized by a nonzero complex parameter $q = e^{2\pi i \hbar}$, such that the lack of commutativity in $\mathcal{R}^q$ is infinitesimally measured in the classical limit $\hbar \to 0$ by the Poisson bracket of the space $\mathcal{R}$. In the case where $\mathcal{R} = \mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S})$ is the character variety, the bracket is provided by the Goldman Poisson structure on $\mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S})$ [15; 16].

The first quantization of the character variety is the $\mathrm{SL}_n(\mathbb{C})$–skein algebra $\mathcal{S}_n^q(\mathfrak{S})$ of the surface $\mathfrak{S}$; see Bullock, Frohman, and Kania-Bartoszyńska [3], Kuperberg [22], Przytycki [27], Sikora [30], Turaev [32], and Witten [33]. The skein algebra is motivated by the classical algebraic geometric approach to studying the character variety $\mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S})$ via its algebra of regular functions $\mathbb{C}[\mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S})]$. An example of a regular function is the trace function $\mathrm{Tr}_\gamma\colon \mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S}) \to \mathbb{C}$ associated to a closed curve $\gamma \in \pi_1(\mathfrak{S})$ sending a representation $\rho\colon \pi_1(\mathfrak{S}) \to \mathrm{SL}_n(\mathbb{C})$ to the trace $\mathrm{Tr}(\rho(\gamma)) \in \mathbb{C}$ of the matrix $\rho(\gamma) \in \mathrm{SL}_n(\mathbb{C})$. A theorem of classical invariant theory, due to Procesi [26], implies that the trace functions Tr_γ generate the algebra of functions $\mathbb{C}[\mathcal{R}_{\mathrm{SL}_n(\mathbb{C})}(\mathfrak{S})]$ as an algebra. According to the philosophy of Turaev and Witten, quantizations of the character variety should be of a 3–dimensional nature. Indeed, knots (or links) K in the thickened surface $\mathfrak{S} \times (0,1)$ represent elements of the skein algebra $\mathcal{S}_n^q(\mathfrak{S})$. The skein algebra $\mathcal{S}_n^q(\mathfrak{S})$ has the advantage of being natural, but can be difficult to study directly.

The second quantization of the $\mathrm{SL}_n(\mathbb{C})$–character variety is the Fock–Goncharov quantum space $\mathcal{T}_n^q(\mathfrak{S})$; see Fock and Goncharov [12], Fock and Chekhov [7], and Kashaev [20]. At the classical level, Fock and Goncharov [9] introduced a framed version $\mathcal{R}_{\mathrm{PSL}_n(\mathbb{C})}(\mathfrak{S})_{\mathrm{fr}}$ (called the $\mathcal{X}$–space) of the $\mathrm{PSL}_n(\mathbb{C})$–character variety, which, roughly speaking, consists of representations $\rho\colon \pi_1(\mathfrak{S}) \to \mathrm{PSL}_n(\mathbb{C})$ equipped with additional linear algebraic data attached to the punctures of $\mathfrak{S}$. Associated to each ideal triangulation λ of the punctured surface $\mathfrak{S}$ is a λ–coordinate chart U_λ for $\mathcal{R}_{\mathrm{PSL}_n(\mathbb{C})}(\mathfrak{S})_{\mathrm{fr}}$ parametrized by N nonzero complex coordinates $X_1, X_2, \dots, X_N$ where the integer N depends only on the topology of the surface $\mathfrak{S}$ and the rank of the Lie group $\mathrm{SL}_n(\mathbb{C})$. These coordinates X_i are computed by taking various generalized cross-ratios of configurations of n–dimensional flags attached to the punctures of $\mathfrak{S}$. When written in terms of these coordinates X_i, the trace functions $\mathrm{Tr}_\gamma = \mathrm{Tr}_\gamma(X_i^{\pm 1/n})$ associated to closed curves γ take the form of Laurent polynomials in n–roots of the variables X_i. At the quantum level, there are q–deformed versions X_i^q of these coordinates, which no longer commute but q–commute with each other. The quantized character variety $\mathcal{T}_n^q(\mathfrak{S})$ is obtained by gluing together quantum tori $\mathcal{T}_n^q(\sigma)$, including one for each triangulation $\sigma = \lambda$ consisting of Laurent polynomials in the quantized Fock–Goncharov coordinates X_i^q. The quantum character variety $\mathcal{T}_n^q(\mathfrak{S})$ has the advantage of being easier to work with than the skein algebra $\mathcal{S}_n^q(\mathfrak{S})$, however it is less intrinsic.

We are interested in studying q–deformed versions Tr_γ^q of the trace functions Tr_γ, associating to a closed curve γ a Laurent polynomial in the quantized Fock–Goncharov coordinates X_i^q. Turaev and Witten's philosophy leads us from the 2–dimensional setting of curves γ on the surface $\mathfrak{S}$ to the 3–dimensional setting of knots K in the thickened surface $\mathfrak{S} \times (0,1)$. In the case of $\mathrm{SL}_2(\mathbb{C})$, such a *quantum trace map* was developed by Bonahon and Wong [1] as an injective algebra homomorphism

$$\mathrm{Tr}^q(\lambda)\colon \mathcal{S}_2^q(\mathfrak{S}) \hookrightarrow \mathcal{T}_2^q(\lambda)$$

from the $\mathrm{SL}_2(\mathbb{C})$–skein algebra to (the λ–quantum torus of) the quantized $\mathrm{SL}_2(\mathbb{C})$–character variety. Their construction is "by hand", but is implicitly related to the theory of the quantum group $\mathrm{U}_q(\mathfrak{sl}_2)$ or, more precisely, of its Hopf dual SL_2^q; see Kassel [21]. Developing a quantum trace map for $\mathrm{SL}_n(\mathbb{C})$ requires a more conceptual approach, making explicit this connection between higher Teichmüller theory and quantum group theory. In a companion paper [6], we make significant progress in this direction. Our goal here is to establish a local building block result that is essential to understanding the quantum trace map more conceptually.

Whereas the classical trace $\mathrm{Tr}_\gamma(\rho) \in \mathbb{C}$ is a number obtained by evaluating the trace of an $\mathrm{SL}_n(\mathbb{C})$–monodromy $\rho(\gamma)$ taken along a curve γ in the surface $\mathfrak{S}$, the quantum trace $\mathrm{Tr}_K(X_i^q) \in \mathcal{T}_n^q(\lambda)$ is a Laurent polynomial obtained from a quantum monodromy associated to a knot K in the thickened surface $\mathfrak{S} \times (0,1)$. This quantum monodromy is essentially constructed by chopping the knot K into little pieces, namely the components C of $K \cap (\lambda_k \times (0,1))$ where the λ_k are the triangles of the ideal triangulation λ, and associating to each piece C a local quantum monodromy matrix $\boldsymbol{M}_C^q \in \mathrm{M}_n(\mathcal{T}_n^q(\lambda_k))$. Here the coefficients of the matrix $\boldsymbol{M}_C^q$ lie in a local quantum torus $\mathcal{T}_n^q(\lambda_k)$ associated to the triangle λ_k, closely associated to the quantum torus $\mathcal{T}_n^q(\lambda)$.

Theorem *When C is an arc on the corner of a triangle λ_k, the Fock–Goncharov quantum matrix $\boldsymbol{M}_C^q \in \mathrm{M}_n(\mathcal{T}_n^q(\lambda_k))$ is a $\mathcal{T}_n^q(\lambda_k)$–point of the quantum special linear group SL_n^q. In other words, each such matrix defines an algebra homomorphism*

$$\varphi(\boldsymbol{M}_C^q)\colon \mathrm{SL}_n^q \to \mathcal{T}_n^q(\lambda_k)$$

by the property that the n^2–many generators of the algebra SL_n^q are sent to the corresponding n^2–many entries of the matrix $\boldsymbol{M}_C^q$ (see Section 2.4.1).

See Theorem 2.8 (and Douglas [5, Theorem 3.10]). Our proof uses a quantum version of the technology of Fock and Goncharov, called snakes.

The main property of the quantum trace $\mathrm{Tr}_K(X_i^q) \in \mathcal{T}_n^q(\lambda)$ is its invariance under isotopy of the knot K. This is equivalent to invariance under a handful of local Reidemeister-like moves in the thickened triangulated surface. These topological moves are independent of n, and can be seen as the oriented versions of the moves depicted in [1, Figures 15–19]. In particular, due to their local nature, these moves have a purely algebraic formulation as equalities involving $n \times n$ matrices with coefficients in the quantum torus. Our main result is essentially equivalent to the algebraic formulation of one of these moves, specifically that depicted in [1, Figure 17]; see also [6, Section 6].

For an independent study of these same algebraic identities underlying the isotopy invariance of the quantum trace map, in the context of integrable systems, see Chekhov and Shapiro [4, Theorems 2.12 and 2.14] (which, in particular, reproduces our main result). This was motivated in part by Schrader and Shapiro [28; 29]; see also Fock and Goncharov [8], Gekhtman, Shapiro, and Vainshtein [14], and Goncharov and Shen [17]. Our work complements that of [4] by focusing attention on a single isotopy move, and conceptualizing the associated quantum phenomenon as arising naturally from the underlying geometry.

Acknowledgements

This work would not have been possible without the constant guidance and support of my PhD advisor Francis Bonahon. We thank Sasha Shapiro for informing us about related research and for enjoyable conversations. We are also grateful to the referee for their helpful comments. This work was partially supported by the US National Science Foundation grants DMS-1406559 and DMS-1711297.

1 Fock–Goncharov snakes

We recall some of the classical (as opposed to the quantum) geometric theory of Fock and Goncharov [9], underlying the quantum theory discussed later on; see also [10; 11]. This section is a condensed version of [5, Chapter 2]. For other references on Fock–Goncharov coordinates and snakes see [19; 13; 24]. When $n = 2$, these coordinates date back to Thurston's shearing coordinates for Teichmüller space [31].

Let $n \in \mathbb{Z}$ for $n \geqslant 2$, and $V = \mathbb{C}^n$ be the standard n–dimensional complex vector space.

1.1 Generic configurations of flags and Fock–Goncharov invariants

A (*complete*) *flag* E in V is a collection of linear subspaces $E^{(a)} \subseteq V$ indexed by $0 \leqslant a \leqslant n$, satisfying the property that each subspace $E^{(a)}$ is properly contained in the subspace $E^{(a+1)}$. In particular, $E^{(a)}$ is a–dimensional, $E^{(0)} = \{0\}$, and $E^{(n)} = V$. Denote the space of flags by $\mathrm{Flag}(V)$.

1.1.1 Generic triples and quadruples of flags There are at least two notions of genericity for a configuration of flags. We will use just one of them, the maximum span property; for a complementary notion, the minimum intersection property see [5, Section 2.10].

Definition 1.1 A flag tuple $(E_1, E_2, \dots, E_k) \in \mathrm{Flag}(V)^k$ satisfies the *maximum span property* if either of the following equivalent conditions are satisfied: for all $0 \leqslant a_1, a_2, \dots, a_k \leqslant n$,

(1) for all $a_1 + a_2 + \dots + a_k = n$, the sum $E_1^{(a_1)} + E_2^{(a_2)} + \dots + E_k^{(a_k)} = E_1^{(a_1)} \oplus E_2^{(a_2)} \oplus \dots \oplus E_k^{(a_k)}$ is direct, and thus the sum is V, or

(2) the dimension formula $\dim(E_1^{(a_1)} + E_2^{(a_2)} + \dots + E_k^{(a_k)})$ equals $\min(a_1 + a_2 + \dots + a_k, n)$.

In the case $n = 3$, such a flag triple $(E, F, G) \in \mathrm{Flag}(V)^3$ is called a *maximum span flag triple*, and in the case $n = 4$, such a flag quadruple $(E, F, G, H) \in \mathrm{Flag}(V)^4$ is called a *maximum span flag quadruple*.

1.1.2 Discrete triangle The *discrete n–triangle* $\Theta_n \subseteq \mathbb{Z}^3_{\geqslant 0}$ is defined by

$$\Theta_n = \{(a, b, c) \in \mathbb{Z}^3_{\geqslant 0} \mid a + b + c = n\}.$$

See Figure 1. The *interior* $\mathrm{int}(\Theta_n) \subseteq \Theta_n$ of the discrete triangle is defined by

$$\mathrm{int}(\Theta_n) = \{(a, b, c) \in \Theta_n \mid a, b, c > 0\}.$$

An element $\nu \in \Theta_n$ is called a *vertex* of Θ_n. Put $\Gamma(\Theta_n) = \{(n, 0, 0), (0, n, 0), (0, 0, n)\} \subseteq \Theta_n$. An element $\nu \in \Gamma(\Theta_n)$ is called a *corner vertex* of Θ_n.

1.1.3 Fock–Goncharov triangle and edge invariants For a maximum span triple of flags $(E, F, G) \in \mathrm{Flag}(V)^3$, Fock and Goncharov assigned to each interior point $(a, b, c) \in \mathrm{int}(\Theta_n)$ a *triangle invariant* $\tau_{abc}(E, F, G) \in \mathbb{C} - \{0\}$, defined by the formula

$$\tau_{abc}(E, F, G) = \frac{e^{(a-1)} \wedge f^{(b+1)} \wedge g^{(c)}}{e^{(a+1)} \wedge f^{(b-1)} \wedge g^{(c)}} \frac{e^{(a)} \wedge f^{(b-1)} \wedge g^{(c+1)}}{e^{(a)} \wedge f^{(b+1)} \wedge g^{(c-1)}} \frac{e^{(a+1)} \wedge f^{(b)} \wedge g^{(c-1)}}{e^{(a-1)} \wedge f^{(b)} \wedge g^{(c+1)}} \in \mathbb{C} - \{0\}.$$

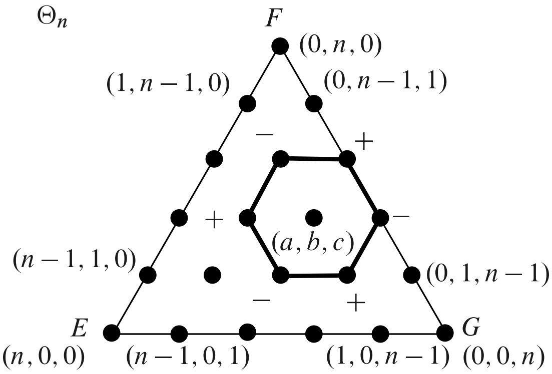


Figure 1: Discrete triangle, and triangle invariants for a generic flag triple.

Here $e^{(a')}$, $f^{(b')}$, and $g^{(c')}$ are choices of generators for the exterior powers $\Lambda^{a'}(E^{(a')}) \subseteq \Lambda^{a'}(V)$, $\Lambda^{b'}(F^{(b')}) \subseteq \Lambda^{b'}(V)$, and $\Lambda^{c'}(G^{(c')}) \subseteq \Lambda^{c'}(V)$, respectively. The maximum span property ensures that each wedge product $e^{(a')} \wedge f^{(b')} \wedge g^{(c')}$ is nonzero in $\Lambda^{a'+b'+c'}(V) = \Lambda^n(V) \cong \mathbb{C}$. Since there are the same number of terms in the numerator as the denominator, $\tau_{abc}(E,F,G)$ is independent of this choice of isomorphism $\Lambda^n(V) \cong \mathbb{C}$. Since each generator $e^{(a')}$, $f^{(b')}$, and $g^{(c')}$ appears exactly once in the numerator and denominator, $\tau_{abc}(E,F,G)$ is independent of the choices of these generators.

The six numerators and denominators appearing in the expression defining $\tau_{abc}(E,F,G)$ can be visualized as the vertices of a hexagon in Θ_n centered at (a,b,c); see Figure 1.

Similarly, for a maximum span quadruple of flags $(E,G,F,F') \in \mathrm{Flag}(V)^4$, Fock and Goncharov assigned to each integer $1 \leqslant j \leqslant n-1$ an *edge invariant* $\epsilon_j(E,G,F,F')$ by

$$\epsilon_j(E,G,F,F') = -\frac{e^{(j)} \wedge g^{(n-j-1)} \wedge f^{(1)}}{e^{(j)} \wedge g^{(n-j-1)} \wedge f'^{(1)}} \frac{e^{(j-1)} \wedge g^{(n-j)} \wedge f'^{(1)}}{e^{(j-1)} \wedge g^{(n-j)} \wedge f^{(1)}} \in \mathbb{C} - \{0\}.$$

The four numerators and denominators appearing in the expression defining $\epsilon_j(E,G,F,F')$ can be visualized as the vertices of a square, which crosses the "common edge" between two "adjacent" discrete triangles $\Theta_n(G,F,E)$ and $\Theta_n(E,F',G)$; see Figure 2.

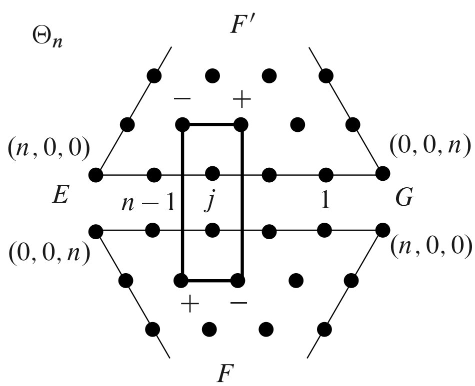


Figure 2: Edge invariants for a generic flag quadruple.

1.1.4 Action of PGL(V) on generic flag triples The action of the general linear group GL(V) on the vector space V induces an action of the projective linear group PGL(V) on the space Flag(V) of flags. The corresponding diagonal action of PGL(V) on Flag$(V)^n$ restricts to generic configurations of flags. By an elementary argument, for $n = 2$ this diagonal action on generic flag pairs (E, F) has a single orbit in Flag$(V)^2$.

Theorem 1.2 (Fock and Goncharov) *Two maximum span flag triples (E, F, G) and (E', F', G') have the same triangle invariants, namely $\tau_{abc}(E, F, G) = \tau_{abc}(E', F', G') \in \mathbb{C} - \{0\}$ for every $(a, b, c) \in \mathrm{int}(\Theta_n)$, if and only if there exists $\varphi \in \mathrm{PGL}(V)$ such that $(\varphi E, \varphi F, \varphi G) = (E', F', G') \in \mathrm{Flag}(V)^3$.*

Conversely, for each choice of nonzero complex numbers $x_{abc} \in \mathbb{C} - \{0\}$ assigned to the interior points $(a, b, c) \in \mathrm{int}(\Theta_n)$, there exists a maximum span flag triple (E, F, G) such that $\tau_{abc}(E, F, G) = x_{abc}$ for all (a, b, c).

Proof See [9, Section 9]. The proof uses the concept of snakes, due to Fock and Goncharov. For a sketch of the proof and some examples see [5, Section 2.19]. □

1.2 Snakes and projective bases

1.2.1 Snakes Snakes are combinatorial objects associated to the $(n-1)$–discrete triangle Θ_{n-1}; see Section 1.1.2. In contrast to Θ_n, we denote the coordinates of a vertex $\nu \in \Theta_{n-1}$ by $\nu = (\alpha, \beta, \gamma)$ corresponding to solutions $\alpha + \beta + \gamma = n - 1$ for $\alpha, \beta, \gamma \in \mathbb{Z}_{\geqslant 0}$.

Definition 1.3 A *snake-head* η is a fixed corner vertex of the $(n-1)$–discrete triangle

$$\eta \in \{(n-1, 0, 0), (0, n-1, 0), (0, 0, n-1)\} = \Gamma(\Theta_{n-1}) \subseteq \Theta_{n-1}.$$

Remark 1.4 In a moment, we will define a snake. The most general definition involves choosing a snake-head $\eta \in \Gamma(\Theta_{n-1})$. For simplicity, we define a snake only in the case $\eta = (n-1, 0, 0)$. The definition for other choices of snake-heads follows by triangular symmetry. We will usually take $\eta = (n-1, 0, 0)$ and will alert the reader if otherwise.

Definition 1.5 A *left n–snake* (for the snake-head $\eta = (n-1, 0, 0) \in \Gamma(\Theta_{n-1})$), or just *snake*, σ is an ordered list $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in (\Theta_{n-1})^n$ of n–many vertices $\sigma_k = (\alpha_k, \beta_k, \gamma_k)$ in the discrete triangle Θ_{n-1}, called *snake-vertices*, satisfying

$$\alpha_k = k - 1, \quad \beta_k \geqslant \beta_{k+1}, \quad \text{and} \quad \gamma_k \geqslant \gamma_{k+1} \quad \text{for } k = 1, 2, \dots, n.$$

See Figure 3. On the right-hand side, we show a snake $\sigma = (\sigma_k)_k$ in the case $n = 5$ (where we have taken some artistic license to assist the reader in locating the snake's head and tail; in Section 3, we will find it useful to split the snake in half down its length, as illustrated in Figure 15). On the left-hand side, we show how the snake-vertices $\sigma_k \in \Theta_{n-1}$ can be pictured as small upward-facing triangles Δ in the n–discrete triangle Θ_n.

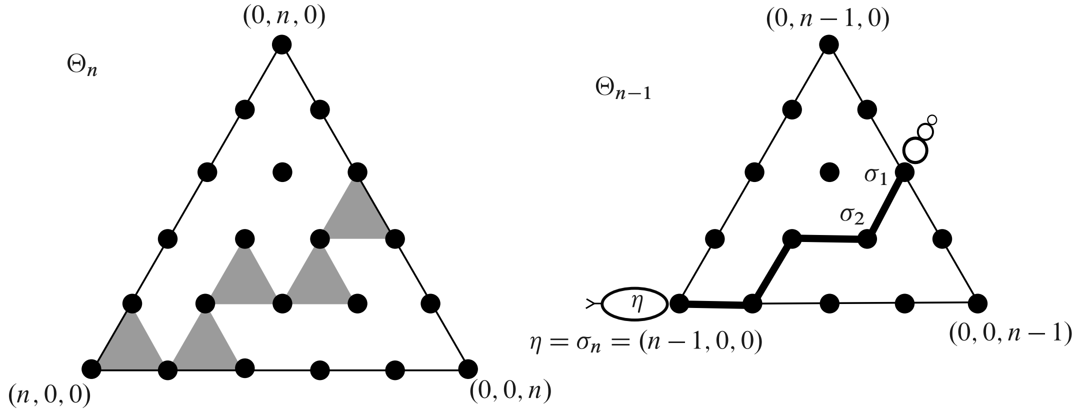


Figure 3: Snake.

1.2.2 Line decomposition of V^* associated to a generic triple of flags and a snake Let $V^* = \{\text{linear map } V \to \mathbb{C}\}$. For a subspace $W \subseteq V$, define $W^\perp = \{u \in V^* \mid u(w) = 0 \text{ for all } w \in W\}$. A *line* in a vector space V' is a 1–dimensional subspace.

Fix a maximum span triple $(E, F, G) \in \mathrm{Flag}(V)^3$. For any vertex $\nu = (\alpha, \beta, \gamma) \in \Theta_{n-1}$,

$$\dim((E^{(\alpha)} \oplus F^{(\beta)} \oplus G^{(\gamma)})^\perp) = 1$$

by the maximum span property, since $\alpha + \beta + \gamma = n - 1$. Consequently, the subspace

$$L_{(\alpha,\beta,\gamma)} := (E^{(\alpha)} \oplus F^{(\beta)} \oplus G^{(\gamma)})^\perp \subseteq V^*$$

is a line for all vertices $(\alpha, \beta, \gamma) \in \Theta_{n-1}$.

If in addition we are given a snake $\sigma = (\sigma_k)_k$, then we may consider the n–many lines

$$L_{\sigma_k} = L_{(\alpha_k,\beta_k,\gamma_k)} \subseteq V^* \quad \text{for } k = 1, \dots, n,$$

where $\sigma_k = (\alpha_k, \beta_k, \gamma_k) \in \Theta_{n-1}$. By genericity, we obtain a direct sum line decomposition

$$V^* = \bigoplus_{k=1}^{n} L_{\sigma_k}.$$

1.2.3 Projective basis of V^* associated to a generic triple of flags and a snake Given a generic flag triple (E, F, G) and a snake σ, Fock and Goncharov construct in addition a projective basis $[\mathcal{U}]$ of V^* adapted to the associated line decomposition. Here $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ is a linear basis of V^* such that $u_k \in L_{\sigma_k}$ for all k, and the *projective basis* $[\mathcal{U}]$ is the equivalence class of $\mathcal{U}$ under the relation $\{u_1, u_2, \dots, u_n\} \sim \{\lambda u_1, \lambda u_2, \dots, \lambda u_n\}$ for all $\lambda \neq 0$.

Put $\sigma_k = (\alpha_k, \beta_k, \gamma_k)$. We begin by choosing a covector u_n in the line $L_{\sigma_n} \subseteq V^*$, called a *normalization*. Having defined covectors $u_n, u_{n-1}, \dots, u_{k+1}$, we will define a covector

$$u_k \in L_{\sigma_k} = (E^{(\alpha_k)} \oplus F^{(\beta_k)} \oplus G^{(\gamma_k)})^\perp \subseteq V^*.$$

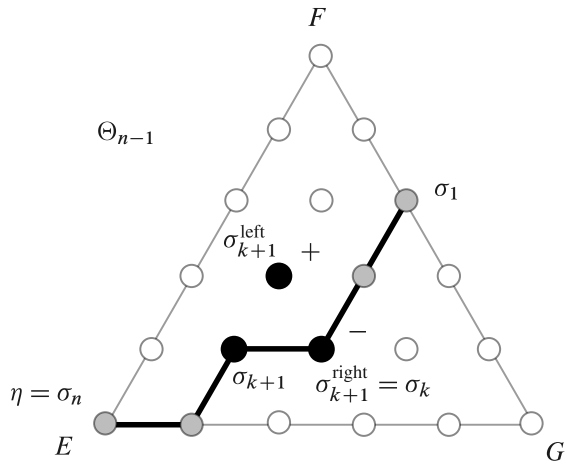


Figure 4: Three coplanar lines involved in the definition of a projective basis. For the meaning of the $+$ and $-$ signs see Definition 1.6.

By the definition of snakes, we see that given σ_{k+1} there are only two possibilities for σ_k, denoted by $\sigma_{k+1}^{\text{left}}$ and $\sigma_{k+1}^{\text{right}}$:

$$\begin{aligned}
\sigma_{k+1}^{\text{left}} &= (\alpha_{k+1}^{\text{left}}, \beta_{k+1}^{\text{left}}, \gamma_{k+1}^{\text{left}}) &&\text{for}\quad \alpha_{k+1}^{\text{left}} = k-1, \quad \beta_{k+1}^{\text{left}} = \beta_{k+1}+1, \quad \gamma_{k+1}^{\text{left}} = \gamma_{k+1},\\
\sigma_{k+1}^{\text{right}} &= (\alpha_{k+1}^{\text{right}}, \beta_{k+1}^{\text{right}}, \gamma_{k+1}^{\text{right}}) &&\text{for}\quad \alpha_{k+1}^{\text{right}} = k-1, \quad \beta_{k+1}^{\text{right}} = \beta_{k+1}, \quad \gamma_{k+1}^{\text{right}} = \gamma_{k+1}+1.
\end{aligned}$$

See Figure 4, in which $\sigma_k = \sigma_{k+1}^{\text{right}}$. Thus the lines $L_{\sigma_{k+1}^{\text{left}}}$ and $L_{\sigma_{k+1}^{\text{right}}}$ can be written

$$\begin{aligned}
L_{\sigma_{k+1}^{\text{left}}} &= (E^{(k-1)} \oplus F^{(\beta_{k+1}+1)} \oplus G^{(\gamma_{k+1})})^{\perp} \subseteq V^*,\\
L_{\sigma_{k+1}^{\text{right}}} &= (E^{(k-1)} \oplus F^{(\beta_{k+1})} \oplus G^{(\gamma_{k+1}+1)})^{\perp} \subseteq V^*.
\end{aligned}$$

It follows by the maximum span property that the three lines $L_{\sigma_{k+1}}$, $L_{\sigma_{k+1}^{\text{left}}}$, and $L_{\sigma_{k+1}^{\text{right}}}$ in V^* are distinct and coplanar. Specifically, they lie in the plane

$$(E^{(k-1)} \oplus F^{(\beta_{k+1})} \oplus G^{(\gamma_{k+1})})^{\perp} \subseteq V^*,$$

which is indeed 2–dimensional, since $(k-1)+\beta_{k+1}+\gamma_{k+1} = (n-1)-1$, as $\alpha_{k+1} = k$. Thus, if u_{k+1} is a nonzero covector in the line $L_{\sigma_{k+1}}$, then there exist unique nonzero covectors u_{k+1}^{left} and u_{k+1}^{right} in the lines $L_{\sigma_{k+1}^{\text{left}}}$ and $L_{\sigma_{k+1}^{\text{right}}}$, respectively, such that

$$u_{k+1} + u_{k+1}^{\text{left}} + u_{k+1}^{\text{right}} = 0 \in V^*.$$

Definition 1.6 Having chosen a normalization $u_n \in L_{\sigma_n} = L_{(n-1,0,0)}$ and having inductively defined $u_{k'} \in L_{\sigma_{k'}}$ for $k' = n, n-1, \dots, k+1$, define $u_k \in L_{\sigma_k}$ by

(1) $u_k = +u_{k+1}^{\text{left}} \in L_{\sigma_{k+1}^{\text{left}}}$ if $\sigma_k = \sigma_{k+1}^{\text{left}}$,

(2) $u_k = -u_{k+1}^{\text{right}} \in L_{\sigma_{k+1}^{\text{right}}}$ if $\sigma_k = \sigma_{k+1}^{\text{right}}$.

See Figure 4, which falls into (2). Note if the initial normalization u_n is replaced by λu_n for some scalar $\lambda \neq 0$, then u_k is replaced by λu_k for all $1 \leqslant k \leqslant n$. Thus this process produces a projective basis $[\mathcal{U}] = [\{u_1, u_2, \dots, u_n\}]$ of V^*, as desired. We call $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ the *normalized projective basis* for V^* depending on the normalization $u_n \in L_{\sigma_n}$.

1.3 Snake moves

1.3.1 Elementary matrices Let $\mathcal{A}$ be a commutative algebra with 1, such as $\mathcal{A} = \mathbb{C}$. Let $X^{1/n}, Z^{1/n} \in \mathcal{A}$, and put $X = (X^{1/n})^n$ and $Z = (Z^{1/n})^n$. Let $\mathrm{M}_n(\mathcal{A})$ (resp. $\mathrm{SL}_n(\mathcal{A})$) denote the ring of $n \times n$ matrices (resp. having determinant equal to 1) over $\mathcal{A}$ (see also Section 2.1.2).

For $k = 1, 2, \dots, n-1$ define the k^{th} *left-elementary matrix* $\boldsymbol{S}_k^{\mathrm{left}}(X) \in \mathrm{SL}_n(\mathcal{A})$ by

$$\boldsymbol{S}_k^{\mathrm{left}}(X) = X^{-(k-1)/n} \begin{pmatrix} X &&&&&&& \\ & \ddots &&&&&& \\ && X &&&&& \\ &&& 1 & 1 &&& \\ &&&& 1 &&& \\ &&&&& 1 && \\ &&&&&& \ddots & \\ &&&&&&& 1 \end{pmatrix} \in \mathrm{SL}_n(\mathcal{A}) \quad \text{where } X \text{ appears } k-1 \text{ times,}$$

and define the k^{th} *right-elementary matrix* $\boldsymbol{S}_k^{\mathrm{right}}(X) \in \mathrm{SL}_n(\mathcal{A})$ by

$$\boldsymbol{S}_k^{\mathrm{right}}(X) = X^{+(k-1)/n} \begin{pmatrix} 1 &&&&&&& \\ & \ddots &&&&&& \\ && 1 &&&&& \\ &&& 1 &&&& \\ &&& 1 & 1 &&& \\ &&&&& X^{-1} && \\ &&&&&& \ddots & \\ &&&&&&& X^{-1} \end{pmatrix} \in \mathrm{SL}_n(\mathcal{A}) \quad \text{where } X \text{ appears } k-1 \text{ times.}$$

Note that $\boldsymbol{S}_1^{\mathrm{left}}(X)$ and $\boldsymbol{S}_1^{\mathrm{right}}(X)$ do not, in fact, involve the variable X, and so we will denote these matrices simply by $\boldsymbol{S}_1^{\mathrm{left}}$ and $\boldsymbol{S}_1^{\mathrm{right}}$, respectively.

For $j = 1, 2, \dots, n-1$ define the j^{th} *edge-elementary matrix* $\boldsymbol{S}_j^{\mathrm{edge}}(Z) \in \mathrm{SL}_n(\mathcal{A})$ by

$$\boldsymbol{S}_j^{\mathrm{edge}}(Z) = Z^{-j/n} \begin{pmatrix} Z &&&&& \\ & \ddots &&&& \\ && Z &&& \\ &&& 1 && \\ &&&& \ddots & \\ &&&&& 1 \end{pmatrix} \in \mathrm{SL}_n(\mathcal{A}) \quad \text{where } Z \text{ appears } j \text{ times.}$$

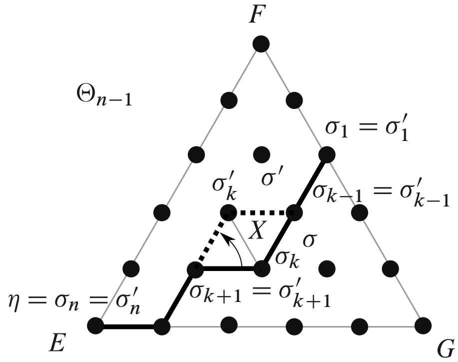


Figure 5: Diamond move.

1.3.2 Adjacent snake pairs

Definition 1.7 We say that an ordered pair (σ, σ') of snakes σ and σ' forms an *adjacent pair of snakes* if it satisfies either of the following conditions:

(1) For some $2 \leqslant k \leqslant n-1$,
  (a) $\sigma_j = \sigma'_j$ for $1 \leqslant j \leqslant k-1$ and $k+1 \leqslant j \leqslant n$,
  (b) $\sigma_k = \sigma_{k+1}^{\text{right}} (= \sigma_{k+1}^{\prime\text{right}})$ and $\sigma'_k = \sigma_{k+1}^{\text{left}} (= \sigma_{k+1}^{\prime\text{left}})$,
  in which case (σ, σ') is called an adjacent pair of *diamond-type*; see Figure 5.

(2) (a) $\sigma_j = \sigma'_j$ for $2 \leqslant j \leqslant n$,
  (b) $\sigma_1 = \sigma_2^{\text{right}} (= \sigma_2^{\prime\text{right}})$ and $\sigma'_1 = \sigma_2^{\text{left}} (= \sigma_2^{\prime\text{left}})$,
  in which case (σ, σ') is called an adjacent pair of *tail-type*; see Figure 6.

1.3.3 Diamond and tail moves Let (σ, σ') be an adjacent pair of snakes of diamond-type, as shown in Figure 5.

Consider the snake-vertices $\sigma_{k+1} (= \sigma'_{k+1})$, σ_k, σ'_k, and $\sigma_{k-1} (= \sigma'_{k-1})$. One checks that

$$\alpha_k = \alpha'_k = k-1, \quad \beta'_k = \beta_{k-1} = \beta_{k+1} + 1, \quad \text{and} \quad \gamma_k = \gamma_{k-1} = \gamma_{k+1} + 1.$$

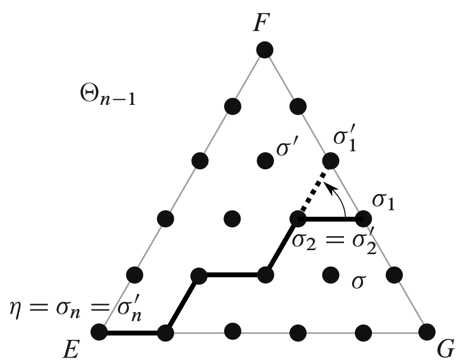


Figure 6: Tail move.

Taken together, these three coordinates form a vertex

$$(a,b,c) = (k-1, \beta_{k+1}+1, \gamma_{k+1}+1) \in \mathrm{int}(\Theta_n)$$

in the interior of the n–discrete triangle Θ_n (not Θ_{n-1}), since $(k-1)+(\beta_{k+1}+1)+(\gamma_{k+1}+1) = (\alpha_{k+1}+\beta_{k+1}+\gamma_{k+1})+1 = n$. The coordinates of this internal vertex (a,b,c) can also be thought of as delineating the boundary of a small downward-facing triangle ∇ in the discrete triangle Θ_{n-1}, whose three vertices are σ_k, σ'_k, and σ_{k-1} (Figure 5). Put $X_{abc} = \tau_{abc}(E,F,G) \in \mathbb{C}-\{0\}$, namely X_{abc} is the Fock–Goncharov triangle invariant (Section 1.1.3) associated to the generic flag triple (E,F,G) and the internal vertex $(a,b,c) \in \mathrm{int}(\Theta_n)$.

The proposition below is the main ingredient going into the proof of Theorem 1.2. First, we set our conventions for change of basis matrices for bases of V^*.

Given any basis $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ of V^*, and given a covector u in V^*, the *coordinate covector* $[u]_{\mathcal{U}}$ of the covector u with respect to the basis $\mathcal{U}$ is the unique row matrix $[u]_{\mathcal{U}} = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$ in $\mathrm{M}_{1,n}(\mathbb{C})$ such that $u = \sum_{i=1}^n y_i u_i$. If $\mathcal{U}' = \{u'_1, u'_2, \dots, u'_n\}$ is another basis for V^*, then the *change of basis matrix* $\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'}$ going from the basis $\mathcal{U}$ to the basis $\mathcal{U}'$ is the unique invertible matrix in $\mathrm{GL}_n(\mathbb{C}) \subseteq \mathrm{M}_n(\mathbb{C})$ satisfying

$$[u]_{\mathcal{U}} \boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} = [u]_{\mathcal{U}'} \in \mathrm{M}_{1,n}(\mathbb{C}) \quad \text{for } u \in V^*.$$

Change of basis matrices satisfy the property

$$\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}''} = \boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} \boldsymbol{B}_{\mathcal{U}'\to\mathcal{U}''} \in \mathrm{GL}_n(\mathbb{C}) \quad \text{for } \mathcal{U}, \mathcal{U}' \text{ and } \mathcal{U}'' \text{ bases for } V^*.$$

Proposition 1.8 (Fock and Goncharov) *Let (E,F,G) be a maximum span flag triple, (σ,σ') an adjacent pair of snakes, and $\mathcal{U}$ and $\mathcal{U}'$ the corresponding normalized projective bases of V^*, satisfying the compatibility condition $u_n = u'_n \in L_{\sigma_n} = L_\eta$.*

If (σ,σ') is of diamond-type, then the change of basis matrix $\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} \in \mathrm{GL}_n(\mathbb{C})$ is

$$\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} = X_{abc}^{+(k-1)/n} \boldsymbol{S}_k^{\mathrm{left}}(X_{abc}) \in \mathrm{GL}_n(\mathbb{C}) \qquad \text{(see Section 1.3.1).}$$

We say this case expresses a diamond move from the snake σ to the adjacent snake σ'.

If (σ,σ') is of tail-type, then the change of basis matrix $\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'}$ equals

$$\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} = \boldsymbol{S}_1^{\mathrm{left}} \in \mathrm{SL}_n(\mathbb{C}) \qquad \text{(see Section 1.3.1).}$$

We say this case expresses a tail move from the snake σ to the adjacent snake σ'.

Proof See [9, Section 9]. We also provide a proof in [5, Section 2.18]. □

1.3.4 Right snakes and right snake moves Our definition of a (left) snake in Section 1.2.1 took the snake-head $\eta = \sigma_n$ to be the n^{th} snake-vertex. There is another possibility, where $\eta = \sigma_1$.

Definition 1.9 A *right n–snake* σ (for the snake-head $\eta = (n-1,0,0) \in \Gamma(\Theta_{n-1})$) is an ordered list $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in (\Theta_{n-1})^n$ of n–many vertices $\sigma_k = (\alpha_k, \beta_k, \gamma_k)$, satisfying

$$\alpha_k = n-k, \quad \beta_k \geqslant \beta_{k-1}, \quad \text{and} \quad \gamma_k \geqslant \gamma_{k-1} \quad \text{for } k = 1,2,\dots,n.$$

Right snakes for other snake-heads $\eta \in \Gamma(\Theta_{n-1})$ are similarly defined by triangular symmetry.

To adjust for using right snakes, the definitions of Sections 1.2.3, 1.3.2, and 1.3.3 need to be modified. Given σ_{k-1}, there are two possibilities for σ_k:

$$\sigma_{k-1}^{\text{left}} = (\alpha_{k-1}^{\text{left}}, \beta_{k-1}^{\text{left}}, \gamma_{k-1}^{\text{left}}) \quad \text{for} \quad \alpha_{k-1}^{\text{left}} = n-k, \quad \beta_{k-1}^{\text{left}} = \beta_{k-1}+1, \quad \gamma_{k-1}^{\text{left}} = \gamma_{k-1},$$
$$\sigma_{k-1}^{\text{right}} = (\alpha_{k-1}^{\text{right}}, \beta_{k-1}^{\text{right}}, \gamma_{k-1}^{\text{right}}) \quad \text{for} \quad \alpha_{k-1}^{\text{right}} = n-k, \quad \beta_{k-1}^{\text{right}} = \beta_{k-1}, \quad \gamma_{k-1}^{\text{right}} = \gamma_{k-1}+1.$$

The algorithm defining the (ordered) projective basis $[\mathcal{U}] = [\{u_1, u_2, \dots, u_n\}]$ becomes

(1) $u_k = -u_{k-1}^{\text{left}} \in L_{\sigma_{k-1}^{\text{left}}}$ if $\sigma_k = \sigma_{k-1}^{\text{left}}$,

(2) $u_k = +u_{k-1}^{\text{right}} \in L_{\sigma_{k-1}^{\text{right}}}$ if $\sigma_k = \sigma_{k-1}^{\text{right}}$.

In particular, the algorithm starts by making a choice of normalization covector $u_1 \in L_{\sigma_1} = L_{(n-1,0,0)}$. Notice that, compared to the setting of left snakes (Definition 1.6 and Figure 4), the signs defining the projective basis have been swapped.

An ordered pair (σ, σ') of right snakes forms an adjacent pair if either:

(1) For some $2 \leqslant k \leqslant n-1$,
  - (a) $\sigma_j = \sigma'_j$ for $1 \leqslant j \leqslant k-1$ and $k+1 \leqslant j \leqslant n$,
  - (b) $\sigma_k = \sigma_{k-1}^{\text{left}} (= \sigma_{k-1}'^{\text{left}})$ and $\sigma'_k = \sigma_{k-1}^{\text{right}} (= \sigma_{k-1}'^{\text{right}})$,

  in which case (σ, σ') is called an adjacent pair of diamond-type.

(2) (a) $\sigma_j = \sigma'_j$ for $1 \leqslant j \leqslant n-1$,
  - (b) $\sigma_n = \sigma_{n-1}^{\text{left}} (= \sigma_{n-1}'^{\text{left}})$ and $\sigma'_n = \sigma_{n-1}^{\text{right}} (= \sigma_{n-1}'^{\text{right}})$,

  in which case (σ, σ') is called an adjacent pair of tail-type.

Given an adjacent pair (σ, σ') of right snakes of diamond-type, there is naturally associated a vertex $(a,b,c) \in \Theta_n$ to which is assigned a Fock–Goncharov triangle invariant X_{abc}.

Proposition 1.10 (Fock and Goncharov) *Let (E, F, G) be a maximum span triple, (σ, σ') an adjacent pair of right snakes, and $\mathcal{U}$ and $\mathcal{U}'$ the corresponding normalized projective bases of V^*, satisfying the compatibility condition $u_1 = u'_1 \in L_{\sigma_1} = L_\eta$.*

If (σ, σ') is of diamond-type, then the change of basis matrix $\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} \in \mathrm{GL}_n(\mathbb{C})$ equals

$$\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} = X_{abc}^{-(k-1)/n} \boldsymbol{S}_k^{\text{right}}(X_{abc}) \in \mathrm{GL}_n(\mathbb{C}) \qquad \text{(see Section 1.3.1).}$$

If (σ, σ') is of tail-type, then the change of basis matrix $\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'}$ equals

$$\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} = \boldsymbol{S}_1^{\text{right}} \in \mathrm{SL}_n(\mathbb{C}) \qquad \text{(see Section 1.3.1).}$$

Proof See [9, Section 9]. This is similar to the proof of Proposition 1.8. □

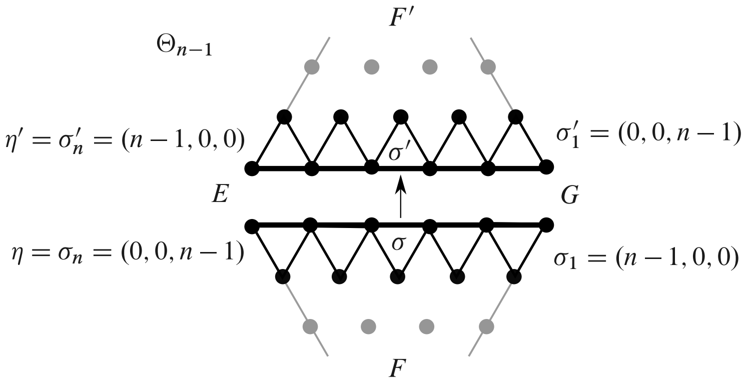


Figure 7: Edge move.

Remark 1.11 From now on, "snake" means "left snake", as in Definition 1.5, and we will say explicitly if we are using right snakes.

1.3.5 Snake moves for edges

Caution 1.12 In this subsubsection, we will consider snake-heads in the set of corner vertices $\Gamma(\Theta_{n-1})$ other than $(n-1,0,0)$, specifically η below; see Remark 1.4.

Let (E,G,F,F') be a maximum span flag quadruple; see Section 1.1.1. By Section 1.1.3, for each $j=1,\dots,n-1$ we may consider the Fock–Goncharov edge invariant $Z_j = \epsilon_j(E,G,F,F') \in \mathbb{C}-\{0\}$ associated to the quadruple (E,G,F,F').

Consider two copies of the discrete triangle; see Figure 7. The bottom triangle $\Theta_{n-1}(G,F,E)$ has a maximum span flag triple (G,F,E) assigned to the corner vertices $\Gamma(\Theta_{n-1})$, and the top triangle $\Theta_{n-1}(E,F',G)$ has assigned to $\Gamma(\Theta_{n-1})$ a maximum span flag triple (E,F',G).

Define (left) snakes σ and σ' in $\Theta_{n-1}(G,F,E)$ and $\Theta_{n-1}(E,F',G)$, respectively, as follows:

$$\sigma_k = (n-k,0,k-1) \in \Theta_{n-1}(G,F,E) \quad \text{for } k=1,\dots,n,$$
$$\sigma'_k = (k-1,0,n-k) \in \Theta_{n-1}(E,F',G) \quad \text{for } k=1,\dots,n.$$

The line decompositions associated to the snakes σ and σ' and their respective triples of flags are the same:

$$L_{\sigma_k} = L_{\sigma'_k} = (E^{(k-1)} \oplus G^{(n-k)})^{\perp} \subseteq V^* \quad \text{for } k=1,\dots,n.$$

Let $\mathcal{U}$ and $\mathcal{U}'$ be the associated normalized projective bases, where the normalizations are chosen in a compatible way, that is, such that $u_n = u'_n$ in $L_{\sigma_n} = L_{\sigma'_n}$.

Proposition 1.13 (Fock and Goncharov) *The change of basis matrix expressing the snake edge move* $\sigma \to \sigma'$ *is*

$$\boldsymbol{B}_{\mathcal{U}\to\mathcal{U}'} = \prod_{j=1}^{n-1} Z_j^{+j/n} \boldsymbol{S}_j^{\mathrm{edge}}(Z_j) \in \mathrm{GL}_n(\mathbb{C}) \qquad \text{(see Section 1.3.1).}$$

Proof See [9, Section 9]. This is similar to the proof of Proposition 1.8; see also [5, Section 2.22]. □

1.4 Classical left, right, and edge matrices

Caution 1.14 We consider snake-heads in the set of corner vertices $\{(n-1,0,0),(0,n-1,0),(0,0,n-1)\}$ other than $(n-1,0,0)$; see Remark 1.4.

We will also consider both (left) snakes and right snakes; see Remark 1.11.

We begin the process of algebraizing the geometry discussed throughout this first section.

1.4.1 Snake sequences

Left setting Define a snake-head $\eta \in \Gamma(\Theta_{n-1})$ and two (left) snakes σ^{bot} and σ^{top}, called the bottom and top snakes, respectively, by

$$\eta = (n-1,0,0), \quad \sigma_k^{\text{bot}} = (k-1,0,n-k), \quad \text{and} \quad \sigma_k^{\text{top}} = (k-1,n-k,0) \quad \text{for } k = 1,\dots,n.$$

Right setting Define η and right snakes σ^{bot} and σ^{top} by

$$\eta = (0,0,n-1), \quad \sigma_k^{\text{bot}} = (k-1,0,n-k), \quad \text{and} \quad \sigma_k^{\text{top}} = (0,k-1,n-k) \quad \text{for } k = 1,\dots,n.$$

In either left or right setting, consider a sequence $\sigma^{\text{bot}} = \sigma^1, \sigma^2, \dots, \sigma^{N-1}, \sigma^N = \sigma^{\text{top}}$ of snakes having the same snake-head η as σ^{bot} and σ^{top}, such that (σ^l, σ^{l+1}) is an adjacent pair; see Figure 8. Note that this sequence of snakes is not in general unique. For the N–many projective bases $[\mathfrak{U}^l] = [\{u_1^l, u_2^l, \dots, u_n^l\}]$ associated to the snakes σ^l, choose a common normalization $u_n^l := u_n \in L_\eta$ (resp. $u_1^l := u_1 \in L_\eta$), where the same u_n (resp. u_1) is used for all l, when working in the left (resp. right) setting. Then, the change of basis matrix $\boldsymbol{B}_{\mathfrak{U}^{\text{bot}} \to \mathfrak{U}^{\text{top}}}$ can be decomposed as (see Section 1.3.3)

$$(*) \qquad \boldsymbol{B}_{\mathfrak{U}^{\text{bot}} \to \mathfrak{U}^{\text{top}}} = \boldsymbol{B}_{\mathfrak{U}^1 \to \mathfrak{U}^2} \boldsymbol{B}_{\mathfrak{U}^2 \to \mathfrak{U}^3} \cdots \boldsymbol{B}_{\mathfrak{U}^{N-1} \to \mathfrak{U}^N} \in \mathrm{GL}_n(\mathbb{C}).$$

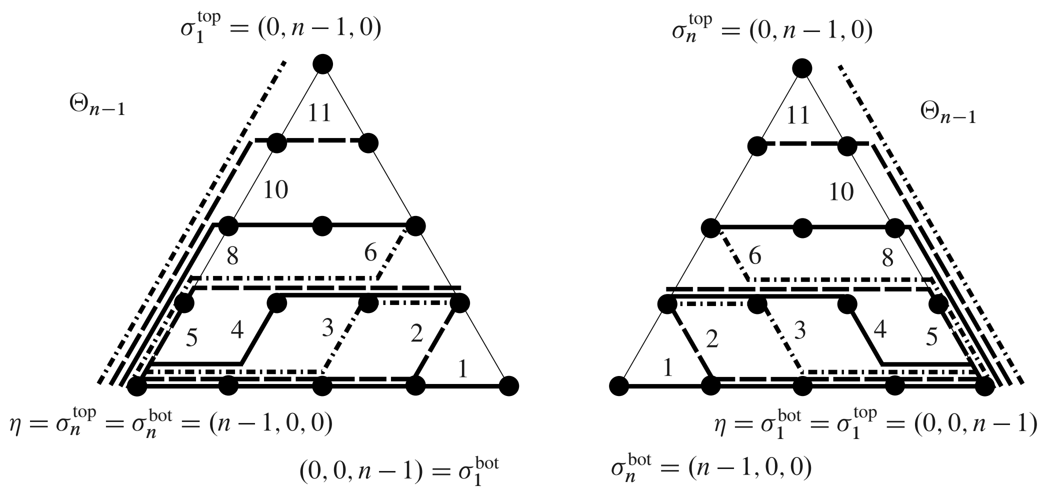


Figure 8: Classical snake sweep for $n = 5$. The preferred choices for the left and right snake sequences are on the left and right, respectively.

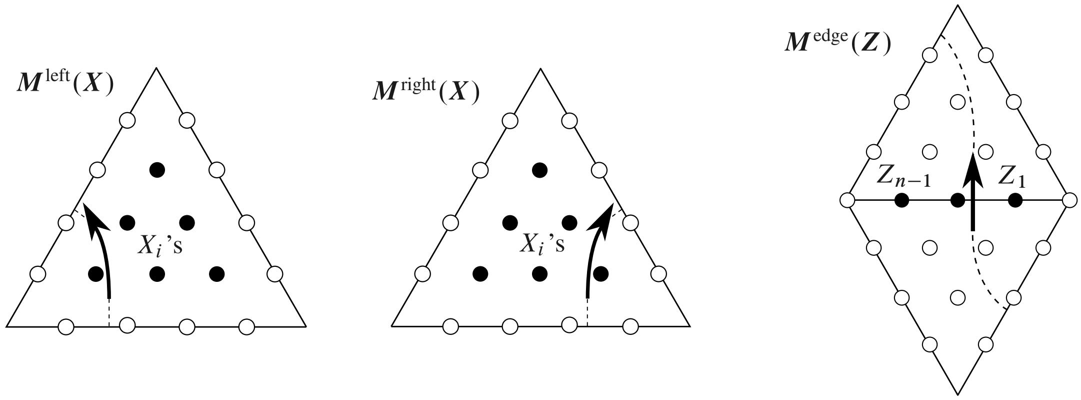


Figure 9: Classical matrices (viewed from the Θ_n–perspective): from left to right, the left, right, and edge matrices.

Here the matrices $\boldsymbol{B}_{\mathfrak{U}^l \to \mathfrak{U}^{l+1}}$ are computed as in Proposition 1.8 (resp. Proposition 1.10) in the left (resp. right) setting, and in particular are completely determined by the Fock–Goncharov triangle invariants $X_{abc} \in \mathbb{C} - \{0\}$ associated to the internal vertices $(a, b, c) \in \operatorname{int}(\Theta_n)$ of the n–discrete triangle.

Note that the matrix $\boldsymbol{B}_{\mathfrak{U}^{\text{bot}} \to \mathfrak{U}^{\text{top}}}$ is, by definition, independent of the choice of snake sequence $(\sigma^l)_l$. For concreteness, throughout we make a preferred choice of such sequence, depending on whether we are in the left or right setting; see Figure 8.

1.4.2 Algebraization Let $\mathcal{A}$ be a commutative algebra (Section 1.3.1). For $i = 1, 2, \dots, \frac{1}{2}(n-1)(n-2)$, let $X_i^{1/n} \in \mathcal{A}$ and put $X_i = (X_i^{1/n})^n$. For $j = 1, 2, \dots, n-1$, let $Z_j^{1/n} \in \mathcal{A}$ and put $Z_j = (Z_j^{1/n})^n$. Note, $\frac{1}{2}(n-1)(n-2)$ is the number of elements $(a, b, c) \in \operatorname{int}(\Theta_n)$, which we arbitrarily enumerate $1, 2, \dots, \frac{1}{2}(n-1)(n-2)$; see Figure 9, left and center. And note that $n-1$ is the number of noncorner vertices of Θ_n lying on a single edge, which we enumerate $1, 2, \dots, n-1$ as shown in Figure 9, right. Let $\boldsymbol{X} = (X_i)_i$ and $\boldsymbol{Z} = (Z_j)_j$ be the corresponding tuples of these elements of $\mathcal{A}$.

As a notational convention, given a family $\boldsymbol{M}_l \in \mathrm{M}_n(\mathcal{A})$ of $n \times n$ matrices, put

$$\prod_{l=m}^{p} \boldsymbol{M}_l = \boldsymbol{M}_m \boldsymbol{M}_{m+1} \cdots \boldsymbol{M}_p, \qquad \prod_{l=p+1}^{m} \boldsymbol{M}_l = 1 \quad \text{for } m \leqslant p,$$

$$\coprod_{l=p}^{m} \boldsymbol{M}_l = \boldsymbol{M}_p \boldsymbol{M}_{p-1} \cdots \boldsymbol{M}_m, \qquad \coprod_{l=m-1}^{p} \boldsymbol{M}_l = 1 \quad \text{for } m \leqslant p.$$

Definition 1.15 The *left matrix* $\boldsymbol{M}^{\text{left}}(\boldsymbol{X})$ in $\mathrm{SL}_n(\mathcal{A})$ is defined by

$$\boldsymbol{M}^{\text{left}}(\boldsymbol{X}) = \coprod_{k=n-1}^{1} \left(\boldsymbol{S}_1^{\text{left}} \prod_{l=2}^{k} \boldsymbol{S}_l^{\text{left}}(X_{(l-1)(n-k)(k-l+1)}) \right) \in \mathrm{SL}_n(\mathcal{A}),$$

where the matrix $\boldsymbol{S}_l^{\text{left}}(X_{abc})$ is the l^{th} left-elementary matrix; see Section 1.3.1.

Similarly, the *right matrix* $\boldsymbol{M}^{\text{right}}(\boldsymbol{X})$ in $\mathrm{SL}_n(\mathcal{A})$ is defined by

$$\boldsymbol{M}^{\text{right}}(\boldsymbol{X}) = \coprod_{k=n-1}^{1} \left(\boldsymbol{S}_1^{\text{right}} \prod_{l=2}^{k} \boldsymbol{S}_l^{\text{right}}(X_{(k-l+1)(n-k)(l-1)}) \right) \in \mathrm{SL}_n(\mathcal{A}),$$

where the matrix $\boldsymbol{S}_l^{\text{right}}(X_{abc})$ is the l^{th} right-elementary matrix; see Section 1.3.1.

Lastly, the *edge matrix* $\boldsymbol{M}^{\text{edge}}(\boldsymbol{Z})$ in $\mathrm{SL}_n(\mathcal{A})$ is defined by

$$\boldsymbol{M}^{\text{edge}}(\boldsymbol{Z}) = \prod_{l=1}^{n-1} \boldsymbol{S}_l^{\text{edge}}(Z_l) \in \mathrm{SL}_n(\mathcal{A}),$$

where the matrix $\boldsymbol{S}_l^{\text{edge}}(Z_l)$ is the l^{th} edge-elementary matrix; see Section 1.3.1. See Figure 9.

Remark 1.16 In the case where $\mathcal{A} = \mathbb{C}$ and the $X_i = \tau_{abc}(E, F, G)$ and $Z_j = \epsilon_j(E, G, F, F')$ in $\mathbb{C} - \{0\}$ are the triangle and edge invariants (as in Sections 1.3.3, 1.3.4, and 1.3.5), then the left and right matrices $\boldsymbol{M}^{\text{left}}(\boldsymbol{X})$ and $\boldsymbol{M}^{\text{right}}(\boldsymbol{X})$ are the normalized change of basis matrix $\boldsymbol{B}_{\mathcal{U}^{\text{bot}} \to \mathcal{U}^{\text{top}}} / \operatorname{Det}^{1/n}$ (see $(*)$) in the left and right settings, respectively, normalized to have determinant 1, and decomposed in terms of our preferred snake sequence (Figure 8). Also, the edge matrix $\boldsymbol{M}^{\text{edge}}(\boldsymbol{Z})$ is the normalization $\boldsymbol{B}_{\mathcal{U} \to \mathcal{U}'} / \operatorname{Det}^{1/n}$ of the change of basis matrix from Proposition 1.13. Note, these normalizations require choosing n–roots of the invariants X_i and Z_j.

2 Quantum matrices

Although we will not use explicitly the geometric results of the previous section, those results motivate the algebraic objects that are our main focus.

Throughout, let $q \in \mathbb{C} - \{0\}$ and $\omega = q^{1/n^2}$ be a n^2–root of q. Technically, also choose $\omega^{1/2}$.

2.1 Quantum tori, matrix algebras, and the Weyl quantum ordering

2.1.1 Quantum tori Let $\boldsymbol{P}$ (for "Poisson") be an integer $N \times N$ antisymmetric matrix.

Definition 2.1 The *quantum torus* (*with n–roots*) $\mathcal{T}^{\omega}(\boldsymbol{P})$ associated to $\boldsymbol{P}$ is the quotient of the free algebra $\mathbb{C}\{X_1^{1/n}, X_1^{-1/n}, \dots, X_N^{1/n}, X_N^{-1/n}\}$ in the indeterminates $X_i^{\pm 1/n}$ by the two-sided ideal generated by the relations

$$X_i^{1/n} X_j^{1/n} = \omega^{\boldsymbol{P}_{ij}} X_j^{1/n} X_i^{1/n} \quad \text{and} \quad X_i^{1/n} X_i^{-1/n} = X_i^{-1/n} X_i^{1/n} = 1.$$

Put $X_i^{\pm 1} = (X_i^{\pm 1/n})^n$. We refer to the $X_i^{\pm 1/n}$ as *generators*, and the X_i as *quantum coordinates*, or just *coordinates*. Define the subset of fractions

$$\mathbb{Z}/n = \left\{ \frac{m}{n} \;\middle|\; m \in \mathbb{Z} \right\} \subseteq \mathbb{Q}.$$

Written in terms of the coordinates X_i and the fractions $r \in \mathbb{Z}/n$, we have the relations

$$X_i^{r_i} X_j^{r_j} = q^{\boldsymbol{P}_{ij} r_i r_j} X_j^{r_j} X_i^{r_i} \quad \text{for } r_i, r_j \in \mathbb{Z}/n.$$

2.1.2 Matrix algebras

Definition 2.2 The *matrix algebra* $\mathrm{M}_n(\mathcal{T})$ with coefficients in a possibly noncommutative algebra $\mathcal{T}$ is the vector space of $n \times n$ matrices, equipped with the usual multiplicative structure. Namely, the product $\boldsymbol{MN}$ of two matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ is defined entrywise by

$$(\boldsymbol{MN})_{ij} = \sum_{k=1}^{n} \boldsymbol{M}_{ik}\boldsymbol{N}_{kj} \in \mathcal{T} \quad \text{for } 1 \leqslant i, j \leqslant n.$$

Here we use the usual convention that the entry $\boldsymbol{M}_{ij}$ of a matrix $\boldsymbol{M}$ is the entry in the i^{th} row and j^{th} column. Note that the order of $\boldsymbol{M}_{ik}$ and $\boldsymbol{N}_{kj}$ in the above equation matters since these elements might not commute in $\mathcal{T}$.

2.1.3 Weyl quantum ordering If $\mathcal{T}$ is a quantum torus, then there is a linear map

$$[-]\colon \mathbb{C}\{X_1^{1/n}, X_1^{-1/n}, \dots, X_N^{1/n}, X_N^{-1/n}\} \to \mathcal{T}$$

from the free algebra to $\mathcal{T}$, called the *Weyl quantum ordering*, defined by the property that a word $X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}$ for $r_a \in \mathbb{Z}/n$ (note that i_a may equal i_b if $a \neq b$) is mapped to

$$[X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}] = (q^{-\frac{1}{2}\sum_{1 \leqslant a < b \leqslant k} \boldsymbol{P}_{i_a i_b} r_a r_b}) X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k},$$

where on the right-hand side we implicitly mean the equivalence class in $\mathcal{T}$. Also, the empty word is mapped to 1. Note that the Weyl ordering $[-]$ depends on the choice of $\omega^{1/2}$; see the beginning of Section 2.

The Weyl ordering is specially designed to satisfy the symmetry

$$[X_{i_1}^{r_1} X_{i_2}^{r_2} \cdots X_{i_k}^{r_k}] = [X_{i_{\sigma(1)}}^{r_{\sigma(1)}} X_{i_{\sigma(2)}}^{r_{\sigma(2)}} \cdots X_{i_{\sigma(k)}}^{r_{\sigma(k)}}]$$

for every permutation σ of $\{1, \dots, k\}$; see [1]. Also, $[X_i^{1/n} X_i^{-1/n}] = 1$. Consequently, a linear map

$$[-]\colon \mathbb{C}[X_1^{\pm 1/n}, \dots, X_N^{\pm 1/n}] \to \mathcal{T}$$

is induced from the commutative Laurent polynomial algebra to $\mathcal{T}$. This determines a linear map of matrix algebras

$$[-]\colon \mathrm{M}_n(\mathbb{C}[X_1^{\pm 1/n}, \dots, X_N^{\pm 1/n}]) \to \mathrm{M}_n(\mathcal{T}) \quad \text{given by } [\boldsymbol{M}]_{ij} = [\boldsymbol{M}_{ij}] \text{ in } \mathcal{T}.$$

2.2 Fock–Goncharov quantum torus for a triangle

Let $\Gamma(\Theta_n)$ denote the set of corner vertices $\Gamma(\Theta_n) = \{(n,0,0), (0,n,0), (0,0,n)\}$ of the discrete triangle Θ_n; see Section 1.1.2.

Define a function

$$\boldsymbol{P}\colon (\Theta_n - \Gamma(\Theta_n)) \times (\Theta_n - \Gamma(\Theta_n)) \to \{-2, -1, 0, 1, 2\}$$

using the *quiver* with vertex set $\Theta_n - \Gamma(\Theta_n)$ illustrated in Figure 10. The function $\boldsymbol{P}$ is defined by sending the ordered tuple (ν_1, ν_2) of vertices of $\Theta_n - \Gamma(\Theta_n)$ to 2 (resp. -2) if there is a solid arrow pointing from ν_1 to ν_2 (resp. ν_2 to ν_1), to 1 (resp. -1) if there is a dotted arrow pointing from ν_1 to ν_2 (resp. ν_2 to ν_1),

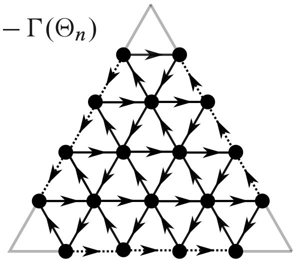


Figure 10: Quiver defining the Fock–Goncharov quantum torus.

and to 0 if there is no arrow connecting ν_1 and ν_2. Note that all of the small downward-facing triangles are oriented clockwise, and all of the small upward-facing triangles are oriented counterclockwise. By labeling the vertices of $\Theta_n - \Gamma(\Theta_n)$ by their coordinates (a,b,c) we may think of the function $\boldsymbol{P}$ as an $N \times N$ antisymmetric matrix $\boldsymbol{P} = (\boldsymbol{P}_{abc,a'b'c'})$, called the *Poisson matrix* associated to the quiver. Here $N = 3(n-1) + \frac{1}{2}(n-1)(n-2)$; compare with Section 1.4.2.

Definition 2.3 Define the *Fock–Goncharov quantum torus*

$$\mathcal{T}_n^\omega = \mathbb{C}[X_1^{\pm 1/n}, X_2^{\pm 1/n}, \dots, X_N^{\pm 1/n}]^\omega$$

associated to the discrete n–triangle Θ_n to be the quantum torus $\mathcal{T}^\omega(\boldsymbol{P})$ defined by the $N \times N$ Poisson matrix $\boldsymbol{P}$, with generators $X_i^{\pm 1/n} = X_{abc}^{\pm 1/n}$ for all $(a,b,c) \in \Theta_n - \Gamma(\Theta_n)$. Note that when $q = \omega = 1$ this recovers the classical Laurent polynomial algebra $\mathcal{T}_n^1 = \mathbb{C}[X_1^{\pm 1/n}, X_2^{\pm 1/n}, \dots, X_N^{\pm 1/n}]$.

As a notational convention, for $j = 1, 2, \dots, n-1$ we write $Z_j^{\pm 1/n}$ (resp. $Z_j'^{\pm 1/n}$ and $Z_j''^{\pm 1/n}$) in place of $X_{j0(n-j)}^{\pm 1/n}$ (resp. $X_{j(n-j)0}^{\pm 1/n}$ and $X_{0j(n-j)}^{\pm 1/n}$); see Figure 11. So, *triangle-coordinates* will be denoted by $X_i = X_{abc}$ for $(a,b,c) \in \operatorname{int}(\Theta_n)$ while *edge-coordinates* will be denoted by Z_j, Z_j', and Z_j''.

2.3 Quantum left and right matrices

2.3.1 Weyl quantum ordering for the Fock–Goncharov quantum torus Let $\mathcal{T} = \mathcal{T}_n^\omega$ be the Fock–Goncharov quantum torus (Section 2.2). Then the Weyl ordering $[-]$ of Section 2.1.3 gives a map

$$[-]\colon \mathrm{M}_n(\mathcal{T}_n^1) \to \mathrm{M}_n(\mathcal{T}_n^\omega),$$

where we have used the identification $\mathcal{T}_n^1 = \mathbb{C}[X_1^{\pm 1/n}, X_2^{\pm 1/n}, \dots, X_N^{\pm 1/n}]$ discussed in Section 2.2.

2.3.2 Quantum left and right matrices For a commutative algebra $\mathcal{A}$, in Section 1.4.2 we defined the classical matrices $\boldsymbol{M}^{\text{left}}(\boldsymbol{X})$, $\boldsymbol{M}^{\text{right}}(\boldsymbol{X})$, and $\boldsymbol{M}^{\text{edge}}(\boldsymbol{Z})$ in $\mathrm{SL}_n(\mathcal{A})$. If $\mathcal{A} = \mathbb{C}[X_1^{\pm 1/n}, \dots, X_N^{\pm 1/n}] = \mathcal{T}_n^1$, we now use these matrices to define the primary objects of study.

Definition 2.4 Put vectors $\boldsymbol{X} = (X_i)$, $\boldsymbol{Z} = (Z_j)$, $\boldsymbol{Z}' = (Z_j')$, $\boldsymbol{Z}'' = (Z_j'')$ as in Figure 11. We define the *quantum left matrix* $\boldsymbol{L}^\omega$ in $\mathrm{M}_n(\mathcal{T}_n^\omega)$ by the formula

$$\boldsymbol{L}^\omega = \boldsymbol{L}^\omega(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{Z}') = [\boldsymbol{M}^{\text{edge}}(\boldsymbol{Z}) \boldsymbol{M}^{\text{left}}(\boldsymbol{X}) \boldsymbol{M}^{\text{edge}}(\boldsymbol{Z}')] \in \mathrm{M}_n(\mathcal{T}_n^\omega),$$

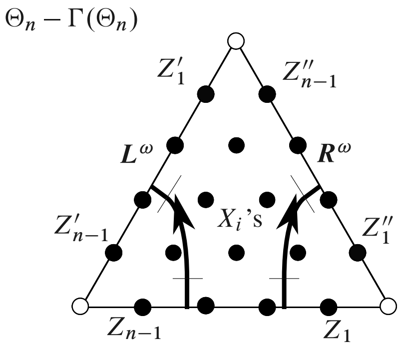


Figure 11: Quantum left and right matrices (compare with Figure 9).

where we have applied the Weyl quantum ordering $[-]$ discussed in Section 2.3.1 to the product $\boldsymbol{M}^{\mathrm{edge}}(\boldsymbol{Z})\boldsymbol{M}^{\mathrm{left}}(\boldsymbol{X})\boldsymbol{M}^{\mathrm{edge}}(\boldsymbol{Z}')$ of classical matrices in $\mathrm{M}_n(\mathcal{T}^1_n)$. In other words, we apply the Weyl ordering to each entry of the classical matrix.

Similarly, as in Figure 11, we define the *quantum right matrix* $\boldsymbol{R}^\omega$ in $\mathrm{M}_n(\mathcal{T}^\omega_n)$ by

$$\boldsymbol{R}^\omega = \boldsymbol{R}^\omega(\boldsymbol{Z},\boldsymbol{X},\boldsymbol{Z}'') = [\boldsymbol{M}^{\mathrm{edge}}(\boldsymbol{Z})\boldsymbol{M}^{\mathrm{right}}(\boldsymbol{X})\boldsymbol{M}^{\mathrm{edge}}(\boldsymbol{Z}'')] \in \mathrm{M}_n(\mathcal{T}^\omega_n).$$

2.4 Main result

2.4.1 Quantum SL_n and its points Let $\mathcal{T}$ be a possibly noncommutative algebra.

Definition 2.5 We say that a 2×2 matrix $\boldsymbol{M} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $\mathrm{M}_2(\mathcal{T})$ is a $\mathcal{T}$*–point of the quantum matrix algebra* M^q_2, denoted by $\boldsymbol{M} \in \mathrm{M}^q_2(\mathcal{T}) \subseteq \mathrm{M}_2(\mathcal{T})$, if

$(**)$ $ba = qab,\quad dc = qcd,\quad ca = qac,\quad db = qbd,\quad bc = cb,\quad da - ad = (q-q^{-1})bc\quad$ in $\mathcal{T}$.

We say that a matrix $\boldsymbol{M} \in \mathrm{M}_2(\mathcal{T})$ is a $\mathcal{T}$*–point of the quantum special linear group* SL^q_2, denoted by $\boldsymbol{M} \in \mathrm{SL}^q_2(\mathcal{T}) \subseteq \mathrm{M}^q_2(\mathcal{T}) \subseteq \mathrm{M}_2(\mathcal{T})$, if $\boldsymbol{M} \in \mathrm{M}^q_2(\mathcal{T})$ and the *quantum determinant*

$$\mathrm{Det}^q(\boldsymbol{M}) = ad - q^{-1}bc = 1 \in \mathcal{T}.$$

These notions are also defined for $n \times n$ matrices, as follows:

Definition 2.6 A matrix $\boldsymbol{M} \in \mathrm{M}_n(\mathcal{T})$ is a $\mathcal{T}$*–point of the quantum matrix algebra* M^q_n, denoted by $\boldsymbol{M} \in \mathrm{M}^q_n(\mathcal{T}) \subseteq \mathrm{M}_n(\mathcal{T})$, if every 2×2 submatrix of $\boldsymbol{M}$ is a $\mathcal{T}$–point of M^q_2. That is,

$$\boldsymbol{M}_{im}\boldsymbol{M}_{ik} = q\boldsymbol{M}_{ik}\boldsymbol{M}_{im},\quad \boldsymbol{M}_{jm}\boldsymbol{M}_{im} = q\boldsymbol{M}_{im}\boldsymbol{M}_{jm},\quad \boldsymbol{M}_{im}\boldsymbol{M}_{jk} = \boldsymbol{M}_{jk}\boldsymbol{M}_{im},$$
$$\boldsymbol{M}_{jm}\boldsymbol{M}_{ik} - \boldsymbol{M}_{ik}\boldsymbol{M}_{jm} = (q-q^{-1})\boldsymbol{M}_{im}\boldsymbol{M}_{jk},$$

for all $i<j$ and $k<m$, where $1 \leqslant i,j,k,m \leqslant n$.

The *quantum determinant* $\mathrm{Det}^q(\boldsymbol{M}) \in \mathcal{T}$ of a matrix $\boldsymbol{M} \in \mathrm{M}_n(\mathcal{T})$ is

$$\mathrm{Det}^q(\boldsymbol{M}) = \sum_{\sigma \in \mathfrak{S}_n} (-q^{-1})^{l(\sigma)} \boldsymbol{M}_{1\sigma(1)}\boldsymbol{M}_{2\sigma(2)}\cdots\boldsymbol{M}_{n\sigma(n)},$$

where the length $l(\sigma)$ of the permutation σ is the minimum number of factors appearing in a decomposition of σ as a product of adjacent transpositions $(i, i+1)$; see, for example, [2, Chapter I.2].

A matrix $\boldsymbol{M} \in \mathrm{M}_n(\mathcal{T})$ is a $\mathcal{T}$–*point of the quantum special linear group* SL_n^q, denoted by $\boldsymbol{M} \in \mathrm{SL}_n^q(\mathcal{T}) \subseteq \mathrm{M}_n^q(\mathcal{T}) \subseteq \mathrm{M}_n(\mathcal{T})$, if both $\boldsymbol{M} \in \mathrm{M}_n^q(\mathcal{T})$ and $\mathrm{Det}^q(\boldsymbol{M}) = 1$.

Remark 2.7 (1) It follows from the definitions that if a $\mathcal{T}$–point $\boldsymbol{M} \in \mathrm{M}_n^q(\mathcal{T}) \subseteq \mathrm{M}_n(\mathcal{T})$ is a triangular matrix, then the diagonal entries $\boldsymbol{M}_{ii} \in \mathcal{T}$ commute and $\mathrm{Det}^q(\boldsymbol{M}) = \prod_i \boldsymbol{M}_{ii} \in \mathcal{T}$.

(2) The subsets $\mathrm{M}_n^q(\mathcal{T}) \subseteq \mathrm{M}_n(\mathcal{T})$ and $\mathrm{SL}_n^q(\mathcal{T}) \subseteq \mathrm{M}_n^q(\mathcal{T})$ are generally not closed under matrix multiplication (see, however, the proof sketch below for a relaxed property).

(3) More abstractly, the *quantum special linear group* SL_n^q is the noncommutative algebra defined as the quotient of the free algebra on generators $\boldsymbol{m}_{ij}$ for $1 \leqslant i, j \leqslant n$ subject to the four relations appearing in Definition 2.6 (with $\boldsymbol{M}_{ij}$ replaced by $\boldsymbol{m}_{ij}$) plus the relation $\mathrm{Det}^q(\boldsymbol{m}) = 1$; see, for example, [2, Chapter I.2]. Note then that a $\mathcal{T}$–point $\boldsymbol{M}$ of SL_n^q is equivalent to an algebra homomorphism $\varphi(\boldsymbol{M})\colon \mathrm{SL}_n^q \to \mathcal{T}$ defined by the property that $\varphi(\boldsymbol{M})(\boldsymbol{m}_{ij}) = \boldsymbol{M}_{ij}$ for all $1 \leqslant i, j \leqslant n$.

2.4.2 Main result Take $\mathcal{T} = \mathcal{T}_n^\omega$ to be the Fock–Goncharov quantum torus for the discrete n–triangle Θ_n; see Section 2.2. Let $\boldsymbol{L}^\omega$ and $\boldsymbol{R}^\omega$ in $\mathrm{M}_n(\mathcal{T}_n^\omega)$ be the quantum left and right matrices, respectively, as defined in Definition 2.4.

Theorem 2.8 *The quantum left and right matrices*

$$\boldsymbol{L}^\omega = \boldsymbol{L}^\omega(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{Z}') \quad \textit{and} \quad \boldsymbol{R}^\omega = \boldsymbol{R}^\omega(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{Z}'') \in \mathrm{M}_n(\mathcal{T}_n^\omega)$$

are $\mathcal{T}_n^\omega$–points of the quantum special linear group SL_n^q. That is, $\boldsymbol{L}^\omega, \boldsymbol{R}^\omega \in \mathrm{SL}_n^q(\mathcal{T}_n^\omega) \subseteq \mathrm{M}_n(\mathcal{T}_n^\omega)$.

The proof, provided in Section 3, uses a quantum version of Fock–Goncharov snakes (Section 1).

Sketch of proof (see Section 3 for more details) In the case $n = 2$, this is an enjoyable calculation. When $n \geqslant 3$, the argument hinges on the following well-known fact (see for example [21, Proposition IV.3.4 and Section IV.10]): if $\mathcal{T}$ is an algebra with subalgebras $\mathcal{T}', \mathcal{T}'' \subseteq \mathcal{T}$ that commute in the sense that $a'a'' = a''a'$ for all $a' \in \mathcal{T}'$ and $a'' \in \mathcal{T}''$, and if $\boldsymbol{M}' \in \mathrm{M}_n(\mathcal{T}') \subseteq \mathrm{M}_n(\mathcal{T})$ and $\boldsymbol{M}'' \in \mathrm{M}_n(\mathcal{T}'') \subseteq \mathrm{M}_n(\mathcal{T})$ are $\mathcal{T}$–points of SL_n^q, then the matrix product (Definition 2.2) $\boldsymbol{M}'\boldsymbol{M}'' \in \mathrm{M}_n(\mathcal{T}'\mathcal{T}'') \subseteq \mathrm{M}_n(\mathcal{T})$ is also a $\mathcal{T}$–point of SL_n^q.

Put $\boldsymbol{M}_{\mathrm{FG}} := \boldsymbol{L}^\omega$, the quantum left matrix, say. The proof is the same for the quantum right matrix. See Definition 2.4. The strategy is to see $\boldsymbol{M}_{\mathrm{FG}} \in \mathrm{M}_n(\mathcal{T}_n^\omega)$ as the product of simpler matrices, over mutually commuting subalgebras, that are themselves points of SL_n^q.

More precisely, for a fixed sequence of adjacent snakes $\sigma^{\mathrm{bot}} = \sigma^1, \sigma^2, \dots, \sigma^N = \sigma^{\mathrm{top}}$ moving left across the triangle from the bottom edge to the top-left edge, we will define for each $i = 1, \dots, N-1$ an auxiliary algebra $\mathcal{S}_{j_i}^\omega$, called a *snake-move algebra*, for $j_i \in \{1, \dots, n-1\}$, corresponding to the adjacent snake pair (σ^i, σ^{i+1}). As a technical step, there is a distinguished subalgebra $\mathcal{T}_L \subseteq \mathcal{T}_n^\omega$ satisfying $\boldsymbol{M}_{\mathrm{FG}} \in \mathrm{M}_n(\mathcal{T}_L) \subseteq \mathrm{M}_n(\mathcal{T}_n^\omega)$. We construct an algebra embedding $\mathcal{T}_L \hookrightarrow \bigotimes_i \mathcal{S}_{j_i}^\omega$. Through this embedding, we may view $\boldsymbol{M}_{\mathrm{FG}} \in \mathrm{M}_n(\mathcal{T}_L) \subseteq \mathrm{M}_n\big(\bigotimes_i \mathcal{S}_{j_i}^\omega\big)$.

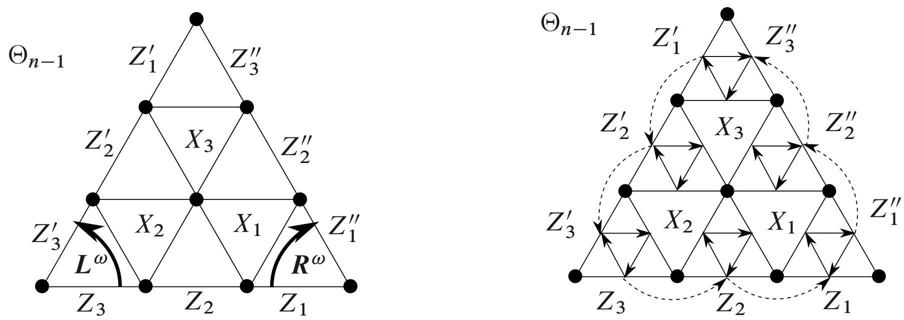


Figure 12: Quantum matrices and quantum torus for $n=4$. Left and right matrices (left) and the quiver (right).

We construct (Proposition 3.3), for each i, a matrix $\boldsymbol{M}_{j_i} \in \mathrm{M}_n(\mathcal{S}^\omega_{j_i}) \subseteq \mathrm{M}_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$ such that $\boldsymbol{M}_{j_i}$ is an $\mathcal{S}^\omega_{j_i}$–point of SL^q_n; in other words $\boldsymbol{M}_{j_i} \in \mathrm{SL}^q_n(\mathcal{S}^\omega_{j_i}) \subseteq \mathrm{SL}^q_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$. Since by definition the subalgebras $\mathcal{S}^\omega_{j_i}, \mathcal{S}^\omega_{j_{i'}} \subseteq \bigotimes_i \mathcal{S}^\omega_{j_i}$ commute if $i \neq i'$ as they constitute different tensor factors of $\bigotimes_i \mathcal{S}^\omega_{j_i}$, it follows from the essential fact mentioned above that $\boldsymbol{M} := \boldsymbol{M}_{j_1}\boldsymbol{M}_{j_2}\cdots\boldsymbol{M}_{j_{N-1}} \in \mathrm{M}_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$ is a $\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$–point of SL^q_n; in other words $\boldsymbol{M} \in \mathrm{SL}^q_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$.

Since this matrix product $\boldsymbol{M}$, as well as the quantum left matrix $\boldsymbol{M}_{\mathrm{FG}}$, is being viewed as an element of $\mathrm{M}_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$, it makes sense to ask whether $\boldsymbol{M}_{\mathrm{FG}} \stackrel{?}{=} \boldsymbol{M} \in \mathrm{M}_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$. We show that this is true, implying that $\boldsymbol{M}_{\mathrm{FG}} \in \mathrm{SL}^q_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$. Since $\boldsymbol{M}_{\mathrm{FG}} \in \mathrm{M}_n(\mathcal{T}_L) \subseteq \mathrm{M}_n\big(\bigotimes_i \mathcal{S}^\omega_{j_i}\big)$, we conclude that $\boldsymbol{M}_{\mathrm{FG}}$ is in $\mathrm{SL}^q_n(\mathcal{T}_L) \subseteq \mathrm{SL}^q_n(\mathcal{T}^\omega_n)$. □

2.5 Example

Consider the case $n=4$; see Figure 12. On the right-hand side is the quiver defining the commutation relations in the quantum torus $\mathcal{T}^\omega_4$, recalling Figure 10, but viewed in Θ_{n-1}. Note that there is a one-to-one correspondence between points $(a,b,c) \in \mathrm{int}(\Theta_n)$ and small downward-facing triangles inside Θ_{n-1}; see Figure 12. In particular, to each downward-facing triangle there is associated a triangle-coordinate X_i.

Some sample commutation relations in $\mathcal{T}^\omega_4$ are

$$X_3 Z_2'' = q^2 X_3 Z_2'', \quad X_3 X_1 = q^{-2} X_1 X_3, \quad Z_3 Z_2 = q Z_2 Z_3, \quad \text{and} \quad Z_3 Z_3' = q^2 Z_3' Z_3.$$

Then, the quantum left and right matrices $\boldsymbol{L}^\omega$ and $\boldsymbol{R}^\omega$ are computed as

$$\boldsymbol{L}^\omega = \left[Z_1^{-\frac14} Z_2^{-\frac24} Z_3^{-\frac34} \begin{pmatrix} Z_1Z_2Z_3 & & & \\ & Z_2Z_3 & & \\ & & Z_3 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} X_1^{-\frac14} \begin{pmatrix} X_1 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 1 \end{pmatrix} X_2^{-\frac24} \begin{pmatrix} X_2 & & & \\ & X_2 & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix} \right.$$
$$\left. \begin{pmatrix} 1 & 1 & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} X_3^{-\frac14} \begin{pmatrix} X_3 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} Z_1'^{-\frac14} Z_2'^{-\frac24} Z_3'^{-\frac34} \begin{pmatrix} Z_1'Z_2'Z_3' & & & \\ & Z_2'Z_3' & & \\ & & Z_3' & \\ & & & 1 \end{pmatrix} \right]$$

and

$$\boldsymbol{R}^{\omega}=\left[Z_1^{-\frac14}Z_2^{-\frac24}Z_3^{-\frac34}\begin{pmatrix}Z_1Z_2Z_3&&&\\&Z_2Z_3&&\\&&Z_3&\\&&&1\end{pmatrix}\begin{pmatrix}1&&&\\&1&&\\&&1&\\&&1&1\end{pmatrix}X_2^{+\frac14}\begin{pmatrix}1&&&\\&1&&\\&1&1&\\&&&X_2^{-1}\end{pmatrix}X_1^{+\frac24}\begin{pmatrix}1&&&\\1&1&&\\&&X_1^{-1}&\\&&&X_1^{-1}\end{pmatrix}\right.$$
$$\left.\begin{pmatrix}1&&&\\&1&&\\&&1&\\&&1&1\end{pmatrix}X_3^{+\frac14}\begin{pmatrix}1&&&\\&1&&\\&1&1&\\&&&X_3^{-1}\end{pmatrix}\begin{pmatrix}1&&&\\&1&&\\&&1&\\&&1&1\end{pmatrix}Z_1''^{-\frac14}Z_2''^{-\frac24}Z_3''^{-\frac34}\begin{pmatrix}Z_1''Z_2''Z_3''&&&\\&Z_2''Z_3''&&\\&&Z_3''&\\&&&1\end{pmatrix}\right].$$

Theorem 2.8 says that these two matrices are elements of $\mathrm{SL}_4^q(\mathcal{T}_4^{\omega})$. For instance, the entries a, b, c, and d of the 2×2 submatrix (arranged as a 4×1 matrix) of $\boldsymbol{L}^{\omega}$

$$\begin{pmatrix}a\\b\\c\\d\end{pmatrix}=\begin{pmatrix}\boldsymbol{L}^{\omega}_{13}\\\boldsymbol{L}^{\omega}_{14}\\\boldsymbol{L}^{\omega}_{23}\\\boldsymbol{L}^{\omega}_{24}\end{pmatrix}=\begin{pmatrix}[Z_3^{\frac14}Z_2^{\frac24}Z_1^{\frac34}Z_3'^{\frac14}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{-\frac14}X_2^{-\frac24}X_3^{-\frac14}]+[Z_3^{\frac14}Z_2^{\frac24}Z_1^{\frac34}Z_3'^{\frac14}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{-\frac14}X_2^{\frac24}X_3^{-\frac14}]\\ \qquad+[Z_3^{\frac14}Z_2^{\frac24}Z_1^{\frac34}Z_3'^{\frac14}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{\frac34}X_2^{\frac24}X_3^{-\frac14}]\\ [Z_3^{\frac14}Z_2^{\frac24}Z_1^{\frac34}Z_3'^{-\frac34}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{-\frac14}X_2^{-\frac24}X_3^{-\frac14}]\\ [Z_3^{\frac14}Z_2^{\frac24}Z_1^{-\frac14}Z_3'^{\frac14}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{-\frac14}X_2^{-\frac24}X_3^{-\frac14}]+[Z_3^{\frac14}Z_2^{\frac24}Z_1^{-\frac14}Z_3'^{\frac14}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{-\frac14}X_2^{\frac24}X_3^{-\frac14}]\\ [Z_3^{\frac14}Z_2^{\frac24}Z_1^{-\frac14}Z_3'^{-\frac34}Z_2'^{-\frac24}Z_1'^{-\frac14}X_1^{-\frac14}X_2^{-\frac24}X_3^{-\frac14}]\end{pmatrix}$$

satisfy $(**)$. For a computer demonstration of this see [6, Appendix B]. We also verify in that appendix that $(**)$ is satisfied by the entries a, b, c, and d of the 2×2 submatrix (arranged as a 4×1 matrix) of $\boldsymbol{R}^{\omega}$:

$$\begin{pmatrix}a\\b\\c\\d\end{pmatrix}=\begin{pmatrix}\boldsymbol{R}^{\omega}_{31}\\\boldsymbol{R}^{\omega}_{32}\\\boldsymbol{R}^{\omega}_{41}\\\boldsymbol{R}^{\omega}_{42}\end{pmatrix}=\begin{pmatrix}[Z_3^{\frac14}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{\frac14}X_1^{\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{\frac34}]\\ [Z_3^{\frac14}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{\frac14}X_1^{-\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{-\frac14}]+[Z_3^{\frac14}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{\frac14}X_1^{\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{-\frac14}]\\ [Z_3^{-\frac34}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{\frac14}X_1^{\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{\frac34}]\\ [Z_3^{-\frac34}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{-\frac34}X_1^{-\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{-\frac14}]\\ \qquad+[Z_3^{-\frac34}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{\frac14}X_1^{-\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{-\frac14}]\\ \qquad\qquad+[Z_3^{-\frac34}Z_2^{-\frac12}Z_1^{-\frac14}X_2^{\frac14}X_1^{\frac12}X_3^{\frac14}Z_3''^{\frac14}Z_2''^{\frac12}Z_1''^{-\frac14}]\end{pmatrix}$$

Remark 2.9 In order for these matrices to satisfy the relations required just to be in $\mathrm{M}_n^q(\mathcal{T}_n^{\omega})$ (let alone $\mathrm{SL}_n^q(\mathcal{T}_n^{\omega})$), they have to be normalized by dividing out their determinants. For example, the above matrix $\boldsymbol{L}^{\omega}$ for $n=4$ would not satisfy the q–commutation relations required to be a point of $\mathrm{M}_4^q(\mathcal{T}_4^{\omega})$ if we had not included the normalizing term $Z_1^{-1/4}Z_2^{-2/4}Z_3^{-3/4}X_1^{-1/4}X_2^{-2/4}X_3^{-1/4}Z_1'^{-1/4}Z_2'^{-2/4}Z_3'^{-3/4}$, as there would be a 1 in the bottom corner.

3 Quantum snakes: proof of Theorem 2.8

Above, we gave a sketch of the proof. We now fill in the details. Our emphasis will be on the left matrix $\boldsymbol{L}^{\omega}$. The proof for the right matrix $\boldsymbol{R}^{\omega}$ is similar, as we will discuss in Section 3.5.

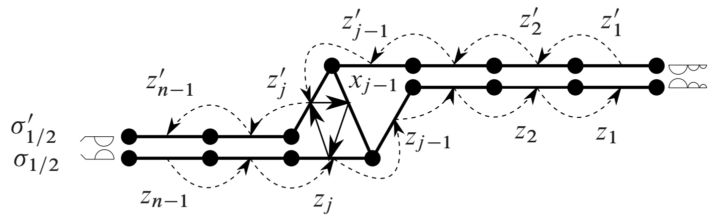


Figure 13: Diamond snake-move algebra for $j = 2, \dots, n-1$.

Fix a sequence $\sigma^{\text{bot}} = \sigma^1, \sigma^2, \dots, \sigma^N = \sigma^{\text{top}}$ of adjacent snakes, as in the left setting; see Section 1.4.1. The proof is valid for any choice of snake sequence, but our demonstrations in figures and examples will be for our preferred snake sequence; see Figure 8. Note that the example quantum matrices in Section 2.5 were presented using this preferred snake sequence.

3.1 Snake-move quantum tori

Definition 3.1 For $j = 1, \dots, n-1$, the j^{th} *snake-move quantum torus* $\mathcal{S}^{\omega}_j = \mathcal{T}(\boldsymbol{P}_j)$ is the quantum torus with Poisson matrix $\boldsymbol{P}_j$ defined by the quiver shown in Figure 13 when $j = 2, \dots, n-1$, and in Figure 14 when $j = 1$. As usual, there is one generator per edge of the quiver, solid arrows carry a weight 2, and dotted arrows carry a weight 1; compare with Section 2.2.

Conceptual Remark 3.2 We provide some guiding intuition for the upcoming constructions; strictly speaking, it is not required for the mathematical progression of the article.

The quiver of Figure 14 for the tail-move quantum torus is divided into a bottom and top side. Similarly, the quiver of Figure 13 for a diamond-move quantum torus has a bottom and top side, connected by a diagonal (where the variable x_{j-1} is located). As illustrated in the figures, we think of the bottom side (with unprimed generators z_j) as the top "snake-half" $\sigma_{1/2}$ of a snake σ that has been "split in half down its length". Similarly, we think of the top side (with primed generators z'_j) as the bottom snake-half $\sigma'_{1/2}$ of a split snake σ'. Compare with Figure 3, which illustrates a classical snake "before splitting".

This snake splitting can be seen more clearly in the *quantum snake sweep* (see Section 3.3 and Figure 15) determined by the sequence of adjacent snakes $\sigma^{\text{bot}} = \sigma^1, \sigma^2, \dots, \sigma^N = \sigma^{\text{top}}$, where each snake σ^i is split in half, so that each snake-half forms a side in one of two adjacent snake-move quantum tori. In the figure, the other halves (colored gray) of the bottom-most and top-most quantum snakes can be thought of as either living in other triangles or not existing at all. Prior to splitting a snake σ in half, the snake

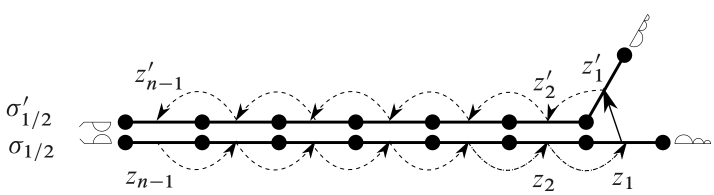


Figure 14: Tail snake-move algebra for $j = 1$.

consists of $n-1$ "vertebrae" connecting the n snake-vertices $\sigma_k \in \Theta_{n-1}$. Upon splitting the snake, the j^{th} vertebra splits into two generators z_j and z'_j living in adjacent snake-move quantum tori.

3.2 Quantum snake-move matrices

We turn to the key observation for the proof.

Proposition 3.3 *For $j = 1, \dots, n-1$, the j^{th} quantum snake-move matrix*

$$\boldsymbol{M}_j := \left[\left(\prod_{k=1}^{n-1} \boldsymbol{S}_k^{\text{edge}}(z_k) \right) \boldsymbol{S}_j^{\text{left}}(x_{j-1}) \left(\prod_{k=1}^{n-1} \boldsymbol{S}_k^{\text{edge}}(z'_k) \right) \right] \in \mathrm{M}_n(\mathcal{S}_j^{\omega})$$

is an $\mathcal{S}_j^{\omega}$–point of the quantum special linear group SL_n^q. That is, $\boldsymbol{M}_j \in \mathrm{SL}_n^q(\mathcal{S}_j^{\omega}) \subseteq \mathrm{M}_n(\mathcal{S}_j^{\omega})$.

Note the use of the Weyl quantum ordering; see Section 2.1.3. Here the matrices $\boldsymbol{S}_k^{\text{edge}}(z)$ and $\boldsymbol{S}_j^{\text{left}}(x)$ for z and x in the commutative algebra $\mathcal{S}_j^1$ are defined as in Section 1.3.1; see also Sections 2.3.1 and 2.3.2. When $j = 1$, the matrix $\boldsymbol{S}_1^{\text{left}}(x_0) = \boldsymbol{S}_1^{\text{left}}$ is well defined, despite x_0 not being defined.

Proposition 3.3 follows from direct calculation. See Section 3.5 for the proof.

For example, in the case $n = 4$ and $j = 3$, the lemma says that the matrix

$$\boldsymbol{M}_3 = \left[z_1^{-\frac{1}{4}} z_2^{-\frac{2}{4}} z_3^{-\frac{3}{4}} \begin{pmatrix} z_1 z_2 z_3 & & & \\ & z_2 z_3 & & \\ & & z_3 & \\ & & & 1 \end{pmatrix} x_2^{-\frac{2}{4}} \begin{pmatrix} x_2 & & & \\ & x_2 & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix} z_1'^{-\frac{1}{4}} z_2'^{-\frac{2}{4}} z_3'^{-\frac{3}{4}} \begin{pmatrix} z'_1 z'_2 z'_3 & & & \\ & z'_2 z'_3 & & \\ & & z'_3 & \\ & & & 1 \end{pmatrix} \right]$$

is in $\mathrm{SL}_4^q(\mathcal{S}_3^{\omega})$.

3.3 Technical step: embedding a distinguished subalgebra $\mathcal{T}_L$ of $\mathcal{T}_n^{\omega}$ into a tensor product $\bigotimes_{i=1}^{N-1} \mathcal{S}_{j_i}^{\omega}$ of snake-move quantum tori

For the snake-sequence $(\sigma^i)_{i=1,\dots,N}$, to each pair (σ^i, σ^{i+1}) of adjacent snakes we associate a snake-move quantum torus $\mathcal{S}_{j_i}^{\omega}$, recalling Figure 15 (see also Conceptual Remark 3.2). Here j_i corresponds to what was called k in Definition 1.7. Recall the Fock–Goncharov quantum torus $\mathcal{T}_n^{\omega}$ (for example, Figure 12).

We now take a technical step. Using the notation of Figures 11 and 12, define $\mathcal{T}_L \subseteq \mathcal{T}_n^{\omega}$ ("L" for "left") to be the subalgebra generated by all the generators (and their inverses) of $\mathcal{T}_n^{\omega}$ except for $Z_1''^{\pm 1/n}, \dots, Z_{n-1}''^{\pm 1/n}$. We claim that the snake-sequence $(\sigma^i)_i$ induces an embedding

$$\mathcal{T}_L \xrightarrow{(\sigma^i)_i} \bigotimes_{i=1}^{N-1} \mathcal{S}_{j_i}^{\omega}$$

of algebras, realizing $\mathcal{T}_L \subseteq \mathcal{T}_n^{\omega}$ as a subalgebra of the tensor product of the snake-move quantum tori $\mathcal{S}_{j_i}^{\omega}$ associated to the adjacent snake pairs (σ^i, σ^{i+1}). Here recall in general that the algebra structure for a tensor product $A \otimes B$ of algebras A and B is defined by $(a \otimes b) \cdot (a' \otimes b') = (a \cdot a') \otimes (b \cdot b')$ for all $a, a' \in A$ and $b, b' \in B$, extended linearly.

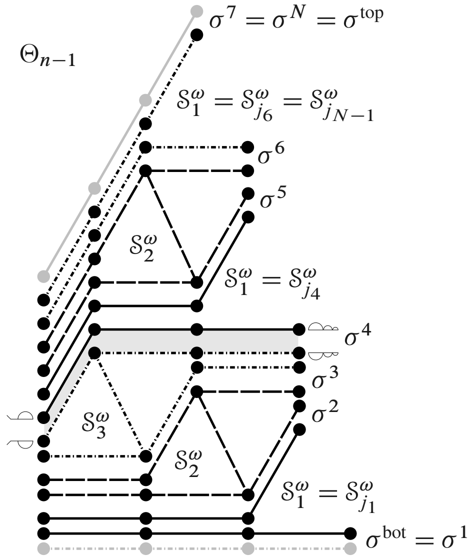


Figure 15: Quantum snake sweep for $n = 4$; compare with Figure 8, left.

A more formal definition of the embedding will be given in Section 3.3.1. We first explain the embedding through an example, in the setting $n = 4$; see Figure 16 (compare with Figure 15).

In this setting, the coordinate X_2, for instance (emphasized in Figure 16), is mapped via

$$X_2 \mapsto 1 \otimes z_2' \otimes z_2 x_2 z_2' \otimes z_2 z_2' \otimes z_2 \otimes 1 \in \mathcal{S}_1^\omega \otimes \mathcal{S}_2^\omega \otimes \mathcal{S}_3^\omega \otimes \mathcal{S}_1^\omega \otimes \mathcal{S}_2^\omega \otimes \mathcal{S}_1^\omega.$$

Similarly, the other coordinates Z_1, Z_3', Z_2, Z_3, X_1, X_1', Z_2', and Z_1' are mapped via

$$\begin{aligned}
Z_1 &\mapsto z_1 \otimes 1 \otimes 1 \otimes 1 \otimes 1 \otimes 1, & Z_3' &\mapsto 1 \otimes 1 \otimes z_3' \otimes z_3 z_3' \otimes z_3 z_3' \otimes z_3 z_3', \\
Z_2 &\mapsto z_2 z_2' \otimes z_2 \otimes 1 \otimes 1 \otimes 1 \otimes 1, & Z_3 &\mapsto z_3 z_3' \otimes z_3 z_3' \otimes z_3 \otimes 1 \otimes 1 \otimes 1, \\
X_1 &\mapsto z_1' \otimes z_1 x_1 z_1' \otimes z_1 z_1' \otimes z_1 \otimes 1 \otimes 1, & X_1' &\mapsto 1 \otimes 1 \otimes 1 \otimes z_1' \otimes z_1 x_1 z_1' \otimes z_1, \\
Z_2' &\mapsto 1 \otimes 1 \otimes 1 \otimes 1 \otimes z_2' \otimes z_2 z_2', & Z_1' &\mapsto 1 \otimes 1 \otimes 1 \otimes 1 \otimes 1 \otimes z_1'.
\end{aligned}$$

Note that the monomials (for instance, $z_2 x_2 z_2'$ or $z_2 z_2'$) appearing in the i^{th} tensor factor of the image of a generator X or Z of the subalgebra $\mathcal{T}_L$ under this mapping consist of mutually commuting generators x's and/or z's of the i^{th} snake-move quantum torus $\mathcal{S}_{j_i}^\omega$, so the order in which they are written is irrelevant. It is clear from Figure 16 that these images satisfy the relations of $\mathcal{T}_L$. In particular, the "interior" dotted arrows lying at each interface between two snake-move quantum tori cancel each other out; note that, in Figure 16, we have omitted drawing some of these dotted arrows. We gather that the mapping is well defined and is an algebra homomorphism. Injectivity follows from the property that every generator (that is, quiver edge) appearing on the right side of Figure 16 corresponds to a unique generator on the left side. Lastly, we technically should have defined the map on the formal n–roots of the coordinates of $\mathcal{T}_L$. This is done in the obvious way; for instance,

$$X_2^{\frac{1}{4}} \mapsto 1 \otimes z_2'^{\frac{1}{4}} \otimes z_2^{\frac{1}{4}} x_2^{\frac{1}{4}} z_2'^{\frac{1}{4}} \otimes z_2^{\frac{1}{4}} z_2'^{\frac{1}{4}} \otimes z_2^{\frac{1}{4}} \otimes 1 \in \mathcal{S}_1^\omega \otimes \mathcal{S}_2^\omega \otimes \mathcal{S}_3^\omega \otimes \mathcal{S}_1^\omega \otimes \mathcal{S}_2^\omega \otimes \mathcal{S}_1^\omega.$$

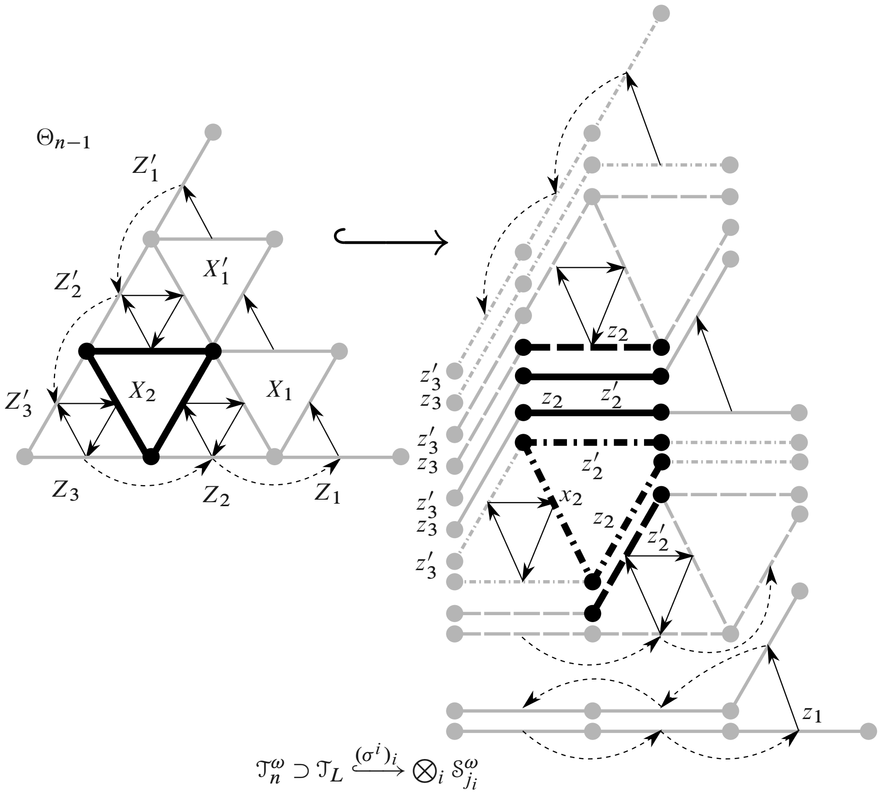


Figure 16: Embedding $\mathcal{T}_L$ in the tensor product of snake-move quantum tori.

3.3.1 Formal definition of the embedding A *segment* $\overline{\mu\nu} = \overline{\nu\mu}$ of the discrete triangle Θ_{n-1} is a line connecting neighboring vertices μ and ν. Segments of the form $\overline{(\alpha,\beta,\gamma),(\alpha-1,\beta,\gamma+1)}$ are called *horizontal*, segments of the form $\overline{(\alpha,\beta,\gamma),(\alpha-1,\beta+1,\gamma)}$ are called *acute*, and segments of the form $\overline{(\alpha,\beta,\gamma),(\alpha,\beta+1,\gamma-1)}$ are called *obtuse*. Let Seg_L denote the set of segments minus the obtuse segments with zero first coordinate. For example, in the case $n=4$, the set Seg_L has 15 elements; see the left-hand side of Figure 16. Let $\mathrm{Coord}_L \subseteq \mathcal{T}_L$ denote the set of coordinates, that is,

$$\mathrm{Coord}_L = \{X_{abc} \mid (a,b,c) \in \mathrm{int}(\Theta_n)\} \cup \{Z_j \mid j = 1,2,\dots,n-1\} \cup \{Z'_j \mid j = 1,2,\dots,n-1\}.$$

Note that the coordinates X_{abc} correspond to the small downward facing triangles in Θ_{n-1}, each of which is a union of a horizontal, acute, and obtuse segment in Seg_L, the coordinates Z_j correspond to horizontal segments with zero second coordinate, the coordinates Z'_j correspond to acute segments with zero third coordinate, and each segment corresponds in this way to a unique coordinate. In particular, there is a canonical surjective function $\pi' \colon \mathrm{Seg}_L \to \mathrm{Coord}_L \subseteq \mathcal{T}_L$. See the left-hand side of Figure 16 for the case $n=4$, where for example the three bold segments constitute the preimage $\pi'^{-1}(X_2)$.

Given a snake sequence $(\sigma^i)_{i=1,2,\dots,N}$, for each $1 \leqslant i \leqslant N-1$ let $\mathrm{Coord}^i \subseteq \mathcal{S}_{j_i}^\omega$ denote the set of coordinates in the snake-move quantum torus $\mathcal{S}_{j_i}^\omega$. There are associated functions $\varphi^i \colon \mathrm{Coord}^i \to \mathrm{Coord}_L$, in general neither surjective nor injective, defined as follows. To each coordinate $z_k \in \mathcal{S}_{j_i}^\omega$ for $k = 1,2,\dots,n-1$ there is associated a segment $\mathrm{seg}(z_k) \in \mathrm{Seg}_L$ of the snake σ^i, namely $\mathrm{seg}(z_k) = \overline{\sigma^i_{k+1}\sigma^i_k}$, to each coordinate $z'_k \in \mathcal{S}^\omega_{j_i}$ for $k = 1, 2, \dots, n-1$ there is associated a segment $\mathrm{seg}(z'_k) \in \mathrm{Seg}_L$ of the snake σ^{i+1}, namely $\mathrm{seg}(z'_k) = \overline{\sigma^{i+1}_{k+1}\sigma^{i+1}_k}$, and to the coordinate $x_{j_i-1} \in \mathcal{S}^\omega_{j_i}$ there is associated an obtuse segment $\mathrm{seg}(x_{j_i-1}) \in \mathrm{Seg}_L$ which is not a segment of a snake, namely $\mathrm{seg}(x_{j_i-1}) = \overline{\sigma^i_{j_i}\sigma^{i+1}_{j_i}}$. Compare with Figure 15. Then φ^i is defined by $\varphi^i(x) = \pi'(\mathrm{seg}(x))$ for all $x \in \mathrm{Coord}^i$.

For example in the case $n = 4$, as illustrated in Figure 16, we have $\varphi^1(z_1) = Z_1$, $\varphi^2(z'_2) = X_2$, $\varphi^3(x_2) = X_2$, $\varphi^4(z'_2) = X_2$, $\varphi^5(z_2) = X_2$, and $\varphi^6(z_3) = Z'_3$.

Finally, define the desired embedding on generators $X^{1/n}$ of $\mathcal{T}_L$, for $X \in \mathrm{Coord}_L$, so that the image of $X^{1/n}$ in the tensor product $\bigotimes_i \mathcal{S}^\omega_{j_i}$ is the pure tensor defined by the property that its i^{th} factor is $\prod_{x \in (\varphi^i)^{-1}(X)} x^{1/n} \in \mathcal{S}^\omega_{j_i}$. This is well defined since the coordinates $x \in (\varphi^i)^{-1}(X)$ in each preimage commute by design. Note, by definition, if $(\varphi^i)^{-1}(X)$ is empty, then the product defining the i^{th} factor is 1.

In Section 3.4, we will make use of the surjective function $\pi\colon \bigcup_{i=1}^{N-1} \mathrm{Coord}^i \to \mathrm{Coord}_L$ defined by $\pi(x) = \varphi^i(x)$ for $x \in \mathrm{Coord}^i$.

3.4 Finishing the proof

Comparing to the sketch of proof given in Section 2.4.2, we gather:

- $\boldsymbol{M}_{\mathrm{FG}} := \boldsymbol{L}^\omega \in \mathrm{M}_n(\mathcal{T}_L) \subseteq \mathrm{M}_n\big(\bigotimes_{i=1}^{N-1} \mathcal{S}^\omega_{j_i}\big)$.
- $\boldsymbol{M} := \boldsymbol{M}_{j_1}\boldsymbol{M}_{j_2}\cdots\boldsymbol{M}_{j_{N-1}} \in \mathrm{SL}^q_n\big(\bigotimes_{i=1}^{N-1} \mathcal{S}^\omega_{j_i}\big) \subseteq \mathrm{M}_n\big(\bigotimes_{i=1}^{N-1} \mathcal{S}^\omega_{j_i}\big)$.

To finish the proof, it remains to show

$$(***) \qquad \boldsymbol{M}_{\mathrm{FG}} \overset{?}{=} \boldsymbol{M} \in \mathrm{M}_n\Big(\bigotimes_{i=1}^{N-1} \mathcal{S}^\omega_{j_i}\Big).$$

The strategy is to commute the many variables (as in the right-hand side of Figure 16) appearing on the right-hand side $\boldsymbol{M} = \prod_i \boldsymbol{M}_{j_i}$ (defined via Proposition 3.3) of $(***)$, until $\boldsymbol{M}$ has been put into the form of the left-hand side $\boldsymbol{M}_{\mathrm{FG}}$ (defined via Definition 2.4 followed by applying the embedding $\mathcal{T}_L \hookrightarrow \bigotimes_i \mathcal{S}^\omega_{j_i}$ of Section 3.3). This is accomplished by applying the following two facts:

Lemma 3.4 (1) *If $\widetilde{\boldsymbol{M}}_1, \widetilde{\boldsymbol{M}}_2, \dots, \widetilde{\boldsymbol{M}}_{N-1}$ are $n \times n$ matrices with coefficients in $(q = \omega = \omega^{1/2} = 1)$–specializations $\mathcal{T}^1_i$ of general quantum tori $\mathcal{T}^\omega_1, \mathcal{T}^\omega_2, \dots, \mathcal{T}^\omega_{N-1}$, viewed as factors in*

$$\mathcal{T}^\omega_1 \otimes \mathcal{T}^\omega_2 \otimes \cdots \otimes \mathcal{T}^\omega_{N-1},$$

then

$$[\widetilde{\boldsymbol{M}}_1][\widetilde{\boldsymbol{M}}_2]\cdots[\widetilde{\boldsymbol{M}}_{N-1}] = [\widetilde{\boldsymbol{M}}_1\widetilde{\boldsymbol{M}}_2\cdots\widetilde{\boldsymbol{M}}_{N-1}] \in \mathrm{M}_n(\mathcal{T}^\omega_1 \otimes \mathcal{T}^\omega_2 \otimes \cdots \otimes \mathcal{T}^\omega_{N-1}).$$

Here we are viewing the tensor product $\mathcal{T}^\omega_1 \otimes \mathcal{T}^\omega_2 \otimes \cdots \otimes \mathcal{T}^\omega_{N-1}$ as a quantum torus in the obvious way, as demonstrated in the proof below.

(2) *For commuting variables z and x, the matrices $\boldsymbol{S}^{\mathrm{edge}}_k(z)$ and $\boldsymbol{S}^{\mathrm{left}}_j(x)$, as in Section 1.3.1, satisfy*

$$\boldsymbol{S}^{\mathrm{edge}}_k(z)\boldsymbol{S}^{\mathrm{left}}_j(x) = \boldsymbol{S}^{\mathrm{left}}_j(x)\boldsymbol{S}^{\mathrm{edge}}_k(z) \quad \textit{if and only if } k \neq j.$$

Proof The proof of (1) is straightforward. To simplify the notation, we demonstrate the calculation for two matrices $\widetilde{\boldsymbol{M}}$ and $\widetilde{\boldsymbol{N}}$ with coefficients in classical tori $\mathcal{T}$ and $\mathcal{U}$ with coordinates $\{X_i\}_{i=1,2,\dots,m}$ and $\{Y_j\}_{j=1,2,\dots,p}$ and quivers ϵ and ζ, respectively, where $\mathcal{T}$ and $\mathcal{U}$ are viewed in $\mathcal{T}\otimes\mathcal{U}$. The proof for finitely many matrices is analogous.

By linearity, it suffices to assume $\widetilde{\boldsymbol{M}}_{ij}\in\mathcal{T}$ and $\widetilde{\boldsymbol{N}}_{kl}\in\mathcal{U}$ are monomials, that is,

$$\widetilde{\boldsymbol{M}}_{ij}=X_1^{a_1^{ij}}X_2^{a_2^{ij}}\cdots X_m^{a_m^{ij}}\quad\text{and}\quad \widetilde{\boldsymbol{N}}_{kl}=Y_1^{b_1^{kl}}Y_2^{b_2^{kl}}\cdots Y_p^{b_p^{kl}}.$$

Recall that, by definition, different tensor factors commute under multiplication in $\mathcal{T}\otimes\mathcal{U}$. We have, for all $1\leqslant i,j\leqslant n$,

$$\begin{aligned}[\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{N}}]_{ij}=[(\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{N}})_{ij}]=\sum_k[\widetilde{\boldsymbol{M}}_{ik}\widetilde{\boldsymbol{N}}_{kj}]&=\sum_k[X_1^{a_1^{ik}}X_2^{a_2^{ik}}\cdots X_m^{a_m^{ik}}Y_1^{b_1^{kj}}Y_2^{b_2^{kj}}\cdots Y_p^{b_p^{kj}}]\\
&=\sum_k q^{-\frac12\kappa}X_1^{a_1^{ik}}X_2^{a_2^{ik}}\cdots X_m^{a_m^{ik}}Y_1^{b_1^{kj}}Y_2^{b_2^{kj}}\cdots Y_p^{b_p^{kj}}\\
&=\sum_k q^{-\frac12\xi}X_1^{a_1^{ik}}X_2^{a_2^{ik}}\cdots X_m^{a_m^{ik}}Y_1^{b_1^{kj}}Y_2^{b_2^{kj}}\cdots Y_p^{b_p^{kj}}\\
&=\sum_k[X_1^{a_1^{ik}}X_2^{a_2^{ik}}\cdots X_m^{a_m^{ik}}][Y_1^{b_1^{kj}}Y_2^{b_2^{kj}}\cdots Y_p^{b_p^{kj}}]=\sum_k[\widetilde{\boldsymbol{M}}_{ik}][\widetilde{\boldsymbol{N}}_{kj}]=([\widetilde{\boldsymbol{M}}][\widetilde{\boldsymbol{N}}])_{ij},\end{aligned}$$

where

$$\kappa=\sum_{1\leqslant\alpha<\beta\leqslant m}(\epsilon\otimes\zeta)_{\alpha\beta}a_\alpha^{ik}a_\beta^{ik}+\sum_{1\leqslant\alpha\leqslant m,1\leqslant\beta\leqslant p}(\epsilon\otimes\zeta)_{\alpha(m+\beta)}a_\alpha^{ik}b_\beta^{kj}+\sum_{1\leqslant\alpha<\beta\leqslant p}(\epsilon\otimes\zeta)_{(m+\alpha)(m+\beta)}b_\alpha^{kj}b_\beta^{kj}$$

and

$$\xi=\sum_{1\leqslant\alpha<\beta\leqslant m}\epsilon_{\alpha\beta}a_\alpha^{ik}a_\beta^{ik}+\sum_{1\leqslant\alpha\leqslant m,1\leqslant\beta\leqslant p}0a_\alpha^{ik}b_\beta^{kj}+\sum_{1\leqslant\alpha<\beta\leqslant p}\zeta_{\alpha\beta}b_\alpha^{kj}b_\beta^{kj}.$$

The proof of (2) is by inspection. □

Proof of Theorem 2.8 By Lemma 3.4(1), it suffices to establish (∗∗∗) when $q=\omega=\omega^{1/2}=1$, in which case we do not need to worry about the Weyl quantum ordering.

It is helpful to introduce a simplifying notation. For coordinates $z_k^{(i)},x_j^{(i)},z_k'^{(i)}\in\mathcal{S}_{j_i}^1$, put

$$\boldsymbol{Z}_k^{(i)}:=\boldsymbol{S}_k^{\text{edge}}(z_k^{(i)}),\quad \boldsymbol{X}_j^{(i)}:=\boldsymbol{S}_{j+1}^{\text{left}}(x_j^{(i)}),\quad\text{and}\quad \boldsymbol{Z}_k'^{(i)}:=\boldsymbol{S}_k^{\text{edge}}(z_k'^{(i)})\in\mathrm{M}_n(\mathcal{S}_{j_i}^1).$$

In this new notation, the matrices $\boldsymbol{M}_{j_i}\in\mathrm{M}_n(\mathcal{S}_{j_i}^1)$ of Proposition 3.3 can be expressed by

$$\boldsymbol{M}_{j_i}=\Bigl(\prod_{k=1}^{n-1}\boldsymbol{Z}_k^{(i)}\Bigr)\boldsymbol{X}_{j_i-1}^{(i)}\Bigl(\prod_{k=1}^{n-1}\boldsymbol{Z}_k'^{(i)}\Bigr)\in\mathrm{M}_n(\mathcal{S}_{j_i}^1),$$

and Lemma 3.4(2) now reads, for any $i_1,i_2\in\{1,2,\dots,N-1\}$,

$$(\dagger)\quad \boldsymbol{Z}_k^{(i_1)}\boldsymbol{X}_j^{(i_2)}=\boldsymbol{X}_j^{(i_2)}\boldsymbol{Z}_k^{(i_1)}\in\mathrm{M}_n\Bigl(\bigotimes_{i=1}^{N-1}\mathcal{S}_{j_i}^1\Bigr)\ \text{ if and only if }\ k\neq j+1\quad(\text{similarly for }\boldsymbol{Z}\to\boldsymbol{Z}').$$

Example ($n=2$) In this case, $N=2$, we have $\mathcal{S}^1_{j_1}=\mathcal{S}^1_1\cong\mathcal{T}_L\subseteq\mathcal{T}^1_n$, and the embedding $\mathcal{T}_L\xhookrightarrow{\sim}\mathcal{S}^1_1$ is the identity, where $Z_1\mapsto z_1^{(1)}$ and $Z'_1\mapsto z_1'^{(1)}$. Equation $(***)$ is also trivial, reading

$$\begin{aligned}\boldsymbol{M}=\boldsymbol{M}_1=\boldsymbol{Z}_1^{(1)}\boldsymbol{X}_0^{(1)}\boldsymbol{Z}_1'^{(1)}&=z_1^{(1)\frac{-1}{2}}\begin{pmatrix}z_1^{(1)}&0\\0&1\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}z_1'^{(1)\frac{-1}{2}}\begin{pmatrix}z_1'^{(1)}&0\\0&1\end{pmatrix}\\&=Z_1^{-\frac{1}{2}}\begin{pmatrix}Z_1&0\\0&1\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}Z_1'^{-\frac{1}{2}}\begin{pmatrix}Z'_1&0\\0&1\end{pmatrix}=\boldsymbol{S}_1^{\mathrm{edge}}(Z_1)\boldsymbol{S}_1^{\mathrm{left}}\boldsymbol{S}_1^{\mathrm{edge}}(Z'_1)=\boldsymbol{M}_{\mathrm{FG}}.\end{aligned}$$

Example ($n=3$) Here $N=4$, the subalgebra $\mathcal{T}_L$ has coordinates Z_1, Z_2, X_1, Z'_1 and Z'_2, and the embedding $\mathcal{T}_L\hookrightarrow\mathcal{S}^1_1\otimes\mathcal{S}^1_2\otimes\mathcal{S}^1_1$ is defined by (compare the $n=4$ case, Figure 16)

$$Z_1\mapsto z_1^{(1)},\quad Z_2\mapsto z_2^{(1)}z_2'^{(1)}z_2^{(2)},\quad X_1\mapsto z_1'^{(1)}z_1^{(2)}x_1^{(2)}z_1'^{(2)}z_1^{(3)},\quad Z'_1\mapsto z_1'^{(3)},\quad\text{and}\quad Z'_2\mapsto z_2'^{(2)}z_2^{(3)}z_2'^{(3)},$$

where we have suppressed the tensor products. Note in this case there is a unique snake-sequence $(\sigma^i)_{i=1,\dots,4}$ so there is only one associated embedding of $\mathcal{T}_L$. Equation $(***)$ reads

$$\begin{aligned}\boldsymbol{M}=\boldsymbol{M}_1\boldsymbol{M}_2\boldsymbol{M}_1&=\underline{\boldsymbol{Z}}_1^{(1)}\underline{\boldsymbol{Z}}_2^{(1)}\underline{\boldsymbol{X}}_0^{(1)}\boldsymbol{Z}_1'^{(1)}\boldsymbol{Z}_2'^{(1)}\cdot\boldsymbol{Z}_1^{(2)}\boldsymbol{Z}_2^{(2)}\underline{\boldsymbol{X}}_1^{(2)}\boldsymbol{Z}_1'^{(2)}\boldsymbol{Z}_2'^{(2)}\cdot\boldsymbol{Z}_1^{(3)}\boldsymbol{Z}_2^{(3)}\underline{\boldsymbol{X}}_0^{(3)}\underline{\boldsymbol{Z}}_1'^{(3)}\underline{\boldsymbol{Z}}_2'^{(3)}\\&=\underline{\boldsymbol{Z}}_1^{(1)}\cdot\underline{\boldsymbol{Z}}_2^{(1)}\boldsymbol{Z}_2'^{(1)}\boldsymbol{Z}_2^{(2)}\cdot\underline{\boldsymbol{X}}_0^{(1)}\cdot\boldsymbol{Z}_1'^{(1)}\boldsymbol{Z}_1^{(2)}\underline{\boldsymbol{X}}_1^{(2)}\boldsymbol{Z}_1'^{(2)}\boldsymbol{Z}_1^{(3)}\cdot\underline{\boldsymbol{X}}_0^{(3)}\cdot\underline{\boldsymbol{Z}}_1'^{(3)}\cdot\boldsymbol{Z}_2'^{(2)}\boldsymbol{Z}_2^{(3)}\underline{\boldsymbol{Z}}_2'^{(3)}\\&=\boldsymbol{S}_1^{\mathrm{edge}}(Z_1)\boldsymbol{S}_2^{\mathrm{edge}}(Z_2)\boldsymbol{S}_1^{\mathrm{left}}\boldsymbol{S}_2^{\mathrm{left}}(X_1)\boldsymbol{S}_1^{\mathrm{left}}\boldsymbol{S}_1^{\mathrm{edge}}(Z'_1)\boldsymbol{S}_2^{\mathrm{edge}}(Z'_2)=\boldsymbol{M}_{\mathrm{FG}},\end{aligned}$$

where for the third equality we have used the reformulation (†) of Lemma 3.4(2) to commute the matrices. Note that the ordering of terms in any of the seven groupings in the fourth expression is immaterial. The fourth equality uses the embedding $\mathcal{T}_L\hookrightarrow\mathcal{S}^1_1\otimes\mathcal{S}^1_2\otimes\mathcal{S}^1_1$.

General case As we saw in the examples, $\boldsymbol{M}=\prod_{i=1}^{N-1}\boldsymbol{M}_{j_i}$ is a product of distinct terms $\boldsymbol{Z}_k^{(i)}$, $\boldsymbol{X}_j^{(i)}$, or $\boldsymbol{Z}_k'^{(i)}$. Let A be the set of terms, that is, $A=\bigcup_{i=1,2,\dots,N-1}\{\boldsymbol{Z}_k^{(i)},\boldsymbol{X}_{j_i-1}^{(i)},\boldsymbol{Z}_k'^{(i)}\mid k=1,2,\dots,n-1\}$. Besides terms of the form $\boldsymbol{X}_0^{(i)}$, there is one term in A for each coordinate $z_k^{(i)}$, $x_j^{(i)}$, and $z_k'^{(i)}$ of $\bigotimes_{i=1}^{N-1}\mathcal{S}^1_{j_i}$. We show that there is an algorithm that commutes these terms into the correct groupings, as in the above examples.

There is a distinguished subset $A_L\subseteq A$, precisely defined in the next paragraph. In the example $n=2$, $A_L=A$, and in the example $n=3$, the terms in A_L are underlined above. All the $\boldsymbol{X}_0^{(i)}$ terms are in A_L. Besides the $\boldsymbol{X}_0^{(i)}$ terms, there is one term in A_L for each coordinate Z_k, X_j, and Z'_k of $\mathcal{T}_L$; see Figure 16. As another example, for $n=4$ and our usual preferred snake sequence $(\sigma^i)_i$, $A_L=\{\boldsymbol{Z}_1^{(1)},\boldsymbol{Z}_2^{(1)},\boldsymbol{Z}_3^{(1)},\boldsymbol{X}_0^{(1)},\boldsymbol{X}_1^{(2)},\boldsymbol{X}_2^{(3)},\boldsymbol{X}_0^{(4)},\boldsymbol{X}_1^{(5)},\boldsymbol{X}_0^{(6)},\boldsymbol{Z}_1'^{(6)},\boldsymbol{Z}_2'^{(6)},\boldsymbol{Z}_3'^{(6)}\}$; see Figures 15 and 16.

More precisely, the general definition of $A_L\subseteq A$, valid for any snake sequence $(\sigma^i)_{i=1,2,\dots,N}$, is as follows. First, $\boldsymbol{Z}_k^{(1)}=\boldsymbol{S}_k^{\mathrm{edge}}(z_k^{(1)})$, $\boldsymbol{Z}_k'^{(N-1)}=\boldsymbol{S}_k^{\mathrm{edge}}(z_k'^{(N-1)})$, and $\boldsymbol{X}_0^{(i)}=\boldsymbol{S}_1^{\mathrm{left}}$ are in A_L for all $k=1,2,\dots,n-1$ and for all $1\leqslant i\leqslant N-1$ such that $j_i-1=0$. And $\boldsymbol{X}_{j_i-1}^{(i)}=\boldsymbol{S}_{j_i}^{\mathrm{left}}(x_{j_i-1}^{(i)})$ is in A_L for all $1\leqslant i\leqslant N-1$ such that $j_i>1$.

Recall that the injectivity of the embedding $\mathcal{T}_L\hookrightarrow\bigotimes_{i=1}^{N-1}\mathcal{S}^1_{j_i}$ followed immediately from the property that every coordinate $z_k^{(i)}$, $x_j^{(i)}$, or $z_k'^{(i)}$ of $\bigotimes_{i=1}^{N-1}\mathcal{S}^1_{j_i}$ corresponds to a unique coordinate Z_k, X_j, or Z'_k of $\mathcal{T}_L$;

see Figure 16. This property thus defines a retraction $r\colon A \twoheadrightarrow A_L$, namely a surjective function restricting to the identity on $A_L \subseteq A$ (by definition, $\boldsymbol{X}_0^{(i)} \mapsto \boldsymbol{X}_0^{(i)}$). See the next paragraph for a precise definition. The retraction r can be visualized as collapsing the right side of Figure 16 to obtain the left side.

More precisely, in the notation of Section 3.3.1, there is a bijection $f\colon A - A_0 \to \bigcup_{i=1}^{N-1} \mathrm{Coord}^i$ defined by $f(\boldsymbol{Z}_k^{(i)}) = z_k^{(i)}$, $f(\boldsymbol{Z}_k'^{(i)}) = z_k'^{(i)}$, and $f(\boldsymbol{X}_{j_i-1}^{(i)}) = x_{j_i-1}^{(i)}$. Here we have put

$$A_0 = \{\boldsymbol{X}_0^{(i)} \mid 1 \leqslant i \leqslant N-1 \text{ such that } j_i - 1 = 0\}.$$

By definition of A_L, the restricted composition g defined by $g = \pi \circ (f|_{A_L - A_0})\colon A_L - A_0 \to \mathrm{Coord}_L$ is a bijection, where $\pi\colon \bigcup_{i=1}^{N-1} \mathrm{Coord}^i \to \mathrm{Coord}_L$ is defined at the end of Section 3.3.1. The retraction $r\colon A \to A_L$ is defined on $A - A_0$ by $r = g^{-1} \circ \pi \circ f$, and as the identity on $A_0 \subseteq A_L$.

The desired algorithm grouping the terms in A, where there is one grouping per term in A_L, is defined by selecting an ungrouped term $a \in A$ and commuting it left or right until it is adjacent to $r(a) \in A_L$. Here the terms are viewed in the expression for $\boldsymbol{M}$. This commutation is possible by Lemma 3.4(2), that is, (†).

More precisely, in the expression for $\boldsymbol{M}$ at step s of the algorithm, for each $a_0 \in A_L$ let $l(a_0, s)$ denote the length of the longest chain of adjacent terms $a \in r^{-1}(a_0)$ such that this chain contains a_0. For instance, in the $n = 3$ example above, for $a_0 = \boldsymbol{X}_1^{(2)}$, initially the length of the chain containing a_0 is 2, while at the end of the algorithm this length is $5 = |r^{-1}(a_0)|$. Assuming for the moment that the algorithm is well defined, that is, that the commutation is possible, we see that $l(a_0, s) \leqslant l(a_0, s+1)$ for all $a_0 \in A_L$ and for all steps s, and moreover that at least one of these inequalities is strict at each step. It follows that the algorithm terminates, at which point the length $l(a_0, s_{\mathrm{term}})$ of the chain containing a_0 is $|r^{-1}(a_0)|$ for all $a_0 \in A_L$. Thus in the expression for $\boldsymbol{M}$ at the end of the algorithm, replacing each string $\prod_{a \in r^{-1}(a_0)} a$ with $\boldsymbol{S}_k^{\mathrm{edge}}(Z_k)$, $\boldsymbol{S}_{j+1}^{\mathrm{left}}(X_j)$, or $\boldsymbol{S}_k^{\mathrm{edge}}(Z_k')$, depending on a_0, completes the proof. It only remains to show that the commutations at each step of the algorithm are possible. Only the diagonal matrices $\boldsymbol{Z}_k^{(i)}$ or $\boldsymbol{Z}_k'^{(i)}$, are "moving" during the commutation, and these matrices commute with each other. So by (†), we just need to argue that, upon commuting $a = \boldsymbol{Z}_k^{(i)}$, say, until it is adjacent to $a_0 = r(a)$, we do not need to commute $\boldsymbol{Z}_k^{(i)}$ past any $\boldsymbol{X}_{k-1}^{(i')}$. For concreteness, assume a_0 is of the form $\boldsymbol{X}_k^{(i'')}$ with $i'' \leqslant i$. The argument is analogous in the cases where a_0 is of the form $\boldsymbol{X}_k^{(i'')}$ with $i'' > i$, or $\boldsymbol{Z}_k^{(1)}$ or $\boldsymbol{Z}_k'^{(N-1)}$. The claim is clear when $i'' = i$, so assume $i'' < i$, so, in particular, $\boldsymbol{Z}_k^{(i)}$ is being commuted to the left until it is just to the right of $\boldsymbol{X}_k^{(i'')}$. Note such a $\boldsymbol{Z}_k^{(i)}$ appears as a horizontal edge lying over the top of the small downward facing triangle corresponding to $\boldsymbol{X}_k^{(i'')}$; compare with Figure 16. In the notation of Section 3.3.1, the horizontal segment $\mathrm{seg}(f(\boldsymbol{Z}_k^{(i)})) \in \mathrm{Seg}_L$ in the discrete triangle Θ_{n-1} is of the form $\overline{(k, \beta, n-1-k-\beta)(k-1, \beta, n-k-\beta)}$. Thus the key observation is that if some snake-move matrix $\boldsymbol{M}_{j_{i'}} = \boldsymbol{M}_k$ contributes $\boldsymbol{X}_{k-1}^{(i')}$ to $\boldsymbol{M}$, then either the bottom snake $\sigma^{i'}$ of the i'–snake-move is later in the snake sequence than the bottom snake σ^i of the i–snake-move, in particular $i \leqslant i'$, or the top snake $\sigma^{i'+1}$ of the i'–snake-move is earlier in the snake sequence than the bottom snake $\sigma^{i''}$ of the i''–snake-move, in particular $i' + 1 \leqslant i''$. In the former case $\boldsymbol{X}_{k-1}^{(i')}$ lies to the right of $\boldsymbol{Z}_k^{(i)}$, and in the latter case $\boldsymbol{X}_{k-1}^{(i')}$ lies to the left of $\boldsymbol{X}_k^{(i'')}$. □

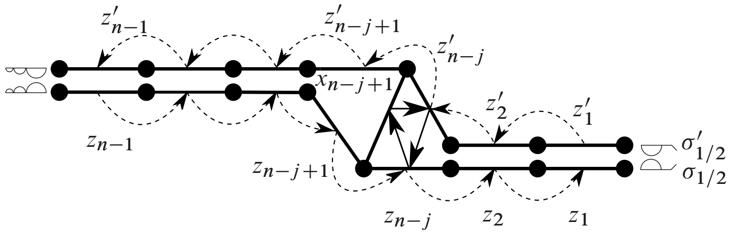


Figure 17: Right diamond snake-move algebra for $j = 2, \dots, n-1$.

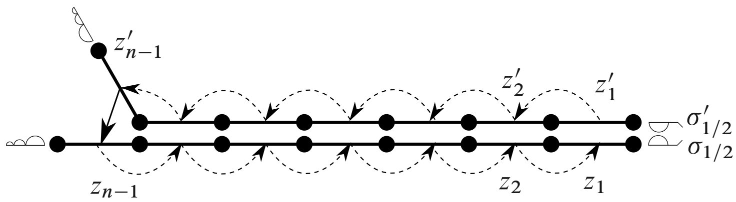


Figure 18: Right tail snake-move algebra for $j = 1$.

3.5 Setup for the quantum right matrix

We end with a few words about the proof for the quantum right matrix $\boldsymbol{M}_{\mathrm{FG}} = \boldsymbol{R}^{\omega}$, which essentially goes the same as for the left matrix.

(1) The right version of the j^{th} snake algebra $\mathcal{S}_j^{\omega}$ for $j = 1, 2, \dots, n-1$ is given by replacing the quivers of Figures 13 and 14 by the quivers shown in Figures 17 and 18.

(2) The j^{th} quantum snake-move matrix $\boldsymbol{M}_j$ of Proposition 3.3 is replaced by

$$\boldsymbol{M}_j := \left[\left(\prod_{k=1}^{n-1} \boldsymbol{S}_k^{\text{edge}}(z_k)\right)\boldsymbol{S}_j^{\text{right}}(x_{n-j+1})\left(\prod_{k=1}^{n-1} \boldsymbol{S}_k^{\text{edge}}(z'_k)\right)\right] \in \mathrm{M}_n(\mathcal{S}_j^{\omega}).$$

Note, when $j = 1$, the matrix $\boldsymbol{S}_1^{\text{right}}(x_n) = \boldsymbol{S}_1^{\text{right}}$ is well defined, despite x_n not being defined.

(3) The subalgebra $\mathcal{T}_R \subseteq \mathcal{T}_n^{\omega}$ is generated by all but the $Z_j'^{\pm 1/n}$; see Figures 11 and 12.

Appendix Proof of Proposition 3.3

Lemma A.1 *If $ZW = q^{\epsilon}WZ$ in some quantum torus $\mathcal{T}$, and if $\sum_{i=1}^{m} r_i = 0$, then*

$$\prod_{i=1}^{m}[Z^{r_i}W^{r_i}] = 1 \in \mathcal{T}.$$

Proof Using $\left(\sum_i r_i\right)^2/2 = \sum_i r_i^2/2 + \sum_{i<j} r_i r_j$, we compute

$$\begin{aligned}\prod_i[Z^{r_i}W^{r_i}] &= q^{-\epsilon\sum_i r_i^2/2}Z^{r_1}W^{r_1}Z^{r_2}W^{r_2}\cdots Z^{r_m}W^{r_m}\\ &= q^{-\epsilon\sum_i r_i^2/2}q^{-\epsilon\sum_{i<j} r_i r_j}Z^{\sum_i r_i}W^{\sum_i r_i} = q^{-\epsilon(\sum_i r_i)^2/2}\cdot Z^0\cdot W^0 = 1.\end{aligned}$$ □

Proof of Proposition 3.3 As a shorthand, put $L_{il} := (\boldsymbol{S}_j^{\mathrm{left}}(x_{j-1}))_{il}$, $\widetilde{E}_{ii} := \prod_{k=1}^{n-1}(\boldsymbol{S}_k^{\mathrm{edge}}(z_k))_{ii}$, and $\widetilde{E}'_{ii} := \prod_{k=1}^{n-1}(\boldsymbol{S}_k^{\mathrm{edge}}(z'_k))_{ii}$. By Definition 2.6, and by the structure of the matrix $\boldsymbol{M}_j$, the following three relations are needed to establish that $\boldsymbol{M}_j$ is in $\mathrm{M}_n^q(\mathcal{S}_j^\omega)$:

(A-1) $$[\widetilde{E}_{jj} L_{j(j+1)} \widetilde{E}'_{(j+1)(j+1)}][\widetilde{E}_{jj} L_{jj} \widetilde{E}'_{jj}] = q[\widetilde{E}_{jj} L_{jj} \widetilde{E}'_{jj}][\widetilde{E}_{jj} L_{j(j+1)} \widetilde{E}'_{(j+1)(j+1)}],$$

(A-2) $$[\widetilde{E}_{(j+1)(j+1)} L_{(j+1)(j+1)} \widetilde{E}'_{(j+1)(j+1)}][\widetilde{E}_{jj} L_{j(j+1)} \widetilde{E}'_{(j+1)(j+1)}] = q[\widetilde{E}_{jj} L_{j(j+1)} \widetilde{E}'_{(j+1)(j+1)}][\widetilde{E}_{(j+1)(j+1)} L_{(j+1)(j+1)} \widetilde{E}'_{(j+1)(j+1)}],$$

(A-3) $$[\widetilde{E}_{ii} L_{ii} \widetilde{E}'_{ii}][\widetilde{E}_{kk} L_{kk} \widetilde{E}'_{kk}] = [\widetilde{E}_{kk} L_{kk} \widetilde{E}'_{kk}][\widetilde{E}_{ii} L_{ii} \widetilde{E}'_{ii}] \quad \text{for } i < k.$$

We begin with (A-1). Note,

$$L_{j(j+1)} = L_{jj} = x_{j-1}^{(1-j)/n} \quad \text{and} \quad [\widetilde{E}_{jj} L_{j(j+1)} \widetilde{E}'_{(j+1)(j+1)}] = [\widetilde{E}_{jj} L_{jj} \widetilde{E}'_{jj} z_j'^{-1}].$$

So it suffices to show that commuting $z_j'^{-1}$ from left to right across $\widetilde{E}_{jj} L_{jj} \widetilde{E}'_{jj}$ contributes a factor q, equivalently, z'_j contributes q^{-1}. Indeed, in $\widetilde{E}_{jj} L_{jj} \widetilde{E}'_{jj}$ we see z'_j only interacts with $x_{j-1}^{(1-j)/n}$ with weight q^2, with $(\boldsymbol{S}_j^{\mathrm{edge}}(z_j))_{jj} = z_j^{(n-j)/n}$ with weight q^{-2}, with $(\boldsymbol{S}_{j+1}^{\mathrm{edge}}(z'_{j+1}))_{jj} = z_{j+1}'^{(n-j-1)/n}$ with weight q, and with $(\boldsymbol{S}_{j-1}^{\mathrm{edge}}(z'_{j-1}))_{jj} = z_{j-1}'^{(1-j)/n}$ with weight q^{-1}. The total exponent of q that z'_j contributes is therefore $(2(1-j) - 2(n-j) + 1(n-j-1) - 1(1-j))/n = -1$.

Next we check (A-2). Note, $L_{(j+1)(j+1)} = L_{j(j+1)} = x_{j-1}^{(1-j)/n}$ and $[\widetilde{E}_{jj} L_{j(j+1)} \widetilde{E}'_{(j+1)(j+1)}] = [z_j \widetilde{E}_{(j+1)(j+1)} L_{(j+1)(j+1)} \widetilde{E}'_{(j+1)(j+1)}]$. So it suffices to show that commuting z_j from right to left across $\widetilde{E}_{(j+1)(j+1)} L_{(j+1)(j+1)} \widetilde{E}'_{(j+1)(j+1)}$ contributes a factor q. Indeed, in

$$\widetilde{E}_{(j+1)(j+1)} L_{(j+1)(j+1)} \widetilde{E}'_{(j+1)(j+1)}$$

we see that z_j only interacts with $x_{j-1}^{(1-j)/n}$ with weight q^2 (because it's moving from right to left), with $(\boldsymbol{S}_j^{\mathrm{edge}}(z'_j))_{(j+1)(j+1)} = z_j'^{-j/n}$ with weight q^{-2}, with $(\boldsymbol{S}_{j+1}^{\mathrm{edge}}(z_{j+1}))_{(j+1)(j+1)} = z_{j+1}^{(n-j-1)/n}$ with weight q, and with $(\boldsymbol{S}_{j-1}^{\mathrm{edge}}(z_{j-1}))_{(j+1)(j+1)} = z_{j-1}^{(1-j)/n}$ with weight q^{-1}. The total exponent of q that z_j contributes is therefore $(2(1-j) - 2(-j) + 1(n-j-1) - 1(1-j))/n = +1$.

Lastly we verify (A-3). Note that the terms in $[\widetilde{E}_{ii} L_{ii} \widetilde{E}'_{ii}]$ appear in the forms $x_{j-1}^{\alpha^i}$ or $z_l^{\beta_l^i} z_l'^{\beta_l^i}$ for $l = 1, 2, \dots, n-1$. We see from the quivers in Figures 13 and 14 that terms of this form mutually commute. So

$$[\widetilde{E}_{ii} L_{ii} \widetilde{E}'_{ii}] = [x_{j-1}^{\alpha^i}] \prod_l [z_l^{\beta_l^i} z_l'^{\beta_l^i}],$$

where the right-hand side is independent of the ordering of the terms. Similarly for $[\widetilde{E}_{kk} L_{kk} \widetilde{E}'_{kk}]$. It follows that $[\widetilde{E}_{ii} L_{ii} \widetilde{E}'_{ii}]$ commutes with $[\widetilde{E}_{kk} L_{kk} \widetilde{E}'_{kk}]$ for all i and k.

It remains to check that the quantum determinant of $\boldsymbol{M}_j$ is equal to $1 \in \mathcal{S}_j^\omega$. Since $\boldsymbol{M}_j$ is in $\mathrm{M}_n^q(\mathcal{S}_j^\omega)$ and is triangular, by Remark 2.7(1) we have $\mathrm{Det}^q(\boldsymbol{M}_j) = \prod_i (\boldsymbol{M}_j)_{ii}$. As the only l such that z_l does not commute with z'_l is $l = j$, the above equation becomes

$$(\boldsymbol{M}_j)_{ii} = [\widetilde{E}_{ii} L_{ii} \widetilde{E}'_{ii}] = [z_j^{\beta_j^i} z_j'^{\beta_j^i}] x_{j-1}^{\alpha^i} \prod_{l \neq j} (z_l z'_l)^{\beta_l^i}.$$

Note, $\sum_i \alpha^i = 0$ and $\sum_i \beta_l^i = 0$ for all $l = 1, 2, \dots, n-1$ by construction of $\boldsymbol{M}_j$ (this is where the normalizing factors come in; compare with the example below Proposition 3.3). It follows that (where the last equality is by Lemma A.1),

$$\mathrm{Det}^q(\boldsymbol{M}_j) = \Bigl(\prod_i [z_j^{\beta_j^i} z_j'^{\beta_j^i}]\Bigr)\Bigl(\prod_i x_{j-1}^{\alpha^i} \prod_{l \neq j} (z_l z_l')^{\beta_l^i}\Bigr) = \Bigl(\prod_i [z_j^{\beta_j^i} z_j'^{\beta_j^i}]\Bigr) 1 = 1. \qquad \square$$

References

[1] **F Bonahon**, **H Wong**, *Quantum traces for representations of surface groups in* $\mathrm{SL}_2(\mathbb{C})$, Geom. Topol. 15 (2011) 1569–1615 MR Zbl

[2] **K A Brown**, **K R Goodearl**, *Lectures on algebraic quantum groups*, Birkhäuser, Basel (2002) MR Zbl

[3] **D Bullock**, **C Frohman**, **J Kania-Bartoszyńska**, *Understanding the Kauffman bracket skein module*, J. Knot Theory Ramifications 8 (1999) 265–277 MR Zbl

[4] **L O Chekhov**, **M Shapiro**, *Log-canonical coordinates for symplectic groupoid and cluster algebras*, Int. Math. Res. Not. 2023 (2023) 9565–9652 MR Zbl

[5] **D C Douglas**, *Classical and quantum traces coming from* $\mathrm{SL}_n(\mathbb{C})$ *and* $\mathrm{U}_q(\mathfrak{sl}_n)$, PhD thesis, University of Southern California (2020) Available at `https://www.proquest.com/docview/2436891478`

[6] **D C Douglas**, *Quantum traces for* $\mathrm{SL}_n(\mathbb{C})$*: The case* $n = 3$, J. Pure Appl. Algebra 228 (2024) art. id. 107652 MR Zbl

[7] **V V Fock**, **L O Chekhov**, *Quantum Teichmüller spaces*, Teoret. Mat. Fiz. 120 (1999) 511–528 MR Zbl In Russian; translated in Theoret. and Math. Phys. 120 (1999) 1245–1259

[8] **V V Fock**, **A B Goncharov**, *Cluster* $\mathcal{X}$*–varieties, amalgamation, and Poisson–Lie groups*, from "Algebraic geometry and number theory" (V Ginzburg, editor), Progr. Math. 253, Birkhäuser, Boston, MA (2006) 27–68 MR Zbl

[9] **V V Fock**, **A B Goncharov**, *Moduli spaces of local systems and higher Teichmüller theory*, Publ. Math. Inst. Hautes Études Sci. 103 (2006) 1–211 MR Zbl

[10] **V V Fock**, **A B Goncharov**, *Dual Teichmüller and lamination spaces*, from "Handbook of Teichmüller theory, I" (A Papadopoulos, editor), IRMA Lect. Math. Theor. Phys. 11, Eur. Math. Soc., Zürich (2007) 647–684 MR Zbl

[11] **V V Fock**, **A B Goncharov**, *Moduli spaces of convex projective structures on surfaces*, Adv. Math. 208 (2007) 249–273 MR Zbl

[12] **V V Fock**, **A B Goncharov**, *Cluster ensembles, quantization and the dilogarithm*, Ann. Sci. Éc. Norm. Supér. 42 (2009) 865–930 MR Zbl

[13] **D Gaiotto**, **G W Moore**, **A Neitzke**, *Spectral networks and snakes*, Ann. Henri Poincaré 15 (2014) 61–141 MR Zbl

[14] **M Gekhtman**, **M Shapiro**, **A Vainshtein**, *Poisson geometry of directed networks in a disk*, Selecta Math. 15 (2009) 61–103 MR Zbl

[15] **W M Goldman**, *The symplectic nature of fundamental groups of surfaces*, Adv. in Math. 54 (1984) 200–225 MR Zbl

[16] **W M Goldman**, *Invariant functions on Lie groups and Hamiltonian flows of surface group representations*, Invent. Math. 85 (1986) 263–302 MR Zbl

[17] **A Goncharov**, **L Shen**, *Quantum geometry of moduli spaces of local systems and representation theory*, preprint (2019) arXiv 1904.10491

[18] **N J Hitchin**, *Lie groups and Teichmüller space*, Topology 31 (1992) 449–473 MR Zbl

[19] **L Hollands**, **A Neitzke**, *Spectral networks and Fenchel–Nielsen coordinates*, Lett. Math. Phys. 106 (2016) 811–877 MR Zbl

[20] **R M Kashaev**, *Quantization of Teichmüller spaces and the quantum dilogarithm*, Lett. Math. Phys. 43 (1998) 105–115 MR Zbl

[21] **C Kassel**, *Quantum groups*, Graduate Texts in Math. 155, Springer (1995) MR Zbl

[22] **G Kuperberg**, *Spiders for rank* 2 *Lie algebras*, Comm. Math. Phys. 180 (1996) 109–151 MR Zbl

[23] **F Labourie**, *Anosov flows, surface groups and curves in projective space*, Invent. Math. 165 (2006) 51–114 MR Zbl

[24] **G Martone**, *Positive configurations of flags in a building and limits of positive representations*, Math. Z. 293 (2019) 1337–1368 MR Zbl

[25] **D Mumford**, **J Fogarty**, **F Kirwan**, *Geometric invariant theory*, 3rd edition, Ergebnisse der Math. 34, Springer (1994) MR Zbl

[26] **C Procesi**, *The invariant theory of $n \times n$ matrices*, Advances in Math. 19 (1976) 306–381 MR Zbl

[27] **J H Przytycki**, *Skein modules of* 3*–manifolds*, Bull. Polish Acad. Sci. Math. 39 (1991) 91–100 MR Zbl

[28] **G Schrader**, **A Shapiro**, *Continuous tensor categories from quantum groups, I: Algebraic aspects*, preprint (2017) arXiv 1708.08107

[29] **G Schrader**, **A Shapiro**, *A cluster realization of $U_q(\mathfrak{sl}_n)$ from quantum character varieties*, Invent. Math. 216 (2019) 799–846 MR Zbl

[30] **A S Sikora**, *Skein theory for* $\mathrm{SU}(n)$*–quantum invariants*, Algebr. Geom. Topol. 5 (2005) 865–897 MR Zbl

[31] **W P Thurston**, *Three-dimensional geometry and topology, 1*, Princeton Mathematical Series 35, Princeton Univ. Press (1997) MR Zbl

[32] **V G Turaev**, *Algebras of loops on surfaces, algebras of knots, and quantization*, from "Braid group, knot theory and statistical mechanics" (C N Yang, M L Ge, editors), Adv. Ser. Math. Phys. 9, World Sci., Teaneck, NJ (1989) 59–95 MR Zbl

[33] **E Witten**, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. 121 (1989) 351–399 MR Zbl

Department of Mathematics, Virginia Tech
Blacksburg, VA, United States

dcdouglas@vt.edu

Received: 14 August 2021 Revised: 3 June 2023

Algebraic & Geometric Topology **24**:5 (2024) 2571–2578
DOI: 10.2140/agt.2024.24.2571
Published: 19 August 2024

Algebraic generators of the skein algebra of a surface

RAMANUJAN SANTHAROUBANE

Let Σ be a surface with negative Euler characteristic, genus at least one and at most one boundary component. We prove that the Kauffman bracket skein algebra of Σ over the field of rational functions can be algebraically generated by a finite number of simple closed curves that are naturally associated to certain generators of the mapping class group of Σ. The action of the mapping class group on the skein algebra gives canonical relations between these generators. From this, we conjecture a presentation for a skein algebra of Σ.

57K31

1 Introduction

1.1 Main results

This paper is focused on finding algebraic generators of the Kauffman bracket skein algebra of a surface. Throughout this paper, we will refer to the Kauffman bracket skein algebra simply as the skein algebra. Let Σ be a compact oriented connected surface of genus at least one and with at most one boundary component. Moreover, we will suppose that Σ has negative Euler characteristic. We denote by $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ the skein module of $\Sigma \times [0,1]$ with coefficients in the field of rational function $\mathbb{Q}(A)$ and by $S(\Sigma)$ the skein module over $\mathbb{Z}[A^{\pm 1}]$. The module $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ is equipped with a natural product given by stacking banded links. For γ is a simple closed curve on Σ, we write γ for the element $\gamma \times \left[\frac{1}{2}, \frac{2}{3}\right]$ in $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ and we denote by t_γ the Dehn twist along γ.

Theorem 1.1 *Let $\{\gamma_j\}_{j \in I}$ be a finite set of nonseparating simple closed curves such that*

(1) *for any $i, j \in I$, the curves γ_i and γ_j intersect at most once*;

(2) *the set $\{t_{\gamma_j}\}_{j \in I}$ generates the mapping class group of Σ.*

Then $\{\gamma_j\}_{j \in I}$ generates $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ as a $\mathbb{Q}(A)$–algebra. Moreover, $\{\gamma_j\}_{j \in I}$ generates

$$S(\Sigma) \otimes \mathbb{Z}\left[A^{\pm 1}, \frac{1}{A^2 - A^{-2}}\right]$$

as a $\mathbb{Z}[A^{\pm 1}, 1/(A^2 - A^{-2})]$–algebra.

We recall that the mapping class group of Σ is $\pi_0(\mathrm{Homeo}^+(\Sigma, \partial\Sigma))$. We will now give an interpretation of some relations that should hold for the generators in the previous theorem. Let us fix $\{\gamma_j\}_{j \in I}$ a set of simple closed curves on Σ satisfying the hypothesis of Theorem 1.1. Let $\mathbb{Q}(A)\langle I \rangle$ be the free noncommutative $\mathbb{Q}(A)$–algebra generated by $\{X_j\}_{j\in I}$. The theorem says that there exists a surjective algebra homomorphism

$$\Psi\colon \mathbb{Q}(A)\langle I\rangle \to \mathcal{S}(\Sigma,\mathbb{Q}(A)) \tag{1}$$

defined by

$$\Psi(X_j)=\frac{\gamma_j}{A^2-A^{-2}} \quad \text{for all } j\in I.$$

Now for $j\in I$ and $\epsilon\in\{-1,1\}$, let $T_j^{\epsilon}\colon \mathbb{Q}(A)\langle I\rangle \to \mathbb{Q}(A)\langle I\rangle$ be the homomorphism of $\mathbb{Q}(A)$–algebra defined by

$$T_j^{\epsilon}(X_k)=\begin{cases} X_k & \text{if } \iota(\gamma_j,\gamma_k)=0,\\ \epsilon(A^{\epsilon}X_jX_k-A^{-\epsilon}X_kX_j) & \text{if } \iota(\gamma_j,\gamma_k)=1.\end{cases} \tag{2}$$

Here ι is the geometric intersection of simple closed curves. With the convention that a Dehn twist always turn to the right with respect to the orientation of the surface, we can check that Ψ exchanges the actions of the T_j and the t_j in the sense that

$$\Psi(T_j^{\epsilon}X)=t_{\gamma_j}^{\epsilon}(\Psi(X)) \quad \text{for all } X\in\mathbb{Q}(A)\langle I\rangle,\ j\in I, \text{ and } \epsilon\in\{-1,1\}.$$

Let $\Gamma(\Sigma)$ be the mapping class group of Σ and let $\overline{\Gamma}(\Sigma)$ be the group $\Gamma(\Sigma)$ modulo its center. Suppose that $I=\{1,\dots,N\}$ and let us denote each t_{γ_j} simply by t_j. Note that the map $t_j^{\epsilon}\mapsto T_j^{\epsilon}$ does not extend to an action of $\overline{\Gamma}(\Sigma)$ on $\mathbb{Q}(A)\langle I\rangle$. Indeed $\mathbb{Q}(A)\langle I\rangle$ is a free noncommutative algebra and the relations in $\overline{\Gamma}(\Sigma)$ satisfied by the Dehn twists $\{t_j^{\pm1}\}_{j\in J}$ have no reason to be satisfied by the operators $\{T_j^{\pm1}\}_{j\in J}$. Hence the relations between the $\{t_j^{\pm1}\}_{j\in J}$ give relations between the generators $\{\gamma_j\}_{j\in J}$ in $\mathcal{S}(\Sigma,\mathbb{Q}(A))$. Suppose that $\overline{\Gamma}(\Sigma)$ has the following presentation with respect to the generators $\{t_j\}_{j\in I}$:

$$\overline{\Gamma}(\Sigma)=\langle t_1,\dots,t_N \mid R_1(t_1,\dots,t_N)=\cdots=R_K(t_1,\dots,t_N)=1\rangle$$

where K is an integer and the $R_k(t_1,\dots,t_N)$ are some words in $\{t_j^{\pm1}\}_{j\in I}$. Let $\mathcal{R}$ be the bi-ideal of $\mathbb{Q}(A)\langle I\rangle$ generated by the elements

$$\begin{aligned} R_k(T_1,\dots,T_N)X_i-X_i &\quad \text{for } 1\le i\le N \text{ and } 1\le k\le K,\\ T_jT_j^{-1}X_i-X_i &\quad \text{for } 1\le i,j\le N,\\ X_iX_j-X_jX_i &\quad \text{for } \iota(\gamma_i,\gamma_j)=0.\end{aligned}$$

We define

$$\mathcal{A}(\Gamma(\Sigma))=\frac{\mathbb{Q}(A)\langle I\rangle}{\mathcal{R}}, \tag{3}$$

which is a quotient of $\mathbb{Q}(A)\langle I\rangle$ on which the actions of the $T_j^{\pm1}$ extend to a canonical action of $\overline{\Gamma}(\Sigma)$. A direct consequence of Theorem 1.1 is the following:

Corollary 1.2 *The canonical map*

$$\mathcal{A}(\Gamma(\Sigma))\to\mathcal{S}(\Sigma,\mathbb{Q}(A))$$

is surjective.

Conjecture 1.3 *There exists a presentation of $\overline{\Gamma}(\Sigma)$ for which $\mathcal{A}(\Gamma(\Sigma))$ is isomorphic to $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ as a noncommutative $\mathbb{Q}(A)$–algebra.*

1.2 Notes and references

Bullock [1999] was the first to find algebraic generators of the skein algebra of a surface. His generators are over $\mathbb{Z}[A^{\pm 1}]$ and not over $\mathbb{Q}(A)$. The number of his generators is exponential in the genus of the surface whereas here we have a linear number (by choosing the right generators of $\Gamma(\Sigma)$).

It was shown in [Przytycki and Sikora 2000] that each $\mathcal{S}(\Sigma)$ has a generating set of cardinality which is cubic in the genus of the surface.

Finite generation was also prove by Abdiel and Frohman [2017, Theorem 3.7]. Frohman and Kania-Bartoszynska [2018] studied the skein algebra when A is evaluated at a root of unity. They proved that it is generated over its center by a pair of subalgebras from pants decomposition. Their generators have some similarities with the one in the current paper.

Presentations of skein algebras of surfaces are only known in genus zero and one. Bullock and Przytycki [2000] found such a presentation for the one-holed torus, the four-holed sphere and two-holed torus. They related some of these algebras to nonstandard deformations of lie algebras.

When A is specialized to -1, it was shown by Bullock [1997] and Przytycki and Sikora [2000] that the skein algebra of a surface is isomorphic to the ring of algebraic functions of the $\mathrm{SL}(2, \mathbb{C})$ character variety of the surface. Moreover, for $A = \sqrt{-1}$, Marché [2011] gave an homological interpretation of the skein algebra of the surface. Note that the map Ψ defined in (1) is not defined if A is specialized to a 4^{th} primitive root of unity. It is possible to see that if we specialize A at a 4^{th} root of unity in the algebra $\mathcal{A}(\Gamma(\Sigma))$ we find something different from the algebras studied by Bullock, Marché, Przytycki and Sikora.

Humphries generators [1979] and Lickorish generators [1964] are examples of generators of the mapping class groups satisfying the hypothesis of Theorem 1.1. Moreover, presentations for both of these generating sets are known; we refer to book of Farb and Margalit [2012] for more details.

We consider $\Gamma(\Sigma)$ quotiented by its center because the center of the mapping class group acts trivially on the skein algebra of the surface.

Acknowledgements

I want to thank C Frohman, J Marché and G Masbaum for helpful conversations.

2 Quick review of the skein algebra

For any compact oriented manifold M (maybe with boundary), we denote by $\mathcal{S}(M)$ the Kauffman bracket skein module with coefficients in $\mathbb{Z}[A^{\pm 1}]$. We recall that it is the free $\mathbb{Z}[A^{\pm 1}]$–module generated by the

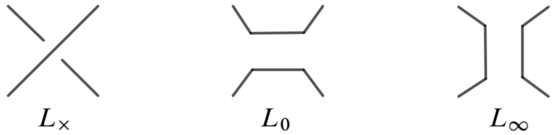


Figure 1: Kauffman triple.

set of isotopy classes of banded links in the interior of M quotiented by the following so-called skein relations. First

$$L_\times = AL_\infty + A^{-1}L_0 \tag{4}$$

where $L_\times$, L_0 and L_∞ are any three banded links in M which are the same outside a small 3–ball but differ inside as in Figure 1. In this case, the triple $L_\times, L_0, L_\infty$ is called a Kauffman triple. The second relation satisfied in $\mathcal{S}(M)$ is

$$L \cup D = -(A^2 + A^{-2})L$$

where L is any link in M and D is a trivial banded knot. We define $\mathcal{S}(M, \mathbb{Q}(A))$ to be the $\mathbb{Q}(A)$–vector space $\mathcal{S}(M) \otimes \mathbb{Q}(A)$ where the tensor product is made over $\mathbb{Z}[A^{\pm 1}]$.

Let Σ be a compact oriented connected surface; we denote by $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ the space $\mathcal{S}(\Sigma \times [0,1], \mathbb{Q}(A))$. Stacking banded links on top of each other gives $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ the structure of a $\mathbb{Q}(A)$–algebra.

A multiloop is a disjoint union of non-null-homotopic simple close curve inside Σ. For $\gamma \subset \Sigma$ a multiloop we write γ for the banded link $\gamma \times \left[\frac{1}{2}, \frac{2}{3}\right]$ in $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ and we will still call this banded link a multiloop. A well-known theorem is the following:

Theorem 2.1 *The set of isotopy classes of multiloops on Σ is a basis of the $\mathbb{Q}(A)$–vector space $\mathcal{S}(\Sigma, \mathbb{Q}(A))$.*

In particular this theorem clearly implies that simple closed curves generate $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ as an algebra. Recall that $\Gamma(\Sigma) = \pi_0(\mathrm{Homeo}^+(\Sigma, \partial\Sigma))$ acts canonically on $\mathcal{S}(\Sigma, \mathbb{Q}(A))$. If $\gamma \subset \Sigma$ is a simple closed curve, we denote by t_γ the Dehn twist along γ. We use the turn right convention for t_γ. More precisely, let $\mathcal{N} \subset \Sigma$ be an annulus neighborhood of γ, we can identify $\mathcal{N}$ with $S^1 \times [0,1]$ using an orientation preserving homeomorphism. Outside $\mathcal{N}$, the map t_γ is defined to be the identity and on $\mathcal{N}$ it is given by the map $(e^{i\theta}, t) \mapsto (e^{i(\theta - 2\pi t)}, t)$. Figure 2 helps to visualize this definition.

The following lemma can be obtained by applying the skein relation (4).

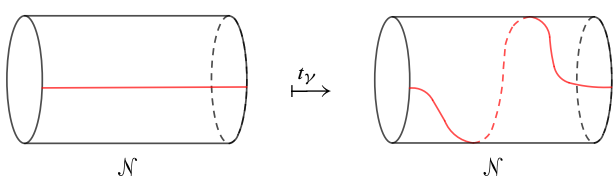


Figure 2

Lemma 2.2 *Let α and β be two simple close curves intersecting once. Then in $\mathcal{S}(\Sigma, \mathbb{Q}(A))$,*

$$t_\alpha^\epsilon(\beta) = \frac{A^\epsilon \alpha\beta - A^{-\epsilon}\beta\alpha}{\epsilon(A^2 - A^{-2})}.$$

3 Proof of Theorem 1.1

Let $\Gamma(\Sigma)$ be the mapping class group of Σ. Let $\{\gamma_j\}_{j\in I}$ be a set of simple closed curves satisfying the hypothesis of the Theorem 1.1 and let $\mathcal{B}$ be the subalgebra of $\mathcal{S}(\Sigma, \mathbb{Q}(A))$ generated by $\{\gamma_j\}_{j\in I}$.

Lemma 3.1 *$\mathcal{B}$ is stable by the action of $\Gamma(\Sigma)$.*

Proof Since $\{t_{\gamma_j}\}_{j\in I}$ generates $\Gamma(\Sigma)$, it enough to prove that for any $j, k \in I$ we have $t_{\gamma_j}^{\pm 1}(\gamma_k) \in \mathcal{B}$. If γ_j does not intersect γ_k then $t_{\gamma_j}^{\pm 1}(\gamma_k) = \gamma_k \in \mathcal{B}$. Now if γ_j intersects γ_k once then, by Lemma 2.2,

$$t_{\gamma_j}^{\pm 1}(\gamma_k) = \frac{A^{\pm 1}\gamma_j\gamma_k - A^{\mp 1}\gamma_k\gamma_j}{\pm(A^2 - A^{-2})} \in \mathcal{B}.$$ □

Lemma 3.2 *If $\mathcal{C}_0$ denotes the set of nonseparating simple closed curves then $\mathcal{C}_0 \subset \mathcal{B}$.*

Proof Let $\gamma \in \mathcal{C}_0$ and $\gamma_0 \in \{\gamma_j\}_{j\in I}$; there exists $\phi \in \Gamma(\Sigma)$ such that $\phi(\gamma_0) = \gamma$. Since γ_0 belongs to $\mathcal{B}$ which is stable by the action of $\Gamma(\Sigma)$ (see the previous lemma), we have $\gamma \in \mathcal{B}$. □

Lemma 3.3 *If $\mathcal{C}_1$ denotes the set of separating simple closed curves then $\mathcal{C}_1 \subset \mathcal{B}$.*

Proof Suppose that the genus of Σ is $g \geq 1$. Let $\delta_1, \dots, \delta_g$ be the curves in Figure 3, where δ_g is trivial when Σ does not have boundary. Let $j \in \{1, \dots, g\}$ and let z_j and z'_j be the two nonseparating curves in the torus with two boundary components defined by δ_j, δ_{j-1}, as shown in Figure 4.

By applying the skein relations, we have

$$z'_j z_j = A^2 x_j x'_j + A^{-2} y_j y'_j + \delta_j + \delta_{j-1}$$

where $\delta_0 = -A^2 - A^{-2}$ and x_j, x'_j, y_j and y'_j are nonseparating curves. By Lemma 3.2, z'_j, z_j, x_j, x'_j, y_j and y'_j are in $\mathcal{B}$, so by an induction on j we can prove that for all $1 \leq j \leq g$ we have $\delta_j \in \mathcal{B}$.

Now if γ is a separating curve, there exists $\phi \in \Gamma(\Sigma)$ and j_0 such that $\gamma = \phi(\delta_{j_0})$. Since $\mathcal{B}$ is stable by the action of $\Gamma(\Sigma)$, we have $\gamma \in \mathcal{B}$. □

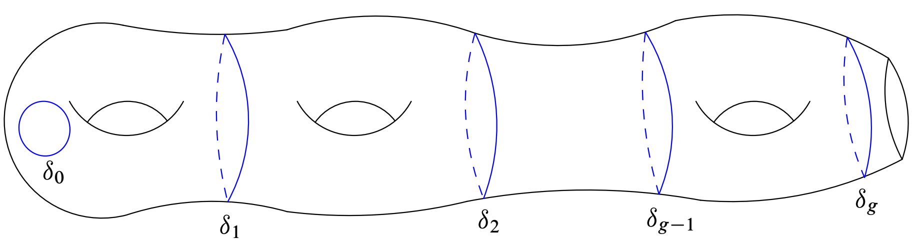


Figure 3

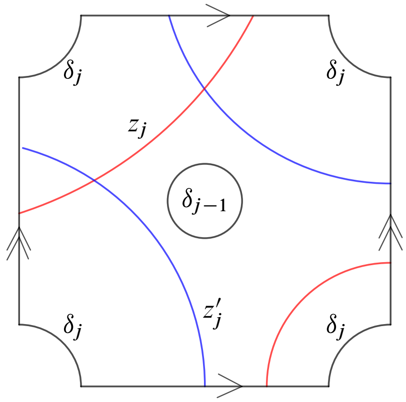


Figure 4

Proof of Theorem 1.1 $\mathscr{S}(\Sigma, \mathbb{Q}(A))$ is algebraically generated by simple closed curves. Hence, combining Lemmas 3.2 and 3.3, we can conclude the proof. Moreover, we can still conclude by Lemmas 3.1, 3.2 and 3.3 that the $\mathbb{Z}[A^{\pm 1}, 1/(A^2 - A^{-2})]$–algebra generated by $\{\gamma_j\}_{j \in I}$ is

$$S(\Sigma) \otimes \mathbb{Z}\Big[A^{\pm 1}, \frac{1}{A^2 - A^{-2}}\Big]. \qquad \square$$

4 Interpretation of the relations in the skein algebra

Let $\{\gamma_j\}_{j \in I}$ be a set of simple closed curves on Σ satisfying the hypothesis of Theorem 1.1. Recall that $\mathbb{Q}(A)\langle I \rangle$ is the free noncommutative $\mathbb{Q}(A)$–algebra generated by $\{X_j\}_{j \in I}$.

Definition 4.1 For $X, Y \in \mathbb{Q}(A)\langle I \rangle$ we define $[X, Y]_A := AXY - A^{-1}YX$.

Recall that the maps $\{T_j^\epsilon\}_{j \in I}$ are defined by (2) in the introduction. Recall also that given a presentation of $\overline{\Gamma}(\Sigma)$ with respected to the generating set $\{t_{\gamma_j}\}_{j \in I}$, we defined $\mathscr{A}(\Gamma(\Sigma))$ via (3) (see the introduction). By definition, any relation satisfied by the $\{t_{\gamma_j}\}_{j \in I}$ (which may not appear in the given presentation) gives some relation in $\mathscr{A}(\Gamma(\Sigma))$. Let us focus on the relations

$$T_j^{-1} T_i^{-1} T_j^{-1} T_i T_j T_i X_a - X_a = 0 \quad \text{for } \iota(\gamma_i, \gamma_j) = 1 \text{ and } a \in I, \tag{5}$$

$$T_j T_j^{-1} X_i - X_i = 0 \quad \text{for } i, j \in I, \tag{6}$$

$$X_i X_j - X_j X_i = 0 \quad \text{for } \iota(\gamma_i, \gamma_j) = 0. \tag{7}$$

Note that these relations hold but are not a complete set of relations in $\mathscr{A}(\Gamma(\Sigma))$. The first relation comes from the braid relations in the mapping class group.

Proposition 4.2 *In* $\mathscr{A}(\Gamma(\Sigma))$, *the relations* (5), (6) *and* (7) *are equivalent to*

$$[[X_j, X_i]_A, X_j]_A = X_i \quad \text{for } \iota(\gamma_i, \gamma_j) = 1, \tag{8}$$

$$X_i X_j - X_j X_i = 0 \quad \text{for } \iota(\gamma_i, \gamma_j) = 0. \tag{9}$$

Proof Let $i, j \in I$. Note that if $\iota(\gamma_i,\gamma_j)=0$, the relation (6) is empty and if $\iota(\gamma_i,\gamma_j)=1$, this relation gives $[[X_j, X_i]_A, X_j]_A = X_i$.

Let $i, j \in I$ such that $\iota(\gamma_i,\gamma_j)=1$. Because of (6), the relation (5) can be rewritten as

$$T_i T_j T_i X_k = T_j T_i T_j X_k$$

for all $k \in I$. It is easy to check that this relation is implied by (8) and (9). □

Remark 4.3 We did not include $T_j^{-1} T_j X_i = X_i$ in the relations defining $\mathscr{A}(\Gamma(\Sigma))$ because they give $[X_j, [X_i, X_j]_A]_A = X_i$ for $\iota(\gamma_i,\gamma_j)=1$ which is the same as $[[X_j, X_i]_A, X_j]_A = X_i$.

4.1 The case of the one-holed torus

Let Σ be a surface of genus one with one boundary component. Its mapping class group is the braid group B_3 whose presentation is $\langle t_1, t_2 \mid t_1t_2t_1 = t_2t_1t_2\rangle$. Here is t_1 is Dehn twist along the canonical meridian of Σ and t_2 is the Dehn twist around the longitude of Σ. Note that these two curves satisfy the hypothesis of Theorem 1.1. The center of this group is the group generated by $(t_1t_2t_1)^2$ and $\overline{\Gamma}(\Sigma)$ is $\mathrm{PSL}_2(\mathbb{Z})$ with presentation

$$\overline{\Gamma}(\Sigma) = \langle t_1, t_2 \mid t_1t_2t_1 = t_2t_1t_2,\ (t_1t_2t_1)^2 = 1\rangle.$$

In this case $\mathscr{A}(\Gamma(\Sigma))$ is a noncommutative algebra generated by X_1 and X_2. Because of Proposition 4.2, the only relations between X_1 and X_2 are

$$[X_1,[X_2,X_1]_A]_A = X_2, \quad [X_2,[X_1,X_2]_A]_A = X_1, \quad (T_1T_2T_1)^2 X_1 = X_1, \quad (T_1T_2T_1)^2 X_2 = X_2.$$

It is easy to check that the two last relations are implied by the two first one. Therefore,

$$\mathscr{A}(\Gamma(\Sigma)) = \langle X_1, X_2 \mid [X_1,[X_2,X_1]_A]_A = X_2,\ [X_2,[X_1,X_2]_A]_A = X_1\rangle.$$

From [Bullock 1999, Theorem 2.1] the skein module of the one-holed torus is isomorphic to $\mathscr{A}(\Gamma(\Sigma))$. Therefore Conjecture 1.3 holds for the one-holed torus.

References

[Abdiel and Frohman 2017] **N Abdiel**, **C Frohman**, *The localized skein algebra is Frobenius*, Algebr. Geom. Topol. 17 (2017) 3341–3373 MR Zbl

[Bullock 1997] **D Bullock**, *Rings of* $\mathrm{SL}_2(\mathbb{C})$*–characters and the Kauffman bracket skein module*, Comment. Math. Helv. 72 (1997) 521–542 MR Zbl

[Bullock 1999] **D Bullock**, *A finite set of generators for the Kauffman bracket skein algebra*, Math. Z. 231 (1999) 91–101 MR Zbl

[Bullock and Przytycki 2000] **D Bullock**, **J H Przytycki**, *Multiplicative structure of Kauffman bracket skein module quantizations*, Proc. Amer. Math. Soc. 128 (2000) 923–931 MR Zbl

[Farb and Margalit 2012] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[Frohman and Kania-Bartoszynska 2018] **C Frohman**, **J Kania-Bartoszynska**, *The structure of the Kauffman bracket skein algebra at roots of unity*, Math. Z. 289 (2018) 889–920 MR Zbl

[Humphries 1979] **S P Humphries**, *Generators for the mapping class group*, from "Topology of low-dimensional manifolds" (R A Fenn, editor), Lecture Notes in Math. 722, Springer (1979) 44–47 MR Zbl

[Lickorish 1964] **W B R Lickorish**, *A finite set of generators for the homeotopy group of a 2–manifold*, Proc. Cambridge Philos. Soc. 60 (1964) 769–778 MR Zbl

[Marché 2011] **J Marché**, *The Kauffman skein algebra of a surface at* $\sqrt{-1}$, Math. Ann. 351 (2011) 347–364 MR Zbl

[Przytycki and Sikora 2000] **J H Przytycki**, **A S Sikora**, *On skein algebras and* $\mathrm{Sl}_2(\mathbb{C})$*–character varieties*, Topology 39 (2000) 115–148 MR Zbl

Laboratoire Mathématiques d'Orsay, CNRS, Université Paris-Saclay
Orsay, France

Current address: *Department of Mathematics, University of Virginia*
Charlottesville, VA, United States

ramanujan.santharoubane@gmail.com

Received: 4 March 2022 Revised: 21 November 2022

Algebraic & Geometric Topology **24**:5 (2024) 2579–2618
DOI: 10.2140/agt.2024.24.2579
Published: 19 August 2024

Bundle transfer of L–homology orientation classes for singular spaces

MARKUS BANAGL

We consider transfer maps on ordinary homology, bordism of singular spaces and homology with coefficients in Ranicki's symmetric L–spectrum, associated to block bundles with closed oriented PL manifold fiber and compact polyhedral base. We prove that if the base polyhedron is a Witt space, for example a pure-dimensional compact complex algebraic variety, then the symmetric L–homology orientation of the base, constructed by Laures, McClure and the author, transfers to the L–homology orientation of the total space. We deduce from this that the Cheeger–Goresky–MacPherson L–class of the base transfers to the product of the L–class of the total space with the cohomological L–class of the stable vertical normal microbundle.

55N33, 55R12, 57N80, 57Q50, 57R20; 57Q20

1 Introduction

To a fiber bundle $p\colon X \to B$ whose structure group is a compact Lie group acting smoothly on the compact smooth d–dimensional manifold fiber F, and whose base space B is a finite complex, Becker and Gottlieb associate in [8] a transfer homomorphism $p^!\colon H_n(B) \to H_{n+d}(X)$. Boardman discusses this transfer and several closely related constructions, such as the Umkehr map and pullback transfers, in [9]. Let $L^*(\alpha)$ denote the cohomological Hirzebruch L–class of a vector bundle α, and for a smooth closed oriented manifold M with tangent bundle TM, let $L_*(M) \in H_*(M;\mathbb{Q})$ denote the Poincaré dual of $L^*(TM)$. Suppose that F is oriented and the structure group of p acts in an orientation-preserving manner. If the base B of the fiber bundle is a smooth closed oriented manifold M, then

$$p^! L_*(M) = L^*(T_p)^{-1} \cap L_*(X), \tag{1}$$

where T_p is the vertical tangent bundle of p. This is a straightforward consequence of the bundle isomorphism $TX \cong p^*TM \oplus T_p$, naturality and the Whitney sum formula for the cohomological L–class, multiplicative properties of the transfer, and the fact that $p^!$ maps the fundamental class of the base to the fundamental class of the total space.

If the base B is a singular pseudomanifold then the above argument does not apply. On the other hand, intersection homology methods still allow for the construction of a homological L–class $L_*(B) \in H_*(B;\mathbb{Q})$ for many types of compact pseudomanifolds B: When B allows for a stratification with only even-codimensional strata, for example a pure-dimensional compact complex algebraic variety, $L_*(B)$ has been defined by Goresky and MacPherson in [21]. This construction has been extended by P Siegel [44] to Witt spaces, ie oriented polyhedral pseudomanifolds that may have strata of odd codimension such that the middle-dimensional middle-perversity rational intersection homology of the corresponding links vanishes. In [1; 2], the author has yet more generally defined $L_*(B)$ for topologically stratified spaces B that allow for Lagrangian structures along strata of odd codimension. A local definition of L–classes for triangulated pseudomanifolds with piecewise-flat metric was given by Cheeger [16] in terms of η–invariants of links. As for manifolds, the L–class of singular spaces plays an important role in the topological classification of such spaces, as shown by Cappell and Weinberger in [13] and by Weinberger in [49].

Let F be a closed oriented d–dimensional PL manifold, B a compact polyhedron and ξ an oriented PL F–block bundle over B; see Casson [15]. Oriented PL F–fiber bundles $p: X \to B$ are a special case. Block bundles, and hence our results here, do not require a locally trivial projection map p. Then ξ still admits a transfer homomorphism

$$\xi^!: H_n(B) \to H_{n+d}(X).$$

See Ebert and Randal-Williams [19] and Section 5. Furthermore, ξ possesses a stable vertical normal PL microbundle ν_ξ; see Hebestreit, Land, Lück and Randal-Williams [25] and Section 2. Here we develop methods that yield, among other results, formula (1) for F–block bundles over Witt spaces B:

Theorem 8.1 *Let B be a closed Witt space (eg a pure-dimensional compact complex algebraic variety) and let F be a closed oriented PL manifold. Let ξ be an oriented PL F–block bundle over B with total space X and oriented stable vertical normal microbundle ν_ξ over X. Then X is a Witt space, and the associated block bundle transfer $\xi^!$ sends the Cheeger–Goresky–MacPherson–Siegel L–class of B to the product*

$$\xi^! L_*(B) = L^*(\nu_\xi) \cap L_*(X). \tag{2}$$

Note that since the cohomological class $L^*(\nu_\xi)$ is invertible, this formula yields a method for computing the Cheeger–Goresky–MacPherson L–class of the total space.

Our method of proof rests on the geometric description of PL cobordism provided by Buoncristiano, Rourke and Sanderson [12] in terms of mock bundles. We construct a transfer $\xi^!\colon E_n(B) \to E_{n+d}(X)$ for any module spectrum E over the Thom spectrum MSPL of oriented PL bundle theory. In addition to the transfer on ordinary homology, this yields transfer homomorphisms on Ranicki's homology with coefficients in the symmetric $\mathbb{L}^\bullet$–spectrum and on Witt bordism theory, Ω_*^{Witt}. We describe the latter transfer geometrically as a pullback transfer and use this, together with mock bundle theory, to show that the Witt bordism transfer sends the fundamental class $[B]_{\mathrm{Witt}} \in \Omega_*^{\mathrm{Witt}}(B)$ to the fundamental class $[X]_{\mathrm{Witt}} \in \Omega_*^{\mathrm{Witt}}(X)$; see Proposition 6.8. Using work of Laures, McClure and the author [7], which provides a map of ring spectra MWITT $\to \mathbb{L}^\bullet(\mathbb{Q})$, where MWITT represents Witt-bordism, as well as a fundamental class $[B]_{\mathbb{L}} \in \mathbb{L}^\bullet(\mathbb{Q})_*(B)$, we then show:

Theorem 7.1 *Let B be a closed Witt space of dimension n and let F be a closed oriented* PL *manifold of dimension d. Let ξ be an oriented* PL *F–block bundle over B with total space X. Then the $\mathbb{L}^\bullet$–homology block bundle transfer*

$$\xi^!\colon \mathbb{L}^\bullet(\mathbb{Q})_n(B) \to \mathbb{L}^\bullet(\mathbb{Q})_{n+d}(X)$$

maps the $\mathbb{L}^\bullet(\mathbb{Q})$–homology fundamental class of B to the $\mathbb{L}^\bullet(\mathbb{Q})$–homology fundamental class of X,

$$\xi^![B]_{\mathbb{L}} = [X]_{\mathbb{L}}.$$

The result on Cheeger–Goresky–MacPherson L–classes is then deduced from an explicit formula for the transfer by tensoring with the rationals. For a PL F–fiber bundle $p\colon X \to B$ over a PL manifold base B, the formula

$$p^![B]_{\mathbb{L}} = [X]_{\mathbb{L}} \in \mathbb{L}^\bullet(\mathbb{Z})_{n+d}(X)$$

was stated by Lück and Ranicki in [32]. The behavior of the L–class for singular spaces under transfers associated to finite-degree covering projections has already been clarified in [4], where we showed that for a closed oriented Whitney stratified pseudomanifold B admitting Lagrangian structures along strata of odd codimension (eg B Witt) and $p\colon X \to B$ an orientation-preserving covering map of finite degree, the L–class of B transfers to the L–class of the cover, ie

$$p^! L_*(B) = L_*(X).$$

For the Witt case, from our perspective this is a special case of (2).

An inclusion $g\colon Y \hookrightarrow X$ of stratified spaces is called normally nonsingular if Y possesses a tubular neighborhood in X that can be equipped with the structure of a real vector bundle; see eg work of Goresky and MacPherson [23] and the author [5]. An oriented normally nonsingular inclusion g of real codimension c has a Gysin map

$$g^!\colon H_*(X;\mathbb{Q}) \to H_{*-c}(Y;\mathbb{Q})$$

on ordinary homology,

$$g^!\colon \mathbb{L}^\bullet(\mathbb{Q})_*(X) \to \mathbb{L}^\bullet_{*-c}(\mathbb{Q})(Y)$$

on $\mathbb{L}^\bullet(\mathbb{Q})$–homology, and

$$g^!\colon \Omega_*^{\mathrm{Witt}}(X) \to \Omega_{*-c}^{\mathrm{Witt}}(Y)$$

on Witt bordism. In [5], we showed that if g is a normally nonsingular inclusion of closed oriented even-dimensional piecewise-linear Witt pseudomanifolds, for example pure-dimensional compact complex algebraic varieties, then

$$g^! L_*(X) = L^*(\nu_g) \cap L_*(Y), \tag{3}$$
$$g^![X]_{\mathbb{L}} = [Y]_{\mathbb{L}}, \quad g^![X]_{\mathrm{Witt}} = [Y]_{\mathrm{Witt}},$$

where ν_g is the normal bundle of g. These formulae have been applied in [5] to compute the Cheeger–Goresky–MacPherson L–class of certain singular Schubert varieties. No previous computations of such classes seem to be available in the literature. Together with the bundle transfer formula (2), this makes it possible to compute the transfer of the Cheeger–Goresky–MacPherson L–class associated to a normally nonsingular map, that is, a map which can be factored as a composition of a normally nonsingular inclusion, followed by the projection of an oriented PL F–fiber bundle ξ with closed PL manifold fiber F; see Section 9.

For complex algebraic, possibly singular, varieties X, Brasselet, Schürmann and Yokura [11] introduced Hodge-theoretic intersection Hirzebruch characteristic classes $\mathrm{IT}_{y,*}(X)$, which agree with $L_*(X)$ for $y = 1$ and X nonsingular or, more generally, a rational homology manifold; see de Bobadilla and Pallarés [10]. Using results of Schürmann [43] and Maxim and Schürmann [34], we established an algebraic version of (3) for $\mathrm{IT}_{1,*}$ in a context of appropriately normally nonsingular regular algebraic embeddings [5, Theorem 6.30]. Similarly, we expect IT_{1*} to satisfy a relation analogous to (2) for smooth algebraic morphisms $p\colon X \to B$, where ν_p would now be inverse to the algebraic relative tangent bundle $T_{X/B}$. Such a relation, together with our results here, then enable further comparison between the Hodge-theoretic class IT_{1*} and the topological class L_*. The aforementioned normally nonsingular maps form a topological parallel to the algebraic concept of a local complete intersection morphism, ie a morphism of varieties that can be factored into a closed regular embedding and a smooth morphism. Hence our results impact the behavior of topological characteristic classes under transfers associated to local complete intersection morphisms.

Acknowledgments This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German research foundation) research grant 495696766.

2 Stable vertical normal block bundles

Block bundles with manifold fiber over compact polyhedra admit stable vertical normal closed disc block bundles; see eg [19; 25], as well as [12, page 83] for the more general mock bundle situation. We will use the vertical normal block bundle later in the description of the Umkehr map, and thus recall the construction in the form we need.

Let F be a closed oriented PL manifold of dimension d and let K be a finite ball complex with associated polyhedron $B = |K|$. (The polyhedron B is not required to be a manifold.) Let ξ be an oriented PL

F–block bundle over K (see Casson [15]) with total space $X = E(\xi)$. Let b denote the dimension of B so that $\dim X = d + b$. The block of ξ over a cell $\sigma \in K$ will be denoted by $\xi(\sigma)$. For every σ, there is a block-preserving PL homeomorphism $\xi(\sigma) \cong F \times \sigma$. Thus the blocks of ξ are compact PL manifolds with boundary

$$\xi(\partial\sigma) := \bigcup_{\tau \in \partial\sigma} \xi(\tau).$$

Over the interiors $\mathring{\sigma}$ of cells, we set $\xi(\mathring{\sigma}) := \xi(\sigma) - \xi(\partial\sigma)$.

In order to construct a stable vertical normal PL block bundle (and hence a stable PL microbundle, since $\mathrm{BSPL} \simeq \mathrm{B}\widetilde{\mathrm{SPL}}$) for ξ, choose a block-preserving PL embedding

$$\theta\colon X \hookrightarrow \mathbb{R}^s \times B$$

for sufficiently large $s > 2d + b + 1$, ie a PL embedding such that

$$\theta(\xi(\mathring{\sigma})) \subset \mathbb{R}^s \times \mathring{\sigma}$$

and

$$\theta|\colon (\xi(\sigma), \xi(\partial\sigma)) \to (\mathbb{R}^s \times \sigma, \mathbb{R}^s \times \partial\sigma)$$

is a locally flat PL embedding of manifolds for every simplex $\sigma \subset K$. One way to obtain such an embedding is to choose first a PL embedding $e\colon X \hookrightarrow \mathbb{R}^s$. By Casson [15, Lemma 6, page 43], ξ can be equipped with a choice of block fibration $p\colon X \to B$. This is a PL map such that $\xi(\sigma) = p^{-1}(\sigma)$ for every cell $\sigma \in K$. Then $\theta := (e, p)\colon X \hookrightarrow \mathbb{R}^s \times B$ is a block-preserving PL embedding. (The local flatness is ensured by requiring the codimension to be at least 3.) Another method to construct a block-preserving embedding θ is by induction over the cells $\sigma \in K$, starting with the 0–cells σ^0 and embeddings $\theta\colon \xi(\sigma^0) \cong F \subset \mathbb{R}^s \times \sigma^0 \cong \mathbb{R}^s$. These are then extended to proper embeddings of manifolds-with-boundary $\theta\colon \xi(\sigma^1) \subset \mathbb{R}^s \times \sigma^1$ for every 1–cell σ^1 in K, etc. As in [41], an embedding $j\colon M \to Q$ of manifolds is *proper* if $j^{-1}(\partial Q) = \partial M$.

Recall that one says that a PL embedding $j\colon A \to P$ of polyhedra possesses a *normal* PL *closed disc block bundle* if there exists a regular neighborhood N of $j(A)$ in P such that N is the total space of a PL closed disc block bundle over $j(A)$ whose zero section embedding agrees with the inclusion $j(A) \subset N$.

Proposition 2.1 *Let ξ be an F–block bundle over a finite cell complex K with polyhedron $B = |K|$, where F is a closed* PL *manifold. A block-preserving* PL *embedding $\theta\colon X \to \mathbb{R}^s \times B$ of the total space X of ξ possesses a normal* PL *closed $(s-d)$–disc block bundle ν_θ over $\theta(X)$. If ξ is oriented, then ν_θ is canonically oriented.*

Proof We begin by constructing a particular regular neighborhood N of $\theta(X)$ in $\mathbb{R}^s \times B$ such that N is compatible with the blocks $\theta(\xi(\sigma))$ and $\mathbb{R}^s \times \sigma$ for all cells $\sigma \in K$. There exists a locally finite simplicial complex L with subcomplexes $T, L_\sigma \subset L$ ($\sigma \in K$) such that

(i) $|L| = \mathbb{R}^s \times B$,

(ii) $\theta(X) = |T|$,

(iii) for every cell σ in K, $\mathbb{R}^s \times \sigma = |L_\sigma|$, and

(iv) each simplex of L meets T in a single face or not at all.

It follows from (ii) and $\theta(\xi(\sigma)) \subset \mathbb{R}^s \times \sigma$ that the compact manifold $M_\sigma := \theta(\xi(\sigma))$ is triangulated by $L_\sigma \cap T$. The boundary of M_σ is triangulated by the subcomplex $L_{\partial\sigma} \cap T$ of L, where $L_{\partial\sigma}$ is the subcomplex of L given by

$$L_{\partial\sigma} = \bigcup_{\tau \in \partial\sigma} L_\tau.$$

Furthermore, (iv) implies that each simplex of L_σ meets $L_\sigma \cap T$ in a single face or not at all. Let $f: L \to [0,1] = \Delta^1$ be the unique simplicial map such that $f^{-1}(0) = |T|$. Then the preimage

$$N := f^{-1}\big[0, \tfrac{1}{2}\big] \subset \mathbb{R}^s \times B$$

is a regular neighborhood of $\theta(X)$ in $\mathbb{R}^s \times B$. The intersection $Q_\sigma := N \cap (\mathbb{R}^s \times \sigma)$ is a regular neighborhood of the manifold M_σ in the manifold $\mathbb{R}^s \times \sigma$. This regular neighborhood meets the boundary $\mathbb{R}^s \times \partial\sigma$ transversely, ie $N \cap (\mathbb{R}^s \times \partial\sigma)$ is a regular neighborhood of $\theta(\xi(\sigma)) \cap (\mathbb{R}^s \times \partial\sigma) = \theta(\xi(\partial\sigma))$ in $\mathbb{R}^s \times \partial\sigma$. The boundary of the compact manifold Q_σ is described by

$$\partial Q_\sigma = \big(f^{-1}\big(\tfrac{1}{2}\big) \cap (\mathbb{R}^s \times \sigma)\big) \cup \bigcup_{\tau \in \partial\sigma} Q_\tau \tag{4}$$

and M_σ is a proper submanifold of Q_σ.

We will construct a PL closed disc block bundle ν_θ over $\theta(X)$ by induction on the cells σ of K. The total space $E(\nu_\theta)$ of ν_θ is given by $E(\nu_\theta) := N$. Given a nonnegative integer n, we set

$$L_n := \bigcup_{\sigma} L_\sigma,$$

where the union is taken over all cells $\sigma \in K$ of dimension at most n. The corresponding polyhedron is $|L_n| = \mathbb{R}^s \times B^n$, where B^n denotes the n–skeleton of B. Set

$$Q_n := \bigcup_{\sigma} Q_\sigma \subset \mathbb{R}^s \times B^n,$$

where the union is taken over all cells $\sigma \in K$ of dimension at most n, so that

$$Q_n = N \cap (\mathbb{R}^s \times B^n).$$

For $n = b = \dim B$ we have $B^n = B$ and thus $Q_b = N$.

Let σ be a 0–cell of K. By [41, Theorem 4.3, page 16], there is a disc block bundle ν_σ over the complex $L_\sigma \cap T$ with total space $E(\nu_\sigma) = Q_\sigma$. Then the collection of blocks $\nu_\sigma(\beta)$, for $\beta \in L_\sigma \cap T$, of the bundles ν_σ endow Q_0 with the structure of a disc block bundle ν_0 over $L_0 \cap T$. Assume inductively that a block bundle ν_{n-1} over $L_{n-1} \cap T$ with total space

$$E(\nu_{n-1}) = Q_{n-1}$$

has been constructed such that for all cells $\sigma \in K$ with $\dim \sigma < n$, the restriction ν_σ of ν_{n-1} to the subcomplex $L_\sigma \cap T \subset L_{n-1} \cap T$ has total space $E(\nu_\sigma) = Q_\sigma$. Let $\sigma \in K$ be an n–cell. The pair $(M_\sigma, \partial M_\sigma)$ is triangulated by $(L_\sigma \cap T, L_{\partial\sigma} \cap T)$. Using the description (4) of ∂Q_σ, we have

$$E(\nu_{n-1}|_{L_{\partial\sigma}\cap T}) = \bigcup_{\tau\in\partial\sigma} E(\nu_{n-1}|_{L_\tau \cap T}) = \bigcup_{\tau} Q_\tau \subset \partial Q_\sigma.$$

Since Q_σ is a regular neighborhood of the compact manifold M_σ, there exists, again by [41, Theorem 4.3], a disc block bundle ν_σ over $L_\sigma \cap T$ with total space $E(\nu_\sigma) = Q_\sigma$ such that

$$\nu_\sigma|_{L_{\partial\sigma}\cap T} = \nu_{n-1}|_{L_{\partial\sigma}\cap T}.$$

Then the collection of blocks $\nu_\sigma(\beta)$, for $\beta \in L_\sigma \cap T$ and $\dim\sigma \le n$, of the bundles ν_σ endow $Q_n = \bigcup_{\dim\sigma\le n} Q_\sigma$ with the structure of a disc block bundle ν_n over $L_n \cap T$. By construction,

$$E(\nu_n|_{L_\sigma\cap T}) = E(\nu_\sigma) = Q_\sigma$$

for all $\sigma \in K$ and $\dim\sigma \le n$. This concludes the inductive step. For $n = b$, $\nu_\theta := \nu_b$ is a PL closed disc block bundle over $L_b \cap T = T$ with total space $E(\nu) = N$.

If P is an oriented codimension-0 submanifold of the boundary ∂M of an oriented manifold M, then the incidence number $\epsilon(P, M)$ is defined to be $+1$ if the orientation on P induced by the orientation of M agrees with the given orientation of P, and -1 otherwise. Suppose that ξ is oriented as an F–block bundle. Then K is an oriented cell complex and each block $\xi(\sigma)$ is oriented (as a manifold) so that $\epsilon(\xi(\tau), \xi(\sigma)) = \epsilon(\tau, \sigma)$ whenever τ is a codimension-1 face of a cell $\sigma \in K$. Requiring θ to be orientation preserving, we obtain orientations of all M_σ such that $\epsilon(M_\tau, M_\sigma) = \epsilon(\tau, \sigma)$. Give every $\mathbb{R}^s \times \sigma$ the product orientation determined by the orientation of the cell σ and the standard orientation of $\mathbb{R}^s$. Then the inclusion embeddings of oriented manifolds $M_\sigma \subset \mathbb{R}^s \times \sigma$ induce unique orientations of the normal bundles ν_σ. The above incidence number relation implies that these orientations fit together to give an orientation of ν_θ. □

Using the PL homeomorphism $\theta\colon X \to \theta(X)$, we may think of ν_θ as a bundle over X.

Proposition 2.2 *For s sufficiently large (compared to the dimension of X), the equivalence class of the disc block bundle ν_θ as constructed in Proposition 2.1 is independent of the choice of blockwise embedding $\theta\colon X \hookrightarrow \mathbb{R}^s \times B$, and thus only depends on the F–block bundle ξ.*

Proof Let $\theta, \theta'\colon X \hookrightarrow \mathbb{R}^s \times B$ be ξ–block-preserving PL embeddings. These give rise to vertical normal disc block bundles ν_θ and $\nu_{\theta'}$. The idea is to construct a $(\xi\times I)$–block-preserving concordance $\bar\theta\colon X \times I \hookrightarrow \mathbb{R}^s \times B \times I$ between θ and θ' and then apply Proposition 2.1 to endow inductively a suitable regular neighborhood of the image of the concordance with the structure of a disc block bundle, extending the disc block bundles ν_θ and $\nu_{\theta'}$. This implies that ν_θ and $\nu_{\theta'}$ are equivalent.

Observe that the equivalence class of the block bundle ν_θ does not change under passage to a simplicial subdivision L_0 of the complex L used in the proof of Proposition 2.1. This is a consequence of M Cohen's uniqueness theorem for regular neighborhoods in general polyhedra [17, Theorem 3.1, page 196] and Rourke and Sanderson's uniqueness theorem for disc block bundle structures [41, Theorem 4.4, page 16].

The cylinder $\overline{B} = B \times I$ is the polyhedron of the product cell complex $\overline{K} = K \times I$, where I carries the minimal cell structure. Let $\overline{X} = X \times I$. The product block bundle $\bar{\xi} := \xi \times I$ is an F–block bundle over the cell complex $\overline{K}$ with total space $E(\xi \times I) = \overline{X}$ and blocks $(\xi \times I)(\sigma \times \tau) = \xi(\sigma) \times \tau$, where σ is a cell in K and τ a cell of I. For sufficiently large s, by induction over the finitely many cells σ in $\overline{K}$, there exists a PL embedding

$$\bar{\theta}\colon \overline{X} \to \mathbb{R}^s \times \overline{B} = (\mathbb{R}^s \times B) \times I$$

such that $\bar{\theta}_0 = \theta \times 0,\ \bar{\theta}_1 = \theta' \times 1,\ \bar{\theta}(\bar{\xi}(\mathring{\sigma})) \subset \mathbb{R}^s \times \mathring{\sigma}$ and

$$\bar{\theta}|\colon (\bar{\xi}(\sigma), \bar{\xi}(\partial\sigma)) \to (\mathbb{R}^s \times \sigma, \mathbb{R}^s \times \partial\sigma)$$

is a locally flat PL embedding of manifolds for every cell $\sigma \subset \overline{K}$. Thus $\bar{\theta}$ is a block-preserving concordance between θ and θ' satisfying

$$\bar{\theta}(X \times I) \cap (\mathbb{R}^s \times B \times 0) = \theta(X) \times 0 \quad \text{and} \quad \bar{\theta}(X \times I) \cap (\mathbb{R}^s \times B \times 1) = \theta'(X) \times 1.$$

There exists a locally finite simplicial complex $\bar{L}$ with subcomplexes $\overline{T}, \bar{L}_\sigma \subset \bar{L}$, for $\sigma \in \overline{K}$, such that

(i) $|\bar{L}| = \mathbb{R}^s \times \overline{B}$,

(ii) the complexes $L \times 0$ and $L' \times 1$ used in constructing ν_θ and $\nu_{\theta'}$ are both subcomplexes of $\bar{L}$ such that

$$|L| = \mathbb{R}^s \times B \times 0 \quad \text{and} \quad |L'| = \mathbb{R}^s \times B \times 1,$$

(iii) $\bar{\theta}(\overline{X}) = |\overline{T}|$,

(iv) for every cell σ in $\overline{K}$,

$$\mathbb{R}^s \times \sigma = |\bar{L}_\sigma|,$$

(v) each simplex of $\bar{L}$ meets $\overline{T}$ in a single face or not at all.

(To achieve the fullness property, (v), it may be necessary to subdivide $L \times 0$ and $L' \times 1$, but we have observed earlier that this does not change the equivalence class of ν_θ or $\nu_{\theta'}$. Thus we may call the subdivisions $L \times 0$ and $L' \times 1$ again.) Let $f\colon L \to [0,1]$ be the unique simplicial map such that $f^{-1}(0) = |T| = \theta(X)$. The disc block bundle ν_θ over T has total space

$$E(\nu_\theta) = N = f^{-1}\big[0, \tfrac{1}{2}\big] \subset \mathbb{R}^s \times B,$$

a regular neighborhood of $\theta(X)$ in $\mathbb{R}^s \times B$. Let $f'\colon L' \to [0,1]$ be the unique simplicial map such that $f'^{-1}(0) = |T'| = \theta'(X)$. The disc block bundle $\nu_{\theta'}$ over T' has total space

$$E(\nu_{\theta'}) = N' = f'^{-1}\big[0, \tfrac{1}{2}\big] \subset \mathbb{R}^s \times B,$$

a regular neighborhood of $\theta'(X)$ in $\mathbb{R}^s \times B$. Let $\bar{f}\colon \bar{L} \to [0,1]$ be the unique simplicial map such that $\bar{f}^{-1}(0) = |\overline{T}| = \bar{\theta}(X \times I)$. By Proposition 2.1 and its proof, the regular neighborhood

$$\overline{N} := \bar{f}^{-1}\big[0, \tfrac{1}{2}\big] \subset \mathbb{R}^s \times B \times I$$

of $\bar{\theta}(X \times I)$ is the total space $E(\nu_{\bar{\theta}}) = \overline{N}$ of a PL disc block bundle $\nu_{\bar{\theta}}$ over $\overline{T}$ such that

$$\nu_{\bar{\theta}}|_{L\times 0} = \nu_\theta \quad \text{and} \quad \nu_{\bar{\theta}}|_{L'\times 1} = \nu_{\theta'}.$$

Thus, pulling back $\nu_{\bar{\theta}}$ to $X \times I$ along $\bar{\theta}$, we obtain a PL disc block bundle over $X \times I$ whose restriction to $X \times 0$ is ν_θ and whose restriction to $X \times 1$ is $\nu_{\theta'}$. This implies that ν_θ and $\nu_{\theta'}$ are equivalent as disc block bundles. □

The oriented normal block bundle ν_θ provided by Proposition 2.1 is classified by a map

$$X \to \mathrm{B}\widetilde{\mathrm{SPL}}_{s-d}\,.$$

If s is sufficiently large, then by Proposition 2.2 the homotopy class of this map does not depend on the choice of blockwise embedding θ. We denote the resulting disc block bundle equivalence class by ν_ξ and refer to it as the *stable vertical normal block bundle* of ξ. The restriction $s > b + 2d + 1$ ensures that the block bundle ν_ξ is in the stable range, there exists a unique (up to equivalence) oriented PL microbundle μ over X whose underlying block bundle is ν_ξ, and this microbundle is also in the stable range: since $\dim X = d + b < (s-d) - 1$, the natural map

$$[X, \mathrm{BSPL}_{s-d}] \cong [X, \mathrm{B}\widetilde{\mathrm{SPL}}_{s-d}]$$

is a bijection. We will refer to μ as the *stable vertical normal microbundle* of ξ.

Example 2.3 For the trivial F–block bundle ξ with total space $X = F \times B$, we may choose a PL embedding $\theta_F\colon F \hookrightarrow \mathbb{R}^s$, for s large, and take $\theta\colon X \hookrightarrow \mathbb{R}^s \times B$ to be $\theta = \theta_F \times \mathrm{id}_B\colon F \times B \hookrightarrow \mathbb{R}^s \times B$, which is ξ–block preserving. Let ν_F be the (stable) normal disc block bundle of θ_F and μ_F its unique lift to a PL microbundle. Then the stable vertical normal block bundle ν_ξ is represented by $\nu_\theta = \mathrm{pr}_1^*\,\nu_F$ and the stable vertical normal microbundle is $\mu = \mathrm{pr}_1^*\,\mu_F$, where $\mathrm{pr}_1\colon F \times B \to F$ is the factor projection.

Example 2.4 If F is a point, then $X = B$ and we may take $\theta\colon X = B \hookrightarrow \mathbb{R}^s \times B$ to be $\theta(x) = (0, x)$. The stable vertical normal block bundle ν_ξ and the stable vertical normal microbundle μ are both trivial.

3 The PL Umkehr map

Given an oriented F–block bundle ξ with nonsingular fiber F over a compact polyhedron and a module spectrum E over the Thom spectrum MSPL, we will construct a transfer homomorphism $\xi^!\colon E_n(B) \to E_{n+d}(X)$. This will be done in Section 5 by composing suspension, the PL Umkehr map $T(\xi)$ and the Thom homomorphism Φ. The Umkehr map will be constructed in the present section, and the Thom homomorphism in the next.

As in Section 2, let F be a closed oriented PL manifold of dimension d and let K be a finite ball complex with associated polyhedron $B = |K|$. Let ξ be an oriented PL F–block bundle over K with total space $X = E(\xi)$. Fix a block-preserving PL embedding $\theta\colon X \hookrightarrow \mathbb{R}^s \times B$ for sufficiently large s, and let us briefly write ν for the vertical normal disc block bundle ν_θ given by Proposition 2.1. As discussed in Section 2, there is a unique PL microbundle μ whose underlying block bundle is ν. The total space $E(\nu) = N$ is a ξ–block-preserving regular neighborhood of $\theta(X)$ in $B \times \mathbb{R}^s$. Let $\dot{\nu}$ denote the sphere block bundle of ν and write ∂N for the total space of $\dot{\nu}$. Let

$$\mathrm{Th}(\nu) := N \cup_{\partial N} \mathrm{cone}(\partial N)$$

be the Thom space of ν. The cone point in $\mathrm{Th}(\nu)$ will be denoted by ∞. Thom spaces of PL microbundles have been constructed by Williamson in [50]. By his construction, we may take $\mathrm{Th}(\mu) = \mathrm{Th}(\nu)$, since the underlying block bundle of μ is ν and the homotopy type of the Thom space depends only on the underlying block bundle (in fact only on the underlying spherical fibration).

We shall construct a PL map

$$T(\xi)\colon S^s B^+ = \mathrm{Th}(\mathbb{R}^s \times B) \to \mathrm{Th}(\nu),$$

called the *Umkehr map*, following the terminology of [8]. Points in $N \subset S^s B^+$ are to be mapped by the identity to points in $N \subset \mathrm{Th}(\nu)$. By Cohen's [17, Theorem 5.3], ∂N is collared in the closure of $(\mathbb{R}^s \times B) - N$. Thus there exists a polyhedral neighborhood V of ∂N in the closure of $(\mathbb{R}^s \times B) - N$ and a PL homeomorphism $h\colon (\partial N) \times [0,1] \cong V$ such that $h(x,0) = x$ for $x \in \partial N$. Now let $T(\xi)$ map those points of V that lie in $h((\partial N) \times \{1\})$ to ∞. Map all points in $S^s B^+ - (N \cup V)$ to ∞. Finally, map the points in V, using the collar coordinate in $[0,1]$, correspondingly along cone lines in $\mathrm{cone}(\partial N) \subset \mathrm{Th}(\nu)$. This concludes the description of the Umkehr map $T(\xi)\colon S^s B^+ \to \mathrm{Th}(\nu)$. Since it sends ∞ to ∞, this is a pointed map.

Example 3.1 We continue Example 2.3 on the trivial F–block bundle ξ. Let $T(F)\colon S^s \to \mathrm{Th}(\nu_F) = \mathrm{Th}(\mu_F)$ be the standard Pontryagin–Thom collapse over a point associated to the embedding $\theta_F\colon F \hookrightarrow \mathbb{R}^s$. The Umkehr map for ξ is given by

$$T(\xi)\colon S^s \wedge B^+ \xrightarrow{T(F)\wedge \mathrm{id}_{B^+}} \mathrm{Th}(\nu_F) \wedge B^+ = \mathrm{Th}(\nu_\theta).$$

If E is any spectrum, then on reduced E–homology the Umkehr map induces a homomorphism

$$T(\xi)_*\colon \widetilde{E}_{n+s}(S^s B^+) \to \widetilde{E}_{n+s}(\mathrm{Th}(\nu)).$$

The suspension isomorphism provides an identification

$$\sigma\colon E_n(B) = \widetilde{E}_n(B^+) \cong \widetilde{E}_{n+s}(S^s B^+).$$

The composition yields a map

$$T(\xi)_* \circ \sigma\colon E_n(B) \to \widetilde{E}_{n+s}(\mathrm{Th}(\nu)) = \widetilde{E}_{n+s}(\mathrm{Th}(\mu)).$$

Example 3.2 We continue Example 3.1 on the trivial F–block bundle ξ. Let E be a commutative ring spectrum and let $[S^s]_E \in \widetilde{E}_s(S^s)$ denote the image of the unit $1 \in \pi_0(E)$ under $\sigma\colon \widetilde{E}_0(S^0) \xrightarrow{\cong} \widetilde{E}_s(S^s)$. Then the above map $T(\xi)_* \circ \sigma$ has the description

$$T(\xi)_*\sigma(a) = (T(F) \wedge \mathrm{id}_{B+})_*\sigma(a) = (T(F) \wedge \mathrm{id}_{B+})_*([S^s]_E \wedge a) = (T(F)_*[S^s]_E) \wedge a,$$

where $a \in E_n(B)$. Setting $[\mathrm{Th}\,\mu_F]_E = T(F)_*[S^s]_E$, we thus arrive at

$$T(\xi)_*\sigma(a) = [\mathrm{Th}\,\mu_F]_E \wedge a.$$

4 The Thom homomorphism, mock bundles and Witt spaces

We recall the Thom homomorphism Φ associated to an oriented PL microbundle μ. This homomorphism will later be used in the definition of the F–block bundle transfer $\xi^!$ with μ the stable vertical normal PL microbundle of ξ. The Thom map is given by taking the cap product with the Thom class of μ. Therefore, we will recall the homotopy-theoretic description $u_{\mathrm{SPL}}(\mu)$ of this class, as well as its geometric description $u_{\mathrm{BRS}}(\mu_{\mathrm{PLB}})$ in terms of mock bundles, as given by Buoncristiano, Rourke and Sanderson [12], where μ_{PLB} denotes the underlying PL closed disc block bundle of μ. In particular, we take the opportunity to provide a brief review of mock bundle theory. Mock bundles over Witt spaces will play an important role later on. One key fact in the subsequent development is that the total space of a mock bundle over a Witt space is again a Witt space.

Let MSPL be the Thom spectrum associated to PL microbundles (or PL $(\mathbb{R}^m, 0)$–bundles; see Kuiper and Lashof [29]). This is a ring spectrum whose homotopy groups can be identified with the bordism groups of oriented PL manifolds via the Pontryagin–Thom isomorphism. Let γ_m^{SPL} denote the universal oriented rank-m PL bundle over the classifying space BSPL_m. An oriented PL microbundle μ of rank m over a compact polyhedron X is classified by a map $X \to \mathrm{BSPL}_m$, which is covered by a bundle map $\mu \to \gamma_m^{\mathrm{SPL}}$. The induced map on Thom spaces yields a homotopy class

$$u_{\mathrm{SPL}}(\mu) \in [\Sigma^\infty \mathrm{Th}(\mu), \Sigma^m \mathrm{MSPL}] = \widetilde{\mathrm{MSPL}}^m(\mathrm{Th}(\mu)).$$

This class $u_{\mathrm{SPL}}(\mu)$ is the *Thom class* of μ in oriented PL cobordism. It is in fact an MSPL–orientation of μ in Dold's sense. Indeed, every $H\mathbb{Z}$–orientable PL bundle is MSPL–orientable; see Hsiang and Wall [26, Lemma 5, page 357] and Switzer [46, page 308].

Buoncristiano, Rourke and Sanderson give a geometric description of MSPL–cobordism in [12], and use it to obtain in particular a geometric description of the Thom class $u_{\mathrm{SPL}}(\mu)$. The geometric cocycles are given by oriented mock bundles, whose definition we recall here.

Definition 4.1 Let K be a finite ball complex and q an integer (possibly negative). A *q–mock bundle* η^q/K with base K and total space $E(\eta)$ consists of a PL map $p\colon E(\eta) \to |K|$ such that, for each $\sigma \in K$, $p^{-1}(\sigma)$ is a compact PL manifold of dimension $q + \dim(\sigma)$, with boundary $p^{-1}(\partial\sigma)$. The preimage $\eta(\sigma) := p^{-1}(\sigma)$ is called the *block* of η over σ.

The empty set is regarded as a manifold of any dimension; thus $\eta(\sigma)$ may be empty for some cells $\sigma \in K$. Note that if σ^0 is a 0–dimensional cell of K, then $\partial\sigma^0 = \varnothing$ and thus $p^{-1}(\partial\sigma) = \varnothing$. Hence the blocks over 0–dimensional cells are *closed* manifolds. Mock bundles over the same complex are *isomorphic* if there exists a block-preserving PL homeomorphism of total spaces. (The homeomorphism is *not* required to preserve the projections.) For our purposes, we need *oriented* mock bundles, which are defined using incidence numbers of cells and blocks: Suppose that $(M^n, \partial M)$ is an oriented PL manifold and $(N^{n-1}, \partial N)$ is an oriented PL manifold with $N \subset \partial M$. Then an incidence number $\epsilon(N, M) = \pm 1$ is defined by comparing the orientation of N with that induced on N from M; $\epsilon(N, M) = +1$ if these orientations agree and -1 if they disagree. An *oriented cell complex* K is a cell complex in which each cell is oriented. We then have the incidence number $\epsilon(\tau, \sigma)$ defined for codimension-1 faces $\tau < \sigma \in K$.

Definition 4.2 An *oriented mock bundle* is a mock bundle η/K over an oriented (finite) ball complex K in which every block is oriented (ie is an oriented PL manifold) such that for each codimension-1 face τ of $\sigma \in K$, $\epsilon(\eta(\tau), \eta(\sigma)) = \epsilon(\tau, \sigma)$.

Using intersection homology, Witt spaces were introduced by Siegel in [44] as a geometric cycle reservoir representing KO–homology at odd primes. Sources on intersection homology include [21; 22; 27; 20; 3].

Definition 4.3 A *Witt space* is an oriented PL pseudomanifold where the links L^{2k} of odd-codimensional PL intrinsic strata have vanishing middle-perversity degree-k rational intersection homology,

$$\mathrm{IH}_k^{\bar{m}}(L; \mathbb{Q}) = 0.$$

For example, pure-dimensional complex algebraic varieties are Witt spaces, since they are oriented pseudomanifolds and possess a Whitney stratification whose strata all have even codimension. The vanishing condition on the intersection homology of links L^{2k} is equivalent to requiring the canonical morphism from lower middle to upper middle-perversity intersection chain sheaves to be an isomorphism in the derived category of sheaf complexes. Consequently, these middle-perversity intersection chain sheaves are Verdier self-dual, and this induces global Poincaré duality for the middle-perversity intersection homology groups of a Witt space. In particular, Witt spaces X have a well-defined bordism invariant signature and L–classes $L_*(X) \in H_*(X; \mathbb{Q})$ which agree with the Poincaré duals of Hirzebruch's tangential L–classes when X is smooth. The notion of *Witt spaces with boundary* can be introduced as pairs $(X, \partial X)$, where X is a PL space and ∂X a stratum-preservingly collared PL subspace of X such that $X - \partial X$ and ∂X are both compatibly oriented Witt spaces. The following result is [5, Lemma 3.11], which is itself an analog of [12, Lemma 1.2, page 21].

Lemma 4.4 *Let (K, L) be a finite ball complex pair such that the polyhedron $|K|$ is an n–dimensional compact Witt space with (possibly empty) boundary $\partial|K| = |L|$. Orient K in such a way that the sum of oriented n–balls is a cycle rel boundary. (This is possible since $|K|$, being a Witt space, is an oriented pseudomanifold with boundary.) Let η/K be an oriented q–mock bundle over K with projection p. Then the total space $E(\eta)$ is an $(n+q)$–dimensional compact Witt space with boundary $p^{-1}(\partial|K|)$.*

Let (K, L) be any finite ball complex pair. Oriented mock bundles η_0 and η_1 over K, both empty over L, are *cobordant*, if there is an oriented mock bundle η over $K \times I$, empty over $L \times I$, such that $\eta|_{K\times 0} \cong \eta_0$ and $\eta|_{K\times 1} \cong \eta_1$. This is an equivalence relation, and we set

$$\Omega^q_{\mathrm{SPL}}(K, L) := \{[\eta^q/K] : \eta|_L = \varnothing\},$$

where $[\eta^q/K]$ denotes the cobordism class of the oriented q–mock bundle η^q/K over K. Then by the duality theorem [12, Theorem II.3.3] of Buoncristiano, Rourke and Sanderson, Spanier–Whitehead duality, together with the Pontryagin–Thom isomorphism, provides an isomorphism

$$\beta\colon \Omega^{-q}_{\mathrm{SPL}}(K, L) \cong \mathrm{MSPL}^q(|K|, |L|) \tag{5}$$

for compact $|K|$ and $|L|$, which is natural with respect to inclusions $(K', L') \subset (K, L)$; see also [12, Remarks(3), top of page 32]. This is the geometric description of oriented PL cobordism that we use here. The functor $\Omega^*_{\mathrm{SPL}}(-)$ gives rise to a functor on the category of compact polyhedral pairs and homotopy classes of continuous maps, which will be denoted by the same symbol [12, Theorem II.1.1].

Let $\alpha\colon |K| \to \mathrm{B}\widetilde{\mathrm{SPL}}_m$ be an oriented PL closed disc block bundle of rank m over a finite complex K. Let N denote the total space of α and ∂N the total space of the sphere block bundle of α. Then α has a Thom class (see [12, page 26])

$$u_{\mathrm{BRS}}(\alpha) \in \Omega^{-m}_{\mathrm{SPL}}(N, \partial N), \tag{6}$$

which we shall call the BRS–*Thom class* of α, given as follows: Let $i\colon K \to N$ be the zero section of α. Endow N with the ball complex structure given by taking the blocks $\alpha(\sigma)$ of the bundle α as balls, together with the balls of a suitable ball complex structure on the total space ∂N of the sphere block bundle of α. Then $i\colon K \to N$ is the projection of an oriented $(-m)$–mock bundle η, and thus determines an element

$$u_{\mathrm{BRS}}(\alpha) = [\eta] \in \Omega^{-m}_{\mathrm{SPL}}(N, \partial N).$$

The block of η over a ball $\alpha(\sigma)$ of N is $\sigma \in K$. The following is [5, Lemma 3.14].

Lemma 4.5 *Let $\alpha\colon |K| \to \mathrm{BSPL}_m$ be an oriented* PL $(\mathbb{R}^m, 0)$*–bundle, with $|K|$ compact. Under the isomorphism β in* (5)*, the* BRS*–Thom class $u_{\mathrm{BRS}}(\alpha_{\mathrm{PLB}})$ of the underlying oriented* PL *block bundle α_{PLB} of α gets mapped to the Thom class $u_{\mathrm{SPL}}(\alpha)$.*

Let E be an MSPL–module spectrum. Then there is a cap product

$$\cap\colon \mathrm{MSPL}^p(X, A) \otimes E_q(X, A) \to E_{q-p}(X).$$

The reduced cobordism group of the Thom space can be expressed as a relative group,

$$\widetilde{\mathrm{MSPL}}^p(\mathrm{Th}(\mu)) \cong \mathrm{MSPL}^p(N, \partial N),$$

where N, as in Section 3, is the total space of the underlying oriented PL closed disc block bundle of μ. Let

$$\rho_*\colon E_*(N) \xrightarrow{\cong} E_*(X)$$

be the inverse of the isomorphism induced on E–homology by the inclusion $X \hookrightarrow N$ of the zero section. Using the cap product

$$\cap\colon \mathrm{MSPL}^m(N, \partial N) \otimes E_q(N, \partial N) \to E_{q-m}(N) \overset{\rho_*}{\cong} E_{q-m}(X),$$

we obtain the *Thom homomorphism*

$$\Phi := \rho_*(u_{\mathrm{SPL}}(\mu) \cap -)\colon \widetilde{E}_q(\mathrm{Th}(\mu)) \cong E_q(N, \partial N) \to E_{q-m}(X).$$

Under suitable conditions this map is an isomorphism, for example if X is connected, E is a ring spectrum and $u_{\mathrm{SPL}}(\mu)$ determines an E–orientation of μ. (See [46, page 309, Theorem 14.6]; recall that our X is a finite complex.)

5 Block bundle transfer

Let E be a module spectrum over the Thom spectrum MSPL of oriented PL bundle theory. As in Section 2, F denotes a closed oriented PL manifold of dimension d and K a finite ball complex with associated polyhedron $B = |K|$ of dimension b. Let ξ be an oriented PL F–block bundle over K with total space $X = E(\xi)$. Following Boardman [9] and Becker and Gottlieb [8], we shall construct a *transfer homomorphism*

(7) $$\xi^!\colon E_n(B) \to E_{n+d}(X).$$

Let μ denote the stable oriented vertical normal PL microbundle of ξ whose underlying disc block bundle is ν_θ, the oriented vertical normal disc block bundle of the F–block bundle ξ, associated to a block-preserving embedding θ for ξ. The rank of μ and ν_θ is $m = s - d$ for $d = \dim F$ and s large. The block bundle transfer is defined to be the composition

$$E_n(B) \xrightarrow{T(\xi)_* \circ \sigma} \widetilde{E}_{n+s}(\mathrm{Th}(\mu)) \xrightarrow{\Phi} E_{n+d}(X),$$

where σ is the suspension isomorphism, $T(\xi)$ is the Umkehr map of Section 3 and Φ is the Thom homomorphism of μ as described in Section 4.

Remark 5.1 The geometric description of block bundle transfer as provided above is serviceable for the subsequent PL geometric arguments concerning orientation classes of PL pseudomanifolds. We are grateful to a referee for pointing out that the geometric context embeds into a more general homotopy theoretic one as follows: A fibration π with Poincaré complex fiber over a base space B possesses a vertical Spivak fibration ν_π and a canonical map of spectra $\Sigma^\infty B^+ \to \mathrm{Th}(\nu_\pi)$, provided one has ∞–categorical functoriality of the Pontryagin–Thom collapse associated to a Poincaré complex; see Carmeli, Cnossen, Ramzi and Yanovski [14] and Klein, Malkiewich and Ramzi [28]. Since the vertical Spivak fibration of an oriented block bundle with manifold blocks has a stable PL bundle reduction (by Hebestreit, Land, Lück and Randal-Williams [25]), it is MSPL–oriented.

We are mainly interested in the case where E is ordinary homology, Ranicki's symmetric $\mathbb{L}^\bullet$–spectrum or Witt bordism. Let us discuss each of these cases in turn.

5.1 Block transfer on ordinary homology

Let $H\mathbb{Z}$ denote the Eilenberg–Mac Lane spectrum of the ring $\mathbb{Z}$. The stable universal Thom class in $H^0(\mathrm{MSPL})$ yields a map $\alpha\colon \mathrm{MSPL} \to H\mathbb{Z}$, and this map is a ring map. Thus α makes the ring spectrum $H\mathbb{Z}$ into an MSPL–module by taking the action map to be

$$\mathrm{MSPL} \wedge H\mathbb{Z} \xrightarrow{\alpha\wedge\mathrm{id}} H\mathbb{Z} \wedge H\mathbb{Z} \xrightarrow{\mu_H} H\mathbb{Z},$$

where μ_H is the product on $H\mathbb{Z}$. The induced map

$$\alpha_*\colon \Omega_n^{\mathrm{SPL}}(Z) \cong \mathrm{MSPL}_n(Z) \to H_n(Z;\mathbb{Z})$$

is the Steenrod–Thom homomorphism sending the bordism class of a singular PL manifold $[f\colon M^n \to Z] \in \Omega_n^{\mathrm{SPL}}(Z)$ to $f_*[M] \in H_n(Z;\mathbb{Z})$. We recall the following standard fact:

Proposition 5.2 (Rudyak [42, Proposition V.1.6]) *Let $\tau\colon D \to E$ be a ring morphism of ring spectra. Let γ be an $(S^n,*)$–fibration equipped with a D–orientation $u_D \in \widetilde{D}^n(\mathrm{Th}\,\gamma)$. Then the image $\tau(u_D) \in \widetilde{E}^n(\mathrm{Th}\,\gamma)$ is an E–orientation of γ.*

We apply this to the ring morphism $\alpha\colon \mathrm{MSPL} \to H\mathbb{Z}$ and to our microbundle μ, which we had already equipped with the MSPL–orientation $u_{\mathrm{SPL}}(\mu)$. By the proposition, the homomorphism

$$\alpha\colon \widetilde{\mathrm{MSPL}}^{s-d}(\mathrm{Th}(\mu)) \to \widetilde{H}^{s-d}(\mathrm{Th}(\mu);\mathbb{Z})$$

sends $u_{\mathrm{SPL}}(\mu)$ to an $H\mathbb{Z}$–orientation

$$u_{\mathbb{Z}}(\mu) := \alpha(u_{\mathrm{SPL}}(\mu)) \in \widetilde{H}^{s-d}(\mathrm{Th}(\mu);\mathbb{Z}). \tag{8}$$

(This is the Thom class of μ in ordinary cohomology.) Another standard fact from stable homotopy theory is:

Lemma 5.3 *Let D and E be ring spectra, and $\tau\colon D \to E$ a ring morphism. We consider E as a D–module via the action map*

$$D \wedge E \xrightarrow{\tau\wedge\mathrm{id}} E \wedge E \xrightarrow{\mu_E} E.$$

This module structure yields a cap product $\cap_{D,E}\colon D^*(X) \otimes E_*(X) \to E_*(X)$. The ring structure on E yields a cap product $\cap_E\colon E^*(X) \otimes E_*(X) \to E_*(X)$. Then the diagram

$$\begin{array}{ccc} D^*(X) \otimes E_*(X) & \xrightarrow{\cap_{D,E}} & E_*(X) \\ \downarrow{\scriptstyle \tau\otimes\mathrm{id}} & & \| \\ E^*(X) \otimes E_*(X) & \xrightarrow{\cap_E} & E_*(X) \end{array}$$

commutes.

By this lemma and (8), the transfer on ordinary homology is given by

$$\begin{aligned} \xi^!(-) &= \rho_*(u_{\mathrm{SPL}}(\mu) \cap_{\mathrm{MSPL},H\mathbb{Z}} T(\xi)_*\sigma(-)) = \rho_*(\alpha(u_{\mathrm{SPL}}(\mu)) \cap_{H\mathbb{Z}} T(\xi)_*\sigma(-)) \\ &= \rho_*(u_{\mathbb{Z}}(\mu) \cap_{H\mathbb{Z}} T(\xi)_*\sigma(-)). \end{aligned}$$

We summarize: the block bundle transfer (7) on ordinary homology $E = H\mathbb{Z}$ is given by

$$\xi^! = \rho_*(u_{\mathbb{Z}}(\mu) \cap T(\xi)_*\sigma(-))\colon H_n(B;\mathbb{Z}) \to H_{n+d}(X;\mathbb{Z}).$$

5.2 Block transfer on Witt bordism

Let $\Omega_*^{\mathrm{Witt}}(-)$ denote Witt bordism theory as defined by Siegel in [44]. Elements of $\Omega_n^{\mathrm{Witt}}(Z)$ are Witt bordism classes of continuous maps $f\colon W^n \to Z$ defined on an n–dimensional closed Witt space W. Let MWITT be the Quinn spectrum associated to the ad-theory of Witt spaces, representing Witt bordism via a natural equivalence

$$\mathrm{MWITT}_*(-) \cong \Omega_*^{\mathrm{Witt}}(-). \tag{9}$$

See Banagl, Laures and McClure [7]. A weakly equivalent spectrum was first considered by Curran [18]. He verified that this spectrum is an MSO–module [18, Theorem 3.6, page 117]. The product of two Witt spaces is again a Witt space. This implies essentially that MWITT is a ring spectrum; for more details see [7]. (There, we focused on the spectrum MIP representing bordism of integral intersection homology Poincaré spaces studied by Goresky and Siegel in [24] and by Pardon in [36], but everything works in an analogous, indeed simpler, manner for $\mathbb{Q}$–Witt spaces.) Every oriented PL manifold is a Witt space. Hence there is a map

$$\phi_W\colon \mathrm{MSPL} \to \mathrm{MWITT},$$

which, using the methods of ad-theories and Quinn spectra employed in [7], can be constructed to be multiplicative. Using this ring map, the spectrum MWITT becomes an MSPL–module with action map

$$\mathrm{MSPL} \wedge \mathrm{MWITT} \to \mathrm{MWITT}$$

given by the composition

$$\mathrm{MSPL} \wedge \mathrm{MWITT} \xrightarrow{\phi_W \wedge \mathrm{id}} \mathrm{MWITT} \wedge \mathrm{MWITT} \to \mathrm{MWITT}.$$

(The product of a Witt space and an oriented PL manifold is again a Witt space.) In particular, there is a cap product

$$\cap\colon \mathrm{MSPL}^j(Z,Y) \otimes \mathrm{MWITT}_n(Z,Y) \to \mathrm{MWITT}_{n-j}(Z) \tag{10}$$

and a transfer

$$\xi^!\colon \mathrm{MWITT}_n(B) \to \mathrm{MWITT}_{n+d}(X),$$

where ξ is our F–block bundle over B and $d = \dim F$.

Let C be any finite ball complex with subcomplex $D \subset C$ and suppose that $Z = |C|$ and $Y = |D|$. By Buoncristiano, Rourke and Sanderson [12], a geometric description of the above cap product (10) is given as follows: One uses the canonical identifications to think of the cap product as a product

$$\cap\colon \Omega_{\mathrm{SPL}}^{-j}(C,D) \otimes \Omega_n^{\mathrm{Witt}}(|C|,|D|) \to \Omega_{n-j}^{\mathrm{Witt}}(|C|).$$

Let us first discuss the absolute case $D = \varnothing$, and then return to the relative one. If C is simplicial, $f: W \to C$ is a simplicial map from an n–dimensional triangulated closed Witt space W to C, and η^q is a q–mock bundle over C (with $q = -j$), then one has (see [12, page 29])

$$[\eta^q/C] \cap [f: W \to |C|] = [g: E(f^*\eta) \to |C|] \in \Omega^{\text{Witt}}_{n-j}(|C|),$$

where g is the diagonal arrow in the cartesian diagram

$$\begin{array}{ccc} E(f^*\eta) & \longrightarrow & E(\eta) \\ \big\downarrow & \overset{g}{\searrow} & \big\downarrow p \\ W & \xrightarrow{\ f\ } & C \end{array}$$

Here, one uses the fact (see [12, II.2, page 23]) that mock bundles over simplicial complexes admit pullbacks under simplicial maps. By Lemma 4.4, $E(f^*\eta)$ is a closed Witt space. For the relative case, we observe that if $(W, \partial W)$ is a compact Witt space with boundary, $f: (W, \partial W) \to (|C|, |D|)$ maps the boundary into $|D|$ and $\eta|_D = \varnothing$, then $f^*\eta|_{\partial W} = \varnothing$ and so $\partial E(f^*\eta) = \varnothing$, ie the Witt space $E(f^*\eta)$ is *closed*. Hence it defines an absolute bordism class.

In Section 6, we provide a more direct geometric description of the Witt bordism transfer

$$\xi^!: \text{MWITT}_n(B) \to \text{MWITT}_{n+d}(X)$$

as a pullback transfer $\xi^!_{\text{PB}}: \Omega^{\text{Witt}}_n(B) \to \Omega^{\text{Witt}}_{n+d}(X)$.

5.3 Block transfer on $\mathbb{L}^\bullet$–homology

We write $\mathbb{L}^\bullet = \mathbb{L}^\bullet(\mathbb{Z}) = \mathbb{L}^\bullet\langle 0\rangle(\mathbb{Z})$ for Ranicki's connected symmetric algebraic L–spectrum with homotopy groups $\pi_n(\mathbb{L}^\bullet) = L^n(\mathbb{Z})$, the symmetric L–groups of the ring of integers; see eg [39]. Technically, we shall use the construction of $\mathbb{L}^\bullet$ as the Quinn spectrum of a suitable ad-theory; see Banagl, Laures and McClure [7]. That construction is weakly equivalent to Ranicki's. Localization $\mathbb{Z} \to \mathbb{Q}$ induces a map $\epsilon_{\mathbb{Q}}: \mathbb{L}^\bullet(\mathbb{Z}) \to \mathbb{L}^\bullet(\mathbb{Q})$, and $\pi_n(\mathbb{L}^\bullet(\mathbb{Q})) = L^n(\mathbb{Q})$ with

$$L^n(\mathbb{Q}) \cong \begin{cases} \mathbb{Z} \oplus (\mathbb{Z}/_2)^\infty \oplus (\mathbb{Z}/_4)^\infty & \text{if } n \equiv 0 \pmod 4, \\ 0 & \text{if } n \not\equiv 0 \pmod 4. \end{cases}$$

The spectra $\mathbb{L}^\bullet(\mathbb{Z})$ and $\mathbb{L}^\bullet(\mathbb{Q})$ are ring spectra. Let MSTOP be the Thom spectrum associated to oriented topological $(\mathbb{R}^n, 0)$–bundles. There is a canonical forget map

$$\phi_F: \text{MSPL} \to \text{MSTOP}.$$

Ranicki [37, page 290] constructed a map

$$\sigma^*: \text{MSTOP} \to \mathbb{L}^\bullet,$$

and in [7], we constructed a map

$$\tau: \text{MWITT} \to \mathbb{L}^\bullet(\mathbb{Q}).$$

(Actually, we even constructed an integral map MIP $\to \mathbb{L}^\bullet$, where MIP represents bordism of integral intersection homology Poincaré spaces, but everything works in the same manner for Witt theory, if one uses the $\mathbb{L}^\bullet$–spectrum with rational coefficients.) This map is multiplicative, ie a ring map, as shown in [7, Section 12], and the diagram

$$\begin{array}{ccccc} & & \mathrm{MSTOP} & \xrightarrow{\sigma^*} & \mathbb{L}^\bullet(\mathbb{Z}) \\ & \overset{\phi_F}{\nearrow} & & & \downarrow{\scriptstyle \epsilon_{\mathbb{Q}}} \\ \mathrm{MSPL} & & & & \\ & \underset{\phi_W}{\searrow} & & & \\ & & \mathrm{MWITT} & \xrightarrow{\tau} & \mathbb{L}^\bullet(\mathbb{Q}) \end{array} \tag{11}$$

homotopy commutes, since it comes from a commutative diagram of ad-theories under applying the symmetric spectrum functor $\boldsymbol{M}$ of Laures and McClure [31]. Using the ring map $\tau\phi_W : \mathrm{MSPL} \to \mathbb{L}^\bullet(\mathbb{Q})$, the spectrum $\mathbb{L}^\bullet(\mathbb{Q})$ becomes an MSPL–module with action map

$$\mathrm{MSPL} \wedge \mathbb{L}^\bullet(\mathbb{Q}) \to \mathbb{L}^\bullet(\mathbb{Q})$$

given by the composition

$$\mathrm{MSPL} \wedge \mathbb{L}^\bullet(\mathbb{Q}) \xrightarrow{(\tau\phi_W)\wedge \mathrm{id}} \mathbb{L}^\bullet(\mathbb{Q}) \wedge \mathbb{L}^\bullet(\mathbb{Q}) \to \mathbb{L}^\bullet(\mathbb{Q}).$$

The associated transfer is

$$\xi^! : \mathbb{L}^\bullet(\mathbb{Q})_n(B) \to \mathbb{L}^\bullet(\mathbb{Q})_{n+d}(X),$$

with ξ our F–block bundle over B and $d = \dim F$.

We shall show that the block bundle transfer $\xi^!$ commutes with the passage, under τ_*, from Witt bordism theory to $\mathbb{L}^\bullet(\mathbb{Q})$–homology. The homotopy commutative diagram

$$\begin{array}{ccc} \mathrm{MSPL} \wedge \mathrm{MWITT} & \xrightarrow{\mathrm{id}\wedge\tau} & \mathrm{MSPL} \wedge \mathbb{L}^\bullet(\mathbb{Q}) \\ {\scriptstyle \phi_W\wedge\mathrm{id}}\downarrow & & \downarrow{\scriptstyle (\tau\phi_W)\wedge\mathrm{id}} \\ \mathrm{MWITT} \wedge \mathrm{MWITT} & \xrightarrow{\tau\wedge\tau} & \mathbb{L}^\bullet(\mathbb{Q}) \wedge \mathbb{L}^\bullet(\mathbb{Q}) \\ \downarrow & & \downarrow \\ \mathrm{MWITT} & \xrightarrow{\tau} & \mathbb{L}^\bullet(\mathbb{Q}) \end{array}$$

shows that $\tau : \mathrm{MWITT} \to \mathbb{L}^\bullet(\mathbb{Q})$ is an MSPL–module morphism. In the proof of Lemma 5.5, we use the following standard fact:

Lemma 5.4 *If E is a ring spectrum, F and F' are module spectra over E and $\phi\colon F \to F'$ is an E–module morphism, then the diagram*

$$\begin{array}{ccc} E^m(X,A) \otimes F_n(X,A) & \xrightarrow{\cap} & F_{n-m}(X) \\ {\scriptstyle \mathrm{id}\otimes\phi_*}\downarrow & & \downarrow{\scriptstyle \phi_*} \\ E^m(X,A) \otimes F'_n(X,A) & \xrightarrow{\cap} & F'_{n-m}(X) \end{array}$$

commutes: if $u \in E^m(X,A)$ and $a \in F_n(X,A)$, then

$$\phi_*(u \cap a) = u \cap \phi_*(a).$$

Lemma 5.5 *The Thom homomorphisms Φ of an oriented* PL *microbundle μ of rank m over a compact polyhedron X commute with the passage from Witt bordism to $\mathbb{L}^\bullet(\mathbb{Q})$–homology, that is, the diagram*

$$\begin{array}{ccc} \widetilde{\mathrm{MWITT}}_n(\mathrm{Th}(\mu)) & \xrightarrow{\Phi} & \mathrm{MWITT}_{n-m}(X) \\ \tau_*\downarrow & & \tau_*\downarrow \\ \widetilde{\mathbb{L}^\bullet}(\mathbb{Q})_n(\mathrm{Th}(\mu)) & \xrightarrow{\Phi} & \mathbb{L}^\bullet(\mathbb{Q})_{n-m}(X) \end{array}$$

commutes.

Proof As τ_* is a natural transformation of homology theories, it commutes with the isomorphism ρ_*. Since τ is an MSPL–module morphism, Lemma 5.4 applies to give

$$\tau_*\Phi = \tau_*\rho_*(u \cap -) = \rho_*\tau_*(u \cap -) = \rho_*(u \cap \tau_*(-)) = \Phi\tau_*,$$

where $u = u_{\mathrm{SPL}}(\mu)$. □

Proposition 5.6 *Let F be a closed oriented d–dimensional* PL *manifold and let ξ be an oriented F–block bundle with total space X over the compact polyhedral base B. Then the diagram*

$$\begin{array}{ccc} \mathrm{MWITT}_n(B) & \xrightarrow{\xi^!} & \mathrm{MWITT}_{n+d}(X) \\ \tau_*\downarrow & & \tau_*\downarrow \\ \mathbb{L}^\bullet(\mathbb{Q})_n(B) & \xrightarrow{\xi^!} & \mathbb{L}^\bullet(\mathbb{Q})_{n+d}(X) \end{array}$$

commutes.

Proof Let μ be the stable vertical normal microbundle of ξ. The right-hand square of the diagram

$$\begin{array}{ccccc} \mathrm{MWITT}_n(B) & \xrightarrow[\sigma]{\cong} & \widetilde{\mathrm{MWITT}}_{n+s}(S^s B^+) & \xrightarrow{T(\xi)_*} & \widetilde{\mathrm{MWITT}}_{n+s}(\mathrm{Th}(\mu)) \\ \tau_*\downarrow & & \tau_*\downarrow & & \downarrow\tau_* \\ \mathbb{L}^\bullet(\mathbb{Q})_n(B) & \xrightarrow[\sigma]{\cong} & \widetilde{\mathbb{L}^\bullet}(\mathbb{Q})_{n+s}(S^s B^+) & \xrightarrow{T(\xi)_*} & \widetilde{\mathbb{L}^\bullet}(\mathbb{Q})_{n+s}(\mathrm{Th}(\mu)) \end{array}$$

commutes, as τ_* is a natural transformation of homology theories. The left-hand square, involving the suspension isomorphism σ, commutes for the same reason. The statement now follows from Lemma 5.5. □

An oriented topological $(\mathbb{R}^m, 0)$–bundle α over a CW complex Z, classified by a map $Z \to \mathrm{BSTOP}_m$, possesses a Thom class

$$u_{\mathrm{STOP}}(\alpha) \in \widetilde{\mathrm{MSTOP}}^m(\mathrm{Th}(\alpha))$$

in oriented topological cobordism. The next auxiliary result on compatibility of Thom classes is standard; see eg [5, Lemma 3.7].

Lemma 5.7 *Let α be an oriented* PL *$(\mathbb{R}^m, 0)$–bundle. On cobordism groups, the homomorphism*

$$\phi_F : \widetilde{\mathrm{MSPL}}^m(\mathrm{Th}(\alpha)) \to \widetilde{\mathrm{MSTOP}}^m(\mathrm{Th}(\alpha_{\mathrm{TOP}}))$$

induced by the canonical map $\phi_F : \mathrm{MSPL} \to \mathrm{MSTOP}$ sends the Thom class of α to the Thom class of the underlying oriented topological $(\mathbb{R}^m, 0)$–bundle α_{STOP},

$$\phi_F(u_{\mathrm{SPL}}(\alpha)) = u_{\mathrm{STOP}}(\alpha_{\mathrm{STOP}}).$$

Following [37, pages 290–291], an oriented topological $(\mathbb{R}^m, 0)$–bundle α has a canonical $\mathbb{L}^\bullet$–cohomology orientation

$$u_{\mathbb{L}}(\alpha) \in \widetilde{\mathbb{L}^\bullet}^m(\mathrm{Th}(\alpha)),$$

which we shall also refer to as the $\mathbb{L}^\bullet$–*cohomology Thom class* of α, defined by

$$u_{\mathbb{L}}(\alpha) := \sigma^*(u_{\mathrm{STOP}}(\alpha)). \tag{12}$$

The morphism of spectra $\epsilon_{\mathbb{Q}}\colon \mathbb{L}^\bullet(\mathbb{Z}) \to \mathbb{L}^\bullet(\mathbb{Q})$ coming from localization induces a homomorphism

$$\widetilde{\mathbb{L}^\bullet}^m(\mathrm{Th}(\alpha)) \to \widetilde{\mathbb{L}^\bullet}(\mathbb{Q})^m(\mathrm{Th}(\alpha)).$$

We denote the image of $u_{\mathbb{L}}(\alpha)$ under this map again by $u_{\mathbb{L}}(\alpha) \in \widetilde{\mathbb{L}^\bullet}(\mathbb{Q})^m(\mathrm{Th}(\alpha))$.

Lemma 5.8 *Let α be an oriented* PL $(\mathbb{R}^m, 0)$*–bundle. The homomorphism*

$$\tau\phi_W\colon \widetilde{\mathrm{MSPL}}^m(\mathrm{Th}(\alpha)) \to \mathbb{L}^\bullet(\mathbb{Q})^m(\mathrm{Th}(\alpha))$$

induced by the ring morphism $\tau\phi_W\colon \mathrm{MSPL} \to \mathbb{L}^\bullet(\mathbb{Q})$ sends the MSPL*–cohomology Thom class of α to the $\mathbb{L}^\bullet$–cohomology Thom class of* (*the underlying topological bundle of*) α,

$$\tau\phi_W(u_{\mathrm{SPL}}(\alpha)) = u_{\mathbb{L}}(\alpha).$$

Proof By Lemma 5.7, Ranicki's definition (12) and the homotopy commutativity of (11),

$$\tau\phi_W(u_{\mathrm{SPL}}(\alpha)) = \epsilon_{\mathbb{Q}}\sigma^*\phi_F(u_{\mathrm{SPL}}(\alpha)) = \epsilon_{\mathbb{Q}}\sigma^*(u_{\mathrm{STOP}}(\alpha_{\mathrm{STOP}})) = u_{\mathbb{L}}(\alpha_{\mathrm{STOP}}). \qquad \square$$

Lemma 5.8, together with Lemma 5.3, implies that the F–block bundle transfer on $\mathbb{L}^\bullet(\mathbb{Q})$–homology is given by

$$\xi^! = \rho_*(u_{\mathbb{L}}(\mu) \cap T(\xi)_*\sigma(-))\colon \mathbb{L}^\bullet(\mathbb{Q})_n(B) \to \mathbb{L}^\bullet(\mathbb{Q})_{n+d}(X).$$

Example 5.9 We continue Example 3.2 and compute the transfer for the trivial F–block bundle ξ with total space $X = F \times B$. Let E be a commutative ring spectrum and $\phi\colon \mathrm{MSPL} \to E$ a morphism of ring spectra, equipping E with the structure of an MSPL–module. Recall that we had chosen a PL embedding $\theta_F\colon F \hookrightarrow \mathbb{R}^s$ with s large enough that θ_F has a tubular neighborhood given by a PL microbundle μ_F which represents the stable normal PL microbundle of F. The stable vertical normal bundle of ξ is then given by $\mu = \mathrm{pr}_1^*\mu_F$. Its Thom class $u_{\mathrm{SPL}}(\mu) \in \widetilde{\mathrm{MSPL}}^{s-d}(\mathrm{Th}(\mu)) = \widetilde{\mathrm{MSPL}}^{s-d}(\mathrm{Th}(\mu_F) \wedge B^+)$ is $u_{\mathrm{SPL}}(\mu) = u_{\mathrm{SPL}}(\mu_F) \wedge 1$, since the bundle map $\mu \to \gamma^{\mathrm{SPL}}_{s-d}$ factors as $\mu \to \mu_F \to \gamma^{\mathrm{SPL}}_{s-d}$, where the first map covers the projection $\mathrm{pr}_1\colon F \times B \to F$ and the second map the classifying map for μ_F. The element $\phi(u_{\mathrm{SPL}}(\mu_F))$ is an E–orientation of μ_F [42, Proposition V.1.6] and thus $[F]_E := \rho_{F*}\big(\phi(u_{\mathrm{SPL}}(\mu_F)) \cap [\mathrm{Th}\,\mu_F]_E\big) \in E_d(F)$, with $\rho_{F*}\colon E_d(N_F) \cong E_d(F)$, is an E–homology orientation for the PL manifold F [42, Proposition V.2.8; 46, page 333, Lemma 14.40]. The transfer $\xi^!\colon E_n(B) \to E_{n+d}(F \times B)$ is then given on $a \in E_n(B)$ by

$$\xi^!(a) = [F]_E \times a, \tag{13}$$

as follows from the calculation

$$\begin{aligned}\xi^!(a) &= \Phi T(\xi)_*\sigma(a) = \Phi([\operatorname{Th}\mu_F]_E \wedge a) = \rho_*(\phi(u_{\mathrm{SPL}}(\mu)) \cap ([\operatorname{Th}\mu_F]_E \wedge a)) \\ &= \rho_*(\phi(u_{\mathrm{SPL}}(\mu_F) \wedge 1) \cap ([\operatorname{Th}\mu_F]_E \wedge a)) = \rho_*((\phi(u_{\mathrm{SPL}}(\mu_F)) \wedge 1) \cap ([\operatorname{Th}\mu_F]_E \wedge a)) \\ &= \rho_*((\phi(u_{\mathrm{SPL}}(\mu_F)) \cap [\operatorname{Th}\mu_F]_E) \times (1 \cap a)) = \rho_{F*}(\phi(u_{\mathrm{SPL}}(\mu_F)) \cap [\operatorname{Th}\mu_F]_E) \times a = [F]_E \times a.\end{aligned}$$

6 Geometric pullback transfer on bordism

As in previous sections, F is a closed d–dimensional oriented PL manifold and ξ is an oriented PL F–block bundle with total space X over a finite ball complex K, where $B = |K|$. We shall geometrically construct a pullback transfer

$$\xi^!_{\mathrm{PB}} \colon \Omega^{\mathrm{Witt}}_n(B) \to \Omega^{\mathrm{Witt}}_{n+d}(X)$$

on Witt bordism. Let $f \colon W \to B$ be a continuous map representing an element $[f]$ of $\Omega^{\mathrm{Witt}}_n(B)$. Choose a PL map $g \colon W \to B$ homotopic to f. We follow Casson's method for pulling back F–block bundles [15]. (Note that the pullback of block and mock bundles is not generally defined through cartesian diagrams.) There is a compact polyhedron V and a factorization

$$\begin{array}{ccc} W & \overset{j}{\hookrightarrow} & B \times V \\ & {\scriptstyle g}\searrow & \downarrow{\scriptstyle \mathrm{pr}_1} \\ & & B \end{array}$$

of g into a PL embedding j followed by a standard projection. Let L be a cell complex with $|L| = V$. The F–block bundle pullback $\mathrm{pr}_1^*\,\xi$ is by definition $\xi \times L$, an F–block bundle over the cell complex $K \times L$ with total space $E(\mathrm{pr}_1^*\,\xi) = X \times V$. Thus the first factor projection $X \times V \to X$ defines a PL map

$$\mathrm{pr}_1 \colon E(\mathrm{pr}_1^*\,\xi) \to X.$$

Let C be the product cell complex $C := K \times L$ and put $\eta := \mathrm{pr}_1^*\,\xi$. Let C' be a subdivision of C such that the subpolyhedron $j(W) \subset V \times B$ is given by $j(W) = |D'|$ for a subcomplex D' of C'. Block bundles can be subdivided, and this does not change the total space [15, page 37]. Let η' over C' be a subdivision of η, $E(\eta') = E(\eta)$. Block bundles can be restricted to subcomplexes. The total space of the restriction is given by the union of the blocks over the cells of the subcomplex. Thus we can restrict η' to the subcomplex D' of C' and obtain an F–block bundle $\eta'|_{D'}$ whose total space is a PL subspace $E(\eta'|_{D'}) \hookrightarrow E(\eta') = E(\eta)$. The composition

$$E(\eta'|_{D'}) \hookrightarrow E(\eta) \to X$$

gives a map

$$\bar{g} \colon E(\eta'|_{D'}) \to X. \tag{14}$$

Let $j^*\eta$ be the F–block bundle over W that corresponds to $\eta'|_{D'}$ under the PL homeomorphism $j \colon W \cong j(W)$. The pullback F–block bundle $g^*\xi$ is then defined to be

$$g^*\xi = j^*\eta = j^*(\mathrm{pr}_1^*\,\xi).$$

Thus (14) is a map

$$\bar{g}\colon E(g^*\xi) \to X.$$

Note that $E(g^*\xi)$ is a compact polyhedron. In the above construction of pullbacks and $\bar{g}$, it was not important that the Witt domain W has empty boundary; everything applies to compact W with boundary as well. Indeed, Casson's pullback applies of course to PL maps with general polyhedral domain. Let ξ and ξ' be F–block bundles over cell complexes K and K' such that $|K| = B = |K'|$. Recall that ξ and ξ' are called *equivalent* if, for some common subdivision K'' of K and K', the subdivision of ξ over K'' is isomorphic to the subdivision of ξ' over K''. (An *isomorphism* of F–block bundles over the same complex is a block-preserving homeomorphism of total spaces.) An equivalence $\phi\colon \xi \cong \xi'$ of F–block bundles over B induces an equivalence

$$g^*\phi\colon g^*\xi \cong g^*\xi'$$

such that

$$\begin{array}{ccc} E(g^*\xi) & \xrightarrow{\bar{g}} & E(\xi) = X \\ {\scriptstyle g^*\phi}\downarrow\cong & & \cong\downarrow{\scriptstyle\phi} \\ E(g^*\xi') & \xrightarrow{\bar{g}'} & E(\xi') = X' \end{array}$$

commutes.

Lemma 6.1 *Let $g\colon W \to B$ be a* PL *map defined on a compact Witt space with possibly nonempty boundary ∂W. Then the total space $E(g^*\xi)$ is a closed Witt space with boundary $E((g^*\xi)|_{\partial W})$.*

Proof An F–block bundle is in particular a mock bundle. Thus $g^*\xi$ is a mock bundle over the Witt space W and the result follows from Lemma 4.4. □

By Lemma 6.1, the map $\bar{g}\colon E(g^*\xi) \to X$ represents an element $[\bar{g}] \in \Omega^{\mathrm{Witt}}_{n+d}(X)$.

For future reference and additional clarity in subsequent arguments, let us record explicitly:

Lemma 6.2 *Let W and W' be closed n–dimensional Witt spaces. If $f \simeq f'\colon W \to X$ are homotopic maps, then $[f] = [f'] \in \Omega^{\mathrm{Witt}}_n(X)$. If $\phi\colon W \xrightarrow{\cong} W'$ is a* PL *homeomorphism, and $f\colon W \to X$ and $f'\colon W' \to X$ maps such that $f' \circ \phi = f$, then $[f] = [f'] \in \Omega^{\mathrm{Witt}}_n(X)$.*

Proof The first statement, asserting homotopy invariance, is part of the fact that Witt bordism constitutes a homology theory and is proven by considering a homotopy as a Witt bordism, noting that the cylinder on a closed Witt space is a Witt space with boundary. The bordism required by the second statement is given by taking a cylinder on the domain of the PL homeomorphism and a cylinder on the target of the PL homeomorphism, and then gluing the two cylinders using the homeomorphism. □

Lemma 6.3 *The class*

$$[\bar{g}\colon E(g^*\xi) \to X] \in \Omega^{\mathrm{Witt}}_{n+d}(X)$$

depends only on the Witt class $[g] \in \Omega^{\mathrm{Witt}}_n(B)$.

Proof Let $g_0\colon W_0 \to B$ and $g_1\colon W_1 \to B$ be PL maps such that $[g_0] = [g_1] \in \Omega_n^{\mathrm{Witt}}(B)$. Let $G\colon W \to B$ be a Witt bordism between g_0 and g_1; we may assume G to be PL. Let $i_j\colon W_j \hookrightarrow W$ denote the boundary inclusions for $j = 0, 1$. Since $g_j = G \circ i_j$, there is an equivalence $g_j^*\xi \cong i_j^* G^*\xi$ such that

$$\begin{array}{ccc} E(g_j^*\xi) & & \\ {\scriptstyle\cong}\downarrow & \searrow^{\bar{g}_j} & \\ E(i_j^*G^*\xi) & \xrightarrow{\overline{Gi}_j} & E(\xi) = X \end{array}$$

commutes. Thus $\bar{g}_j$ and $\overline{Gi}_j$ are Witt bordant, for $j = 0, 1$, by Lemma 6.2. According to Lemma 6.1, $E(G^*\xi)$ is a compact Witt space with boundary $E(i_0^*G^*\xi) \sqcup E(i_1^*G^*\xi)$. The diagram

$$\begin{array}{ccc} E(i_j^*G^*\xi) & & \\ \downarrow & \searrow^{\overline{Gi}_j} & \\ E(G^*\xi) & \xrightarrow{\overline{G}} & E(\xi) = X \end{array}$$

commutes for $j = 0, 1$. Hence, $\overline{G}$ is a Witt bordism between $\overline{Gi}_0$ and $\overline{Gi}_1$. □

We define the *geometric transfer* (or *pullback transfer*)

$$\xi^!_{\mathrm{PB}}\colon \Omega_n^{\mathrm{Witt}}(B) \to \Omega_{n+d}^{\mathrm{Witt}}(X)$$

by

$$\xi^!_{\mathrm{PB}}[g\colon W \to B] = [\bar{g}\colon E(g^*\xi) \to X],$$

where g is a PL representative of the bordism class. Let

$$\xi^!_{\mathrm{BRS}}\colon \Omega_n^{\mathrm{Witt}}(B) \to \Omega_{n+d}^{\mathrm{Witt}}(X)$$

be the map

$$\xi^!_{\mathrm{BRS}}[g] := \rho_*(u_{\mathrm{BRS}}(\nu) \cap T(\xi)_*\sigma[g]),$$

where $\nu = \nu_\xi\colon X \to \mathrm{B}\widetilde{\mathrm{SPL}}_{s-d}$ represents the stable vertical normal PL disc block bundle of ξ. This is a technical intermediary; in terms of their respective definitions, the difference between $\xi^!_{\mathrm{BRS}}$ and $\xi^!$ is that the former uses the Thom class $u_{\mathrm{BRS}}(\nu)$, while the latter uses the Thom class $u_{\mathrm{SPL}}(\mu)$. We will eventually see that $\xi^!_{\mathrm{PB}} = \xi^!_{\mathrm{BRS}} = \xi^!$ on Witt bordism. Towards that goal, let us first investigate the behavior of both the pullback transfer and the BRS–transfer under standard factor projections.

Proposition 6.4 *Let B and D be compact polyhedra. Let $\xi \times D$ denote the F–block bundle over $B \times D$ obtained by pulling back ξ under the projection $\mathrm{pr}_1\colon B \times D \to B$. Then the diagrams*

$$\begin{array}{ccc} \Omega_n^{\mathrm{Witt}}(B) & \xrightarrow{\xi^!_{\mathrm{PB}}} & \Omega_{n+d}^{\mathrm{Witt}}(X) \\ {\scriptstyle \mathrm{pr}_{1*}}\uparrow & & \uparrow{\scriptstyle \mathrm{pr}_{1*}} \\ \Omega_n^{\mathrm{Witt}}(B \times D) & \xrightarrow{(\xi\times D)^!_{\mathrm{PB}}} & \Omega_{n+d}^{\mathrm{Witt}}(X \times D) \end{array} \qquad (15)$$

and

$$(16)\qquad \begin{array}{ccc} \Omega_n^{\mathrm{Witt}}(B) & \xrightarrow{\xi^!_{\mathrm{BRS}}} & \Omega_{n+d}^{\mathrm{Witt}}(X) \\ \uparrow{\scriptstyle \mathrm{pr}_{1*}} & & \uparrow{\scriptstyle \mathrm{pr}_{1*}} \\ \Omega_n^{\mathrm{Witt}}(B \times D) & \xrightarrow{(\xi\times D)^!_{\mathrm{BRS}}} & \Omega_{n+d}^{\mathrm{Witt}}(X \times D) \end{array}$$

commute.

Proof We will first establish the commutativity of (15) involving the pullback transfers. Recall that the F–block bundle ξ is given over a cell complex K with $|K| = B$. Let J be a cell complex with polyhedron $|J| = D$. Then $\xi \times D$ is an F–block bundle over the cell complex $K \times J$. Let $[g] \in \Omega_n^{\mathrm{Witt}}(B \times D)$ be any element, represented by a PL map $g\colon W \to B \times D$. Choose a compact polyhedron V and a factorization of g as

$$(17)\qquad \begin{array}{ccc} W & \overset{j}{\hookrightarrow} & (B \times D) \times V \\ & {\scriptstyle g}\searrow & \downarrow{\scriptstyle \mathrm{pr}_{B\times D}} \\ & & B \times D \end{array}$$

Let L be a cell complex with $|L| = V$. We will compute $\xi^!_{\mathrm{PB}}\,\mathrm{pr}_{1*}[g]$. The element $\mathrm{pr}_{1*}[g]$ is represented by $\mathrm{pr}_1 \circ g$ with factorization

$$\begin{array}{ccc} W & \overset{j}{\hookrightarrow} & B \times D \times V \\ & {\scriptstyle \mathrm{pr}_1\circ g}\searrow & \downarrow{\scriptstyle \mathrm{pr}_B} \\ & & B \end{array}$$

The pullback $\mathrm{pr}_B^*\,\xi = \xi \times J \times L$ has total space $E(\mathrm{pr}_B^*\,\xi) = X \times D \times V$ which projects to X via

$$\mathrm{pr}_X\colon E(\mathrm{pr}_B^*\,\xi) = X \times D \times V \to X = E(\xi).$$

Let C be the cell complex $C = K \times J \times L$ and let C' be a subdivision of C such that $j(W)$ is given by $j(W) = |D'|$ for some subcomplex D' of C'. Let $(\mathrm{pr}_B^*\,\xi)'$ be the block bundle over C' obtained by subdivision of $\mathrm{pr}_B^*\,\xi$. Then $(\mathrm{pr}_B^*\,\xi)'$ can be restricted to D', and the total space of this restriction $(\mathrm{pr}_B^*\,\xi)'|_{D'}$ is a subspace of $E((\mathrm{pr}_B^*\,\xi)') = E(\mathrm{pr}_B^*\,\xi)$. The composition of the subspace inclusion with pr_X yields a map

$$\overline{\mathrm{pr}_1 \circ g}\colon E((\mathrm{pr}_B^*\,\xi)'|_{D'}) \subset E(\mathrm{pr}_B^*\,\xi) = X \times D \times V \xrightarrow{\mathrm{pr}_X} X$$

such that

$$\xi^!_{\mathrm{PB}}[\mathrm{pr}_1 \circ g] = [\overline{\mathrm{pr}_1 \circ g}].$$

Let us compute $(\xi \times D)^!_{\mathrm{PB}}[g]$. The relevant factorization is (17); the pullback

$$\mathrm{pr}_{B\times D}^*(\xi \times D) = (\xi \times D) \times L = \xi \times J \times L = \mathrm{pr}_B^*\,\xi$$

has total space $E(\mathrm{pr}_{B\times D}^*(\xi \times D)) = X \times D \times V$ which projects to $X \times D$ via

$$\mathrm{pr}_{X\times D}\colon E(\mathrm{pr}_{B\times D}^*(\xi \times D)) = (X \times D) \times V \to X \times D = E(\xi \times D).$$

Let $(\mathrm{pr}^*_{B\times D}(\xi\times D))'$ be the block bundle over C' obtained by subdivision of $\mathrm{pr}^*_{B\times D}(\xi\times D)$. Then $(\mathrm{pr}^*_{B\times D}(\xi\times D))' = (\mathrm{pr}^*_B\,\xi)'$, and thus

$$(\mathrm{pr}^*_{B\times D}(\xi\times D))'|_{D'} = (\mathrm{pr}^*_B\,\xi)'|_{D'}.$$

Consider the commutative diagram

$$\begin{array}{ccccc} E((\mathrm{pr}^*_{B\times D}(\xi\times D))'|_{D'}) & \hookrightarrow & E(\mathrm{pr}^*_{B\times D}(\xi\times D)) & \xrightarrow{\mathrm{pr}_{X\times D}} & X\times D \\ \| & & \| & & \downarrow \mathrm{pr}_1 \\ E((\mathrm{pr}^*_B\,\xi)'|_{D'}) & \hookrightarrow & E(\mathrm{pr}^*_B\,\xi) & \xrightarrow{\mathrm{pr}_X} & X \end{array}$$

The upper horizontal composition is a map $\bar g$ such that

$$(\xi\times D)^!_{\mathrm{PB}}[g] = [\bar g],$$

and the lower horizontal composition is $\overline{\mathrm{pr}_1\circ g}$. Therefore,

$$\mathrm{pr}_{1*}(\xi\times D)^!_{\mathrm{PB}}[g] = [\mathrm{pr}_1\circ\bar g] = [\overline{\mathrm{pr}_1\circ g}] = \xi^!_{\mathrm{PB}}[\mathrm{pr}_1\circ g] = \xi^!_{\mathrm{PB}}\,\mathrm{pr}_{1*}[g].$$

Thus (15) commutes, as claimed.

It remains to establish the commutativity of (16). Let $\nu_\xi = \nu_\theta$ denote the stable vertical normal PL disc block bundle associated to a particular choice of blockwise embedding $\theta\colon X = E(\xi)\hookrightarrow \mathbb{R}^s\times B$ by Proposition 2.1. The PL embedding

$$E(\xi\times D) = X\times D \xrightarrow{\theta\times\mathrm{id}_D} (\mathbb{R}^s\times B)\times D = \mathbb{R}^s\times(B\times D)$$

is block preserving over $B\times D$ with respect to the F–blocks $(\xi\times D)(\sigma\times\tau) = \xi(\sigma)\times\tau$ of $\xi\times D$ for $\sigma\in K$ and $\tau\in J$, as

$$(\theta\times\mathrm{id}_D)(\xi(\sigma)\times\tau) = \theta(\xi(\sigma))\times\tau \subset (\mathbb{R}^s\times\sigma)\times\tau = \mathbb{R}^s\times(\sigma\times\tau).$$

Thus the stable vertical normal disc-block bundle of $\xi\times D$ can be computed from the embedding $\theta\times\mathrm{id}_D$, which yields

$$\nu_{\xi\times D} = \nu_{\theta\times\mathrm{id}_D} = \nu_\theta\times D = \nu_\xi\times D,$$

a disc block bundle over $X\times D$. Recall that the Thom space of the block bundle ν_ξ is

$$\mathrm{Th}(\nu_\xi) = N\cup_{\partial N}\mathrm{cone}(\partial N),$$

with $N = E(\nu_\xi)$. Thus, with $N' := E(\nu_{\xi\times D}) = N\times D$, we have

$$\mathrm{Th}(\nu_{\xi\times D}) = N'\cup_{\partial N'}\mathrm{cone}(\partial N').$$

Here $\partial N'$ denotes the total space of the sphere bundle of $\xi\times D$, $\partial N' = (\partial N)\times D$. The projection $\mathrm{pr}_1\colon N' = N\times D\to N$ induces a map

$$\mathrm{Th}(\mathrm{pr}_1)\colon (N\times D)\cup_{(\partial N)\times D}\mathrm{cone}((\partial N)\times D)\to N\cup_{\partial N}\mathrm{cone}(\partial N),$$

ie a map

$$\mathrm{Th}(\mathrm{pr}_1)\colon \mathrm{Th}(\nu_{\xi\times D})\to\mathrm{Th}(\nu_\xi).$$

The suspension of the projection $\mathrm{pr}_1 : B \times D \to B$ is a map $S^s\,\mathrm{pr}_1 : S^s(B\times D)^+ \to S^s B^+$. The F–block bundle $\xi \times D$ has its Umkehr map

$$T(\xi\times D)\colon S^s(B\times D)^+ \to \mathrm{Th}(\nu_{\xi\times D})$$

such that the diagram

$$\begin{array}{ccc} S^s B^+ & \xrightarrow{T(\xi)} & \mathrm{Th}(\nu_\xi) \\ {\scriptstyle S^s\,\mathrm{pr}_1}\uparrow & & \uparrow{\scriptstyle \mathrm{Th}(\mathrm{pr}_1)} \\ S^s(B\times D)^+ & \xrightarrow{T(\xi\times D)} & \mathrm{Th}(\nu_{\xi\times D}) \end{array}$$

commutes up to homotopy. The induced diagram

$$\begin{array}{ccc} \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(S^s B^+) & \xrightarrow{T(\xi)_*} & \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}\,\nu_\xi) \\ {\scriptstyle (S^s\,\mathrm{pr}_1)_*}\uparrow & & \uparrow{\scriptstyle \mathrm{Th}(\mathrm{pr}_1)_*} \\ \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(S^s(B\times D)^+) & \xrightarrow{T(\xi\times D)_*} & \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}\,\nu_{\xi\times D}) \end{array}$$

on reduced Witt bordism commutes. The diagram

$$\begin{array}{ccc} \Omega^{\mathrm{Witt}}_{n}(B) & \underset{\cong}{\xrightarrow{\sigma}} & \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(S^s B^+) \\ {\scriptstyle \mathrm{pr}_{1*}}\uparrow & & \uparrow{\scriptstyle (S^s\,\mathrm{pr}_1)_*} \\ \Omega^{\mathrm{Witt}}_{n}(B\times D) & \underset{\cong}{\xrightarrow{\sigma}} & \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(S^s(B\times D)^+) \end{array}$$

commutes by the naturality of the suspension isomorphism σ.

It remains to show that

$$\begin{array}{ccccc} \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}\,\nu_\xi) & \xrightarrow{u_{\mathrm{BRS}}(\nu_\xi)\cap-} & \Omega^{\mathrm{Witt}}_{n+d}(E\nu_\xi) & \underset{\cong}{\xrightarrow{\rho_*}} & \Omega^{\mathrm{Witt}}_{n+d}(X) \\ {\scriptstyle \mathrm{Th}(\mathrm{pr}_1)_*}\uparrow & & {\scriptstyle \mathrm{pr}_{1*}}\uparrow & & \uparrow{\scriptstyle \mathrm{pr}_{1*}} \\ \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}\,\nu_{\xi\times D}) & \xrightarrow{u_{\mathrm{BRS}}(\nu_{\xi\times D})\cap-} & \Omega^{\mathrm{Witt}}_{n+d}(E\nu_{\xi\times D}) & \underset{\cong}{\xrightarrow{\rho_*}} & \Omega^{\mathrm{Witt}}_{n+d}(X\times D) \end{array} \tag{18}$$

commutes. The right-hand side commutes, since the zero section embedding $X\times D \hookrightarrow E\nu_{\xi\times D} = N\times D$ of $\nu_{\xi\times D}$ is given by $i\times\mathrm{id}_D$, where i is the zero section embedding $i\colon X \hookrightarrow E\nu_\xi = N$ of ν_ξ, so that

$$\begin{array}{ccc} N = E\nu_\xi & \xleftarrow{\;i\;} & X \\ {\scriptstyle \mathrm{pr}_1}\uparrow & & \uparrow{\scriptstyle \mathrm{pr}_1} \\ N\times D = E\nu_{\xi\times D} & \xleftarrow{i\times\mathrm{id}_D} & X\times D \end{array}$$

commutes. We will prove that the left-hand side commutes as well. The map

$$\mathrm{Th}(\mathrm{pr}_1)\colon \mathrm{Th}(\nu_{\xi\times D}) \to \mathrm{Th}(\nu_\xi)$$

induces a homomorphism

$$\mathrm{Th}(\mathrm{pr}_1)^*\colon \Omega^{d-s}_{\mathrm{SPL}}(E\nu_\xi, \dot{E}\nu_\xi) \to \Omega^{d-s}_{\mathrm{SPL}}(E\nu_{\xi\times D}, \dot{E}\nu_{\xi\times D}),$$

which agrees with the homomorphism

$$\mathrm{pr}_1^*\colon \Omega^{d-s}_{\mathrm{SPL}}(N,\partial N) \to \Omega^{d-s}_{\mathrm{SPL}}((N,\partial N)\times D)$$

induced by the map of pairs $\mathrm{pr}_1\colon (N,\partial N)\times D\to (N,\partial N)$. By the naturality of the BRS–Thom class [12, top of page 27], this homomorphism maps the BRS–Thom class of ν_ξ to the BRS–Thom class of $\mathrm{pr}_1^*\,\nu_\xi=\nu_{\xi\times D}$,

$$\mathrm{pr}_1^*(u_{\mathrm{BRS}}(\nu_\xi))=u_{\mathrm{BRS}}(\nu_{\xi\times D}).$$

Given any element

$$[g]\in\widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}\,\nu_{\xi\times D})=\Omega^{\mathrm{Witt}}_{n+s}((N,\partial N)\times D),$$

the computation

$$u_{\mathrm{BRS}}(\nu_\xi)\cap\mathrm{pr}_{1*}[g]=\mathrm{pr}_{1*}(\mathrm{pr}_1^*\,u_{\mathrm{BRS}}(\nu_\xi)\cap[g])=\mathrm{pr}_{1*}(u_{\mathrm{BRS}}(\nu_{\xi\times D})\cap[g])$$

shows that the left-hand side of (18) commutes. □

The pullback transfer $\xi^!_{\mathrm{PB}}$ on Witt bordism agrees with the transfer $\xi^!_{\mathrm{BRS}}$:

Proposition 6.5 *The diagram*

$$\begin{array}{ccccc}
\widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(S^sB^+) & \xrightarrow{T(\xi)_*} & \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}(\nu)) & \xrightarrow{u_{\mathrm{BRS}}(\nu)\cap-} & \Omega^{\mathrm{Witt}}_{n+d}(E(\nu)) \\
\sigma\uparrow\cong & & & & \cong\downarrow\rho_* \\
\Omega^{\mathrm{Witt}}_n(B) & & \xrightarrow{\quad\xi^!_{\mathrm{PB}}\quad} & & \Omega^{\mathrm{Witt}}_{n+d}(X)
\end{array}$$

commutes, that is, $\xi^!_{\mathrm{PB}}=\xi^!_{\mathrm{BRS}}$.

Proof Let $h\colon W^n\to B$ be a continuous map from a closed n–dimensional Witt space W to B, representing an element $[h]\in\Omega^{\mathrm{Witt}}_n(B)$. By simplicial approximation, we may assume that h is PL. We begin by observing that, by Proposition 6.4, it suffices to prove the statement for the case where $h\colon W\to B$ is a PL embedding: Given any PL map $h\colon W\to B$, consider the graph embedding

$$(h,\mathrm{id}_W)\colon W\to B\times W$$

which factors h as

$$\begin{array}{ccc}
W & \xhookrightarrow{(h,\mathrm{id}_W)} & B\times W \\
 & {\scriptstyle h}\searrow & \downarrow \mathrm{pr}_1 \\
 & & B
\end{array}$$

Let $\xi\times W$ denote the F–block bundle over $B\times W$ obtained by pulling back ξ under the projection $\mathrm{pr}_1\colon B\times W\to B$. If the statement is known for embeddings, then

$$(\xi\times W)^!_{\mathrm{PB}}[(h,\mathrm{id}_W)]=(\xi\times W)^!_{\mathrm{BRS}}[(h,\mathrm{id}_W)].$$

Hence by Proposition 6.4 with $D=W$,

$$\begin{aligned}
\xi^!_{\mathrm{PB}}[h]=\xi^!_{\mathrm{PB}}[\mathrm{pr}_1\circ(h,\mathrm{id}_W)]&=\xi^!_{\mathrm{PB}}\,\mathrm{pr}_{1*}[(h,\mathrm{id}_W)]=\mathrm{pr}_{1*}(\xi\times W)^!_{\mathrm{PB}}[(h,\mathrm{id}_W)]\\
&=\mathrm{pr}_{1*}(\xi\times W)^!_{\mathrm{BRS}}[(h,\mathrm{id}_W)]=\xi^!_{\mathrm{BRS}}\,\mathrm{pr}_{1*}[(h,\mathrm{id}_W)]=\xi^!_{\mathrm{BRS}}[h].
\end{aligned}$$

Consequently, it remains to prove the equality $\xi^!_{\mathrm{PB}}=\xi^!_{\mathrm{BRS}}$ on Witt bordism classes that are represented by PL embeddings.

As in the construction of the Umkehr map $T(\xi)$ in Section 3, let N denote the total space $E(\nu)$ of the stable vertical normal closed disc block bundle $\nu = \nu_\theta$ of ξ associated to a choice of block-preserving embedding $\theta\colon X \hookrightarrow \mathbb{R}^s \times B$, where X is the total space of the given F–block bundle ξ. Thus N is a ξ–block-preserving regular neighborhood of $\theta(X)$ in $\mathbb{R}^s \times B$. Recall that ∂N denotes the total space of the sphere block bundle of ν. Let $D^s \subset \mathbb{R}^s$ be a closed PL ball about the origin which is large enough that $(D^s - \partial D^s) \times B$ contains $N \cup V \subset \mathbb{R}^s \times B$, where V is the outside collar to ∂N used in the construction of the Umkehr map; such a ball exists by compactness of X.

Let $h\colon W \hookrightarrow B$ be a PL embedding of a closed Witt space into B. Recall that K is a cell complex with polyhedron $|K| = B$ and ξ is given over K. By subdivision of K and ξ, we may assume that $h(W) = |K_W|$ for a subcomplex K_W of K. Let L_S be a finite simplicial complex such that

(i) $|L_S| = S^s B^+$,

(ii) there is a subcomplex L of L_S such that $|L| = D^s \times B$,

(iii) for every simplex $\sigma \in K$, there is a subcomplex L_σ of L such that
$$|L_\sigma| = D^s \times \sigma,$$

(iv) there exists a subcomplex L_θ of L such that $|L_\theta| = \theta(X)$,

(v) the stable vertical normal bundle ν is a (disc) block bundle over the complex L_θ such that
$$E(\nu) \cap (D^s \times \sigma) = \bigcup_{\tau \in L_\sigma \cap L_\theta} \nu(\tau),$$
where $\nu(\tau)$ is the disc-block of ν over the simplex τ.

Property (iii) implies that L_τ is a subcomplex of L_σ for every face τ of $\sigma \in K$. Furthermore,
$$D^s \times h(W) = D^s \times |K_W| = \bigcup_{\sigma \in K_W} D^s \times \sigma = \bigcup_{\sigma \in K_W} |L_\sigma| = \Big| \bigcup_{\sigma \in K_W} L_\sigma \Big|,$$
so that
$$L_W := \bigcup_{\sigma \in K_W} L_\sigma$$
is a simplicial subcomplex of L with $D^s \times h(W) = |L_W|$. Since the embedding $\theta\colon X \hookrightarrow \mathbb{R}^s \times B$ is block preserving with respect to the F–blocks of ξ, we have $\theta(\xi(\sigma)) = (D^s \times \sigma) \cap \theta(X)$ for all $\sigma \in K$. So by (iv),
$$\theta(\xi(\sigma)) = |L_\sigma| \cap |L_\theta| = |L_\sigma \cap L_\theta|.$$

Thus the embedded F–blocks $\theta(\xi(\sigma))$ are triangulated by the subcomplex $L_\sigma \cap L_\theta$ of L.

The image $\sigma[h]$ under the suspension isomorphism
$$\sigma\colon \Omega_n^{\mathrm{Witt}}(B) \xrightarrow{\cong} \widetilde{\Omega}_{n+s}^{\mathrm{Witt}}(S^s B^+) = \Omega_{n+s}^{\mathrm{Witt}}((D^s, \partial D^s) \times B)$$
is represented by the closed product PL embedding
$$\mathrm{id} \times h\colon (D^s \times W, \partial(D^s \times W)) \hookrightarrow (D^s \times B, (\partial D^s) \times B).$$

The Umkehr map is a PL map

$$T(\xi)\colon S^s B^+ = \mathrm{Th}(\mathbb{R}^s \times B) = \frac{D^s \times B}{(\partial D^s)\times B} \to \mathrm{Th}(\nu),$$

which is the identity near $\theta(X)$. Composing with it, we obtain a PL map

$$f = T(\xi)\circ(\mathrm{id}\times h)\colon (D^s \times W, \partial(D^s\times W)) \to (\mathrm{Th}(\nu),\infty).$$

Let A be the ball complex with $|A| = N = E(\nu)$ whose balls include the blocks of ν. The rest of the balls come from the sphere block bundle of ν. The BRS–Thom class $u_{\mathrm{BRS}}(\nu)\in\Omega^{-(s-d)}_{\mathrm{SPL}}(N,\partial N)$ is represented by the mock bundle η with projection given by the zero section $i\colon \theta(X)\to A$. Thus the total space of η is $E(\eta)=\theta(X)$. The mock bundle η is an embedded mock bundle in the sense of Buoncristiano, Rourke and Sanderson [12, page 34]: the restriction $i|\colon \eta(\sigma)\to\sigma$ for a ball $\sigma=\nu(\tau)\in A$ is the inclusion $\tau\hookrightarrow\nu(\tau)$, which is locally flat by definition of a disc block bundle. Furthermore, $i|\colon\eta(\sigma)\to\sigma$ is proper, ie $i|^{-1}(\partial\sigma)=\partial\eta(\sigma)$. We wish to compute the cap product

$$u_{\mathrm{BRS}}(\nu)\cap[f\colon (D^s,\partial D^s)\times W\to(\mathrm{Th}(\nu),\infty)]\in\Omega^{\mathrm{Witt}}_{n+d}(E(\nu)).$$

The base complex of η is only known to be a ball complex, not a simplicial complex as required for pulling back a mock bundle via a cartesian square. Thus we need to subdivide simplicially. Let L' be a simplicial subdivision of L and let A' be a simplicial subdivision of A such that A' is a subcomplex of L'. Thus,

$$|A'| = E(\nu) \quad\text{and}\quad |L'| = D^s\times B.$$

The complex L' contains a (simplicial) subcomplex L'_W given by

$$L'_W = \{\tau\in L' : \tau\subset\sigma \text{ for some } \sigma\in L_W\}.$$

This is a subdivision of $L_W\subset L$, and

$$|L'_W| = |L_W| = D^s\times h(W).$$

So the inclusion

$$|L'_W| = D^s\times h(W)\hookrightarrow D^s\times B = |L'|$$

is a *simplicial* map

$$L'_W\hookrightarrow L'.$$

By [12, Theorem 2.1, page 23] (see also [40, subdivision theorem, page 128]), mock bundles can be subdivided: if α is a mock bundle over a ball complex D with total space $E(\alpha)$ and projection $p\colon E(\alpha)\to D$, and D' is a subdivision of D, then there exists a mock bundle α' over D' together with a PL homeomorphism $\phi\colon E(\alpha)\xrightarrow{\cong} E(\alpha')$ which preserves α–blocks over D, and a homotopy

$$F\colon E(\alpha)\times I\to|D|=|D'| \quad\text{with}\quad F_0=p \ \text{ and } \ F_1=p'\phi,$$

which respects the α–blocks over D. (Here $p'\colon E(\alpha')\to|D'|$ is the projection of α'.) Moreover, if α is an embedded mock bundle, then the subdivision theorem yields again an embedded mock bundle and the homotopy can be taken to be an isotopy which is covered by an ambient isotopy. We apply this to the zero section mock bundle η over A: Since A' is a (simplicial) subdivision of A, there thus exists a correspondingly subdivided mock bundle η' over A'. Since η is an embedded mock bundle $i\colon \theta(X) = E(\eta) \hookrightarrow E(\nu)$, so is η'. Thus the projection map i' of η' may be taken to be a PL embedding $i'\colon E(\eta') \hookrightarrow E(\nu)$. As the zero section i does not touch the sphere bundle of ν (ie η has empty blocks over ∂N), the same is true for the perturbation i'. There exists a PL homeomorphism $\phi\colon \theta(X) = E(\eta) \xrightarrow{\cong} E(\eta')$ which preserves η–blocks over the ball complex A. The maps i and $i'\phi$ are isotopic via an isotopy

$$F\colon \theta(X) \times I \to E(\nu) \times I \quad \text{with } F_0 = i \text{ and } F_1 = i'\phi.$$

This isotopy is covered by an ambient isotopy

$$H\colon E(\nu) \times I \to E(\nu) \times I \quad \text{with } H_0 = \mathrm{id}$$

such that

$$\begin{array}{ccc} \theta(X) \times I & \xrightarrow{F_0 \times \mathrm{id} = i \times \mathrm{id}} & E(\nu) \times I \\ & {\scriptstyle F} \searrow \quad \swarrow {\scriptstyle H} & \\ & E(\nu) \times I & \end{array}$$

commutes. This implies

$$H_1 \circ i = F_1 = i' \circ \phi. \tag{19}$$

By an induction on the cells $\sigma \in K$, starting with the 0–dimensional cells, F and H can be constructed to preserve blocks over K. More precisely: Let ν_σ denote the restriction of ν to the embedded F–block $\theta(\xi(\sigma)) = \theta(X) \cap (D^s \times \sigma)$. Since ν is a block bundle over the complex L_θ and $\theta(\xi(\sigma))$ is triangulated by $L_\sigma \cap L_\theta$, the total space of ν_σ is given by $E(\nu_\sigma) = \bigcup_\tau \nu(\tau)$, where τ ranges over all simplices of $L_\sigma \cap L_\theta$. Thus by (v) above,

$$E(\nu_\sigma) = E(\nu) \cap (D^s \times \sigma). \tag{20}$$

Then H can be inductively arranged to satisfy

$$H_t(E(\nu_\sigma)) = E(\nu_\sigma) \tag{21}$$

for all $\sigma \in K$ and all $t \in [0, 1]$, as follows: Recall that Buoncristiano, Rourke and Sanderson's construction of H in their proof of the mock bundle subdivision theorem proceeds inductively over cells of the base, starting with the 0–cells. In the present context, one organizes their induction as follows: Start with the 0–skeleton A^0 of A. For every 0–cell σ^0 of K, subdivide η over $A^0 \cap D^s \times \sigma^0$ within the manifold $E(\nu) \cap D^s \times \sigma^0$. Extend this subdivision for every 1–cell σ^1 of K to a subdivision over $A^0 \cap D^s \times \sigma^1$ within the manifold $E(\nu) \cap D^s \times \sigma^1$. Continue in this way with 2–cells σ^2, etc, until all cells of K have been used. Then move on to the 1–skeleton A^1 of A. For every 0–cell σ^0 of K, extend the subdivision to a subdivision over $A^1 \cap D^s \times \sigma^0$ within the manifold $E(\nu) \cap D^s \times \sigma^0$. Extend this subdivision for every 1–cell σ^1 of K to a subdivision over $A^1 \cap D^s \times \sigma^1$ within the manifold $E(\nu) \cap D^s \times \sigma^1$, and so on.

The mock bundle η' is defined over the simplicial complex A' with polyhedron $|A'| = E(\nu)$, but using the canonical inclusions $E(\nu) \subset D^s \times B$ and $E(\nu) \subset \mathrm{Th}(\nu)$ we may regard η' as a mock bundle over $D^s \times B$, and as a mock bundle over $\mathrm{Th}(\nu)$. In more detail, the composition

$$E(\eta') \overset{i'}{\hookrightarrow} E(\nu) = |A'| \hookrightarrow D^s \times B = |L'|$$

is the projection of a mock bundle over the complex L', whose blocks over simplices in A' are the blocks of η' and blocks over simplices not in A' are taken to be empty. (Here, we are using that η' has empty blocks over the sphere bundle ∂N.) Similarly, after extending the triangulation A' to a triangulation T' of $\mathrm{Th}(\nu)$ by coning off simplices of A' that are in ∂N (and adding the cone point ∞ as a 0–simplex), the composition

$$E(\eta') \overset{i'}{\hookrightarrow} E(\nu) = |A'| \hookrightarrow \mathrm{Th}(\nu) = |T'|$$

is the projection of a mock bundle over the complex T', whose blocks over simplices in A' are the blocks of η' and blocks over simplices not in A' are again taken to be empty. In view of the commutative diagram

$$\begin{array}{ccc} E(\eta') & & \\ \big\downarrow{\scriptstyle i'} & & \\ E(\nu) & \searrow & \\ \big\downarrow & & \\ D^s \times B & \xrightarrow{T(\xi)} & \mathrm{Th}(\nu) \end{array}$$

the pullback $T(\xi)^*(\eta'/_{T'})$ under the Umkehr map is precisely $\eta'/_{L'}$. Therefore, the mock bundle pullback $f^*(\eta')$ is given by

$$f^*(\eta') = (\mathrm{id} \times h)^* T(\xi)^*(\eta'/_{T'}) = (\mathrm{id} \times h)^*(\eta'/_{L'}).$$

The mock bundle η' (contrary to η, possibly) is defined over a *simplicial* complex L' and, as pointed out above, the inclusion $D^s \times h(W) \hookrightarrow D^s \times B = |L'|$ is a *simplicial* map

$$L'_W \hookrightarrow L'.$$

Therefore, the mock bundle pullback $f^*(\eta') = (\mathrm{id} \times h)^*(\eta')$ is given by the *cartesian* diagram

$$\begin{array}{ccc} E((\mathrm{id} \times h)^*\eta') & \longrightarrow & E(\eta') \\ \big\downarrow & \overset{g}{\searrow} & \big\downarrow{\scriptstyle i'} \\ D^s \times h(W) = |L'_W| & \hookrightarrow & D^s \times B = |L'| \end{array}$$

It follows that the cap product of the BRS–Thom class with $[f]$ is given by the diagonal arrow

$$u_{\mathrm{BRS}}(\nu) \cap [f] = [g] \in \Omega^{\mathrm{Witt}}_{n+d}(E(\nu)),$$

the total space of the pullback is given by

$$E((\mathrm{id} \times h)^*\eta') = (D^s \times h(W)) \cap i'E(\eta')$$

and g is the subspace inclusion

$$g\colon (D^s \times h(W)) \cap i'E(\eta') \subset i'E(\eta') \subset E(\nu).$$

We show next that the final stage $H_1\colon E(\nu)\to E(\nu)$ of the ambient isotopy H induces a homeomorphism

$$H_1\colon E(\xi|_{K_W}) \xrightarrow{\cong} (D^s\times h(W))\cap i'E(\eta'), \tag{22}$$

where we use θ to identify $X=E(\xi)$ and $\theta(X)$, and to identify $E(\xi|_{K_W})$ and $\theta(X)\cap(D^s\times|K_W|)$. The homeomorphism H_1 restricts to a homeomorphism

$$H_1|\colon \theta(X)\cap(D^s\times|K_W|)\xrightarrow{\cong} H_1(\theta(X)\cap(D^s\times|K_W|)),$$

whose target we shall now compute:

$$\begin{aligned}
H_1(\theta(X)\cap(D^s\times|K_W|)) &= H_1(\theta(X)\cap E(\nu)\cap(D^s\times|K_W|)) = H_1(\theta(X))\cap H_1(E(\nu)\cap D^s\times|K_W|)\\
&= H_1(\theta(X))\cap H_1\Big(E(\nu)\cap\bigcup_{\sigma\in K_W} D^s\times\sigma\Big)\\
&= H_1(\theta(X))\cap H_1\Big(\bigcup_{\sigma\in K_W} E(\nu)\cap(D^s\times\sigma)\Big)\\
&= H_1(\theta(X))\cap\bigcup_{\sigma\in K_W} H_1(E(\nu)\cap(D^s\times\sigma))\\
&= H_1(\theta(X))\cap\bigcup_{\sigma\in K_W}(E(\nu)\cap(D^s\times\sigma)) && \text{(by (20) and (21))}\\
&= H_1(\theta(X))\cap E(\nu)\cap\bigcup_{\sigma\in K_W}(D^s\times\sigma) = H_1(\theta(X))\cap\bigcup_{\sigma\in K_W}(D^s\times\sigma)\\
&= H_1iE(\eta)\cap\bigcup_{\sigma\in K_W}|L_\sigma|\\
&= i'\phi E(\eta)\cap|L_W| && \text{(by (19))}\\
&= i'E(\eta')\cap(D^s\times h(W)).
\end{aligned}$$

Thus we obtain the homeomorphism (22). In the diagram

$$\begin{array}{ccc}
E(\xi|_{K_W}) \hookrightarrow E(\xi)=\theta(X) & \xrightarrow{i} & E(\nu)\\
{\scriptstyle H_1|}\downarrow\cong & & \cong\downarrow{\scriptstyle H_1}\\
i'E(\eta')\cap(D^s\times h(W)) & \xrightarrow{g} & E(\nu)
\end{array}$$

all the horizontal arrows are subspace inclusions and thus the diagram commutes. By Lemma 6.2 applied to the PL homeomorphism $H_1|$,

$$[g]=[g\circ H_1|]\in\Omega^{\mathrm{Witt}}_{n+d}(E(\nu)).$$

By commutativity of the diagram,

$$[g\circ H_1|]=[E(\xi|_{K_W})\subset\theta(X)\xrightarrow{H_1 i}E(\nu)]=[E(\xi|_{K_W})\subset\theta(X)\xrightarrow{i'\phi}E(\nu)].$$

By restriction, the isotopy F gives rise to an isotopy

$$\widehat{F}\colon E(\xi|_{K_W})\times I\subset\theta(X)\times I\xrightarrow{F}E(\nu)$$

from

$$\widehat{F}_0 = E(\xi|_{K_W}) \subset \theta(X) \xrightarrow{F_0 = i} E(\nu)$$

to

$$\widehat{F}_1 = E(\xi|_{K_W}) \subset \theta(X) \xrightarrow{F_1 = i'\phi} E(\nu).$$

By Lemma 6.2,

$$[\widehat{F}_0] = [\widehat{F}_1] \in \Omega^{\mathrm{Witt}}_{n+d}(E(\nu)).$$

Therefore,

$$[g] = [g \circ H_1|] = [\widehat{F}_1] = [\widehat{F}_0] \in \Omega^{\mathrm{Witt}}_{n+d}(E(\nu)).$$

Now the geometric pullback transfer of $[h\colon W \hookrightarrow B]$ is given by

$$\xi^!_{\mathrm{PB}}[h\colon W \hookrightarrow B] = [E(\xi|_{K_W}) \subset E(\xi) = \theta(X)].$$

Hence

$$i_*\xi^!_{\mathrm{PB}}[h\colon W \hookrightarrow B] = [\widehat{F}_0].$$

Finally, since i_* and ρ_* are inverses of each other,

$$\begin{aligned}\xi^!_{\mathrm{PB}}[h\colon W \hookrightarrow B] = \rho_*[\widehat{F}_0] = \rho_*[g] = \rho_*(u_{\mathrm{BRS}}(\nu) \cap [f]) &= \rho_*(u_{\mathrm{BRS}}(\nu) \cap [T(\xi) \circ (\mathrm{id} \times h)]) \\ &= \rho_*(u_{\mathrm{BRS}}(\nu) \cap T(\xi)_*\sigma[h]) = \xi^!_{\mathrm{BRS}}[h],\end{aligned}$$

as was to be shown. □

We will refer to the map $\rho_*(u_{\mathrm{BRS}}(\nu) \cap -)$ as the *geometric Thom homomorphism*.

Proposition 6.6 *The homotopy-theoretic Thom homomorphism* Φ *agrees with the geometric Thom homomorphism*, *that is*, *the diagram*

$$\begin{array}{ccc} \widetilde{\mathrm{MWITT}}_{n+s}(\mathrm{Th}(\mu)) & \xrightarrow{\Phi} & \mathrm{MWITT}_{n+d}(X) \\ \cong\downarrow & & \downarrow\cong \\ \widetilde{\Omega}^{\mathrm{Witt}}_{n+s}(\mathrm{Th}(\nu)) & \xrightarrow{\rho_*(u_{\mathrm{BRS}}(\nu)\cap -)} & \Omega^{\mathrm{Witt}}_{n+d}(X) \end{array}$$

commutes.

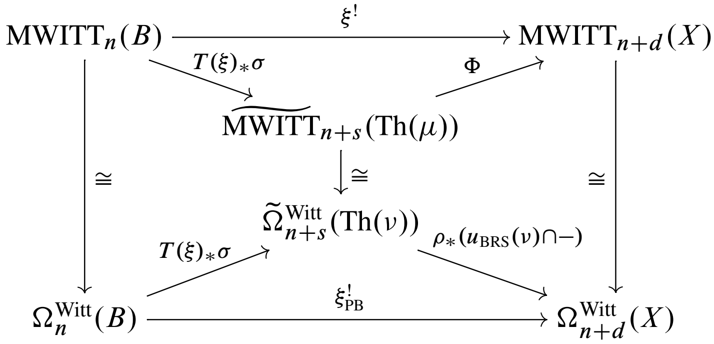
Proof Recall that Φ is given by $\Phi = \rho_*(u_{\mathrm{SPL}}(\mu) \cap -)$. The result follows from Lemma 4.5 applied to μ with underlying oriented block bundle $\mu_{\mathrm{PLB}} = \nu$, together with the geometric description of the cap product given in [12]. □

Proposition 6.7 *Manifold-block bundle transfer on* MWITT*–homology and geometric pullback transfer on Witt bordism agree*, *that is*, *the diagram*

$$\begin{array}{ccc} \mathrm{MWITT}_n(B) & \xrightarrow{\xi^!} & \mathrm{MWITT}_{n+d}(X) \\ \downarrow\cong & & \cong\downarrow \\ \Omega^{\mathrm{Witt}}_n(B) & \xrightarrow{\xi^!_{\mathrm{PB}}} & \Omega^{\mathrm{Witt}}_{n+d}(X) \end{array}$$

commutes.

Proof We must show that the outer square of the diagram



commutes. The upper part commutes by definition of the F–block bundle transfer $\xi^!$. The left-hand part commutes as the vertical arrows are given by a natural isomorphism of homology theories, while the right-hand part commutes by Proposition 6.6. The lower part of the diagram, involving the pullback transfer $\xi^!_{\mathrm{PB}}$, commutes according to Proposition 6.5. □

A closed n–dimensional Witt space W has a *fundamental class*

$$[W]_{\mathrm{Witt}} \in \Omega^{\mathrm{Witt}}_n(W)$$

in Witt bordism represented by the identity map, $[W]_{\mathrm{Witt}} = [\mathrm{id}\colon W \to W]$. This class corresponds to a unique class $[W]_{\mathrm{Witt}} \in \mathrm{MWITT}_n(W)$ under the natural identification (9).

Proposition 6.8 *Suppose B is a closed Witt space of dimension n. Then the total space X of the oriented F–block bundle ξ over B is a closed Witt space and the geometric pullback transfer*

$$\xi^!_{\mathrm{PB}}\colon \Omega^{\mathrm{Witt}}_n(B) \to \Omega^{\mathrm{Witt}}_{n+d}(X)$$

maps the Witt fundamental class of B to the Witt fundamental class of X,

$$\xi^!_{\mathrm{PB}}[B]_{\mathrm{Witt}} = [X]_{\mathrm{Witt}}.$$

Proof If the base B is Witt, then the total space X is Witt by Lemma 6.1. The Witt fundamental class $[B]_{\mathrm{Witt}}$ is represented by the identity map $g = \mathrm{id}_B\colon B \to B$ (which is PL). Pulling back under this identity map, the map $\bar{g}\colon E(\mathrm{id}^*\xi) \to X$ is the identity $\mathrm{id}\colon E(\mathrm{id}^*\xi) = X \to X$. Therefore,

$$\xi^!_{\mathrm{PB}}[\mathrm{id}\colon B \to B] = [\bar{g}\colon E(\mathrm{id}^*\xi) \to X] = [\mathrm{id}_X] = [X]_{\mathrm{Witt}}. \qquad \square$$

Example 6.9 We continue our previous examples on the trivial F–block bundle ξ with total space $X = F \times B$ for B any compact polyhedron. The geometric pullback transfer $\xi^!_{\mathrm{PB}}\colon \Omega^{\mathrm{Witt}}_n(B) \to \Omega^{\mathrm{Witt}}_{n+d}(F \times B)$ is then by construction $\xi^!_{\mathrm{PB}}[g\colon W \to B] = [\mathrm{id}_F \times g\colon F \times W \to F \times B]$. The Witt bordism $\times$–product

$$\times\colon \Omega^{\mathrm{Witt}}_d(F) \times \Omega^{\mathrm{Witt}}_n(B) \to \Omega^{\mathrm{Witt}}_{d+n}(F \times B), \quad [h] \times [g] = [h \times g],$$

can be used to decompose the class $[\mathrm{id}_F \times g]$ as $[F]_{\mathrm{Witt}} \times [g]$. We thus find that

$$\xi^!_{\mathrm{PB}}[g] = [F]_{\mathrm{Witt}} \times [g],$$

which agrees with (13). If $B = W$ is an n–dimensional closed Witt space and g the identity, then $\xi^!_{\mathrm{PB}}[B]_{\mathrm{Witt}} = [F]_{\mathrm{Witt}} \times [B]_{\mathrm{Witt}} = [F \times B]_{\mathrm{Witt}}$, in agreement with Proposition 6.8.

7 Transfer of the $\mathbb{L}^\bullet$–homology fundamental class

In [7], we constructed a canonical $\mathbb{L}^\bullet(\mathbb{Q})$*–homology fundamental class*

$$[X]_{\mathbb{L}} \in \mathbb{L}^\bullet(\mathbb{Q})_n(X)$$

for closed n–dimensional Witt spaces X using the morphism $\tau\colon \mathrm{MWITT} \to \mathbb{L}^\bullet(\mathbb{Q})$ of ring spectra. This class is the image of the Witt theory fundamental class $[X]_{\mathrm{Witt}}$ under the map

$$\tau_*\colon \Omega_n^{\mathrm{Witt}}(X) \cong \mathrm{MWITT}_n(X) \to \mathbb{L}^\bullet(\mathbb{Q})_n(X),$$

ie $[X]_{\mathbb{L}} = \tau_*[X]_{\mathrm{Witt}}$.

Theorem 7.1 *Suppose B is a closed Witt space of dimension n. Then the total space X of the oriented F–block bundle ξ over B is a closed Witt space and the block bundle transfer*

$$\xi^!\colon \mathbb{L}^\bullet(\mathbb{Q})_n(B) \to \mathbb{L}^\bullet(\mathbb{Q})_{n+d}(X)$$

maps the $\mathbb{L}^\bullet(\mathbb{Q})$–homology fundamental class of B to the $\mathbb{L}^\bullet(\mathbb{Q})$–homology fundamental class of X,

$$\xi^![B]_{\mathbb{L}} = [X]_{\mathbb{L}}.$$

Proof By Proposition 6.8, $\xi^!_{\mathrm{PB}}[B]_{\mathrm{Witt}} = [X]_{\mathrm{Witt}}$ for the pullback transfer. Thus, using Proposition 6.7 on the compatibility of block bundle transfer and pullback transfer,

$$\xi^![B]_{\mathrm{Witt}} = \xi^!_{\mathrm{PB}}[B]_{\mathrm{Witt}} = [X]_{\mathrm{Witt}}.$$

Finally, by Proposition 5.6,

$$\xi^![B]_{\mathbb{L}} = \xi^!\tau_*[B]_{\mathrm{Witt}} = \tau_*\xi^![B]_{\mathrm{Witt}} = \tau_*[X]_{\mathrm{Witt}} = [X]_{\mathbb{L}}. \qquad \square$$

Example 7.2 We describe the $\mathbb{L}^\bullet(\mathbb{Q})$–homology transfer and illustrate Theorem 7.1 for the trivial F–block bundle ξ with total space $X = F \times B$. We use the notation of the earlier examples on this special case. By Lemma 5.8, $u_{\mathbb{L}}(\mu_F) = \tau\phi_W(u_{\mathrm{SPL}}(\mu_F))$. Hence, using [38, page 552, Proposition 7.1.2],

$$[F]_{\mathbb{L}} = \rho_{F*}\big(\tau\phi_W(u_{\mathrm{SPL}}(\mu_F)) \cap [\mathrm{Th}\,\mu_F]_{\mathbb{L}}\big) \in \mathbb{L}^\bullet(\mathbb{Q})_d(F).$$

See also [39, page 186, Proposition 16.16(c)]. Consequently, (13) applies to yield the description

$$\xi^!(a) = [F]_{\mathbb{L}} \times a$$

for the transfer $\xi^!\colon \mathbb{L}^\bullet(\mathbb{Q})_n(B) \to \mathbb{L}^\bullet(\mathbb{Q})_{n+d}(F \times B)$. When B is a closed n–dimensional Witt space, we obtain

$$\xi^![B]_{\mathbb{L}} = [F]_{\mathbb{L}} \times [B]_{\mathbb{L}} = [F \times B]_{\mathbb{L}}$$

(where the second equality has been established in [7, Theorem 13.1]), in agreement with Theorem 7.1.

8 Behavior of the Cheeger–Goresky–MacPherson L–class under transfer

Rationally, Theorem 7.1 leads to a formula describing the behavior of the Cheeger–Goresky–MacPherson L–class under block bundle transfer.

Theorem 8.1 *Let B be a closed Witt space and let F be a closed oriented* PL *manifold. Let ξ be an oriented* PL *F–block bundle over B with total space X and oriented stable vertical normal* PL *microbundle μ over X. Then the associated block bundle transfer $\xi^!$ sends the Cheeger–Goresky–MacPherson L–class of B to the product*

$$\xi^! L_*(B) = L^*(\mu) \cap L_*(X).$$

Proof By Theorem 7.1, the $\mathbb{L}^\bullet$–homology transfer $\xi^!$ of ξ sends the $\mathbb{L}^\bullet(\mathbb{Q})$–homology fundamental class of B to the $\mathbb{L}^\bullet(\mathbb{Q})$–homology fundamental class of X: $\xi^![B]_{\mathbb{L}} = [X]_{\mathbb{L}}$. It remains to analyze what this equation means after we tensor with $\mathbb{Q}$, ie after we apply the localization morphism

$$\mathbb{L}^\bullet(\mathbb{Q}) \to \mathbb{L}^\bullet(\mathbb{Q})_{(0)} = \bigvee_i S^i H(L^i(\mathbb{Q}) \otimes \mathbb{Q}) = \bigvee_j S^{4j} H\mathbb{Q},$$

which is a ring morphism of ring spectra. By [7, Lemma 11.1],

$$[B]_{\mathbb{L}} \otimes \mathbb{Q} = L_*(B) \quad \text{and} \quad [X]_{\mathbb{L}} \otimes \mathbb{Q} = L_*(X).$$

Let $\Delta_{\mathrm{SPL}} \colon \mathrm{MSPL} \to \mathrm{KO}\left[\frac{1}{2}\right]$ be the Sullivan orientation [45]. Using work of Land and Nikolaus [30], we construct in [6, Proposition 2.1] a particular equivalence of $\mathbb{E}_\infty$–ring spectra

$$\kappa \colon \mathrm{KO}\left[\tfrac{1}{2}\right] \xrightarrow{\simeq} \mathbb{L}^\bullet(\mathbb{R})\left[\tfrac{1}{2}\right] = \mathbb{L}^\bullet(\mathbb{Z})\left[\tfrac{1}{2}\right]$$

and show that the composition

$$\mathrm{MSPL} \xrightarrow{\Delta_{\mathrm{SPL}}} \mathrm{KO}\left[\tfrac{1}{2}\right] \xrightarrow{\kappa} \mathbb{L}^\bullet(\mathbb{Z})\left[\tfrac{1}{2}\right]$$

is homotopic to Ranicki's orientation σ^* [6, Proposition 3.3]. Furthermore, using work of Taylor and Williams [47] as well as of Morgan and Sullivan [35], we describe in [6] a particular equivalence

$$\mathbb{L}^\bullet(\mathbb{Z})_{(0)} \xrightarrow{\simeq} \bigoplus_{i \in \mathbb{Z}} H\mathbb{Q}[4i]$$

such that the diagram

$$\begin{array}{ccccccc}
\mathrm{MSPL} & \xrightarrow{\Delta_{\mathrm{SPL}}} & \mathrm{KO}\left[\frac{1}{2}\right] & \xrightarrow{\mathrm{loc}} & \mathrm{KO}_{(0)} & \xrightarrow[\simeq]{\mathrm{ph}} & \bigoplus_{i\in\mathbb{Z}} H\mathbb{Q}[4i] \\
 & \searrow^{\sigma^*} & \kappa \downarrow \simeq & & \kappa_{(0)} \downarrow \simeq & & \| \\
 & & \mathbb{L}^\bullet\left[\frac{1}{2}\right] & \xrightarrow{\mathrm{loc}} & \mathbb{L}^\bullet_{(0)} & \xrightarrow{\simeq} & \bigoplus_{i\in\mathbb{Z}} H\mathbb{Q}[4i]
\end{array}$$

commutes up to homotopy, where ph denotes the Pontryagin character. Now, it is well known that the Pontryagin character of the Sullivan orientation is given by

$$\mathrm{ph}(\mathrm{loc}\,\Delta_{\mathrm{SPL}}) = L^{-1} \cup u \in H^*(\mathrm{MSPL}; \mathbb{Q}),$$

where L is the universal PL L–class $L \in H^*(\mathrm{BSPL}; \mathbb{Q})$ and u the stable Thom class

$$u \in H^0(\mathrm{MSPL}; \mathbb{Z}) = \mathbb{Z}.$$

(See Madsen and Milgram [33, Corollary 5.4, page 102].) Thus by commutativity of the diagram, the rational localization of σ^* is given by $L^{-1} \cup u$. Hence, for our PL microbundle $\mu\colon X \to \mathrm{BSPL}$,

$$u_{\mathbb{L}}(\mu) \otimes \mathbb{Q} = \rho^* L^*(\mu)^{-1} \cup u_{\mathbb{Q}}(\mu),$$

where $u_{\mathbb{Q}}(\mu) \in \widetilde{H}^{s-d}(\mathrm{Th}(\mu); \mathbb{Q})$ is the Thom class of μ in ordinary rational cohomology. (See also Ranicki's [39, Remark 16.2, page 176] for topological block bundles. Note that a PL microbundle has an underlying topological block bundle by composition with $\mathrm{BSPL} \to \mathrm{B}\widetilde{\mathrm{SPL}} \to \mathrm{B}\widetilde{\mathrm{STOP}}$, and that Ranicki omits cupping with $u_{\mathbb{Q}}(\mu)$ in his notation.) Thus

$$\begin{aligned} L_*(X) = [X]_{\mathbb{L}} \otimes \mathbb{Q} &= (\xi^![B]_{\mathbb{L}}) \otimes \mathbb{Q} = \rho_*(u_{\mathbb{L}}(\mu) \cap T(\xi)_*\sigma[B]_{\mathbb{L}}) \otimes \mathbb{Q} \\ &= \rho_*\big(u_{\mathbb{L}}(\mu) \otimes \mathbb{Q} \cap T(\xi)_*\sigma([B]_{\mathbb{L}} \otimes \mathbb{Q})\big) = \rho_*\big((\rho^* L^*(\mu)^{-1} \cup u_{\mathbb{Q}}(\mu)) \cap T(\xi)_*\sigma L_*(B)\big) \\ &= \rho_*\big(\rho^* L^*(\mu)^{-1} \cap (u_{\mathbb{Q}}(\mu) \cap T(\xi)_*\sigma L_*(B))\big) = L^*(\mu)^{-1} \cap \rho_*\big(u_{\mathbb{Q}}(\mu) \cap T(\xi)_*\sigma L_*(B)\big) \\ &= L^*(\mu)^{-1} \cap \xi^! L_*(B). \end{aligned}$$

□

If t is a stable inverse for μ, then t has the interpretation of a stable vertical tangent bundle for ξ, and by Theorem 8.1, the following formula holds:

$$\xi^! L_*(B) = L^*(t)^{-1} \cap L_*(X).$$

Example 8.2 We discuss Theorem 8.1 vis-à-vis (13) in the situation of a trivial F–block bundle ξ over B, using the notation of earlier examples on this case. Let $[F]_{\mathbb{Q}} \in H_d(F; \mathbb{Q})$ denote the fundamental class of the oriented PL manifold F in ordinary rational homology. By (13),

$$\xi^!(a) = [F]_{\mathbb{Q}} \times a$$

for $a \in H_n(B; \mathbb{Q})$. For a closed Witt space B, we obtain in particular

$$\xi^! L_*(B) = [F]_{\mathbb{Q}} \times L_*(B). \tag{23}$$

Let TF denote the tangent PL microbundle of the PL manifold F. Then $\mu_F \oplus TF$ is the trivial microbundle, and hence $L^*(\mu_F)L^*(TF) = L^*(\mu_F \oplus TF) = 1$. Furthermore, the Hirzebruch signature theorem holds for PL manifolds and $L_*(F) = L^*(TF) \cap [F]_{\mathbb{Q}}$; see Madsen and Milgram [33, Chapter 4C] and Thom [48]. According to Theorem 8.1,

$$\begin{aligned} \xi^! L_*(B) = L^*(\mu) \cap L_*(X) &= (L^*(\mu_F) \times 1) \cap (L_*(F) \times L_*(B)) = (L^*(\mu_F) \cap L_*(F)) \times (1 \cap L_*(B)) \\ &= (L^*(\mu_F) \cap L^*(TF) \cap [F]_{\mathbb{Q}}) \times L_*(B) = [F]_{\mathbb{Q}} \times L_*(B), \end{aligned}$$

confirming (23). It is perhaps worthwhile to emphasize that transfer does not in general commute with localization of spectra: if $\xi^!_{\mathbb{Q}}$ denotes the transfer on ordinary rational homology and $\xi^!_{\mathbb{L}}$ the transfer on $\mathbb{L}^\bullet(\mathbb{Q})$–homology, then generally $\xi^!_{\mathbb{Q}}(- \otimes \mathbb{Q}) \neq \xi^!_{\mathbb{L}}(-) \otimes \mathbb{Q}$. For example,

$$\xi^!_{\mathbb{Q}}([B]_{\mathbb{L}} \otimes \mathbb{Q}) = \xi^!_{\mathbb{Q}}(L_*(B)) = [F]_{\mathbb{Q}} \times L_*(B),$$

which contains less information than

$$(\xi^!_{\mathbb{L}}[B]_{\mathbb{L}}) \otimes \mathbb{Q} = [F \times B]_{\mathbb{L}} \otimes \mathbb{Q} = L_*(F \times B) = L_*(F) \times L_*(B).$$

9 Normally nonsingular maps

Let $f: Y \to X$ be a PL map of closed Witt spaces which is the composition

$$\begin{array}{ccc} Y & \overset{g}{\hookrightarrow} & Z \\ & {\scriptstyle f}\searrow & \downarrow{\scriptstyle p} \\ & & X \end{array}$$

of an oriented normally nonsingular inclusion g with normal bundle ν_g, followed by the projection p of an oriented PL F–fiber bundle ξ with closed PL manifold fiber F and stable vertical normal bundle ν_ξ. Then f is a *normally nonsingular map* in the sense of [22, Definition 5.4.3]. Let c be the codimension of g and d the dimension of F. The bundle transfer $\xi^!$ and the Gysin restriction $g^!$ compose to give a transfer homomorphism

$$H_n(X;\mathbb{Q}) \xrightarrow{\xi^!} H_{n+d}(Z;\mathbb{Q}) \xrightarrow{g^!} H_{n+d-c}(Y;\mathbb{Q}),$$

with $c-d$ the *relative dimension* of f. Combining Theorem 8.1 with [5, Theorem 3.18], we obtain

$$\begin{aligned} g^!\xi^! L_*(X) &= g^!(L^*(\nu_\xi) \cap L_*(Z)) = g^* L^*(\nu_\xi) \cap g^! L_*(Z) = g^* L^*(\nu_\xi) \cap (L^*(\nu_g) \cap L_*(Y)) \\ &= L^*(g^*\nu_\xi \oplus \nu_g) \cap L_*(Y), \end{aligned}$$

at least when Y and Z have even dimensions.

References

[1] **M Banagl**, *Extending intersection homology type invariants to non-Witt spaces*, Mem. Amer. Math. Soc. 760 (2002) MR Zbl

[2] **M Banagl**, *The L–class of non-Witt spaces*, Ann. of Math. 163 (2006) 743–766 MR Zbl

[3] **M Banagl**, *Topological invariants of stratified spaces*, Springer (2007) MR Zbl

[4] **M Banagl**, *Topological and Hodge L–classes of singular covering spaces and varieties with trivial canonical class*, Geom. Dedicata 199 (2019) 189–224 MR Zbl

[5] **M Banagl**, *Gysin restriction of topological and Hodge-theoretic characteristic classes for singular spaces*, New York J. Math. 26 (2020) 1273–1337 MR Zbl

[6] **M Banagl**, *Transfer and the spectrum-level Siegel–Sullivan* KO*–orientation for singular spaces*, preprint (2023) arXiv 2305.06045

[7] **M Banagl**, **G Laures**, **J E McClure**, *The L–homology fundamental class for* IP*–spaces and the stratified Novikov conjecture*, Selecta Math. 25 (2019) art. id. 7 MR Zbl

[8] **J C Becker**, **D H Gottlieb**, *The transfer map and fiber bundles*, Topology 14 (1975) 1–12 MR Zbl

[9] **J M Boardman**, *Stable homotopy theory*, *V*: *Duality and Thom spectra*, mimeographed notes, University of Warwick (1966)

[10] **J F de Bobadilla**, **I Pallarés**, *The Brasselet–Schürmann–Yokura conjecture on L–classes of singular varieties*, preprint (2020) arXiv 2007.11537

[11] **J-P Brasselet**, **J Schürmann**, **S Yokura**, *Hirzebruch classes and motivic Chern classes for singular spaces*, J. Topol. Anal. 2 (2010) 1–55 MR Zbl

[12] **S Buoncristiano**, **C P Rourke**, **B J Sanderson**, *A geometric approach to homology theory*, London Math. Soc. Lecture Note Ser. 18, Cambridge Univ. Press (1976) MR Zbl

[13] **S Cappell**, **S Weinberger**, *Classification de certains espaces stratifiés*, C. R. Acad. Sci. Paris Sér. I Math. 313 (1991) 399–401 MR Zbl

[14] **S Carmeli**, **B Cnossen**, **M Ramzi**, **L Yanovski**, *Characters and transfer maps via categorified traces*, preprint (2023) arXiv 2210.17364v2

[15] **A J Casson**, *Generalisations and applications of block bundles*, from "The Hauptvermutung book", *K*–Monogr. Math. 1, Kluwer Academic, Dordrecht (1996) 33–67 MR Zbl

[16] **J Cheeger**, *Spectral geometry of singular Riemannian spaces*, J. Differential Geom. 18 (1983) 575–657 MR Zbl

[17] **M M Cohen**, *A general theory of relative regular neighborhoods*, Trans. Amer. Math. Soc. 136 (1969) 189–229 MR Zbl

[18] **S J Curran**, *Intersection homology and free group actions on Witt spaces*, Michigan Math. J. 39 (1992) 111–127 MR Zbl

[19] **J Ebert**, **O Randal-Williams**, *Generalised Miller–Morita–Mumford classes for block bundles and topological bundles*, Algebr. Geom. Topol. 14 (2014) 1181–1204 MR Zbl

[20] **G Friedman**, *Singular intersection homology*, New Math. Monogr. 33, Cambridge Univ. Press (2020) MR Zbl

[21] **M Goresky**, **R MacPherson**, *Intersection homology theory*, Topology 19 (1980) 135–162 MR Zbl

[22] **M Goresky**, **R MacPherson**, *Intersection homology, II*, Invent. Math. 72 (1983) 77–129 MR Zbl

[23] **M Goresky**, **R MacPherson**, *Stratified Morse theory*, Ergebnisse der Math. 14, Springer (1988) MR Zbl

[24] **M Goresky**, **P Siegel**, *Linking pairings on singular spaces*, Comment. Math. Helv. 58 (1983) 96–110 MR Zbl

[25] **F Hebestreit**, **M Land**, **W Lück**, **O Randal-Williams**, *A vanishing theorem for tautological classes of aspherical manifolds*, Geom. Topol. 25 (2021) 47–110 MR Zbl

[26] **W C Hsiang**, **C T C Wall**, *Orientability of manifolds for generalised homology theories*, Trans. Amer. Math. Soc. 118 (1965) 352–359 MR Zbl

[27] **F Kirwan**, **J Woolf**, *An introduction to intersection homology theory*, 2nd edition, Chapman & Hall, Boca Raton, FL (2006) MR Zbl

[28] **J R Klein**, **C Malkiewich**, **M Ramzi**, *On the multiplicativity of the Euler characteristic*, Proc. Amer. Math. Soc. 151 (2023) 4997–5006 MR Zbl

[29] **N H Kuiper**, **R K Lashof**, *Microbundles and bundles, I: Elementary theory*, Invent. Math. 1 (1966) 1–17 MR Zbl

[30] **M Land**, **T Nikolaus**, *On the relation between K– and L–theory of C^*–algebras*, Math. Ann. 371 (2018) 517–563 MR Zbl

[31] **G Laures**, **J E McClure**, *Multiplicative properties of Quinn spectra*, Forum Math. 26 (2014) 1117–1185 MR Zbl

[32] **W Lück**, **A Ranicki**, *Surgery obstructions of fibre bundles*, J. Pure Appl. Algebra 81 (1992) 139–189 MR Zbl

[33] **I Madsen**, **R J Milgram**, *The classifying spaces for surgery and cobordism of manifolds*, Ann. of Math. Stud. 92, Princeton Univ. Press (1979) MR Zbl

[34] **L Maxim**, **J Schürmann**, *Characteristic classes of mixed Hodge modules and applications*, from "Schubert varieties, equivariant cohomology and characteristic classes" (J Buczyński, M Michałek, E Postinghel, editors), Eur. Math. Soc., Zürich (2018) 159–202 MR Zbl

[35] **J W Morgan**, **D P Sullivan**, *The transversality characteristic class and linking cycles in surgery theory*, Ann. of Math. 99 (1974) 463–544 MR Zbl

[36] **W L Pardon**, *Intersection homology Poincaré spaces and the characteristic variety theorem*, Comment. Math. Helv. 65 (1990) 198–233 MR Zbl

[37] **A Ranicki**, *The total surgery obstruction*, from "Algebraic topology" (J L Dupont, I H Madsen, editors), Lecture Notes in Math. 763, Springer (1979) 275–316 MR Zbl

[38] **A Ranicki**, *Exact sequences in the algebraic theory of surgery*, Math. Notes 26, Princeton Univ. Press (1981) MR Zbl

[39] **A A Ranicki**, *Algebraic L–theory and topological manifolds*, Cambridge Tracts in Math. 102, Cambridge Univ. Press (1992) MR Zbl

[40] **C P Rourke**, *Block structures in geometric and algebraic topology*, from "Actes du Congrès International des Mathématiciens" (M Berger, J Dieudonné, J Leray, J-L Lions, P Malliavin, J-P Serre, editors), volume II, Gauthier-Villars, Paris (1971) 127–132 MR Zbl

[41] **C P Rourke**, **B J Sanderson**, *Block bundles, I*, Ann. of Math. 87 (1968) 1–28 MR Zbl

[42] **Y B Rudyak**, *On Thom spectra, orientability, and cobordism*, Springer (1998) MR Zbl

[43] **J Schürmann**, *Specialization of motivic Hodge–Chern classes*, preprint (2009) arXiv 0909.3478

[44] **P H Siegel**, *Witt spaces: a geometric cycle theory for KO–homology at odd primes*, Amer. J. Math. 105 (1983) 1067–1105 MR Zbl

[45] **D Sullivan**, *Geometric topology*, *I*: *Localization*, *periodicity*, *and Galois symmetry*, Massachusetts Institute of Technology, Cambridge, MA (1970) MR Zbl

[46] **R M Switzer**, *Algebraic topology: homotopy and homology*, Grundl. Math. Wissen. 212, Springer (1975) MR Zbl

[47] **L Taylor**, **B Williams**, *Surgery spaces: formulae and structure*, from "Algebraic topology" (P Hoffman, V Snaith, editors), Lecture Notes in Math. 741, Springer (1979) 170–195 MR Zbl

[48] **R Thom**, *Les classes caractéristiques de Pontrjagin des variétés triangulées*, from "Symposium internacional de topología algebraica", Universidad Nacional Autónoma de México and UNESCO, México (1958) 54–67 MR Zbl

[49] **S Weinberger**, *The topological classification of stratified spaces*, Univ. of Chicago Press (1994) MR Zbl

[50] **R E Williamson, Jr**, *Cobordism of combinatorial manifolds*, Ann. of Math. 83 (1966) 1–33 MR Zbl

Institut für Mathematik, Ruprecht-Karls-Universität Heidelberg
Heidelberg, Germany

banagl@mathi.uni-heidelberg.de

Received: 8 April 2022 Revised: 9 August 2023 msp

Algebraic & Geometric Topology **24**:5 (2024) 2619–2654
DOI: 10.2140/agt.2024.24.2619
Published: 19 August 2024

A reduction of the string bracket to the loop product

KATSUHIKO KURIBAYASHI
TAKAHITO NAITO
SHUN WAKATSUKI
TOSHIHIRO YAMAGUCHI

The negative cyclic homology for a differential graded algebra over the rational field has a quotient of the Hochschild homology as a direct summand if the S–action is trivial. With this fact, we show that the string bracket in the sense of Chas and Sullivan is reduced to the loop product followed by the BV operator on the loop homology provided the given manifold is *BV-exact*. The reduction is indeed derived from the equivalence between the BV-exactness and the triviality of the S–action. Moreover, it is proved that a Lie bracket on the loop cohomology of the classifying space of a connected compact Lie group possesses the same reduction. By using these results, we consider the nontriviality of string brackets. We also show that a simply connected space with positive weights is BV-exact. Furthermore, the *higher BV-exactness* is discussed featuring the cobar-type Eilenberg–Moore spectral sequence.

55P35, 55P50, 55T20

1 Introduction

Let LM be the free loop space, namely, the space of continuous maps from the circle S^1 to a space M with compact-open topology. The rotation on the domain space S^1 of LM induces an S^1–action on LM. Then we have the S^1–equivariant homology $H_*^{S^1}(LM) = H_*(ES^1 \times_{S^1} LM)$ for a space M. The *string bracket* is a Lie bracket on the S^1–equivariant homology of the free loop space LM of an orientable closed manifold M, which is introduced by Chas and Sullivan in [8]. The bracket is defined by using the loop product on the loop homology $H_*(LM)$ and maps in the Gysin exact sequence of the S^1–principal bundle

$$S^1 \to ES^1 \times LM \to ES^1 \times_{S^1} LM. \tag{1.1}$$

In particular, the Batalin–Vilkovisky (BV) identity of the BV operator on the loop homology induces the Jacobi identity for the string bracket; see the proof of [8, Theorem 6.1].

As for computations of the string brackets, Basu [2] and Félix, Thomas and Vigué-Poirrier [19] have determined explicitly the rational string bracket of the product of spheres. For a simply connected closed manifold M whose rational cohomology is generated by a single element, the rational string bracket is trivial though the rational loop product of M is highly nontrivial; see [2, Theorem 3.4] and [19, Section 5.2, Example 1]. On the other hand, a result due to Tabing [43] shows that the integral string bracket of the sphere is nontrivial.

The loop homology of the classifying space BG of a connected compact Lie group G in the sense of Chataur and Menichi [9] admits the BV algebra structure; see also [27, Theorem C.1]. Therefore, the same argument as that about manifolds allows us to deduce that the string cohomology of BG is endowed with a graded Lie algebra structure; see Proposition 2.3 and Chen, Eshmatov and Liu [11, Theorem 1.1].

The aim of this article is to investigate general methods for computing the rational string brackets for a manifold and the classifying space of a connected compact Lie group. The key strategy is to use Jones' isomorphisms

$$H^*(LM;\mathbb{Q}) \cong \mathrm{HH}_*(A_{\mathrm{PL}}(M)) \quad \text{and} \quad H^*_{S^1}(LM;\mathbb{Q}) \cong \mathrm{HC}^-_*(A_{\mathrm{PL}}(M)),$$

where $A_{\mathrm{PL}}(M)$ is the polynomial de Rham algebra over $\mathbb{Q}$ of a simply connected space M and the right-hand sides of the isomorphisms denote the Hochschild homology and the negative cyclic homology of the complex, respectively; see Section 3 for more details. Furthermore, the decomposition theorem of the negative cyclic homology and the cyclic homology (additive K–theory in the sense of Feigin and Tsygan [20]) in Vigué-Poirrier [46] and Kuribayashi and Yamaguchi [29] is applied in the computation; see Theorem 2.15. It turns out that for a simply connected closed manifold M, the rational string bracket for M is reduced to the loop product of M followed by the BV operator provided the manifold possesses the exactness of the operator; see Definition 2.9.

Assertion 1.2 *Let M be a simply connected closed manifold. Suppose further that M is **BV-exact**. Then the string bracket in the string homology $H_*^{S^1}(LM;\mathbb{Q})$ is regarded as the **loop bracket** in the loop homology $H_*(LM;\mathbb{Q})$ up to isomorphism, and hence the string bracket is determined by the Gerstenhaber bracket in the Hochschild cohomology of the polynomial de Rham algebra $A_{\mathrm{PL}}(M)$ of M.*

The detail is described in Corollary 2.16. In particular, the nilpotency of the string bracket is equivalent to that of the Gerstenhaber bracket. We stress that the Gerstenhaber algebra in Assertion 1.2 is considered with the Lie model for M without using the loop product; see Félix, Menichi and Thomas [16]. It is worth mentioning that the BV-exactness, which is introduced to consider the reduction of the string brackets, is a new homotopy invariant deeply related to other traditional rational homotopy invariants for spaces. We discuss and summarize this topic in Assertion 1.3 below and several paragraphs before the assertion.

Félix, Thomas and Vigué-Poirrier [19] have given an explicit description of the rational string bracket of M with its Sullivan model. On the other hand, our method for computing the string bracket is formulated with the loop product and the BV operator on the loop homology. Moreover, the BV-exactness is also described in terms of the loop homology. Therefore, it is possible to make a computation of the dual to the string bracket on the equivariant homology $H_*^{S^1}(LM;\mathbb{Q})$ by considering *only* behavior of the BV operator on the loop homology $H_*(LM;\mathbb{Q})$; see Remark 2.14 for more details. This is an advantage of our result.

In the case of the classifying space, the same strategy as above is applicable in the computation of the string bracket. In fact, for the classifying space BG of *every* compact connected Lie group G, the rational string bracket for BG is described as the BV operator followed by the dual loop coproduct; see Theorems 2.7(i) and 2.8(i). As for general properties of the string brackets, the theorems allow us to deduce that the Lie bracket on the string cohomology $H^*_{S^1}(LBG;\mathbb{Q})$ is highly nontrivial even if $\operatorname{rank} G = 1$; see Proposition 5.2. Moreover, Propositions 5.3 and 5.4 assert that the loop homology endowed with the string bracket of a simply connected Lie group G is nilpotent if and only if $\operatorname{rank} G = 1$.

The notion of a *Gorenstein space* due to Félix, Halperin and Thomas [14] enables us to deal with a manifold and the classifying space of a Lie group simultaneously. As a consequence, with the influence of string topology on Gorenstein spaces (see Félix and Thomas [18]), we have Theorems 2.7, 2.8 and 2.15 mentioned above.

We moreover propose a method for computing the string bracket of a non-BV-exact space M. To this end, we introduce a bracket on the cobar-type Eilenberg–Moore spectral sequence (EMSS) converging to $H^*_{S^1}(LM;\mathbb{Q})$ which is compatible with the string bracket of the target; see Theorem 7.7. Moreover, the EMSS carries a decomposition compatible with the Hodge decomposition of the target; see Remark 7.3. While there is no computational example obtained by applying the spectral sequence, in future work, it is expected that the EMSS is applicable in computing the string bracket explicitly; see Section 1.1 problems.

As described above, the BV-exactness is a key to computing string brackets on Gorenstein spaces. Moreover, it is worthwhile mentioning that the BV-exactness for a space M is equivalent to the triviality of the S–action in Connes' exact sequence; see Theorem 2.11. In fact, the new invariant is only described in terms of the Hochschild homology while the S–action is defined on the negative cyclic homology. A deep consideration due to Vigué-Poirrier in [45; 46] shows that the S–action on the negative cyclic homology is trivial if M is formal. Thus we see that the class of BV-exact spaces contains that of formal spaces; see Corollary 2.13.

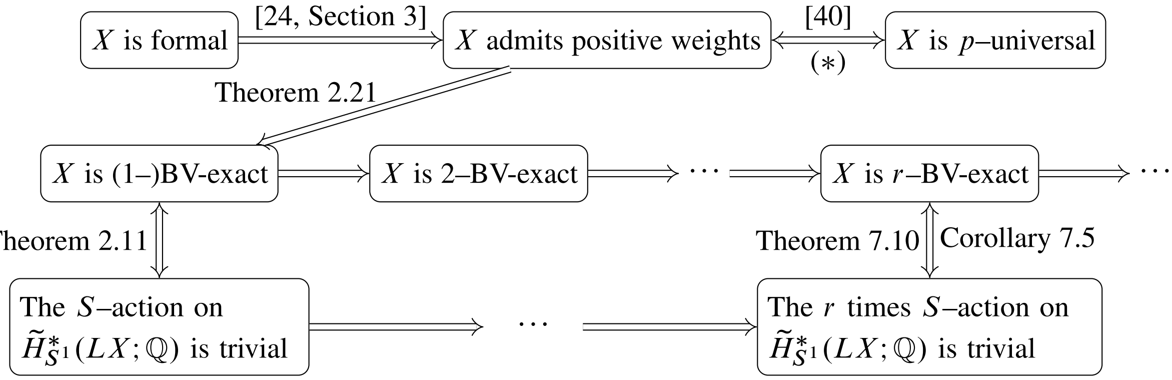
With historical perspectives, we comment on relationships among notions of *p–universality* in Mimura, O'Neill and Toda [36], *positive weights* in Body and Douglas [4], the BV-exactness and its variants; see Definition 2.20 for positive weights.

By definition, simply connected spaces X and Y are said to be *p–equivalent* if there is a map $f: X \to Y$ which induces $H^*(X;\mathbb{Z}/p) \cong H^*(Y;\mathbb{Z}/p)$, where p is a prime or zero and $\mathbb{Z}/0 = \mathbb{Q}$. In [41], Serre raised the so-called symmetry question whether the existence of a p–equivalence $X \to Y$ implies the existence of a p–equivalence in the reverse direction $Y \to X$. However, in general, the p–equivalence does not satisfy the symmetricity.

Mimura, O'Neil and Toda [36] defined the notion of a *p–universal* space and proved that in the full subcategory of p–universal spaces of the category of simply connected spaces whose homotopy types are those of finite CW complexes, the p–equivalence is indeed an equivalence relation. We observe that the p–universality does not depend on p or 0; see [36, Proposition 2.9]. Afterward, Body and Douglas [4] defined the concept of positive weights for Sullivan minimal models. Scheerer's result [40, Theorem 2], in turn, yields that the two notions of p–universality and positive weights are equivalent.

By using the EMSS mentioned above, we also introduce the notion of *r–BV-exactness*; see Definition 7.11. The r–BV-exactness for a simply connected space M is equivalent to the collapsing at the E_{r+1}–term of the EMSS for M; see Corollary 7.5. The decomposition of the EMSS allows us to deduce that the notion of BV-exactness is indeed equivalent to that of 1–BV-exactness; see Theorem 7.10. Thus r–BV-exactness is regarded as a higher version of BV-exactness. We summarize important relationships among invariants mentioned above.

Assertion 1.3 *The following implications concerning rational homotopy invariants hold for a simply connected space X:*



Here the reduced cohomology $\widetilde{H}^*_{S^1}(LX;\mathbb{Q})$ is the cokernel of the map $H^*_{S^1}(*;\mathbb{Q}) \to H^*_{S^1}(LX;\mathbb{Q})$ induced by the trivial map, and the S–action on $\widetilde{H}^*_{S^1}(LX;\mathbb{Q})$ is defined by the multiplication of the generator of $\widetilde{H}^*(BS^1;\mathbb{Q})$ with the map induced by the projection q of the fibration

$$LX \to ES^1 \times_{S^1} LX \xrightarrow{q} BS^1.$$

Observe that the equivalence $(*)$ holds if X has the homotopy type of a finite CW complex.

As mentioned above, a simply connected space admitting positive weights is BV-exact. Proposition 6.1 gives an example of a nonformal BV-exact manifold. Moreover, we obtain an elliptic and non-BV-exact space in Appendix A.

This manuscript is organized as follows. In Section 2, our results are stated in detail. In Section 3, we recall the Hochschild homology, the cyclic homology and Connes' exact sequences. Moreover, the Gorenstein space in the sense of Félix, Halperin and Thomas [14] is also recalled. Section 4 provides the proofs of our results described in Section 2. Section 5 discusses the nilpotency of the string homology of a Lie group and the classifying space of a Lie group. In Section 6, the BV-exactness for a nonformal manifold of dimension 11 is considered. Thanks to the reduction for computing the bracket described in Section 2, we determine explicitly the dual string bracket for the manifold; see Theorem 6.6. We believe that the result gives the first example which computes the string bracket of a non formal space. Section 7 considers the cobar-type Eilenberg–Moore spectral sequence (EMSS) for computing string brackets of non-BV-exact manifolds.

In Appendix A, we obtain an example of an elliptic and non-BV-exact space. Appendix B describes the Gysin exact sequence associated with the principal bundle $S^1 \to ES^1 \times LM \xrightarrow{p} ES^1 \times_{S^1} LM$ for a simply connected space M in terms of Sullivan models; see Whitehead [52, (5.12) Theorem] for the exact sequence.

Finally, on page 2651 there is a list of symbols used repeatedly in this article.

1.1 Problems

We propose questions and problems on topics in this article.

(P1) If a space is BV-exact, then does it admit positive weights?

(P2) For each $r > 1$, is there an r–BV-exact space which is not $(r-1)$–BV-exact?

(P3) Is a space r–BV-exact for some $r < \infty$?

(P4) By making use of the EMSS in Section 7, compute explicitly the string brackets of a non-BV-exact manifold.

2 String brackets described in terms of the Hochschild homology

While the underlying field in Proposition 2.3 below is of *arbitrary* characteristic, other results described in this section hold for a field of characteristic zero.

Let $\mathbb{K}$ be a field and denote the singular homology and cohomology with coefficients in $\mathbb{K}$ by $H_*(-)$ and $H^*(-)$, respectively. For an orientable closed manifold M of dimension d, the Chas and Sullivan loop product $\bullet$ on the shifted homology $\mathbb{H}_*(LM) := H_{*+d}(LM)$ is unital, associative and graded commutative; see [8]. Consider the principal bundle $S^1 \to ES^1 \times LM \xrightarrow{p} ES^1 \times_{S^1} LM$. The bundle gives rise to the homology Gysin sequence

$$\cdots \to \mathbb{H}_{*-d}(LM) \xrightarrow{p_*} H_*^{S^1}(LM) \xrightarrow{c} H_{*-2}^{S^1}(LM) \xrightarrow{\mathrm{M}} \mathbb{H}_{*-d-1}(LM) \to \cdots .$$

The *string bracket* $[\ ,\ ]$ on $H_*^{S^1}(LM)$ is defined by

$$[a,b] := (-1)^{|a|-d} p_*(\mathrm{M}(a) \bullet \mathrm{M}(b)) \quad \text{for } a,b \in H_*^{S^1}(LM). \tag{2.1}$$

The bracket is of degree $2-d$ and gives a Lie algebra structure to the equivariant homology of LM.

Let G be a connected compact Lie group of dimension d. We write $\mathbb{H}^*(LBG) := H^{*+d}(LBG)$ and $\mathcal{H}^*(LBG) := H_{S^1}^{*+d+1}(LBG)$. With this notation, the cohomology Gysin sequence associated with the principal bundle $S^1 \to ES^1 \times LBG \xrightarrow{p} ES^1 \times_{S^1} LBG$ induces an exact sequence of the form

$$\cdots \to \mathcal{H}^{*-2}(LBG) \xrightarrow{S} \mathcal{H}^*(LBG) \xrightarrow{\pi:=p^*} \mathbb{H}^{*+1}(LBG) \xrightarrow{\beta} \mathcal{H}^{*-1}(LBG) \to \cdots .$$

Chataur and Menichi [9] have proved that there exists an associative and graded commutative multiplication $\odot$ on $\mathbb{H}^*(LBG)$ which is induced by the dual loop coproduct with an appropriate sign; see [27, Corollary B.3] and also Section 3. Then the *dual string cobracket* $[\ ,\ ]$ on $\mathcal{H}^*(LBG)$ is defined by

$$[x,y] := (-1)^{\|x\|} \beta(\pi(x) \odot \pi(y)) \quad \text{for } x,y \in \mathcal{H}^*(LBG). \tag{2.2}$$

Here the notation $\|x\|$ means the degree of x as an element in the shifted cohomology.

Proposition 2.3 *Let G be a connected compact Lie group of dimension d and $\mathbb{K}$ a field of arbitrary characteristic. Then the dual string cobracket gives $\mathcal{H}^*(LBG)$ a graded Lie algebra structure.*

Remark 2.4 Proposition 2.3 is a particular case of [9, Theorem 65] and [11, Theorem 1.1]. The result [9, Theorem 65] shows the Lie algebra structure on a homological conformal field theory. The result [11, Theorem 1.1] describes a gravity algebra structure on the negative cyclic homology of a mixed complex; see [21] for a gravity algebra. We give an elementary proof of this proposition by taking care of sign convention in Section 4.

We relate the string brackets (ie the string bracket (2.1) and the dual string cobracket (2.2)) above to the Hochschild homology and the cyclic homology. Let Ω be a connected differential graded algebra (DGA) over a field $\mathbb{K}$ of arbitrary characteristic. A DGA Ω is called a *cochain algebra* if the differential is of degree $+1$. If the differential of a DGA Ω decreases degree by one, we call the DGA Ω a *chain algebra*. Let Ω be a chain algebra, which is nonpositive; that is, $\Omega = \bigoplus_{i \le 0} \Omega_i$. We recall Connes' exact sequences [30, Theorem 2.2.1 and Proposition 5.1.5] for the Hochschild homology, cyclic homology and the negative cyclic homology of Ω, which are of the form

$$\begin{aligned} &\cdots \to \mathrm{HH}_{n+1}(\Omega) \xrightarrow{I} \mathrm{HC}_{n+1}(\Omega) \xrightarrow{S'} \mathrm{HC}_{n-1}(\Omega) \xrightarrow{B_{\mathrm{HH}}} \mathrm{HH}_n(\Omega) \to \cdots, \\ &\cdots \to \mathrm{HC}^-_{n+2}(\Omega) \xrightarrow{S = \times u} \mathrm{HC}^-_n(\Omega) \xrightarrow{\pi} \mathrm{HH}_n(\Omega) \xrightarrow{\beta} \mathrm{HC}^-_{n+1}(\Omega) \to \cdots, \\ &\cdots \to \mathrm{HC}^-_{n+1}(\Omega) \xrightarrow{\times u} \mathrm{HC}^{\mathrm{per}}_{n-1}(\Omega) \xrightarrow{\tilde{\pi}} \mathrm{HC}_{n-1}(\Omega) \xrightarrow{B_{\mathrm{HC}}} \mathrm{HC}^-_n(\Omega) \to \cdots. \end{aligned} \tag{2.5}$$

Here S denotes the S–action and the maps B_{HH}, β and B_{HC} are induced by Connes' B–map B; see Section 3.1 for more details. The reduced versions of the Hochschild homology and the negative cyclic homology of Ω are denoted by $\widetilde{\mathrm{HH}}_*(\Omega)$ and $\widetilde{\mathrm{HC}}^-_*(\Omega)$, respectively; see Section 3.1.

Remark 2.6 Following Jones [26], we define the Hochschild homology and the cyclic homology for a chain algebra but not a cochain algebra. For a cochain algebra Ω, we define a chain algebra $\Omega_\sharp$ by $(\Omega_\sharp)_{-i} = \Omega^i$ for i. Thus, for a nonnegative cochain algebra $\mathcal{M}$, we have a nonpositive chain algebra $\mathcal{M}_\sharp$. The Hochschild homology and the negative cyclic homology of $\mathcal{M}$ are defined by $\mathrm{HH}_*(\mathcal{M}_\sharp)$ and $\mathrm{HC}^-_*(\mathcal{M}_\sharp)$, respectively. By abuse of notation, we may write $\mathrm{HH}_*(\mathcal{M})$ and $\mathrm{HC}^-_*(\mathcal{M})$ for $\mathrm{HH}(\mathcal{M}_\sharp)$ and $\mathrm{HC}^-_*(\mathcal{M}_\sharp)$, respectively.

The constructions of the string brackets above are generalized with *Gorenstein spaces*. An orientable manifold and the classifying space of a connected Lie group are typical examples of Gorenstein spaces; see Section 3 for the definition and fundamental properties of a Gorenstein space. For a Gorenstein space M of dimension d, we define a comultiplication $\bullet^\vee$ and a multiplication $\odot$ on the cohomology $H^*(LM;\mathbb{K})$, which are called the *dual loop product* and the *dual loop coproduct*, respectively; see Section 3. Therefore, by using the formulae (2.1) and (2.2) above, we have the string bracket and the dual string cobracket for a Gorenstein space M with $\bullet := (\bullet^\vee)^\vee$ and $\odot$, respectively; see Theorems 2.7 and 2.8 below for more details. We do not know the string brackets satisfy the Jacobi identity for general Gorenstein spaces. However, as seen in Theorem 2.8, these constructions indeed give generalizations of brackets (2.1) on manifolds and (2.2) on classifying spaces.

The following theorem asserts that the dual to the string bracket in the sense of Chas and Sullivan for a manifold is the dual loop product followed by the BV operator. Moreover, we see that the string bracket in Proposition 2.3 is described as the BV operator followed by the dual loop coproduct.

In the rest of this section, we further assume that $\mathbb{K}$ is a field of characteristic zero and a DGA Ω is locally finite; that is the homology $H_i(\Omega)$ is finite-dimensional for each $i \le 0$.

Theorem 2.7 *Let M be a simply connected Gorenstein space and Ω the chain algebra $A_{\mathrm{PL}}(M)_\sharp \otimes_{\mathbb{Q}} \mathbb{K}$. Suppose that the S–action on the reduced negative cyclic homology $\widetilde{\mathrm{HC}}^-_*(\Omega)$ is trivial. Then:*

(i) *There is a commutative diagram*

$$\begin{array}{ccc}
((\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\Delta)\oplus\mathbb{K}[u])^{\otimes 2} & \xrightarrow[\cong]{\Xi\otimes\Xi} & \mathrm{HC}^-_*(\Omega)^{\otimes 2} \\
{\scriptstyle \Delta\otimes\Delta}\big\downarrow & & {\scriptstyle \pi\otimes\pi}\big\downarrow \\
\mathrm{HH}_*(\Omega)^{\otimes 2} & & \mathrm{HH}_*(\Omega)^{\otimes 2} \\
{\scriptstyle \odot}\big\downarrow & & {\scriptstyle \odot}\big\downarrow \\
\mathrm{HH}_*(\Omega) & & \mathrm{HH}_*(\Omega) \\
{\scriptstyle \mathrm{Cokernel}}\big\downarrow & & {\scriptstyle \beta}\big\downarrow \\
(\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\Delta)\oplus\mathbb{K}[u] & \xrightarrow[\cong]{\Xi} & \mathrm{HC}^-_*(\Omega)
\end{array}$$

Here $\Delta = B_{\mathrm{HH}} \circ I \colon \mathrm{HH}_*(\Omega) \to \mathrm{HH}_*(\Omega)$ is the "BV operator", $\odot$ is the product described in Section 3.3, Cokernel *is defined by* (projection on the cokernel, 0)*, and the horizontal isomorphism Ξ is defined by the composite*

$$(\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\Delta)\oplus\mathbb{K}[u] \xrightarrow[\cong]{I} \widetilde{\mathrm{HC}}_*(\Omega)\oplus\mathbb{K}[u] \xrightarrow[\cong]{B_{\mathrm{HC}}} \widetilde{\mathrm{HC}}^-_*(\Omega)\oplus\mathbb{K}[u] \xrightarrow[\cong]{\mathrm{sp}} \mathrm{HC}^-_*(\Omega),$$

with the map sp *in Remark 3.1 below.*

(ii) *There is a commutative diagram*

$$\begin{array}{ccc}
((\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\Delta)\oplus\mathbb{K}[u])^{\otimes 2} & \xrightarrow[\cong]{\Xi\otimes\Xi} & \mathrm{HC}^-_*(\Omega)^{\otimes 2} \\
{\scriptstyle \mathrm{Cokernel}\otimes\mathrm{Cokernel}}\big\uparrow & & {\scriptstyle \beta\otimes\beta}\big\uparrow \\
\mathrm{HH}_*(\Omega)^{\otimes 2} & & \mathrm{HH}_*(\Omega)^{\otimes 2} \\
{\scriptstyle \bullet^\vee}\big\uparrow & & {\scriptstyle \bullet^\vee}\big\uparrow \\
\mathrm{HH}_*(\Omega) & & \mathrm{HH}_*(\Omega) \\
{\scriptstyle \Delta}\big\uparrow & & {\scriptstyle \pi}\big\uparrow \\
(\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\Delta)\oplus\mathbb{K}[u] & \xrightarrow[\cong]{\Xi} & \mathrm{HC}^-_*(\Omega).
\end{array}$$

Here $\Delta = B_{\mathrm{HH}}\circ I$ is the BV operator of the BV algebra $\mathrm{HH}_*(\Omega)$, and the horizontal isomorphism Ξ is the one defined in (i).

We call the right-hand vertical composites in Theorem 2.7(i) and (ii) the *dual string cobracket* and the *dual string bracket*, respectively.

Note that the condition on the S–action can be replaced with BV-exactness; see Definition 2.9 and Remark 2.14 for details. It is also worth mentioning that the composite $B_{\mathrm{HH}} \circ I$ is nothing but the cohomological Batalin–Vilkovisky (BV) operator Δ on the Hochschild homology of a DGA Ω if Ω is the polynomial de Rham algebra of a manifold or the classifying space of a connected Lie group. By abuse of terminology, we may call $B_{\mathrm{HH}} \circ I$ the BV operator in general.

As mentioned above, under the isomorphisms Θ_1 and Θ_2 due to Jones [26, Theorem A], the loop cohomology $H^*(LM)$ and the string cohomology $H^*_{S^1}(LM)$ are identified with the Hochschild homology and the negative cyclic homology of $A_{\mathrm{PL}}(M)$, respectively. Thus, we have:

Theorem 2.8 (i) *The dual string cobracket for BG described in Proposition 2.3 coincides with that in Theorem 2.7*(i) *up to isomorphisms Θ_1 and Θ_2.*

(ii) *Let M be a simply connected closed manifold of dimension d. Then the dual $[\ ,\ ]^\vee$ to the string bracket in the sense of Chas and Sullivan on M coincides with the dual string bracket in Theorem 2.7*(ii) *up to isomorphisms Θ_1 and Θ_2.*

In view of [27, Theorem 4.1], Theorem 2.7(i) and Theorem 2.8(i) allow us to compute the dual string cobracket on $H^*_{S^1}(LBG;\mathbb{K})$ explicitly if $\mathbb{K}$ is a field of characteristic zero; see Section 5. We observe that the classifying space BG is formal and then the S–action is trivial; see Corollary 2.13 below.

Moreover, by dualizing Theorem 2.7(ii) and Theorem 2.8(ii), we have Theorem 2.15, described below, for computing the string bracket of a manifold. It turns out that, in the rational case, the original string bracket can be formulated as the loop product followed by the BV operator on the loop homology. Before describing our main result concerning a manifold, we need a notion of the Batalin–Vilkovisky exactness.

Definition 2.9 A DGA Ω is *Batalin–Vilkovisky exact* (*BV-exact*) if $\operatorname{Im}\widetilde{B} = \operatorname{Ker}\widetilde{B}$, where the reduced operator $\widetilde{B}:\widetilde{\mathrm{HH}}_*(\Omega)\to\widetilde{\mathrm{HH}}_*(\Omega)$, is a restriction of Connes' B–operator $B:=\pi\circ\beta:\mathrm{HH}_*(\Omega)\to\mathrm{HH}_*(\Omega)$. We say that a simply connected space M is *BV-exact* if the polynomial de Rham algebra $A_{\mathrm{PL}}(M)$ of M is.

Remark 2.10 Let M be a simply connected closed manifold. The result [17, Proposition 2] implies that the dual of the BV operator $\Delta'\colon H_*(LM)\to H_{*+1}(LM)$ is identified with the operator B in Definition 2.9 under the isomorphism Θ_1 mentioned above. Then, it follows that a manifold M is BV-exact if and only if $\operatorname{Im}\widetilde{\Delta}' = \operatorname{Ker}\widetilde{\Delta}'$ for the reduced BV operator $\widetilde{\Delta}'\colon \widetilde{H}_*(LM)\to\widetilde{H}_{*+1}(LM)$.

Theorem 2.11 *A simply connected DGA Ω is BV-exact if and only if the reduced S–action on $\widetilde{\mathrm{HC}}^-_*(\Omega)$ is trivial.*

We refer the reader to Theorem 7.10 for a generalization of the result. An important example with trivial reduced S–action is given by the following proposition due to Vigué-Poirrier.

Proposition 2.12 [46, Proposition 5] *If a simply connected DGA Ω is formal, then the reduced S–action on $\widetilde{\mathrm{HC}}^-_*(\Omega)$ is trivial.*

By combining Theorem 2.11 and Proposition 2.12, we have:

Corollary 2.13 *If a simply connected DGA Ω is formal, then it is BV-exact. As a consequence, a simply connected manifold whose rational cohomology is generated by a single element and the classifying space of a compact connected Lie group are BV-exact.*

We also have a generalization of the corollary; see Theorem 2.21.

Remark 2.14 It follows from Theorem 2.11 that the condition on the S–action in Theorems 2.7 and 2.8 may be replaced with the BV-exactness. This implies that the string brackets are determined exactly with the loop (co)products and the BV operator on the Hochschild homology of a DGA Ω without dealing with the cyclic homology of Ω itself provided Ω is BV-exact. There is an isomorphism

$$\widetilde{\Delta}\colon \widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\widetilde{\Delta} = \widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Ker}\,\widetilde{\Delta} \xrightarrow{\cong} \mathrm{Im}\,\widetilde{\Delta} = \mathrm{Ker}\,\widetilde{\Delta}.$$

Dualizing Theorems 2.7(ii) and 2.8(ii), we have:

Theorem 2.15 *Let M be a simply connected closed manifold and $\mathbb{K}$ a field of characteristic zero. Assume further that M is BV-exact. Then there exists a commutative diagram*

$$\begin{array}{ccccc}
H_*^{S^1}(LM;\mathbb{K})^{\otimes 2} & \xrightarrow[\cong]{\Phi\otimes\Phi} & (\mathrm{Ker}\,\widetilde{\Delta}'\oplus\mathbb{K}[u])^{\otimes 2} & \xrightarrow{\mathrm{inc}\oplus 0} & H_*(LM;\mathbb{K})^{\otimes 2} \\
{\scriptstyle [\ ,\ ]}\downarrow{\scriptstyle \text{the string bracket}} & & & & {\scriptstyle \text{the loop product}}\downarrow{\scriptstyle \bullet} \\
H_*^{S^1}(LM;\mathbb{K}) & \xrightarrow[\cong]{\Phi} & (\mathrm{Ker}\,\widetilde{\Delta}'\oplus\mathbb{K}[u]) & \xleftarrow{\Delta'} & H_*(LM;\mathbb{K})
\end{array}$$

Here $\widetilde{\Delta}'\colon \widetilde{H}_*(LM;\mathbb{K}) \to \widetilde{H}_{*+1}(LM;\mathbb{K})$ denotes the reduced BV operator on the homology, and Φ is the dual of the composite of the isomorphisms Θ_2 and Ξ described in Theorem 2.7.

The shifted homology $\mathbb{H}_*(LM) := H_{*+d}(LM)$ for an orientable closed manifold M of dimension d admits a BV algebra structure with the loop product $\bullet$ and the BV operator Δ'; see [8]. It turns out that the homology is endowed with a Gerstenhaber algebra structure whose Lie bracket (loop bracket) $\{\ ,\ \}$ is given by

$$\{a,b\} = (-1)^{|a|}(\Delta'(a\bullet b) - (\Delta' a)\bullet b - (-1)^{|a|} a\bullet(\Delta' b)) \quad \text{for } a,b\in\mathbb{H}_*(LM).$$

If a and b are in the kernel of Δ', then $\{a,b\} = (-1)^{|a|}\Delta'(a\bullet b)$. Therefore, by virtue of Theorem 2.15, we have:

Corollary 2.16 *Under the same assumption and notations as in Theorem 2.15, the rational string bracket of the loop space LM is regarded as a restriction of the loop bracket up to the isomorphism Φ.*

Remark 2.17 (i) Proposition 2.12 implies that Theorems 2.7, 2.8 and 2.15 are applicable to a formal simply connected closed manifold.

(ii) It follows from [10, Theorem 8.5] that the loop homology of an orientable closed manifold admits a gravity algebra structure extending the Lie algebra structure on the string homology. Theorem 2.15 may enable us to determine a gravity algebra structure on the string homology of a BV-exact manifold M; see Example 5.6.

Remark 2.18 In general, the cyclic homology (additive K–theory [20]) for a DGA does *not* appear as the singular homology of any topological space because the homology is of $\mathbb{Z}$–grading. We stress that, however, the cyclic homology is used to investigate the string brackets for a manifold and the classifying space of a Lie group. In fact, the horizontal isomorphism Ξ in Theorem 2.7 factors through the cyclic homology of $A_{\mathrm{PL}}(M)_\sharp$.

Remark 2.19 By using the description of the dual loop product Dlp in [28, Theorem 2.3] and Theorem 2.7, we may relate the dual of the string bracket to the cup product on $H^*(LM;\mathbb{K})$ for a manifold M. In fact, the isomorphism Ξ in Theorem 2.7 is a morphism of algebras if the S–action is trivial; see [29, Theorem 2.5]. We observe that the additive K–theory $K^+(\Omega) := \mathrm{HC}_{*-1}(\Omega)$ for a chain algebra Ω is a graded algebra with the Loday–Quillen $*$–product in [31]; see [29, Proposition 1.1].

We relate the BV-exactness to a more familiar rational homotopy invariant.

Definition 2.20 A simply connected space X admits *positive weights* if the Sullivan minimal model $(\wedge V, d)$ for X has a direct sum decomposition $V = \bigoplus_{i>0} V_{(i)}$ satisfying $d(V_{(i)}) \subset (\wedge V)_{(i)}$. A nonzero element in $V_{(i)}$ is said to have *weight* i, and the weight on V is extended in a multiplicative way to $\wedge V$. For $x \in (\wedge V)_{(i)}$, its weight is written by $\mathrm{wt}(x) = i$.

Many spaces admit positive weights.

(1) The Sullivan minimal model $\mathcal{M}(X)$ of a formal space X is given by the bigraded model $(\Lambda V, d)$ of its cohomology algebra $H^*(X;\mathbb{Q})$ [24, Section 3], whose lower degree is given by $dV_p \subset (\Lambda V)_{p-1}$ for $p > 0$ and $dV_0 = 0$. Then the space X admits positive weights defined by $\mathrm{wt}(v) := |v| + p$ for $v \in V_p$.

(2) If a space X has a two stage Sullivan minimal model $\mathcal{M}(X) = (\Lambda(V_0 \oplus V_1), d)$ with $dV_0 = 0$ and $dV_1 \subset \Lambda V_0$, then X admits positive weights defined by $\mathrm{wt}(v) := |v| + i$ for $v \in V_i$. For example, a homogeneous space is such a space even if it is not formal; see also Section 6 for such a manifold.

(3) It is known that smooth complex algebraic varieties admit positive weights coming from its mixed Hodge structure [38]. In the paper, the Sullivan minimal models are discussed over $\mathbb{C}$, but admitting positive weights is reduced to that over $\mathbb{Q}$; see [5, Theorem 2.7].

Theorem 2.21 *A simply connected space X admitting positive weights is BV-exact.*

A simply connected space does not necessarily admit positive weights. In fact, there exist a four cell complex [37, Section 4] and elliptic spaces [1, Section 5] not admitting positive weights; see also Appendix A. It is worth mentioning that every finite group is realized as the group of self-homotopy equivalences of a rationalized elliptic space which does not admit positive weights; see [13].

3 Preliminaries

In this section, we recall the Hochschild homology and the cyclic homology together with relationships between them and the loop homology.

3.1 Hochschild and cyclic homology

In this section we recall the definitions of the Hochschild chain complex and the cyclic bar complex in [22] and [23]. Let Ω be a connected commutative DGA over a field $\mathbb{K}$ of arbitrary characteristic endowed with a differential d of degree -1. We call a DGA Ω *nonpositive* if $\Omega = \bigoplus_{i\leq 0}\Omega_i$. In what follows, it is assumed that a DGA is nonpositively graded algebra with the properties above unless otherwise stated. The degree of a homogeneous element x of a graded algebra is denoted by $|x|$.

First we recall the Hochschild chain complex together with the Connes' B–operator. Write $\bar{\Omega} = \Omega/\mathbb{K}$ and $C(\Omega) = \sum_{k=0}^{\infty}\Omega\otimes\bar{\Omega}^{\otimes k}$. We define $\mathbb{K}$–linear maps $b, B\colon C(\Omega)\to C(\Omega)$ of degrees -1 and 1 by

$$
\begin{aligned}
b(w_0,\dots,w_k) = {}& \\
&-\sum_{i=0}^{k}(-1)^{\epsilon_{i-1}}(w_0,\dots,w_{i-1},dw_i,w_{i+1},\dots,w_k) \\
&\qquad -\sum_{i=0}^{k-1}(-1)^{\epsilon_i}(w_0,\dots,w_{i-1},w_iw_{i+1},w_{i+2},\dots,w_k)+(-1)^{(|w_i|-1)\epsilon_{k-1}}(w_kw_0,\dots,w_{k-1}),
\end{aligned}
$$

$$
B(w_0,\dots,w_k) = \sum_{i=0}^{k}(-1)^{(\epsilon_{i-1}+1)(\epsilon_k-\epsilon_{i-1})}(1,w_i,\dots,w_k,w_0,\dots,w_{i-1}).
$$

Here $\deg(w_0,\dots,w_k) = |w_0|+\cdots+|w_k|+k$ for $(w_0,\dots,w_k)\in C(\Omega)$, $\epsilon_i = |w_0|+\cdots+|w_i|-i$ and $|u| = -2$. Note that the formulae $bB+Bb = 0$ and $b^2 = B^2 = 0$ hold. The chain complex $(C(\Omega), b)$ is called the *Hochschild chain complex*. The *Hochschild homology* $\mathrm{HH}_*(\Omega)$ and the *reduced Hochschild homology* $\widetilde{\mathrm{HH}}_*(\Omega)$ are the homologies of the complexes $(C(\Omega), b)$ and $(C(\Omega)/\mathbb{K}, b)$, respectively.

The *cyclic bar complex* is the complex $(C(\Omega)[u^{-1}], b+uB)$, where b and B are regarded as $\mathbb{K}[u^{-1}]$–linear maps extending b and B on $C(\Omega)$. Its homology is denoted by $\mathrm{HC}_*(\Omega)$ and called the *cyclic homology*. The *negative cyclic homology* $\mathrm{HC}^-_*(\Omega)$, the *reduced negative cyclic homology* $\widetilde{\mathrm{HC}}^-_*(\Omega)$ and the *periodic cyclic homology* $\mathrm{HC}^{\mathrm{per}}_*(\Omega)$ of a DGA Ω are defined as the homologies of the complexes $(C(\Omega)[\![u]\!], b+uB)$, $((C(\Omega)/\mathbb{K})[\![u]\!], b+uB)$ and $(C(\Omega)[\![u,u^{-1}], b+uB)$, respectively. Since a DGA in our case has negative degree, the power series algebra $C(\Omega)[\![u]\!]$ coincides with the polynomial algebra $C(\Omega)[u]$; similarly, $(C(\Omega)/\mathbb{K})[\![u]\!] = (C(\Omega)/\mathbb{K})[u]$ and $C(\Omega)[\![u,u^{-1}] = C(\Omega)[u,u^{-1}]$.

We recall Connes' exact sequences (2.5). The projection of the cyclic complex onto itself gives rise to the map S'. More precisely, we have $S'\bigl(\sum_{i\geq 0}x_iu^{-i}\bigr) = \sum_{i\geq 0}x_{i+1}u^{-i}$. Observe that the cyclic homology $\mathrm{HC}_*(\Omega)$ and the negative cyclic homology $\mathrm{HC}^-_*(\Omega)$ are $\mathbb{K}[u]$–modules, where $|u| = -2$. The multiplication $S = \times u\colon \mathrm{HC}^-_{n+2}(\Omega)\to\mathrm{HC}^-_n(\Omega)$ is called the *S–action* on the negative cyclic homology.

For the connecting homomorphism β in Connes' exact sequence (2.5), we see that $\beta([a_0]) = [B(a_0)]$. Moreover, we have $B_{\mathrm{HH}}\big(\big[\big(\sum_{i\geq 0} x_i u^{-i}\big)\big]\big) = [B(x_0)]$ and $B_{\mathrm{HC}}\big(\big[\big(\sum_{i\geq 0} x_i u^{-i}\big)\big]\big) = [B(x_0)]$.

Remark 3.1 Under the same notation as above, the unit and augmentation of Ω yield a split exact sequence of $\mathbb{K}[u]$–modules of the form $0 \to C(\mathbb{K})[u] \to C(\Omega)[u] \to (C(\Omega)/\mathbb{K})[u] \to 0$. Then the splitting map $s' : \widetilde{\mathrm{HC}}^-_*(\Omega) \to \mathrm{HC}^-_*(\Omega)$ gives rise to an isomorphism $\mathrm{sp}: \widetilde{\mathrm{HC}}^-_*(\Omega) \oplus \mathbb{K}[u] \xrightarrow{\cong} \mathrm{HC}^-_*(\Omega)$. We observe that $C(\mathbb{K})[u] = \mathbb{K}[u] = \mathrm{HC}^-_*(\mathbb{K})$.

3.2 Sullivan minimal models

Let $\mathcal{M}(Z) = (\wedge V, d)$ be the Sullivan minimal model of a nilpotent CW complex Z of finite type [15]. It is a free $\mathbb{Q}$–commutative DGA with a $\mathbb{Q}$–graded vector space $V = \bigoplus_{i\geq 1} V^i$, where $\dim V^i < \infty$, and a decomposable differential in the sense that $d(V^i) \subset (\wedge^+ V \cdot \wedge^+ V)^{i+1}$ and $d \circ d = 0$. Here $\wedge^+ V$ denotes the ideal of $\wedge V$ generated by elements of positive degree. Observe that $\mathcal{M}(Z)$ determines the rational homotopy type of Z; that is, the spatial realization $\|\mathcal{M}(Z)\|$ is homotopy equivalent to Z_0, the rationalization of Z. In particular, we see that

$$V^n \cong \mathrm{Hom}(\pi_n(Z), \mathbb{Q}) \quad \text{and} \quad H^*(\wedge V, d) \cong H^*(Z; \mathbb{Q}),$$

the second isomorphism being of graded algebras. A space X is said to be *formal* if there exists a quasi-isomorphism $\rho: \mathcal{M}(X) \to (H^*(X; \mathbb{Q}), 0)$ of DGA's. We refer the reader to [15] for more details.

In what follows, let $\mathbb{K}$ be a field of characteristic zero unless otherwise specifically mentioned. Let $\mathcal{M}$ be a free DGA $(\wedge V, d)$ with $V = \bigoplus_{i>1} V^i$ over $\mathbb{K}$. We denote by $(\mathcal{L}, \delta, s)$ the double complex defined in [7]. Namely, $\mathcal{L} = \wedge(V \oplus \overline{V})$, s is the unique derivation of degree -1 defined by $s(v) = \overline{v}$, $s(\overline{v}) = 0$ and δ is the unique derivation of degree $+1$ which satisfies $\delta|_V = d$ and $\delta s + s\delta = 0$. Here $\overline{V}$ is the suspension of V; that is, $\overline{V}^n = V^{n+1}$. By [7, Theorem 2.4(i)], we see that the map $\Theta: C(\mathcal{M}) \to \mathcal{L}$ defined by $\Theta(a_0, a_1, \dots, a_p) = (1/p!)a_0 s(a_1) \cdots s(a_p)$ is a chain map between the double complexes $(C(\mathcal{M}), b, B)$ and $(\mathcal{L}, \delta, s)$. Moreover, it follows from [7, Theorem 2.4(ii)–(iii)] that the map Θ induces isomorphisms $H(\Theta): \mathrm{HH}_*(\mathcal{M}) = H_*(C(\mathcal{M}), b) \cong H_*(\mathcal{L}, \delta)$ and $H(\Theta \otimes 1): \mathrm{HC}^-_*(\mathcal{M}) = H_*(C(\mathcal{M})[u], b + uB) \cong H_*(\mathcal{L}[u], \delta + u \cdot s)$.

Remark 3.2 As mentioned in Section 3.1, the connecting homomorphism β in Connes' exact sequence (2.5) is given by $\beta([a_0]) = [B(a_0)]$. It follows that $\beta([a_0]) = [s(a_0)]$ up to the isomorphism $H(\Theta)$; see again [7, Theorem 2.4(i)].

Let X be a simply connected space of finite type and LX the free loop space of X. Then the Sullivan minimal model of LX over $\mathbb{K}$, $\mathcal{M}(LX)$, is given by $(\mathcal{L}, \delta)$ (see [48]), and the Sullivan minimal model of the orbit space $ES^1 \times_{S^1} LX$, $\mathcal{M}(ES^1 \times_{S^1} LX)$, is given by $(\mathcal{E}, D) := (\mathcal{L}[u], \delta + u \cdot s)$ (see [47, Theorem A]). Thus we have isomorphisms $\mathrm{HH}_*(\mathcal{M}(X)) \cong H^{-*}(LX; \mathbb{K})$ and $\mathrm{HC}^-_*(\mathcal{M}(X)) \cong H^{-*}(ES^1 \times_{S^1} LX; \mathbb{K})$ by composing Θ_1 and Θ_2 with $H(\Theta)$ and $H(\Theta \otimes 1)$, respectively.

3.3 Loop product and coproduct on Gorenstein spaces

In order to introduce uniformly the loop product due to Chas and Sullivan and the dual loop coproduct due to Chataur and Menichi, we recall the notion of a Gorenstein DGA introduced by Félix, Halperin and Thomas in [14].

Let A be an augmented DGA over $\mathbb{K}$. We call A a *Gorenstein algebra* of dimension d if

$$\dim \operatorname{Ext}_A^*(\mathbb{K}, A) = \begin{cases} 0 & \text{if } * \neq d, \\ 1 & \text{if } * = d. \end{cases}$$

Here Ext is defined by using semifree resolutions; see [14, Appendix] for details. A path-connected space M is called a *Gorenstein space* of dimension d if the polynomial de Rham algebra $A_{\mathrm{PL}}(M)$ is a Gorenstein algebra of dimension d.

The result [14, Theorem 3.1] implies that a simply connected Poincaré duality space, for example a simply connected closed orientable manifold of dimension d, is a Gorenstein space of dimension d. It follows from [14, Proposition 3.2] that the classifying space BG of a connected compact Lie group G is also a Gorenstein space of dimension $-\dim G$. The following result due to Félix and Thomas is a key to defining the loop product and the loop coproduct on the loop homology of a Gorenstein space.

Theorem 3.3 [20, Theorem 12] *Let M be a simply connected Gorenstein space of dimension d whose cohomology with coefficients in $\mathbb{Q}$ is of finite type. Then, for any integer k,*

$$\operatorname{Ext}^k_{A_{\mathrm{PL}}(M^n)}(A_{\mathrm{PL}}(M), A_{\mathrm{PL}}(M^n)) \cong H^{k-(n-1)d}(M; \mathbb{Q}),$$

where $A_{\mathrm{PL}}(M)$ is considered an $A_{\mathrm{PL}}(M^n)$–module via the diagonal map $\mathrm{Diag}\colon M \to M^n$.

For a Gorenstein space M as in Theorem 3.3, let $D(\text{Mod–}A_{\mathrm{PL}}(M^n))$ be the derived category of right $A_{\mathrm{PL}}(M^n)$–modules. In the category, we define $\mathrm{Diag}^!$ by the map which corresponds to a generator of the one-dimensional vector space $H^0(M; \mathbb{Q})$ under the isomorphism $\operatorname{Ext}^{(n-1)d}_{A_{\mathrm{PL}}(M^n)}(A_{\mathrm{PL}}(M), A_{\mathrm{PL}}(M^n)) \cong H^0(M)$. Moreover, for a homotopy fiber square

$$\begin{array}{ccc} E' & \xrightarrow{q} & E \\ {\scriptstyle p'}\downarrow & & \downarrow{\scriptstyle p} \\ M & \xrightarrow{\mathrm{Diag}} & M^n \end{array}$$

there exists a unique map $q^!$ in $\operatorname{Ext}^{(n-1)d}_{A_{\mathrm{PL}}(E)}(A_{\mathrm{PL}}(E'), A_{\mathrm{PL}}(E))$ which fits into the commutative diagram in $D(\text{Mod–}A_{\mathrm{PL}}(M^n))$

$$\begin{array}{ccc} A^*_{\mathrm{PL}}(E') & \xrightarrow{q^!} & A^{*+(n-1)d}_{\mathrm{PL}}(E) \\ {\scriptstyle (p')^*}\uparrow & & \uparrow{\scriptstyle p^*} \\ A^*_{\mathrm{PL}}(M) & \xrightarrow{\mathrm{Diag}^!} & A^{*+(n-1)d}_{\mathrm{PL}}(M^n) \end{array}$$

The result follows from the same proof as that of [18, Theorems 1 and 2].

We recall the definition of the loop product on a simply connected Gorenstein space M. Consider the diagram

$$\begin{array}{ccccc} LM & \xleftarrow{\text{Comp}} & LM \times_M LM & \xrightarrow{q} & LM \times LM \\ & & \downarrow & & \downarrow{\scriptstyle (\mathrm{ev}_0,\mathrm{ev}_0)} \\ & & M & \xrightarrow{\text{Diag}} & M \times M \end{array} \tag{3.4}$$

where the right-hand square is the pull-back of the evaluation map $(\mathrm{ev}_0, \mathrm{ev}_0)$ defined by $\mathrm{ev}_0(\gamma) = \gamma(0)$ along the diagonal map Diag, and Comp denotes the concatenation of loops. By definition, the composite

$$q^! \circ (\mathrm{Comp})^* \colon A_{\mathrm{PL}}(LM) \to A_{\mathrm{PL}}(LM \times_M LM) \to A_{\mathrm{PL}}(LM \times LM)$$

induces Dlp, the dual to the loop product on $H^*(LM;\mathbb{Q})$; see [18, Introduction].

We define a product $\bullet$ on $\mathbb{H}_*(LM) := H_{*+d}(LM)$, which is called the *loop product*, by

$$a \bullet b = (-1)^{d(|a|+d)}((\mathrm{Dlp})^\vee)(a \otimes b)$$

for a and $b \in \mathbb{H}_*(LM)$; see [12, Proposition 4] and [44, Definition 3.2].

In order to recall the loop coproduct for a Gorenstein space M, we consider the commutative diagram

$$\begin{array}{ccccc} LM \times LM & \xleftarrow{q} & LM \times_M LM & \xrightarrow{\text{Comp}} & LM \\ & & \downarrow & & \downarrow{\scriptstyle l} \\ & & M & \xrightarrow{\text{Diag}} & M \times M \end{array}$$

where $l\colon LM \to M \times M$ is a map defined by $l(\gamma) = \big(\gamma(0), \gamma\big(\tfrac{1}{2}\big)\big)$. By definition, the composite

$$\mathrm{Comp}^! \circ q^* \colon A_{\mathrm{PL}}(LM \times LM) \to A_{\mathrm{PL}}(LM \times_M LM) \to A_{\mathrm{PL}}(LM)$$

induces the dual to the loop coproduct Dlcop on $H^*(LM)$. We define a product $\odot$ on the shifted cohomology $\mathbb{H}^*(LM) = H^{*-d}(LM)$, called the *dual loop coproduct*, by

$$a \odot b = (-1)^{d(d-|a|)}\,\mathrm{Dlcop}(a \otimes b) \quad \text{for } a \otimes b \in H^*(LM) \otimes H^*(LM).$$

Remark 3.5 The product $\bullet$ on $\mathbb{H}_*(LM)$ is associative and graded commutative if M is a simply connected Poincaré duality space; see [28, Proposition 2.7]. So is the product $\odot$ on $\mathbb{H}^*(LM)$ if M is the classifying space BG of a connected Lie group G; see [9] and [27, Theorem B.1]. Moreover, so are both of $\bullet$ and $\odot$ if M is a Gorenstein space with $\dim\big(\bigoplus_n \pi_n(M) \otimes \mathbb{Q}\big) < \infty$; see [39, Theorem 1.1] and [50, Theorem 1.5].

Remark 3.6 By the same fashion as above, a Gorenstein space is defined on an arbitrary field $\mathbb{K}$. Then Theorem 3.3 remains true after replacing $A_{\mathrm{PL}}(X)$ with the singular cochain algebra of X with coefficients in $\mathbb{K}$. That is the original assertion in [18]. Moreover, the constructions of the loop product and the loop coproduct are applicable to the Gorenstein space M; that is, those products are defined on the singular cohomology of LM with coefficient in $\mathbb{K}$; see [18]. However, we only use such an algebra defined on a field of characteristic zero for our purpose.

We conclude this section with the definition of a BV algebra. In the next section, the notion plays an important role in defining the dual string cobracket of the classifying space of a Lie group.

Definition 3.7 A graded algebra $(\mathbb{H}^*, \odot)$ equipped with an operator Δ on $\mathbb{H}^*$ of degree -1 is a *BV algebra* if $\Delta \circ \Delta = 0$ and the *Batalin–Vilkovisky identity* holds; that is, for any elements a, b and c in $\mathbb{H}^*$,

$$\Delta(a \odot b \odot c) = \Delta(a \odot b) \odot c + (-1)^{\|a\|} a \odot \Delta(b \odot c) + (-1)^{\|b\|\|a\|+\|b\|} b \odot \Delta(a \odot c) - \Delta(a) \odot b \odot c - (-1)^{\|a\|} a \odot \Delta(b) \odot c - (-1)^{\|a\|+\|b\|} a \odot b \odot \Delta(c),$$

where $\|\alpha\|$ stands for the degree of an element α in $\mathbb{H}^*$.

4 Proofs of assertions

The strategy of the proof of Proposition 2.3 is exactly that of [8, Theorem 6.2]. In order to make the sign computation more clear in our setting, we give the proof.

Proof of Proposition 2.3 It is readily seen that the dual string cobracket satisfies skew-symmetry since the multiplication m is commutative. Indeed, we have

$$[y, x] = (-1)^{\|y\|} \beta(\pi(y) \odot \pi(x)) = (-1)^{\|x\|(\|y\|+1)+1} \beta(\pi(x) \odot \pi(y)) = -(-1)^{\|x\|\|y\|}[x, y].$$

Let $\Delta\colon \mathbb{H}^*(LBG) \to \mathbb{H}^{*-1}(LBG)$ be the cohomological BV operator stated in [27, Appendix E]. Observe that Δ coincides with the composite $\pi\beta$. It follows from [27, Corollary C.3] that the triple $(\mathbb{H}^*(LBG), \odot, \Delta)$ is a BV algebra; hence the bracket

$$\{a, b\} := (-1)^{\|a\|} \Delta(a \odot b) - (-1)^{\|a\|} \Delta(a) \odot b - a \odot \Delta(b)$$

satisfies the Poisson identity

$$\{a, b \odot c\} = \{a, b\} \odot c + (-1)^{(\|a\|-1)\|b\|} b \odot \{a, c\}. \tag{4.1}$$

In the case where $a = \pi(x)$, $b = \pi(y)$ and $c = \pi(z)$, applying β to (4.1) we see that $\beta\{\pi(x), \pi(y) \odot \pi(z)\}$ coincides with

$$\beta(\{\pi(x), \pi(y)\} \odot \pi(z) + (-1)^{(\|\pi(x)\|-1)\|\pi(y)\|} \pi(y) \odot \{\pi(x), \pi(z)\}).$$

Since $\Delta\pi = 0$ and $\beta\Delta = 0$, it follows that

$$\{\pi(x), \pi(y)\} = (-1)^{\|x\|-1} \Delta(\pi(x) \odot \pi(y)) = -\pi[x, y],$$

$$\beta\{\pi(x), \pi(y) \odot \pi(z)\} = -(-1)^{\|y\|} \beta(\pi(x) \odot \pi[y, z]) = -(-1)^{\|x\|+\|y\|}[x, [y, z]].$$

Therefore, by combining the formulae, we see that

$$\begin{aligned} -(-1)^{\|x\|+\|y\|}[x, [y, z]] &= -\beta(\pi[x, y] \odot \pi(z)) - (-1)^{(\|\pi(x)\|-1)\|\pi(y)\|} \beta(\pi(y) \odot \pi[x, z]) \\ &= -(-1)^{\|x\|+\|y\|}[[x, y], z] - (-1)^{(\|\pi(x)\|-1)\|\pi(y)\|+\|y\|}[y, [x, z]] \\ &= (-1)^{\|x\|+\|y\|+(\|x\|+\|y\|)\|z\|}[z, [x, y]] + (-1)^{\|x\|(\|y\|+\|z\|)+\|x\|+\|y\|}[y, [z, x]]. \end{aligned}$$

Multiplying the both sides of the above equality by $(-1)^{\|x\|+\|y\|+1+\|x\|\|z\|}$, we have

$$(-1)^{\|x\|\|z\|}[x,[y,z]] = -(-1)^{\|y\|\|z\|}[z,[x,y]] - (-1)^{\|x\|\|y\|}[y,[z,x]],$$

which is indeed the Jacobi identity. This completes the proof. □

Proof of Theorem 2.7 We will first prove (i). Recall the homomorphisms $B_{\mathrm{HC}}\colon \mathrm{HC}_{n-1}(\Omega) \to \mathrm{HC}^-_n(\Omega)$ and $B_{\mathrm{HH}}\colon \mathrm{HC}_{n-1}(\Omega) \to \mathrm{HH}_n(\Omega)$ in Connes' exact sequence in Section 3, which are defined by $B_{\mathrm{HC}}\big(\sum_{i\geq 0} x_i u^{-i}\big) = Bx_0$ and $B_{\mathrm{HH}}\big(\sum_{i\geq 0} x_i u^{-i}\big) = Bx_0$. The result [29, Theorem 2.5(i)] implies that B is an isomorphism. By assumption, the S–action is trivial. Then, by [29, Theorems 2.5(ii)–(iii)], the map I is an isomorphism. By a direct calculation, we see that $\pi \circ \Xi = B_{\mathrm{HH}} \circ I$ and $\Xi \circ \mathrm{Cokernel} = \beta$. The same consideration as above enables us to obtain the result (ii). □

Proof of Theorem 2.8 The assertions (i) and (ii) follow from [26, Theorem A]; see also [10, Theorem 8.3]. In fact, the dual of the homology Gysin exact sequence for the fibration $S^1 \to ES^1 \times LM \to ES^1 \times_{S^1} LM$ is identified with the Connes exact sequence under isomorphisms Θ_1 and Θ_2 mentioned in the sentence before Theorem 2.8; see [6, Theorem B] and Appendix B for a description of the Gysin sequence in terms of rational models. With those isomorphisms, we compare the dual to string bracket for a manifold and the dual string cobracket for BG with the dual string bracket and the dual string cobracket in Theorem 2.7, respectively.

To this end, we recall that a simply connected closed manifold M of dimension d is a Gorenstein space of dimension d. Moreover, the classifying space BG of a connected compact Lie group G is a Gorenstein space of dimension $d = -\dim G$; see [14]. Thus the result [18, Theorem A] and observations in [18, pages 419–420] yield that the dual loop product $\bullet^\vee$ for the manifold M and the dual loop coproduct $\odot$ for the classifying space BG are nothing but the dual to the loop product and the dual to the loop coproduct, respectively. It turns out that the bracket on $H_*(LM;\mathbb{K})$ for the manifold M and the dual string cobracket on $H^*(LBG;\mathbb{K})$ coincide with the original string brackets (2.1) and (2.2), respectively. Thus, we have the results. □

Proof of Theorem 2.15 Let Ω be the DGA $\Omega = A_{\mathrm{PL}}(M)_\sharp \otimes_{\mathbb{Q}} \mathbb{K}$ for M. We observe that the dual of the BV operator $\Delta'\colon H_*(LM;\mathbb{K}) \to H_*(LM;\mathbb{K})$ on the homology is regarded as the BV operator $\Delta\colon \mathrm{HH}_*(\Omega) \to \mathrm{HH}_*(\Omega)$ in Theorem 2.7; see Remark 2.10.

Let $\widetilde{\mathrm{HH}}_*$ denote the reduced Hochschild homology $\widetilde{\mathrm{HH}}_*(\Omega)$. Dualizing the reduced BV operator $\widetilde{\Delta}'\colon \widetilde{H}_*(LM) \to \widetilde{H}_*(LM)$, we have an exact sequence $(*)$:

$$\widetilde{\mathrm{HH}}_* \xrightarrow{\widetilde{\Delta}'^\vee} \widetilde{\mathrm{HH}}_* \xrightarrow{\pi} \widetilde{\mathrm{HH}}_*/\mathrm{Im}\,\widetilde{\Delta}'^\vee \to 0.$$

Observe that $\widetilde{\Delta}'^\vee = B_{\mathrm{HH}} \circ I = \Delta$. By considering the dual exact sequence of $(*)$, we see that π gives rise to the isomorphism $\pi^\vee\colon \mathrm{Ker}\,\widetilde{\Delta}' = \mathrm{Ker}(\Delta^\vee) \xrightarrow{\cong} (\widetilde{\mathrm{HH}}_*/\mathrm{Im}\,\Delta)^\vee$. Theorem 2.8(ii) yields the result. □

In the rest of the section, we prove Theorems 2.11 and 2.21. First we prove the "if" part of Theorem 2.11.

Proof of the "if" part of Theorem 2.11 Let Ω be a simply connected DGA such that the reduced S–action on $\widetilde{\mathrm{HC}}^-_*(\Omega)$ is trivial. Consider the reduced version of Connes' exact sequence

$$\cdots \to \widetilde{\mathrm{HC}}^-_{n+2}(\Omega) \xrightarrow{S=0} \widetilde{\mathrm{HC}}^-_n(\Omega) \xrightarrow{\pi} \widetilde{\mathrm{HH}}_n(\Omega) \xrightarrow{\beta} \widetilde{\mathrm{HC}}^-_{n+1}(\Omega) \to \cdots,$$

which splits into a short exact sequence

$$0 \to \widetilde{\mathrm{HC}}^-_n(\Omega) \xrightarrow{\pi} \widetilde{\mathrm{HH}}_n(\Omega) \xrightarrow{\beta} \widetilde{\mathrm{HC}}^-_{n+1}(\Omega) \to 0.$$

By definition, there is a decomposition $\widetilde{B} = \pi \circ \beta \colon \widetilde{\mathrm{HH}}_*(\Omega) \to \widetilde{\mathrm{HH}}_*(\Omega)$ and hence the above short exact sequence implies $\operatorname{Ker} \widetilde{B} = \operatorname{Ker} \beta = \operatorname{Im} \pi = \operatorname{Im} \widetilde{B}$. □

In order to prove the "only if" part of Theorem 2.11, we recall the notion of the proper exactness of a sequence of complexes defined in [42].

Definition 4.2 Let $M_1 \to M_2 \to M_3$ be a sequence of complexes and chain maps (of arbitrary degrees).

(i) The sequence is H–exact at M_2 if the sequence of cohomology $H(M_1) \to H(M_2) \to H(M_3)$ is exact.

(ii) The sequence is Z–exact at M_2 if the sequence of modules of cycles $Z(M_1) \to Z(M_2) \to Z(M_3)$ is exact.

(iii) The sequence is *proper exact* (at M_2) if the sequence is exact (as a sequence of underlying graded modules), H–exact and Z–exact [42].

(iv) The sequence is *weakly proper exact* at M_2 if the sequence is exact and H–exact.

The following lemma is useful to prove the proper exactness from the weak proper exactness of a given sequence.

Lemma 4.3 *Let $M_0 \xrightarrow{f_0} M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3 \xrightarrow{f_3} M_4$ be a sequence of complexes which is proper exact at M_2 and weakly proper exact at M_1 and M_3. Then it is proper exact also at M_3.*

Proof For simplicity, we assume that the degrees of the chain maps are zero. We show that $\operatorname{Ker} Z(f_3) \subset \operatorname{Im} Z(f_2)$. For any x_3 in $\operatorname{Ker} Z(f_3)$, there exists an element $y_2 \in M_2$ such that $f_2(y_2) = x_3$ by the exactness at M_3. By the proper exactness at M_2, we see that $dy_2 = f_1(y_1)$ for some $y_1 \in Z(M_1)$. Since $H(f_1)[y_1] = [dy_2] = 0$, it follows from the H–exactness at M_1 that $y_1 - f_0 y_0 = dz$ for some $[y_0] \in H(M_0)$ and $z \in M_1$. It is readily seen that $f_2(y_2 - f_1 z) = f_2 y_2 = x_3$ and $d(y_2 - f_1 z) = 0$. We have the result. □

It is proved that the weak proper exactness for a long sequence yields the proper exactness.

Proposition 4.4 *A weakly proper exact sequence* $0 \to M_0 \to M_1 \to M_2 \to \cdots$ *starting from* 0 *is always proper exact.*

Proof Since the sequence $0 \to 0 \to 0 \to M_0 \to M_1$ is weakly proper exact at M_0 and proper exact at 0, it follows from Lemma 4.3 that the sequence is proper exact at M_0. Similarly, the sequence $0 \to 0 \to M_0 \to M_1 \to M_2$ gives proper exactness at M_1. By repeating this argument, we can prove the proper exactness at M_n for all n. □

Remark 4.5 By the same argument as in the proof above, we can also prove the dual of Proposition 4.4, which asserts that a weakly proper exact sequence ending with 0 is always proper exact.

Next we give a key lemma for proving Theorem 2.11.

Lemma 4.6 *Let* $M_0 \to M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} M_3$ *be a proper exact sequence. Then one has*

$$\operatorname{Im} d \cap \operatorname{Ker} f_2 = d(\operatorname{Ker} f_2).$$

Proof The Z–exactness at M_2 and the H–exactness at M_2 give the result. The details are left to the reader. □

Note that the consequence in Lemma 4.6 is equivalent to the exactness of the sequence of modules of coboundaries.

Now we begin the proof of the "only if" part of Theorem 2.11. Let $(\wedge V, d)$ be a Sullivan model of the DGA Ω with $V = V^{\geq 2}$. Define $(\widetilde{\mathscr{L}}, \delta) = (\wedge^+(V \oplus \overline{V}), \delta)$ and $(\widetilde{\mathscr{E}}, D) = (\wedge u \otimes \widetilde{\mathscr{L}}, D)$; see Section 3. Then $(\widetilde{\mathscr{L}}, \delta)$ and $(\widetilde{\mathscr{E}}, D)$ are chain models for the reduced Hochschild homology and the reduced negative cyclic homology of Ω, respectively. Let $\widetilde{s} \colon \widetilde{\mathscr{L}} \to \widetilde{\mathscr{L}}$ be the derivation defined by $\widetilde{s}(v) = \overline{v}$ and $\widetilde{s}(\overline{v}) = 0$ for $v \in V$. Now we have a direct sum decomposition $(\widetilde{\mathscr{L}}, \delta) = \bigoplus_n (\widetilde{\mathscr{L}}^{(n)}, \delta)$ of complexes, where $\widetilde{\mathscr{L}}^{(n)} = \widetilde{\mathscr{L}} \cap (\wedge V \otimes \wedge^n \overline{V})$. Then $\widetilde{s}$ decomposes into a sequence $0 \to \widetilde{\mathscr{L}}^{(0)} \to \widetilde{\mathscr{L}}^{(1)} \to \widetilde{\mathscr{L}}^{(2)} \to \cdots$ of complexes.

Lemma 4.7 *The sequence* $0 \to \widetilde{\mathscr{L}}^{(0)} \to \widetilde{\mathscr{L}}^{(1)} \to \widetilde{\mathscr{L}}^{(2)} \to \cdots$ *is exact; that is,* $\operatorname{Ker} \widetilde{s} = \operatorname{Im} \widetilde{s}$ *in* $\widetilde{\mathscr{L}}$.

Proof Take a basis $\{v_\lambda\}_\lambda$ of V. Then we have $(\mathscr{L}, s) \cong \bigotimes_\lambda (\wedge(v_\lambda, \overline{v}_\lambda), s)$ and hence $H(\mathscr{L}, s) \cong \mathbb{Q}$, which is equivalent to $H(\widetilde{\mathscr{L}}, \widetilde{s}) \cong 0$. □

Remark 4.8 The operator $\widetilde{B} \colon \widetilde{\mathrm{HH}}_*(\Omega) \to \widetilde{\mathrm{HH}}_*(\Omega)$ is nothing but the homomorphism $H(\widetilde{s})$ up to the isomorphism $H(\Theta)$. This follows from the definition of the map B in Section 3 and Remark 3.2.

Now we recall a result of Vigué-Poirrier which gives a description of the cyclic homology in terms of $\widetilde{\mathscr{L}}$. Here we give a proof for the convenience of the reader.

Lemma 4.9 [45, Lemma 2] *The canonical inclusion* $\Phi\colon (\operatorname{Ker}\widetilde{s}, d) \to (\widetilde{\mathscr{E}}, D)$ *is a quasi-isomorphism.*

Proof Define bounded double complexes $\{K^{p,q}\}$ and $\{\widetilde{\mathscr{E}}^{p,q}\}$ by $K^{p,0} = (\operatorname{Ker}\widetilde{s})^p$ and $K^{p,q} = 0$ for $q \neq 0$, and $\widetilde{\mathscr{E}}^{p,q} = \wedge^q u \otimes \widetilde{\mathscr{L}}^{p-q}$. Then their total chain complexes are $(\operatorname{Ker}\widetilde{s}, \delta)$ and $(\widetilde{\mathscr{E}}, D)$, respectively, and the inclusion Φ gives rise to a morphism of double complexes. Now consider the filtration with respect to p. By Lemma 4.7, we have $E_1^{p,0}K = E_1^{p,0}\widetilde{\mathscr{E}} = (\operatorname{Ker}\widetilde{s})^p$ and $E_1^{p,q}K = E_1^{p,q}\widetilde{\mathscr{E}} = 0$ for $q \neq 0$. Hence $E_1\Phi$ is an isomorphism and so is $H\Phi$ by the convergence of the spectral sequences. □

Now we describe the S–action $S = u \times (-)\colon H(\widetilde{\mathscr{E}}) \to H(\widetilde{\mathscr{E}})$ in terms of $\operatorname{Ker}\widetilde{s}$. By Lemma 4.7, we have an exact sequence $0 \to \operatorname{Ker}\widetilde{s} \to \widetilde{\mathscr{L}} \xrightarrow{\widetilde{s}} \operatorname{Ker}\widetilde{s} \to 0$, and its connecting homomorphism $c\colon H(\operatorname{Ker}\widetilde{s}) \to H(\operatorname{Ker}\widetilde{s})$ is given by $c([\widetilde{s}\alpha]) = [\delta\alpha]$. Note that any element in $H(\operatorname{Ker}\widetilde{s})$ can be written as $[\widetilde{s}\alpha]$ for some $\alpha \in \widetilde{\mathscr{L}}$ with $\delta\widetilde{s}\alpha = 0$, since $\operatorname{Ker}\widetilde{s} = \operatorname{Im}\widetilde{s}$ by Lemma 4.7. By a straightforward computation, we have:

Lemma 4.10 *The map* c *coincides with* S *through* $H\Phi$ *up to sign, ie* $S \circ H\Phi = -H\Phi \circ c$.

We are ready to prove the "only if" part of Theorem 2.11.

Proof of the "only if" part of Theorem 2.11 By Lemmas 4.9 and 4.10, in order to prove the assertion it suffices to show that the connecting homomorphism c is trivial. To this end, we show that $[\delta\alpha] = 0$ in $H(\operatorname{Ker}\widetilde{s})$ for any $\alpha \in \widetilde{\mathscr{L}}$ with $\delta\widetilde{s}\alpha = 0$; see the argument before Lemma 4.10. Remark 4.8 yields that the BV-exactness of the DGA Ω is equivalent to the condition that the sequence $(*)$,

$$0 \to \widetilde{\mathscr{L}}^{(0)} \to \widetilde{\mathscr{L}}^{(1)} \to \widetilde{\mathscr{L}}^{(2)} \to \cdots,$$

is weakly proper exact. Thus, by Proposition 4.4, we see that the sequence $(*)$ is proper exact. Moreover, Lemma 4.6 implies that $\operatorname{Ker}\widetilde{s} \cap \operatorname{Im}\delta = \delta(\operatorname{Ker}\widetilde{s})$. Therefore, it follows that $\delta\alpha \in \operatorname{Ker}\widetilde{s} \cap \operatorname{Im}\delta = \delta(\operatorname{Ker}\widetilde{s})$ for any $\alpha \in \widetilde{\mathscr{L}}$ with $\delta\widetilde{s}\alpha = 0$. We have the result. □

We conclude this section proving Theorem 2.21. The proof is given by slightly modifying the proof of [46, Proposition 5].

Proof of Theorem 2.21 Recall that $(\widetilde{\mathscr{L}}, \delta) = (\wedge^+(V \oplus \overline{V}), \delta)$ is a model of the Hochschild complex. For a derivation $\theta\colon \wedge V \to \wedge V$ of degree 0 with $\theta d = d\theta$ and $\theta(V) \subset \wedge^+ V$, define derivations $L_\theta, e_\theta\colon \widetilde{\mathscr{L}} \to \widetilde{\mathscr{L}}$ by $L_\theta(v) = \theta v$, $L_\theta(\overline{v}) = \widetilde{s}\theta v$, $e_\theta(v) = 0$ and $e_\theta(\overline{v}) = \theta v$. Then, as derivations on $\widetilde{\mathscr{L}}$, we have $[L_\theta, \widetilde{s}] = [L_\theta, \delta] = [e_\theta, \delta] = 0$ and $[e_\theta, \widetilde{s}] = L_\theta$. Hence L_θ induces $H(L_\theta)\colon H(\operatorname{Ker}\widetilde{s}) \to H(\operatorname{Ker}\widetilde{s})$ and it follows that $H(L_\theta) \circ c = 0\colon H(\operatorname{Ker}\widetilde{s}) \to H(\operatorname{Ker}\widetilde{s})$ by a straightforward computation from the above equations.

Now we let θ be the derivation defined by $\theta(x) = \mathrm{wt}(x)x$ for weight-homogeneous elements $x \in \wedge V$. Then for any weight-homogeneous element $\alpha \in H(\operatorname{Ker}\widetilde{s})$, we have $0 = H(L_\theta) \circ c(\alpha) = \mathrm{wt}(\alpha)c(\alpha)$, where the weight on $\widetilde{\mathscr{L}}$ is defined as an extension of that on $\wedge V$ with $\mathrm{wt}(\overline{v}) = \mathrm{wt}(v)$ for $v \in V$. By the positivity of the weight, we have $c(\alpha) = 0$ and hence $c = 0$. Therefore, Lemmas 4.9 and 4.10 imply the triviality of the reduced S–action, which is equivalent to the BV-exactness by Theorem 2.11. □

5 The string brackets for formal spaces

In this section, we consider string brackets for formal spaces as an application of Theorem 2.7.

5.1 Dual string cobrackets for classifying spaces

We begin by considering the string bracket for the classifying space of a connected Lie group of rank one.

Example 5.1 The result [27, Theorem 4.1] enables us to compute the dual loop coproduct on the loop cohomology $\mathbb{H}^*(LBG;\mathbb{Q})$ for every compact connected Lie group G. Thus, in particular, by Theorem 2.7, we determine explicitly the Lie algebra structure of $\mathcal{H}^*(LB\mathrm{SU}(2)) := H^{*+3+1}_{S^1}(LB\mathrm{SU}(2);\mathbb{Q})$ endowed with the dual string cobracket. In fact, we see that

$$\begin{aligned}\mathcal{H}^* := \mathcal{H}^*(LB\mathrm{SU}(2)) &\cong (\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\Delta)_{-*-3-2} \oplus (\mathbb{Q}[u])_{-*-3-1} \\ &\cong \mathbb{Q}\{x, x^2, \ldots, x^n, \ldots\} \oplus \mathbb{Q}\{1, u, u^2, \ldots, u^k, \ldots\}\end{aligned}$$

as vector spaces, where Ω denotes the Sullivan minimal model for SU(2). Observe that $|x^n| = 4n-5$ and $|1| = -4$ for x^n and $1 \in \mathcal{H}^*(LB\mathrm{SU}(2))$. The formula in [27, Theorem 4.1] for the loop product $\odot$ yields that $\Delta(x^n)\odot 1 = nx^{n-1}$, $\Delta(x^n)\odot\Delta(x^m) = \pm nm\Delta(x)x^{n+m-2}$ and $\Delta(1) = 0$ in $\mathbb{H}^*(LB\mathrm{SU}(2))$. Therefore, we see that $[1,1] = 0$, $[x^n, x^m] = 0$ for $m, n \geq 1$, $[u^l, \alpha] = 0$ for every $\alpha \in \mathcal{H}^*$, $l \geq 1$ and $[x^n, 1] = -nx^{n-1}$ for $n \geq 1$.

Next we consider the dual string cobracket for the classifying space of G with arbitrary rank.

Proposition 5.2 *For each n, the n–fold dual string cobracket $[\mathcal{H}, [\mathcal{H}, \ldots, [\mathcal{H}, \mathcal{H}]\cdots]]$ is nontrivial on $\mathcal{H}^* := H^{*+\dim G+1}_{S^1}(LBG;\mathbb{K})$.*

Proof For the case rank $G = 1$, Example 5.1 implies the result. Assume that $N := \mathrm{rank}\, G \geq 2$. Recall the result [27, Theorem 4.3], which asserts that the loop cohomology $\mathbb{H}^*(LBG) := H^{*+\dim G}(LBG)$ is isomorphic to the tensor product of algebra

$$H^*(BG)\otimes H_{-*}(G) = \mathbb{K}[y_1, \ldots, y_N]\otimes\wedge(x_1^\vee, \ldots, x_N^\vee)$$

equipped with the BV operator Δ given by $\Delta(x_i^\vee x_j^\vee) = \Delta(y_i y_j) = \Delta(x_i^\vee) = \Delta(y_i) = 0$ and

$$\Delta(y_i x_j^\vee) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Thus, an induction argument with the BV identity enables us to deduce that

$$\Delta(y_1^{k_1}\cdots y_N^{k_N} x_{i_1}^\vee \cdots x_{i_s}^\vee) = \sum_{1\leq j\leq s} (-1)^{d_j} k_{i_j}\, y_1^{k_1}\cdots y_{i_j}^{k_{i_j}-1}\cdots y_N^{k_N} x_{i_1}^\vee\cdots\widehat{x_{i_j}^\vee}\cdots x_{i_s}^\vee,$$

where $\widehat{\cdot}$ denotes omission and $d_j = |x_{i_1}^\vee| + \cdots + |x_{i_{j-1}}^\vee|$. Therefore, it follows that

$$\Delta(y_2 x_2^\vee x_1^\vee)\odot\Delta(y_1^l x_1^\vee x_2^\vee) = x_1^\vee \odot l y_1^{l-1} x_2^\vee = l y_1^{l-1} x_1^\vee x_2^\vee.$$

Moreover, we see that $\Delta(y_1^{l-1}x_1^\vee x_2^\vee) \neq 0$ for $l \geq 2$. Then the element $y^{l-1}x_1^\vee x_2^\vee$ is not in $\mathrm{Im}\,\Delta$. Observe that $\Delta^2 = 0$. We consider an n–fold bracket of the form

$$\alpha := [y_2 x_2^\vee x_1^\vee, [y_2 x_2^\vee x_1^\vee, \dots, [y_2 x_2^\vee x_1^\vee, y_1^l x_1^\vee x_2^\vee] \cdots]] \quad \text{for } l > n.$$

It turns out that

$$\alpha = l(l-1)\cdots(l-(n-1))\, y_1^{l-n} x_1^\vee x_2^\vee \neq 0$$

in the codomain $(\widetilde{\mathrm{HH}}_*(\Omega)/\mathrm{Im}\,\widetilde{\Delta}) \oplus \mathbb{K}[u]$ of the dual string cobracket. Theorem 2.7(i) allows us to obtain the result. □

5.2 String brackets for manifolds

As an application of Theorem 2.8 (or Theorem 2.15), we give another proof of the first half of the result [2, Theorem 3.4] due to Basu and [19, Example 5.2] due to Félix, Thomas and Vigué-Poirrier.

Proposition 5.3 *For a simply connected closed manifold M such that $H^*(M;\mathbb{Q})$ is generated by a single element, the string bracket is trivial.*

Proof The result [17, Theorem 1] implies that the loop homology of M is isomorphic to the Hochschild cohomology of $A_{\mathrm{PL}}(M)$ endowed with the BV algebra structure due to Menichi [33]. We observe that M is formal. Therefore, Theorem 2.15 and explicit computations in [34, Theorem 16] and [53, Main Theorem] yield the result. In fact, for elements α_1 and α_2 in $\mathrm{Im}\,\widetilde{\Delta} = \mathrm{Ker}\,\widetilde{\Delta}$, we have $\Delta(\alpha_1 \bullet \alpha_2) = 0$; see Theorem 2.11 and Remark 2.10. In particular, we observe the case where $H^*(M) \cong H^*(S^n)$ with n odd. Then the generator a_{-n} of the loop homology $\mathbb{H}_*(LM) := H_{*+n}(LM)$ with odd degree is in $H_0(LM)$. Then the generator a_{-n} is not in $\mathrm{Ker}\,\widetilde{\Delta}$; see Theorem 2.15. □

The result [35, Theorem 39] due to Menichi gives an explicit form of the BV operator on the rational loop homology of a connected compact Lie group. We can also apply the result in our computation. In particular, the behavior of the string bracket as seen in Proposition 5.3 changes drastically in case of a Lie group with rank greater than one.

Proposition 5.4 (cf [19, Example 5.2]) *Let G be a simply connected Lie group with rank greater than one. The Lie algebra $\mathcal{H}_* = H^{S^1}_{*-\dim G+2}(LG;\mathbb{Q})$ endowed with the string bracket is nonnilpotent. More precisely, for any n, the n–fold bracket $[\mathcal{H}, [\mathcal{H}, \dots, [\mathcal{H}, \mathcal{H}] \cdots]]$ is nontrivial.*

Proof We first observe that a simply connected Lie group is formal. Indecomposable elements $x_1, \dots, x_N$ in $\mathbb{H}_*(G)$ are in the reduced homology $\widetilde{H}_*(LG)$ because $N := \mathrm{rank}\, G > 1$. Thus, it follows from [35, Theorem 39] and [25, Theorem 1] that x_i and $(s^{-1}x_j)^k$ are in $\mathrm{Ker}\,\widetilde{\Delta}$. Moreover, there exists a nontrivial n–fold string bracket. For example, for $k > n$, we see that on $\mathcal{H}_*$,

$$[x_j, [x_j, \dots, [x_j, (s^{-1}x_j)^k] \cdots]] = \pm k(k-1)\cdots(k-(n-1))(s^{-1}x_j)^{k-n} \neq 0.$$

This follows from the explicit formula of the BV operator in [35, Theorem 39] and Theorem 2.15. Observe that x_j is in $\mathrm{Ker}\,\widetilde{\Delta}$ if $\mathrm{rank}\, G > 1$. We have the result. □

5.3 Gravity algebras

The *gravity algebra* with higher Lie brackets was introduced by Getzler [21]. We consider a gravity algebra structure which appears on the string homology of a manifold and the classifying space of a Lie group; see, for example, [10, Definition 8.1] for the definition of the gravity algebra.

Example 5.5 The result [11, Theorem 1.1] due to Chen, Eshmatov and Liu shows that the negative cyclic homology of a DGA Ω admits a gravity algebra structure if the Hochschild homology of Ω has a BV algebra structure. The higher Lie bracket $[\ ,\dots,\ ]\colon (\mathrm{HC}^-_*(\Omega))^{\otimes n} \to \mathrm{HC}^-_*(\Omega)$ is defined for $n \geq 2$ by

$$[x_1,\dots,x_n] = (-1)^{(n-1)|x_1|+(n-2)|x_2|+\cdots+|x_{n-1}|}\beta(\pi(x_1)\odot\pi(x_2)\odot\cdots\odot\pi(x_n)),$$

where $\odot$ denotes the dual loop coproduct on the Hochschild homology.

Let G be a connected Lie group. We see that all higher Lie brackets are nontrivial for the classifying space BG. For the case where $\operatorname{rank} G \geq 2$, it follows from Theorem 2.7 that

$$[y_i x_i^\vee,\dots,y_i x_i^\vee, y_2 x_2^\vee x_1^\vee, y_1^l x_1^\vee x_2^\vee] = \pm 1\odot\cdots\odot 1\odot\Delta(y_2 x_2^\vee x_1^\vee)\odot\Delta(y_1^l x_1^\vee x_2^\vee) = l y_1^{l-1} x_1^\vee x_2^\vee \neq 0$$

with the same notation as in the proof of Proposition 5.2. Suppose that $\operatorname{rank} G = 1$. Then, with the same notation as in Example 5.1, we see that $[x^2,\dots,x^2] = \pm\mathrm{Coker}(\Delta(x^2)\odot\cdots\odot\Delta(x^2)\odot\Delta(x^2)\odot 1) = \pm\mathrm{Coker}(\Delta(x^2)\odot\cdots\odot\Delta(x^2)\odot 2x) = \pm 2^{n-1}x^{n-1} \neq 0$ for the higher Lie bracket of rank n.

Example 5.6 In [10], Chen has proved that the string homology of an orientable closed manifolds admits a gravity algebra structure extending the Lie algebra structure; see [10, Theorem 8.5] for more details. Let G be a simply connected Lie group. We see that all higher Lie brackets in the string homology of G are nontrivial if and only if $\operatorname{rank} G > 1$. In fact, in case of $\operatorname{rank} G > 1$, by applying Theorem 2.15 to the higher Lie bracket of G, we have $[x_j, s^{-1}x_j,\dots,s^{-1}x_j] = \pm k(s^{-1}x_j)^{k-1}$ in $H^{S^1}_*(LG;\mathbb{Q})$ with the same notation as in Proposition 5.4. If $\operatorname{rank} G = 1$, the only generator x_1 of odd degree is not in $\operatorname{Ker}\widetilde{\Delta}$ and then all higher Lie brackets are trivial; see the computation in the proof of Proposition 5.4.

6 Computation of the string bracket for a nonformal space

In this section, we consider the string bracket of a nonformal and BV-exact manifold. We begin recalling a nonformal manifold in [19, 6.4 Example].

Let $UTS^6 \to S^6$ be the unit tangent bundle over S^6. Then, we have a simply connected 11–dimensional manifold M which fits in the pullback diagram

$$\begin{array}{ccc} M & \longrightarrow & UTS^6 \\ \downarrow & & \downarrow{\scriptstyle p} \\ S^3\times S^3 & \xrightarrow{\ f\ } & S^6 \end{array}$$

where $f: S^3 \times S^3 \to S^6$ is a smooth map homotopic to the map defined by collapsing the 3–skeleton into a point. Since the Euler class of the unit tangent bundle mentioned above is nontrivial, it follows that the minimal model of M has the form $\mathcal{M} = (\wedge(x,y,z), d)$, where $d(x) = 0 = d(y)$, $d(z) = xy$, $|x| = |y| = 3$ and $|z| = 5$. It is readily seen that M is nonformal since the Massey product $\langle x, x, y\rangle$ does not vanish; see [24, page 277]. Moreover, we have:

Proposition 6.1 *The* 11*–dimensional manifold* M *is* BV*-exact.*

Proposition 6.1 is proved by computing the Hochschild homology explicitly. To this end, we recall the minimal model $\mathcal{M}$ for M mentioned above. The Hochschild homology of $\mathcal{M}$ is the homology of the Sullivan algebra $\mathcal{L} = (\wedge(x,y,z,\bar{x},\bar{y},\bar{z}), d)$, where $d(\bar{x}) = 0 = d(\bar{y})$, $d(\bar{z}) = -\bar{x}y + x\bar{y}$; see Section 3. To compute $H(\mathcal{L})$, we define its subcomplex $\mathcal{L}'$ by $\wedge(x,y,\bar{x},\bar{y},\bar{z})$. By a simple calculation, we have the following lemma.

Lemma 6.2 *The set* $\{\bar{x}^p\bar{y}^q, x\bar{x}^p\bar{y}^q, y\bar{y}^q, xy\bar{z}^r \mid p,q,r \geq 0\}$ *forms a basis of* $H(\mathcal{L}')$.

Next we compute $H(\mathcal{L})$ by comparing with $H(\mathcal{L}')$ and $\mathcal{L}/\mathcal{L}'$.

Proposition 6.3 *The following set forms a basis of the Hochschild homology* $H(\mathcal{L})$:

$$\{\bar{x}^p\bar{y}^q, x\bar{x}^p\bar{y}^q, y\bar{y}^q, xy\bar{z}^{r+1}, xz\bar{x}^p\bar{y}^q, yz\bar{y}^q, xyz\bar{z}^r, z\bar{x}^{p+1}\bar{y}^q - x\bar{x}^p\bar{y}^q\bar{z}, z\bar{y}^{q+1} - y\bar{y}^q\bar{z}\},$$

where p, q *and* r *run over all nonnegative integers.*

Proof Since there is an isomorphism of complexes $(\mathcal{L}/\mathcal{L}', d) \cong (\mathbb{Q}\{z\}, 0) \otimes (\mathcal{L}', d)$, Lemma 6.2 implies that the set $\{z\bar{x}^p\bar{y}^q, xz\bar{x}^p\bar{y}^q, yz\bar{y}^q, xyz\bar{z}^r \mid p,q,r \geq 0\}$ forms a basis of $H(\mathcal{L}/\mathcal{L}')$. Consider the long exact sequence associated with the short exact sequence $0 \to \mathcal{L}' \to \mathcal{L} \to \mathcal{L}/\mathcal{L}' \to 0$. The connecting homomorphism $H(\mathcal{L}/\mathcal{L}') \to H(\mathcal{L})$ sends z to xy and the other basis elements to zero. Hence each basis element of $H(\mathcal{L}')$ or $H(\mathcal{L}/\mathcal{L}')$ corresponds to a basis element of $H(\mathcal{L})$, except for z and xy. By lifting basis elements of $H(\mathcal{L}/\mathcal{L}')$ to cocycles in $\mathcal{L}$, we get the above basis. □

Proof of Proposition 6.1 Let $\widetilde{\mathcal{L}}$ be the reduced complex $\wedge^+(x,y,z,\bar{x},\bar{y},\bar{z})$. Recall that the reduced operation $\widetilde{B}$ in Definition 2.9 is modeled by the map $H\tilde{s}: H(\widetilde{\mathcal{L}}) \to H(\widetilde{\mathcal{L}})$ induced by the derivation $\tilde{s}: \widetilde{\mathcal{L}} \to \widetilde{\mathcal{L}}$, $v \mapsto \bar{v}$ for $v = x, y, z$; see Remark 3.2.

By using the basis given in Proposition 6.3, we see that

$$H\tilde{s}(x\bar{x}^p\bar{y}^q) = \bar{x}^{p+1}\bar{y}^q, \quad H\tilde{s}(y\bar{y}^q) = \bar{y}^{q+1}, \quad H\tilde{s}(xz\bar{x}^p\bar{y}^q) = z\bar{x}^{p+1}\bar{y}^q - x\bar{x}^p\bar{y}^q\bar{z},$$
$$H\tilde{s}(yz\bar{y}^q) = z\bar{y}^{q+1} - y\bar{y}^q\bar{z}, \quad H\tilde{s}(xyz\bar{z}^r) = \frac{r+2}{r+1}xy\bar{z}^{r+1}.$$

This proves $\operatorname{Ker} H\tilde{s} = \operatorname{Im} H\tilde{s}$. □

Remark 6.4 A program [51] on a personal computer for computing the homology of a DGA helps us in proving Proposition 6.3. In fact, the computer calculation shows the basis in the proposition, while our proof is *by hand*.

Remark 6.5 In the minimal model $\mathcal{M}$, we define weights of x, y and z by 1, 1 and 2, respectively. Then, it is readily seen that the model $\mathcal{M}$ for the manifold M admits positive weights. Therefore, Theorem 2.21 enables us to conclude that M is BV-exact. However, the explicit generators of the Hochschild homology of $\mathcal{M}$ represented in Proposition 6.3 are used in the computation below of the string bracket of M. We adhere to the proof of Proposition 6.1.

The negative cyclic homology of $\mathcal{M}$ is isomorphic to the homology of $\mathscr{E} = (\wedge(u)\otimes\wedge(x,y,z,\overline{x},\overline{y},\overline{z}), D)$; see Section 3. Here, the differential D is given by

$$D(x) = u\overline{x}, \quad D(y) = u\overline{y}, \quad D(z) = xy + u\overline{z}, \quad D(\overline{x}) = D(\overline{y}) = 0, \quad D(\overline{z}) = -\overline{x}y + x\overline{y}.$$

Then the morphism β in Theorem 2.7 is induced by the derivation $s\colon \mathscr{L}\to\mathscr{E}$.

It follows from the BV-exactness of the manifold M that $\mathrm{HC}_*^-(\mathcal{M})$ decomposes into a direct sum $\mathbb{Q}[u]\oplus \operatorname{Im}\widetilde{\beta}$, where $\widetilde{\beta}$ is a morphism induced by the map $\widetilde{s}\colon\widetilde{\mathscr{L}}\to\widetilde{\mathscr{E}}$ on the reduced complexes. Hence, by applying $\widetilde{\beta}$ to the basis except for 1 in Proposition 6.3, we see that $\operatorname{Im}\widetilde{\beta}$ is spanned by the homology classes

$$\zeta_{p,q} := \frac{1}{p!q!}\overline{x}^p\overline{y}^q, \qquad \eta_{p,q} := \begin{cases} \dfrac{1}{p!q!}(z\overline{x}^p\overline{y}^q - x\overline{x}^{p-1}\overline{y}^q\overline{z}) & \text{if } p\neq 0,\\ \dfrac{1}{q!}(z\overline{y}^q - y\overline{y}^{q-1}\overline{z}) & \text{if } p=0,\end{cases} \qquad \theta_r := \frac{r+1}{r}xy\overline{z}^r,$$

for $p,q\geq 0$, $r\geq 1$ with $(p,q)\neq(0,0)$. We also put $\zeta_{0,0}=1$ for convenience. Denote by Dsb the dual string bracket $[\ ,\ ]^\vee$ over $\mathbb{Q}$ stated in Theorem 2.8.

Theorem 6.6 *For the dual string bracket* Dsb *over* $\mathbb{Q}$ *of the* 11*–dimensional manifold* M, *one has*

$$\mathrm{Dsb}(\zeta_{p,q}) = \sum_{i=0}^{p+1}\sum_{j=0}^{q+1}\{i(q+1)-j(p+1)\}(\zeta_{i,j}\otimes\eta_{p+1-i,q+1-j} - \eta_{p+1-i,q+1-j}\otimes\zeta_{i,j}),$$

$$\mathrm{Dsb}(\eta_{p,q}) = \theta_2\otimes\zeta_{p,q} - \zeta_{p,q}\otimes\theta_2 - \sum_{i=0}^{p+1}\sum_{j=0}^{q+1}\{i(q+1)-j(p+1)\}\eta_{i,j}\otimes\eta_{p+1-i,q+1-j},$$

$$\mathrm{Dsb}(\theta_r) = 0.$$

Proof We first compute the dual loop product Dlp by the rational model described in [19]. Let $\mathcal{M}=\wedge V$ be the minimal model for M, $\mathcal{P} = (\wedge V)^{\otimes 2}\otimes\wedge\overline{V}$ the Sullivan model for the free path space stated in [15, Section 15] and $\varepsilon_{\mathcal{P}}\colon\mathcal{P}\to\wedge V$ the $(\wedge V)^{\otimes 2}$–semifree resolution of $\wedge V$ which is given by the multiplication of $\wedge V$ and the canonical augmentation of $\wedge\overline{V}$.

By virtue of [19, Lemma 1], we see that a DGA morphism $\mathcal{P}\to\mathcal{P}\otimes_{\wedge V}\mathcal{P}$ defined by $v_1\otimes v_2\mapsto v_1\otimes 1\otimes v_2$, $\overline{x}\mapsto 1\otimes\overline{x}+\overline{x}\otimes 1$, $\overline{y}\mapsto 1\otimes\overline{y}+\overline{y}\otimes 1$, $\overline{z}\mapsto 1\otimes\overline{z}+\overline{z}\otimes 1-\frac{1}{2}\overline{x}\otimes\overline{y}+\frac{1}{2}\overline{y}\otimes\overline{x}$ for $v_i\in V$ is a Sullivan representative for the composition of free paths. This induces a Sullivan representative $\mathcal{M}_{\mathrm{Comp}}\colon\mathscr{L}\to\mathscr{L}\otimes_{\wedge V}\mathscr{L}$ for Comp in (3.4) which has formulae

$$\mathcal{M}_{\mathrm{Comp}}(v) = v, \qquad \mathcal{M}_{\mathrm{Comp}}(\overline{x}) = 1\otimes\overline{x}+\overline{x}\otimes 1,$$
$$\mathcal{M}_{\mathrm{Comp}}(\overline{y}) = 1\otimes\overline{y}+\overline{y}\otimes 1, \quad \mathcal{M}_{\mathrm{Comp}}(\overline{z}) = 1\otimes\overline{z}+\overline{z}\otimes 1-\tfrac{1}{2}\overline{x}\otimes\overline{y}+\tfrac{1}{2}\overline{y}\otimes\overline{x},$$

where $v \in V$. Recall the morphism $\varepsilon_{\mathcal{P}} \otimes 1 : \mathcal{P} \otimes_{(\wedge V)^{\otimes 2}} \mathcal{L}^{\otimes 2} \to \wedge V \otimes_{(\wedge V)^{\otimes 2}} \mathcal{L}^{\otimes 2}$ appeared in the model for Dlp. A section σ of the morphism $\varepsilon_{\mathcal{P}} \otimes 1$ is given by

$$\sigma(v) = v \otimes 1, \quad \sigma(\bar{x} \otimes 1) = 1 \otimes (\bar{x} \otimes 1), \quad \sigma(1 \otimes \bar{x}) = 1 \otimes (1 \otimes \bar{x}), \quad \sigma(\bar{y} \otimes 1) = 1 \otimes (\bar{y} \otimes 1),$$
$$\sigma(1 \otimes \bar{y}) = 1 \otimes (1 \otimes \bar{y}), \quad \sigma(\bar{z} \otimes 1) = 1 \otimes (\bar{z} \otimes 1), \quad \sigma(1 \otimes \bar{z}) = 1 \otimes (1 \otimes \bar{z}) - \bar{x} \otimes (1 \otimes \bar{y}) + \bar{y} \otimes (1 \otimes \bar{x}).$$

Define a $(\wedge V)^{\otimes 2}$–morphism $\mathrm{Diag}^! : \mathcal{P} \to (\wedge V)^{\otimes 2}$ of degree 11 by

$$\mathrm{Diag}^!(1) = (-x \otimes 1 + 1 \otimes x)(-y \otimes 1 + 1 \otimes y)(-z \otimes 1 + 1 \otimes z), \quad \mathrm{Diag}^!|_{\wedge^+ \bar{V}} \equiv 0,$$

which gives a representative of a nonzero element in $\mathrm{Ext}^{11}_{(\wedge V)^{\otimes 2}}(\wedge V, \wedge V)$; see [49, Section 5] for the detail about a construction of the shriek map $\mathrm{Diag}^!$. Then, the result [19, Theorem A] yields that the composite

$$\mathcal{L} \xrightarrow{\mathcal{M}_{\mathrm{Comp}}} \mathcal{L} \otimes_{\wedge V} \mathcal{L} \cong \wedge V \otimes_{(\wedge V)^{\otimes 2}} \mathcal{L}^{\otimes 2} \xrightarrow{\sigma} \mathcal{P} \otimes_{(\wedge V)^{\otimes 2}} \mathcal{L}^{\otimes 2} \xrightarrow{\mathrm{Diag}^! \otimes 1} \mathcal{L}^{\otimes 2}$$

induces the dual loop product Dlp on homology. This rational model and a straightforward computation enable us to compute Dlp explicitly. In fact, we have

$$\begin{aligned}
\mathrm{Dlp}(\bar{x}^p \bar{y}^q) &= xyz \otimes \bar{x}^p \bar{y}^q + \bar{x}^p \bar{y}^q \otimes xyz \\
&\quad + \sum_{i=0}^{p} \sum_{j=0}^{q} \binom{p}{i} \binom{q}{j} \big(-x\bar{x}^i \bar{y}^j \otimes yz\bar{x}^{p-i} \bar{y}^{q-j} + y\bar{x}^i \bar{y}^j \otimes xz\bar{x}^{p-i} \bar{y}^{q-j} \\
&\qquad - xz\bar{x}^i \bar{y}^j \otimes y\bar{x}^{p-i} \bar{y}^{q-j} + yz\bar{x}^i \bar{y}^j \otimes x\bar{x}^{p-i} \bar{y}^{q-j} \big),
\end{aligned}$$

$$\begin{aligned}
\mathrm{Dlp}(z\bar{x}^p \bar{y}^q - x\bar{x}^{p-1} \bar{y}^q \bar{z}) &= xyz \otimes (z\bar{x}^p \bar{y}^q - x\bar{x}^{p-1} \bar{y}^q \bar{z}) + (z\bar{x}^p \bar{y}^q - x\bar{x}^{p-1} \bar{y}^q \bar{z}) \otimes xyz \\
&\quad + xyz\bar{z} \otimes x\bar{x}^{p-1} \bar{y}^q + x\bar{x}^{p-1} \bar{y}^q \otimes xyz\bar{z} + xy\bar{z} \otimes xz\bar{x}^{p-1} \bar{y}^q - xz\bar{x}^{p-1} \bar{y}^q \otimes xy\bar{z} \\
&\quad - \sum_{i=0}^{p} \sum_{j=0}^{q} \binom{p}{i} \binom{q}{j} (xz\bar{x}^i \bar{y}^j \otimes yz\bar{x}^{p-i} \bar{y}^{q-j} - yz\bar{x}^i \bar{y}^j \otimes xz\bar{x}^{p-i} \bar{y}^{q-j}),
\end{aligned}$$

$$\begin{aligned}
\mathrm{Dlp}(z\bar{y}^q - y\bar{y}^{q-1}\bar{z}) &= -xyz \otimes (z\bar{y}^q - y\bar{y}^{q-1}\bar{z}) - (z\bar{y}^q - y\bar{y}^{q-1}\bar{z}) \otimes xyz + xyz\bar{z} \otimes y\bar{y}^{q-1} + y\bar{y}^{q-1} \otimes xyz\bar{z} \\
&\quad + xy\bar{z} \otimes yz\bar{y}^{q-1} - yz\bar{y}^{q-1} \otimes xy\bar{z} - \sum_{j=0}^{q} \binom{q}{j} (xz\bar{y}^j \otimes yz\bar{y}^{q-j} - yz\bar{y}^j \otimes xz\bar{y}^{q-j}),
\end{aligned}$$

$$\mathrm{Dlp}(xy\bar{z}^r) = \sum_{i=0}^{r} \binom{r}{i} (-xyz\bar{z}^i \otimes xy\bar{z}^{r-i} + xy\bar{z}^i \otimes xyz\bar{z}^{r-i}).$$

It follows from Theorem 2.8(ii) that

$$\mathrm{Dsb}(\zeta_{p,q}) = \frac{1}{p!q!} (\beta \otimes \beta) \bullet^\vee (\bar{x}^p \bar{y}^q), \qquad \mathrm{Dsb}(\eta_{p,q}) = \frac{1}{p!q!} (\beta \otimes \beta) \bullet^\vee (z\bar{x}^p \bar{y}^q - x\bar{x}^{p-1} \bar{y}^q \bar{z}),$$
$$\mathrm{Dsb}(\eta_{0,q}) = \frac{1}{q!} (\beta \otimes \beta) \bullet^\vee (z\bar{y}^q - y\bar{y}^{q-1}\bar{z}), \qquad \mathrm{Dsb}(\theta_r) = \frac{r+1}{r} (\beta \otimes \beta) \bullet^\vee (xy\bar{z}^r).$$

Therefore, by these formulae and the computations of Dlp above, we have the result. □

7 The cobar-type EMSS and r–BV-exactness

Let X be a simply connected space. We define a cobracket on the cobar-type Eilenberg–Moore spectral sequence converging to the rational equivariant cohomology of the free loop space LX, which is compatible with the dual to the string bracket in the sense of Chas and Sullivan [8] if X is a simply connected closed manifold.

We begin by recalling the spectral sequence associated with a filtered complex (A, F, d). Consider the submodules $Z_r^{p,q}$ and $B_r^{p,q}$ defined by

$$Z_r^{p,q} := F^p A^{p+q} \cap d^{-1}(F^{p-r} A^{p+q+1}) \quad \text{and} \quad B_r^{p,q} := F^p A^{p+q} \cap d(F^{p-r} A^{p+q-1}). \tag{7.1}$$

With the submodules of A, we have a spectral sequence $\{E_r, d_r\}$ whose E_r–term is defined by $E_r^{p,q} := Z_r^{p,q}/(Z_{r-1}^{p+1,q-1} + B_{r-1}^{p,q})$; see [32, Proof of Theorem 2.6].

We use the same notation as that in Section 2. In particular, for a cochain algebra A, we define a chain algebra $A_\sharp$ by $(A_\sharp)_{-i} = A^i$ for i. The converse is also considered; that is, for a chain algebra Ω, we have a cochain algebra $\Omega^\sharp$ defined by $((\Omega)^\sharp)^i = \Omega_{-i}$ for i; see Remark 2.6.

Let $(\wedge V, d)$ be a Sullivan model of a simply connected commutative cochain algebra A. Define $(\mathscr{L}, \delta) = (\wedge(V \oplus \overline{V}), \delta)$ and $(\mathscr{E}, D) = (\wedge u \otimes \mathscr{L}, D)$; see Section 3. Then complexes $(\mathscr{L}, \delta)$ and $(\mathscr{E}, D)$ compute the Hochschild homology and the negative cyclic homology of $A_\sharp$, respectively. Thus we have the cobar-type Eilenberg–Moore spectral sequence (the EMSS for short) $\{E_r^{*,*}, d_r\}$ converging to $\mathrm{HC}_*^-(A) := (\mathrm{HC}_*^-(A_\sharp))^\sharp$ as an algebra with

$$E_2^{*,*} \cong \mathrm{Cotor}^{*,*}_{\wedge(t)}(\mathrm{HH}_*(A), \mathbb{Q})$$

as a bigraded algebra, where $|t| = 1$ and the $\wedge(t)$–comodule structure on the Hochschild homology $\mathrm{HH}_*(A) := (\mathrm{HH}_*(A_\sharp))^\sharp$ is induced by the derivation s in the cyclic complex $(\mathscr{E}, D)$. In fact, the $\wedge(t)$–comodule structure $\nabla\colon \mathscr{L} \to \mathscr{L} \otimes \wedge(t)$ on $(\mathscr{L}, \delta)$ is given by $\nabla(\alpha) = \gamma(\alpha) \otimes t + \alpha \otimes 1$, where $\gamma(\alpha) = (-1)^{|\alpha|} s(\alpha)$. A map assigning the element au^n in the cyclic complex $\mathscr{E}$ to an element $a[t|\cdots|t]$ in the n^{th} cobar complex gives rise to an isomorphism of complexes. As a consequence, we have isomorphisms

$$\mathrm{Cotor}^*_{\wedge(t)}(\mathscr{L}, \mathbb{Q}) \cong H(\mathscr{E}, D) \cong \mathrm{HC}_*^-(A_\sharp)^\sharp.$$

Remark 7.2 The isomorphisms above allow us to work in the category of $\wedge(t)$–comodule in order to investigate the negative cyclic homology of a DGA.

We observe that, by construction, there is an isomorphism $E_1^{0,*} \cong \mathrm{HH}_*(A)$. In particular, when we choose the polynomial de Rham algebra $A_{\mathrm{PL}}(M)$ for a simply connected space M as the DGA A, the spectral sequence converges to the S^1–equivariant cohomology $\mathrm{HC}_*^-(A) \cong H^*_{S^1}(LM; \mathbb{Q})$, with

$$E_2^{*,*} \cong \mathrm{Cotor}^{*,*}_{H^*(S^1;\mathbb{Q})}(H^*(LM; \mathbb{Q}), \mathbb{Q}).$$

One has a direct sum decomposition $(\widetilde{\mathscr{L}}, \delta) = \bigoplus_n (\widetilde{\mathscr{L}}^{(n)}, \delta)$ of complexes, where $\widetilde{\mathscr{L}}^{(n)} = \widetilde{\mathscr{L}} \cap (\wedge V \otimes \wedge^n \overline{V})$ with $\mathscr{L} = \widetilde{\mathscr{L}} \oplus \mathbb{Q}$. Then the reduced derivation $\widetilde{s}$ decomposes $(\widetilde{\mathscr{L}}, \delta)$ into a sequence $0 \to \widetilde{\mathscr{L}}^{(0)} \to \widetilde{\mathscr{L}}^{(1)} \to \widetilde{\mathscr{L}}^{(2)} \to \cdots$ of complexes. Thus it follows that the EMSS $\{E_r^{*,*}, d_r\}$ is decomposed as

$$\{E_r^{*,*}, d_r\} = \bigoplus_{N \in \mathbb{Z}} \{{}_{(N)}E_r^{*,*}, d_r\} \oplus \{\mathbb{Q}[u], 0\},$$

where $\operatorname{bideg} u = (1,1)$, each spectral sequence $\{{}_{(N)}E_r^{*,*}, d_r\}$ for $N \geq 0$ is constructed by the double complex

$${}_{(N)}\mathscr{K}\colon 0 \to \widetilde{\mathscr{L}}^{(N)} \to \widetilde{\mathscr{L}}^{(N+1)} \otimes \mathbb{Q}\{u\} \to \widetilde{\mathscr{L}}^{(N+2)} \otimes \mathbb{Q}\{u^2\} \to \cdots$$

and for $N < 0$, the spectral sequence $\{{}_{(N)}E_r^{*,*}, d_r\}$ is obtained by the double complex

$${}_{(N)}\mathscr{K}\colon 0 \to 0 \to \cdots \to 0 \to \widetilde{\mathscr{L}}^{(0)} \otimes \mathbb{Q}\{u^{-N}\} \to \widetilde{\mathscr{L}}^{(1)} \otimes \mathbb{Q}\{u^{-N+1}\} \to \cdots .$$

Here, the double complex ${}_{(N)}\mathscr{K}$ is regarded as a filtered complex associated with the horizontal degrees. Thus, in the spectral sequence $\{{}_{(N)}E_r^{*,*}, d_r\}$ for $N < 0$, we have ${}_{(N)}E_r^{i,*} = 0$ for $i < -N$. We observe that each spectral sequence $\{{}_{(N)}E_r^{*,*}, d_r\}$ converges to the target as an algebra.

Remark 7.3 The direct sum of the targets of the spectral sequences $\{{}_{(N)}E_r^{*,*}, d_r\}$ is nothing but the *Hodge decomposition* of $\mathrm{HC}_*^-(A)$; that is, we have $\widetilde{\mathrm{HC}}_*^-(A) = \bigoplus_{N \geq 0} H(\mathscr{K}_{(N)})$; see [6, Section 2]. If A is the polynomial de Rham algebra $A_{\mathrm{PL}}(X)$ for a simply connected space X, then the direct summands in the Hodge decomposition are identified with the eigenspaces of the Adams operation on $\widetilde{H}_{S^1}^*(LX; \mathbb{Q})$; see [6, Theorem 3.2] for the identification. We refer the reader to [30, 4.5.4] for the operation. The result [3, Theorem 1.1] shows that the string bracket respects the Hodge decomposition in some sense. Thus, we are also interested in computations of string brackets, as described in Section 1.1, together with the consideration of the Hodge decomposition.

Proposition 7.4 *If the spectral sequence $\{{}_{(0)}E_r, d_r\}$ collapses at E_r–term, then so does $\{{}_{(N)}E_r, d_r\}$ for each integer N, and then* $\operatorname{Tot} E_r \cong H_{S^1}^*(LX)$ *as a vector space.*

Thus, it is readily seen that the collapsing of the EMSS is governed by that of the zeroth spectral sequence.

Corollary 7.5 *The spectral sequence $\{{}_{(0)}E_r, d_r\}$ collapses at the E_r–term if and only if so does $\{E_r^{*,*}, d_r\}$.*

Lemma 7.6 *For $l \geq r-1$ and $N \in \mathbb{Z}$, we have ${}_{(0)}E_r^{l+N,*+N} \cong {}_{(N)}E_r^{l,*}$.*

Proof For $N < 0$, the multiplication $u^{-N}\times\colon {}_{(0)}E_r^{*,*} \to {}_{(N)}E_r^{*-N,*-N}$ gives an isomorphism. Assume that $N \geq 0$. By definition, we see that

$$\begin{aligned} {}_{(N)}E_r^{l,*} &= {}_{(N)}Z_r^{l,*}/({}_{(N)}Z_{r-1}^{l+1,*-1} + {}_{(N)}B_{r-1}^{l,*}), \\ {}_{(0)}E_r^{l+N,*+N} &= {}_{(0)}Z_r^{l+N,*+N}/({}_{(0)}Z_{r-1}^{l+N+1,*+N-1} + {}_{(0)}B_{r-1}^{l+N,*+N}), \end{aligned}$$

where ${}_{(N)}Z$ and ${}_{(N)}B$ denote the subcomplexes of ${}_{(N)}\mathscr{K}$ defined in (7.1) for the filtered complex $A = {}_{(N)}\mathscr{K}$. Moreover, we have ${}_{(N)}Z_r^{l,*} \cong {}_{(0)}Z_r^{l+N,*+N}$ and ${}_{(N)}Z_{r-1}^{l+1,*-1} \cong {}_{(0)}Z_{r-1}^{l+N+1,*+N-1}$. Since $l \geq r-1$, it follows that ${}_{(N)}B_{r-1}^{l,*} \cong {}_{(0)}B_{r-1}^{l+N,*+N}$. Then the multiplication $u^N \times \colon {}_{(N)}E_r^{l,*} \cong {}_{(0)}E_r^{l+N,*+N}$ is an isomorphism. □

Proof of Proposition 7.4 Lemma 4.7 yields that the spectral sequence $\{{}_{(0)}E_r^{*,*}, d_r\}$ converges to 0, the trivial module. The assumption and Lemma 7.6 imply that ${}_{(N)}E_r^{l,*} = 0$ for $l \geq r-1$ and N. □

Theorem 7.7 *Let M be a simply connected closed manifold, and A the polynomial de Rham algebra $A_{\mathrm{PL}}(M)$ of M. Then the map $[\ ,\ ]_r^\vee \colon E_r^{p,*} \to (E_r^{*,*} \otimes E_r^{*,*})^{p,*+d-2}$ defined by $[\ ,\ ]_r^\vee \equiv 0$ for $p > 0$ and for $p = 0$, the composite*

$$E_r^{0,*} = \operatorname{Ker} d_{r-1} \xrightarrow{i} \mathrm{HH}_*(A) \xrightarrow{\bullet^\vee} \mathrm{HH}_*(A)^{\otimes 2} \xrightarrow{\Delta \otimes \Delta} E_r^{0,*} \otimes E_r^{0,*}$$

gives rise to a cobracket on the spectral sequence, where i denotes the inclusion. That is, it is compatible with the differentials and $H([\ ,\ ]_r^\vee) = [\ ,\ ]_{r+1}^\vee$. Moreover, the cobracket $[\ ,\ ]_\infty^\vee$ is compatible with the dual to the string bracket on $H_*^{S^1}(LM)$ at the E_∞–term in the sense that the composite

$$H_{S^1}^*(LM) \xrightarrow{\pi} E_\infty^{0,*} \xrightarrow{i} \mathrm{HH}_*(A) \xrightarrow{\bullet^\vee} \mathrm{HH}_*(A)^{\otimes 2} \xrightarrow{\Delta \otimes \Delta} E_\infty^{0,*} \otimes E_\infty^{0,*}$$

coincides with the dual to the string bracket modulo $F^1 H_{S^1}^*(LM)$. Here π is the projection and $\{F^l H_{S^1}^*(LM)\}_{l \geq 0}$ is the decreasing filtration associated with the spectral sequence.

Proof By dimensional reasons, it is readily seen that $(d_r \otimes 1 \pm 1 \otimes d_r) \circ [\ ,\ ]_r^\vee = 0 = [\ ,\ ]_r^\vee \circ d_r$ for $p > 0$. Moreover, we see that every element in the image of Δ in $E_r^{0,*}$ is a permanent cocycle. In fact, for $w \in \operatorname{Im} \Delta$, we have $Dw = (\delta + us)w = 0$. Then, it follows that $(d_r \otimes 1 \pm 1 \otimes d_r) \circ [\ ,\ ]_r^\vee = 0 = [\ ,\ ]_r^\vee \circ d_r$ in $E_r^{0,*}$. By the definition of the cobrackets, we have $H([\ ,\ ]_r^\vee) = [\ ,\ ]_{r+1}^\vee$. In fact, the left-hand side is the restriction of $[\ ,\ ]_r^\vee$ in the nontrivial case.

Consider the compatibility of the cobracket at the E_∞–term. We have a commutative diagram

$$\begin{array}{ccccccc}
\mathrm{HC}_*^-(A) & \xrightarrow{\qquad\pi\qquad} & \mathrm{HH}_*(A) & \xrightarrow{\bullet^\vee} & \mathrm{HH}_*(A)^{\otimes 2} & \xrightarrow{\beta\otimes\beta} & \mathrm{HC}_*^-(A)^{\otimes 2} \\
\big\downarrow{\scriptstyle\cong} & & \Big\| & & & \nearrow{\scriptstyle\beta\otimes\beta} & \big\downarrow{\scriptstyle\mathrm{pr}} \\
\mathrm{HC}_*^-(A)/F^1 \oplus \quad F^1/F^2 \oplus \cdots \oplus F^* & & & & & & (\mathrm{HC}_*^-(A)/F^1)^{\otimes 2} \\
\big\downarrow{\scriptstyle\cong} & & & & & & \big\downarrow{\scriptstyle\cong} \\
\bigoplus_{p+q=*} E_\infty^{p,q} \xrightarrow{\mathrm{pr}} E_\infty^{0,*} = \operatorname{Ker} d_{*+1} & \xrightarrow{i} & \mathrm{HH}_*(A) & \xrightarrow{\bullet^\vee} & \mathrm{HH}_*(A)^{\otimes 2} & \xrightarrow{\Delta\otimes\Delta} & E_\infty^{0,*} \otimes E_\infty^{0,*}
\end{array}$$

where $\{F^l\}_{l \geq 0}$ denotes the decreasing filtration of $\mathrm{HC}_*^-(A) \cong H_{S^1}^*(LM)$ associated with the spectral sequence. In fact, the commutativity of the left-hand side square and the right-hand side triangle follows from the construction of the spectral sequence; see for example [32, Proof of Theorem 2.6]. Theorem 2.8(ii) implies the upper sequence is the dual of the string bracket. We have the result. □

Proposition 7.8 *For each N, the S–action $u\times$ on $(\mathscr{E}, D)$ gives rise to a map*

$$S\colon \{{}_{(N)}E_r^{*,*}, d_r\} \to \{{}_{(N-1)}E_r^{*+1,*+1}, d_r\}$$

on the spectral sequence which is compatible with the S–action on the negative cyclic homology $\mathrm{HC}_*^-(A)$.

Proof The S–action on $(\mathscr{E}, D)$ gives rise to a map ${}_{(N)}\mathscr{K} \to {}_{(N-1)}\mathscr{K}$ which increases the filtration degree by $+1$ and is compatible with the differential. Then the map induces the action S on the spectral sequence. □

Lemma 7.9 *Suppose that the E_r–term ${}_{(0)}E_r^{p,q}$ in $\{{}_{(0)}E_r^{*,*}, d_r\}$ is trivial for any (p,q). Then the $(r-1)$ times S–action $S^{r-1}\colon \widetilde{\mathrm{HC}}_*^-(A) \to \widetilde{\mathrm{HC}}_*^-(A)$ is trivial.*

Proof Let x be an element in $\widetilde{\mathrm{HC}}_*^-(A)$. Then x is in $\widetilde{\mathrm{HC}}_*^{-,(n)}(A)$ for some $n \geq 0$ and then it is represented by an element α in ${}_{(n)}E_\infty^{t,*}$ for some $t \geq 0$. Thus the element $S^{r-1}x$ is represented by $S^{r-1}\alpha \in {}_{(n-(r-1))}E_\infty^{t+(r-1),*}$. By assumption, it follows from Lemma 7.6 that ${}_{(N)}E_r^{l,*} = 0$ for $l \geq r-1$ and $N \in \mathbb{Z}$. This implies that $S^{r-1}\alpha = 0$ in the E_∞–term and that there is no extension problem; that is, $S^{r-1}x = S^{r-1}\alpha = 0$ in $\mathrm{HC}_*^{-,(n-(r-1))}(A) \subset \mathrm{HC}_*^-(A)$. This completes the proof. □

Moreover, we have:

Theorem 7.10 *The E_r–term ${}_{(0)}E_r^{p,q}$ in $\{{}_{(0)}E_r^{*,*}, d_r\}$ is trivial for any (p,q) if and only if the $(r-1)$ times S–action S^{r-1} on $\widetilde{\mathrm{HC}}_*^-(A)$ is.*

Proof The "only if" part follows from Lemma 7.9. To prove the "if" part, we assume that S^{r-1} is trivial on $\widetilde{\mathrm{HC}}_*^-(A)$. Take any element $x = x_p \otimes u^p + x_{p+1} \otimes u^{p+1} + \cdots \in {}_{(0)}Z_r^{p,*}$, where $x_i \in \widetilde{\mathscr{L}}^{(i)}$ is zero for sufficiently large i. By the definition of ${}_{(0)}Z_r^{p,*}$, the total differential increases the filtration degree of x by r, ie we have $dx_p = 0$ and $\tilde{s}x_i + dx_{i+1} = 0$ for $p \leq i \leq p+r-2$. Now we have an element $[\tilde{s}x_{p+r-1}] \in H(\mathrm{Ker}\,\tilde{s})$ and the above equation implies that $[dx_{p+1}] = S^{r-1}[sx_{p+r-1}] = 0 \in H(\mathrm{Ker}\,\tilde{s})$ by Lemma 4.10 and the assumption of triviality of S^{r-1}. By Lemma 4.7, we see that $\mathrm{Ker}\,\tilde{s} = \mathrm{Im}\,\tilde{s}$. Thus, there is an element $v_p \in \widetilde{\mathscr{L}}^{(p)}$ with $d\tilde{s}v_p = dx_{p+1}$. By using these elements, we define $y = (x_{p+1} - \tilde{s}v_p) \otimes u^{p+1} + x_{p+2} \otimes u^{p+2} + x_{p+3} \otimes u^{p+3} + \cdots \in {}_{(0)}Z_{r-1}^{p+1,*}$. Then we can show $x - y = x_p \otimes u^p + \tilde{s}v_p \otimes u^{p+1} \in {}_{(0)}B_{r-1}^{p,*}$ by the same argument as above. It follows that $x = y + (x-y) \in {}_{(0)}Z_{r-1}^{p+1,*} + {}_{(0)}B_{r-1}^{p,*}$ and hence $[x] = 0 \in {}_{(0)}Z_r^{p,*}/{}_{(0)}Z_{r-1}^{p+1,*} + {}_{(0)}B_{r-1}^{p,*} = {}_{(0)}E_r^{p,*}$. Since x is an arbitrary element of ${}_{(0)}Z_r^{p,*}$, this proves the "if" part. □

The BV-exactness of a space is equivalent to the condition that the E_2–term of the spectral sequence $\{{}_{(0)}E_r^{*,*}, d_r\}$ is trivial. Then Theorem 7.10 gives another proof of Theorem 2.11. This consideration allows us to propose a *higher version* of the BV-exactness.

Definition 7.11 A simply connected space X is *r–BV-exact* if the E_{r+1}–term ${}_{(0)}E_{r+1}^{p,q}$ in the spectral sequence $\{{}_{(0)}E_r^{*,*}, d_r\}$ associated with X is trivial for any (p,q).

Indeed, there exists a non-BV-exact space in the class of rational elliptic spaces; see Appendix A below. While we are interested in the hierarchy of rational spaces defined by the r–BV-exactness as seen in Section 1.1, we do not pursue the topic in this manuscript.

Appendix A A non-BV-exact space

We describe an example of a non-BV-exact space. Let $(\wedge V, d)$ be the minimal model

$$(\wedge(x_1, x_2, y_1, y_2, y_3, z), d)$$

of an elliptic space X of dimension 228, given in [1, Example 5.2]. The degrees are given by $|x_1| = 10$, $|x_2| = 12$, $|y_1| = 41$, $|y_2| = 43$, $|y_3| = 45$ and $|z| = 119$. The differential is as follows:

$$dx_1 = 0, \quad dx_2 = 0, \qquad dy_1 = x_1^3 x_2, \quad dy_2 = x_1^2 x_2^2, \quad dy_3 = x_1 x_2^3, \qquad dz = x_2(y_1x_2 - x_1y_2)(y_2x_2 - x_1y_3) + x_1^{12} + x_2^{10}.$$

Note that X does not admit positive weights. Indeed, let $\mathrm{wt}(x_1) = i$ and $\mathrm{wt}(x_2) = j$. Then $\mathrm{wt}(y_1) = 3i + j$, $\mathrm{wt}(y_2) = 2i + 2j$ and $\mathrm{wt}(y_3) = i + 3j$. By dz, we have the equations $5i + 6j = 12i = 10j$ induced from $\mathrm{wt}(x_2(y_1x_2 - x_1y_2)(y_2x_2 - x_1y_3)) = \mathrm{wt}(x_1^{12}) = \mathrm{wt}(x_2^{10})$. Thus we obtain $i = j = 0$.

Let $\omega = x_1^{14}y_2y_3 - x_1^{13}x_2y_1y_3 + x_1^{12}x_2^2y_1y_2$ be the representing cocycle of the fundamental class of the manifold, which is considered as an element of $\wedge^+ V = \widetilde{\mathscr{L}}^{(0)} \subset \widetilde{\mathscr{L}}$. Then $[\omega] \notin \mathrm{Im}\big(H(\widetilde{s}) : 0 \to H(\widetilde{\mathscr{L}}^{(0)})\big) = 0$. On the other hand, we have $[\omega] \in \mathrm{Ker}(H(\widetilde{s}) \colon H(\widetilde{\mathscr{L}}^{(0)}) \to H(\widetilde{\mathscr{L}}^{(1)}))$, since $\widetilde{s}(\omega) = \delta(\alpha)$ for the element α defined by

$$\begin{aligned}\alpha = &-1380x_1^{11}x_2^6\bar{y}_3 - 5290x_1^{11}x_2^5y_3\bar{x}_2 - 114x_1^{10}y_1y_2\bar{y}_2 + 114x_1^{10}y_1y_3\bar{y}_1 \\ &- 114x_1^9x_2y_1y_2\bar{y}_1 + \tfrac{93}{2}x_1^2y_2y_3\bar{z} + x_1^2y_2z\bar{y}_3 - x_1^2y_3z\bar{y}_2 + 114x_1x_2^7y_2y_3\bar{y}_3 \\ &- \tfrac{93}{2}x_1x_2y_1y_3\bar{z} - x_1x_2y_1z\bar{y}_3 - 114x_1x_2y_2z\bar{y}_2 + 115x_1x_2y_3z\bar{y}_1 + 113x_1y_1y_3z\bar{x}_2 \\ &+ 572x_1y_2y_3z\bar{x}_1 + 115x_2^9\bar{z} - 114x_2^8y_1y_3\bar{y}_3 + 114x_2^8y_2y_3\bar{y}_2 + 1150x_2^8z\bar{x}_2 \\ &+ \tfrac{93}{2}x_2^2y_1y_2\bar{z} + 115x_2^2y_1z\bar{y}_2 - 115x_2^2y_2z\bar{y}_1 - 340x_2y_1y_2z\bar{x}_2 - 229x_2y_1y_3z\bar{x}_1.\end{aligned}$$

Hence we have $\mathrm{Im}\, H(\widetilde{s}) \subsetneq \mathrm{Ker}\, H(\widetilde{s})$, ie $(\wedge V, d)$ is not BV-exact. Note that we have found the element α by using the program [51] mentioned in Remark 6.4, while the equality is also checked by hand.

Finally we consider the differentials in the spectral sequence $\{{}_{(0)}E_r, d_r\}$ defined in Section 7. Since $d_1[\omega] = H(\widetilde{s})[\omega] = 0$, the cocycle ω defines an element $[\omega] \in {}_{(0)}E_2$, where ${}_{(0)}E_2$ is considered as a subquotient of ${}_{(0)}E_1 = H(\widetilde{\mathscr{L}}, \delta)$. Then the equality $\widetilde{s}(\omega) = \delta(\alpha)$ enables us to compute $d_2[\omega] = [\widetilde{s}\alpha] \neq 0 \in {}_{(0)}E_2$, where the nontriviality is proved by using the program [51]. Thus this Sullivan algebra gives an example such that $d_2 \neq 0$ on ${}_{(0)}E_2$. Note that it is currently unknown whether ${}_{(0)}E_3 = 0$ (ie 2–BV-exact) or not.

Appendix B Connes' B–map in the Gysin exact sequence

In this section, by giving precisely a rational model for the *integration over the fiber* $\beta\colon H^{*+1}(LX;\mathbb{Q}) \to H^*_{S^1}(LX;\mathbb{Q})$, we describe the Gysin exact sequence of the S^1–principal bundle

$$S^1 \to ES^1 \times LX \xrightarrow{p} ES^1 \times_{S^1} LX$$

in terms of Sullivan models for LX and $ES^1 \times_{S^1} LX$. As a consequence, the Gysin sequence is identified with Connes' exact sequence under the isomorphisms $\mathrm{HH}_*(\mathcal{M}(X)) \cong H^{-*}(LX;\mathbb{Q})$ and $\mathrm{HC}^-_*(\mathcal{M}(X)) \cong H^{-*}(ES^1 \times_{S^1} LX;\mathbb{Q})$ described in Section 3.

The cohomology Gysin sequence associated with the bundle has the form

$$\cdots \to H^{*-2}_{S^1}(LX) \xrightarrow{S} H^*_{S^1}(LX) \xrightarrow{p^*} H^*(LX) \xrightarrow{\beta} H^{*-1}_{S^1}(LX) \to \cdots,$$

in which S is defined by the cup product with the Euler class $q^*(u)$, where u is the generator of $H^2(BS^1)$. The S^1–principal bundle above fits in the pullback diagram

$$\begin{array}{ccccc}
 & & S^1 & = & S^1 \\
 & & \downarrow & & \downarrow \\
LX & \longrightarrow & ES^1 \times LX & \longrightarrow & ES^1 \\
\| & & \downarrow p & & \downarrow \\
LX & \longrightarrow & ES^1 \times_{S^1} LX & \xrightarrow{q} & BS^1
\end{array}$$

in which the lower sequence is the fiber bundle associated with the universal bundle $ES^1 \to BS^1$.

We recall the Sullivan models $\mathscr{L}$ and $\mathscr{E}$ defined in Section 3. In the model $\mathscr{E}$ for $ES^1 \times_{S^1} LX$, we write u for the Euler class $q^*(u)$. Thus the map S in the Gysin sequence is regarded as the multiplication by u in the models, namely the S–action in Connes' exact sequence. In order to obtain rational models for p and β, we here consider the relative Sullivan algebra $\mathscr{L}^\wedge := (\mathscr{E} \otimes \wedge(e), \widehat{\delta})$ with base $\mathscr{E}$, where $\widehat{\delta}(e) = u$ and $|e| = 1$.

Lemma B.1 *The canonical projection $\rho\colon \mathscr{L}^\wedge \to \mathscr{L}$ is a homotopy equivalence.*

Proof We define a DGA morphism $\iota\colon \mathscr{L} \to \mathscr{L}^\wedge$ by $\iota(\alpha) = \alpha + (-1)^{|\alpha|}s(\alpha)e$ for $\alpha \in \mathscr{L}$, where s is the derivation on $\mathscr{L}$ stated in Section 3. Then, we have $\rho \circ \iota = 1$ by definition. Moreover, a homotopy $\mathscr{L}^\wedge \to \mathscr{L}^\wedge \otimes \wedge(t, dt)$ defined by $e \mapsto et$, $u \mapsto ut - e\,dt$ and $\alpha \mapsto \alpha + (-1)^{|\alpha|}s(\alpha)e(1-t)$ for $\alpha \in \mathscr{E}$ implies that $\iota \circ \rho$ is homotopic to 1. This completes the proof. □

Proposition B.2 *The derivation $s\colon \mathscr{L} \to \mathscr{E}$ is a rational model for β.*

Proof From Lemma B.1, the inclusion $\mathscr{E} \hookrightarrow \mathscr{L}^\wedge$ is a rational model for the principal bundle

$$p\colon ES^1 \times LM \to ES^1 \times_{S^1} LM.$$

Since β is the fiber integration associated with the principal bundle, it is modeled by a map $\int_e\colon \mathscr{L}^\wedge \to \mathscr{E}$ defined by $\int_e(\alpha_0 + \alpha_1 e) = \alpha_1$ for $\alpha_i \in \mathscr{E}$. Therefore, the result follows since the composite $\int_e \iota$ coincides with the derivation s. □

List of symbols

symbol	meaning	page
$\bullet$	the loop product	2633
$\odot$	the dual loop coproduct	2633
$[\,,\,]$	the string bracket, dual string cobracket	2624
Δ	the BV operator on the Hochschild homology of a differential graded algebra	2626
Δ'	the BV operator on the homology of LM	2627
$\mathrm{HH}_*(\Omega)$	the Hochschild homology of a DGA Ω	2630
$\widetilde{\mathrm{HH}}_*(\Omega)$	the reduced Hochschild homology, $\mathrm{HH}_*(\Omega) \cong \widetilde{\mathrm{HH}}_*(\Omega) \oplus \mathbb{K}$	2630
$\mathrm{HC}^-_*(\Omega)$	the negative cyclic homology of a DGA Ω	2630
$\widetilde{\mathrm{HC}}^-_*(\Omega)$	the reduced negative cyclic homology, $\mathrm{HC}^-_*(\Omega) \cong \widetilde{\mathrm{HC}}^-_*(\Omega) \oplus \mathbb{K}[u]$	2630
S	the S–action on the negative cyclic homology	2630
$\mathscr{L}$, $(\mathscr{L}, \delta)$	the Sullivan minimal model for the free loop space LM (and the Hochschild homology)	2631
$\mathscr{E}$, $(\mathscr{E}, D)$	the Sullivan minimal model for the Borel construction $ES^1 \times_{S^1} LM$ (and the negative cyclic homology)	2631
$(\widetilde{\mathscr{L}}, \delta)$	the reduced version of $(\mathscr{L}, \delta)$	2637
$(\widetilde{\mathscr{L}}^{(n)}, \delta)$	a direct summand of $(\widetilde{\mathscr{L}}, \delta) = \bigoplus_n (\widetilde{\mathscr{L}}^{(n)}, \delta)$	2637
s	a derivation on $\mathscr{L}$, which is a chain model of Δ	2631

Acknowledgements The authors thank Jean-Claude Thomas and Luc Menichi for comments on the first draft of this manuscript. The authors are grateful to the referee for a very careful reading of the manuscript, valuable suggestions and comments.

Kuribayashi was partially supported by a Grant-in-Aid for Scientific Research (B) 21H00982 from Japan Society for the Promotion of Science. Naito was supported by JSPS KAKENHI grant JP18K13403. Wakatsuki was supported by JSPS KAKENHI grant 20J00404. Yamaguchi was partially supported by JSPS KAKENHI grant 20K03591.

References

[1] **M Arkowitz**, **G Lupton**, *Rational obstruction theory and rational homotopy sets*, Math. Z. 235 (2000) 525–539 MR Zbl

[2] **S Basu**, *Transversal string topology & invariants of manifolds*, PhD thesis, State University of New York at Stony Brook (2011) MR Available at `https://www.proquest.com/docview/898888965`

[3] **Y Berest**, **A C Ramadoss**, **Y Zhang**, *Hodge decomposition of string topology*, Forum Math. Sigma 9 (2021) art. id. e33 MR Zbl

[4] **R Body**, **R Douglas**, *Unique factorization of rational homotopy types*, Pacific J. Math. 90 (1980) 21–26 MR Zbl

[5] **R Body**, **M Mimura**, **H Shiga**, **D Sullivan**, *p–universal spaces and rational homotopy types*, Comment. Math. Helv. 73 (1998) 427–442 MR Zbl

[6] **D Burghelea**, **Z Fiedorowicz**, **W Gajda**, *Adams operations in Hochschild and cyclic homology of de Rham algebra and free loop spaces*, K–Theory 4 (1991) 269–287 MR Zbl

[7] **D Burghelea**, **M Vigué-Poirrier**, *Cyclic homology of commutative algebras, I*, from "Algebraic topology: rational homotopy" (Y Félix, editor), Lecture Notes in Math. 1318, Springer (1988) 51–72 MR Zbl

[8] **M Chas**, **D Sullivan**, *String topology*, preprint (1999) arXiv math/9911159

[9] **D Chataur**, **L Menichi**, *String topology of classifying spaces*, J. Reine Angew. Math. 669 (2012) 1–45 MR Zbl

[10] **X Chen**, *An algebraic chain model of string topology*, Trans. Amer. Math. Soc. 364 (2012) 2749–2781 MR Zbl

[11] **X Chen**, **F Eshmatov**, **L Liu**, *Gravity algebra structure on the negative cyclic homology of Calabi–Yau algebras*, J. Geom. Phys. 147 (2020) 103522, 19 MR Zbl

[12] **R L Cohen**, **J D S Jones**, **J Yan**, *The loop homology algebra of spheres and projective spaces*, from "Categorical decomposition techniques in algebraic topology" (G Arone, J Hubbuck, R Levi, M Weiss, editors), Progr. Math. 215, Birkhäuser, Basel (2004) 77–92 MR Zbl

[13] **C Costoya**, **A Viruel**, *Every finite group is the group of self-homotopy equivalences of an elliptic space*, Acta Math. 213 (2014) 49–62 MR Zbl

[14] **Y Félix**, **S Halperin**, **J-C Thomas**, *Gorenstein spaces*, Adv. in Math. 71 (1988) 92–112 MR Zbl

[15] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Math. 205, Springer (2001) MR Zbl

[16] **Y Félix**, **L Menichi**, **J-C Thomas**, *Gerstenhaber duality in Hochschild cohomology*, J. Pure Appl. Algebra 199 (2005) 43–59 MR Zbl

[17] **Y Félix**, **J-C Thomas**, *Rational BV–algebra in string topology*, Bull. Soc. Math. France 136 (2008) 311–327 MR Zbl

[18] **Y Félix**, **J-C Thomas**, *String topology on Gorenstein spaces*, Math. Ann. 345 (2009) 417–452 MR Zbl

[19] **Y Félix**, **J-C Thomas**, **M Vigué-Poirrier**, *Rational string topology*, J. Eur. Math. Soc. 9 (2007) 123–156 MR Zbl

[20] **B L Feĭgin**, **B L Tsygan**, *Additive K–theory*, from "K–theory, arithmetic and geometry" (Y I Manin, editor), Lecture Notes in Math. 1289, Springer (1987) 67–209 MR Zbl

[21] **E Getzler**, *Two-dimensional topological gravity and equivariant cohomology*, Comm. Math. Phys. 163 (1994) 473–489 MR Zbl

[22] **E Getzler**, **J D S Jones**, *A_∞–algebras and the cyclic bar complex*, Illinois J. Math. 34 (1990) 256–283 MR Zbl

[23] **E Getzler**, **J D S Jones**, **S Petrack**, *Differential forms on loop spaces and the cyclic bar complex*, Topology 30 (1991) 339–371 MR Zbl

[24] **S Halperin**, **J Stasheff**, *Obstructions to homotopy equivalences*, Adv. in Math. 32 (1979) 233–279 MR Zbl

[25] **R A Hepworth**, *String topology for Lie groups*, J. Topol. 3 (2010) 424–442 MR Zbl

[26] **J D S Jones**, *Cyclic homology and equivariant homology*, Invent. Math. 87 (1987) 403–423 MR Zbl

[27] **K Kuribayashi**, **L Menichi**, *The Batalin–Vilkovisky algebra in the string topology of classifying spaces*, Canad. J. Math. 71 (2019) 843–889 MR Zbl

[28] **K Kuribayashi**, **L Menichi**, **T Naito**, *Derived string topology and the Eilenberg–Moore spectral sequence*, Israel J. Math. 209 (2015) 745–802 MR Zbl

[29] **K Kuribayashi**, **T Yamaguchi**, *On additive K–theory with the Loday–Quillen *–product*, Math. Scand. 87 (2000) 5–21 MR Zbl

[30] **J-L Loday**, *Cyclic homology*, Grundl. Math. Wissen. 301, Springer (1992) MR Zbl

[31] **J-L Loday**, **D Quillen**, *Cyclic homology and the Lie algebra homology of matrices*, Comment. Math. Helv. 59 (1984) 569–591 MR Zbl

[32] **J McCleary**, *A user's guide to spectral sequences*, 2nd edition, Cambridge Studies in Advanced Mathematics 58, Cambridge Univ. Press (2001) MR Zbl

[33] **L Menichi**, *Batalin–Vilkovisky algebra structures on Hochschild cohomology*, Bull. Soc. Math. France 137 (2009) 277–295 MR Zbl

[34] **L Menichi**, *String topology for spheres*, Comment. Math. Helv. 84 (2009) 135–157 MR Zbl

[35] **L Menichi**, *A Batalin–Vilkovisky algebra morphism from double loop spaces to free loops*, Trans. Amer. Math. Soc. 363 (2011) 4443–4462 MR Zbl

[36] **M Mimura**, **R C O'Neill**, **H Toda**, *On p–equivalence in the sense of Serre*, Jpn. J. Math. 40 (1971) 1–10 MR Zbl

[37] **M Mimura**, **H Toda**, *On p–equivalences and p–universal spaces*, Comment. Math. Helv. 46 (1971) 87–97 MR Zbl

[38] **J W Morgan**, *The algebraic topology of smooth algebraic varieties*, Inst. Hautes Études Sci. Publ. Math. 48 (1978) 137–204 MR Zbl

[39] **T Naito**, *String operations on rational Gorenstein spaces*, preprint (2013) arXiv 1301.1785

[40] **H Scheerer**, *Fibrewise P–universal nilpotent fibrations*, Proc. Roy. Soc. Edinburgh Sect. A 98 (1984) 89–104 MR Zbl

[41] **J-P Serre**, *Groupes d'homotopie et classes de groupes abéliens*, Ann. of Math. 58 (1953) 258–294 MR Zbl

[42] **L Smith**, *Homological algebra and the Eilenberg–Moore spectral sequence*, Trans. Amer. Math. Soc. 129 (1967) 58–93 MR Zbl

[43] **F Y Tabing**, *String homology and Lie algebra structures*, PhD thesis, University of California, Santa Cruz (2015) MR arXiv 1610.03933

[44] **H Tamanoi**, *Cap products in string topology*, Algebr. Geom. Topol. 9 (2009) 1201–1224 MR Zbl

[45] **M Vigué-Poirrier**, *Sur l'algèbre de cohomologie cyclique d'un espace* 1*–connexe applications à la géometrie des variétés*, Illinois J. Math. 32 (1988) 40–52 MR Zbl

[46] **M Vigué-Poirrier**, *Homologie cyclique des espaces formels*, J. Pure Appl. Algebra 91 (1994) 347–354 MR Zbl

[47] **M Vigué-Poirrier**, **D Burghelea**, *A model for cyclic homology and algebraic K–theory of* 1*–connected topological spaces*, J. Differential Geom. 22 (1985) 243–253 MR Zbl

[48] **M Vigué-Poirrier**, **D Sullivan**, *The homology theory of the closed geodesic problem*, J. Differential Geometry 11 (1976) 633–644 MR Zbl

[49] **S Wakatsuki**, *Description and triviality of the loop products and coproducts for rational Gorenstein spaces*, preprint (2016) arXiv 1612.03563

[50] **S Wakatsuki**, *Coproducts in brane topology*, Algebr. Geom. Topol. 19 (2019) 2961–2988 MR Zbl

[51] **S Wakatsuki**, *Kohomology*: *cohomology calulator for Sullivan algebras* (2024) Available at `https://github.com/shwaka/kohomology/index.html`

[52] **G W Whitehead**, *Elements of homotopy theory*, Graduate Texts in Math. 61, Springer (1978) MR Zbl

[53] **T Yang**, *A Batalin–Vilkovisky algebra structure on the Hochschild cohomology of truncated polynomials*, Topology Appl. 160 (2013) 1633–1651 MR Zbl

Department of Mathematical Sciences, Faculty of Science, Shinshu University
Matsumoto, Nagano, Japan

Nippon Institute of Technology
Gakuendai, Miyashiro-machi, Minamisaitama-gun, Saitama, Japan

Department of Mathematical Sciences, Faculty of Science, Shinshu University
Matsumoto, Nagano, Japan

Faculty of Education, Kochi University
Akebono-cho, Kochi, Japan

kuri@math.shinshu-u.ac.jp, naito.takahito@nit.ac.jp,
shun.wakatsuki@math.nagoya-u.ac.jp, tyamag@kochi-u.ac.jp

Received: 12 May 2022 Revised: 11 May 2023

Algebraic & Geometric Topology **24**:5 (2024) 2655–2672
DOI: 10.2140/agt.2024.24.2655
Published: 19 August 2024

Asymptotic dimensions of the arc graphs and disk graphs

KOJI FUJIWARA
SAUL SCHLEIMER

We give quadratic upper bounds for the asymptotic dimensions of the arc graphs and disk graphs.

20F65, 20F69, 57K20; 57M60

1 Introduction

The asymptotic dimension, denoted by $\mathsf{dim}_{\mathsf{asym}} X$, of a metric space X was introduced by Gromov [1993, page 29] as a large-scale analogue of the covering dimension. The curve graph, $\mathscr{C}(S)$, for a surface $S = S_{g,b}$ was introduced by Harvey [1981, page 246] as a sort of Bruhat–Tits building for Teichmüller space. It has since been generalised in many ways. Bell and Fujiwara [2008, Corollary 1] first proved that the asymptotic dimension of $\mathscr{C}(S)$ is finite. More recently, Bestvina and Bromberg [2019, Corollary 1.1] proved that $\mathsf{dim}_{\mathsf{asym}} \mathscr{C}(S_g) \le 4g - 4$ (when $g > 1$) and that $\mathsf{dim}_{\mathsf{asym}} \mathscr{C}(S_{g,b}) \le 4g - 3 + b = \xi'(S_{g,b})$ when $g > 0$ and $b > 0$ (or $g = 0$ and $b > 2$).

Here we combine the machineries of [Bestvina et al. 2015; Masur and Schleimer 2013] to produce a quasi-isometric embedding of the arc graph $\mathscr{A}(S, \Delta)$ into a finite product of quasitrees of curve complexes. From this we deduce the following:

Corollary 3.11 *Suppose that $S = S_{g,b}$ has nonempty boundary. Suppose that $\Delta \subset \partial S$ is a nonempty union of components. Finally, suppose that $\xi'(S) \ge 1$. Then*

$$\mathsf{dim}_{\mathsf{asym}} \mathscr{A}(S, \Delta) \le \tfrac{1}{2}(4g + b)(4g + b - 3) - 2.$$

Sisto (private communication, 2022) suggests that the machineries of [Behrstock et al. 2017, Theorem 5.2; Vokes 2022, Theorem A.2] can be combined to obtain a similar result.

We also obtain the following result for the disk graph $\mathscr{D}(M, S)$ of a compression body:

Corollary 4.18 *Suppose M is a nontrivial spotless compression body with upper boundary $S = S_{g,b}$. Suppose that $\xi'(S) \ge 1$. Then*

$$\mathsf{dim}_{\mathsf{asym}} \mathscr{D}(M, S) \le \tfrac{1}{2}(4g + b)(4g + b - 3) - 2.$$

Hamenstädt [2019, Theorem 3.6] has obtained a similar result when M is a handlebody; see Remark 4.20.

Obtaining lower bounds better than linear, if even possible, would require new ideas. Therefore, we end our introduction with the following:

Question 1.1 How tight are the upper bounds of Corollaries 3.11 and 4.18? ◁

2 Background

Suppose that (X, d_X) is a metric space. Suppose that U and V are nonempty subsets of X with bounded diameter. Then we define their distance as

$$d_X(U, V) = \mathrm{diam}_X(U \cup V).$$

This notation is taken from [Masur and Minsky 2000, Formula (2.1), page 916].

We write $p =_{\mathsf{C}} q$ if, for nonnegative numbers p, q and C, we have both $q \le \mathsf{C}p + \mathsf{C}$ and $p \le \mathsf{C}q + \mathsf{C}$. Also, we use the *cut-off* function: $[p]_{\mathsf{C}}$ is equal to p if $p \ge \mathsf{C}$ and is zero otherwise.

We now more-or-less follow the conventions of [Bell and Dranishnikov 2008]. Suppose that X and Y are metric spaces. A relation $f : X \to Y$ is a *coarse map* if there is a constant C such that, for all $x \in X$, the image $f(x)$ is nonempty and has $\mathrm{diam}_Y(f(x)) \le \mathsf{C}$. A coarse map $f : X \to Y$ is a *coarse embedding* if there are functions $F, G : \mathbb{R} \to \mathbb{R}$ such that

- $\lim_{t\to\infty} F(t) = \lim_{t\to\infty} G(t) = \infty$, and
- for all $x, y \in X$, we have

$$F(d_X(x, y)) \le d_Y(f(x), f(y)) \le G(d_X(x, y)).$$

A coarse map $f : X \to Y$ is *coarsely onto* if there is a constant $\mathsf{C} > 0$ such that, for all $y \in Y$, there is a point $x \in X$ with $d_Y(f(x), y) < \mathsf{C}$.

A coarse map $f : X \to Y$ is *coarsely Lipschitz* if there is a constant $\mathsf{C} > 0$ such that, for all $x, y \in X$, we have

$$d_Y(f(x), f(y)) \le \mathsf{C} \cdot d_X(x, y) + \mathsf{C}.$$

That is, we have no lower bound, but we require the upper bound to be affine.

A coarse embedding f is a *quasi-isometric embedding* if the functions F and G are both affine (with positive coefficients). The more usual definition is to require a constant $\mathsf{C} > 0$ such that $d_X(x, y) =_{\mathsf{C}} d_Y(f(x), f(y))$ for all $x, y \in X$. A quasi-isometric embedding f is a *quasi-isometry* if f is also coarsely onto.

2.1 Asymptotic dimension

We now follow [Gromov 1993, Section 1.E]. A metric space X has *asymptotic dimension* $\mathsf{dim}_{\mathsf{asym}}(X)$ at most n if, for every $R > 0$, there is a $\mathsf{D} > 0$ and a cover $\mathcal{U}$ of X such that

- for all $U \in \mathcal{U}$, we have $\mathrm{diam}_X(U) \le \mathsf{D}$, and
- every metric R–ball in X intersects at most $n + 1$ sets in $\mathcal{U}$.

For example, trees have asymptotic dimension at most one. More generally, following [Bell and Dranishnikov 2008, page 1272], a collection of metric spaces $\mathscr{X}$ has asymptotic dimension at most n, *uniformly*, if for every R there is a constant D such that every $X \in \mathscr{X}$ has a cover as above.

If two metric spaces X and Y are quasi-isometric and $\mathsf{dim}_{\mathsf{asym}}(X) \le n$, then $\mathsf{dim}_{\mathsf{asym}}(Y) \le n$. In view of this, we say a finitely generated group G has asymptotic dimension at most n if some (and thus all) of its Cayley graphs have asymptotic dimension at most n.

We list two well-known facts about asymptotic dimension.

Fact 2.2 [Bell and Dranishnikov 2008, Theorem 32] *Suppose that U and V are metric spaces. We give $U \times V$ the ℓ^1 product metric. Then*

$$\mathsf{dim}_{\mathsf{asym}}\, U \times V \le \mathsf{dim}_{\mathsf{asym}}\, U + \mathsf{dim}_{\mathsf{asym}}\, V. \qquad \square$$

Fact 2.3 [Bell and Dranishnikov 2008, Proof of Proposition 22] *If U coarsely embeds into V, then*

$$\mathsf{dim}_{\mathsf{asym}}\, U \le \mathsf{dim}_{\mathsf{asym}}\, V. \qquad \square$$

2.4 Quasitrees of metric graphs

We quickly review the machinery of *quasitrees of metric spaces*, as introduced in [Bestvina et al. 2015, Section 4]. Suppose that $\mathscr{F}$ is a collection of metric graphs. Suppose also that we have, for every pair of distinct graphs $A, B \in \mathscr{F}$, a nonempty subset $\pi_B(A) \subset B$. Also, fix a sufficiently large constant $\mathsf{k} > 0$.

With respect to the data $(\mathscr{F}, \pi, \mathsf{k})$ we require the following axioms:

Axiom 2.5 (bounded projections) *For distinct $A, B \in \mathscr{F}$, we have*

$$\mathrm{diam}_B(\pi_B(A)) \le \mathsf{k}. \qquad \square$$

For $A, B, C \in \mathscr{F}$, and if $A \ne B$ and $B \ne C$, we adopt the shorthand

$$d_B(A, C) = d_B(\pi_B(A), \pi_B(C)).$$

Axiom 2.6 (Behrstock inequality) *For distinct $A, B, C \in \mathscr{F}$, at most one of the following is greater than k:*

$$d_A(B, C), \quad d_B(A, C), \quad d_C(A, B). \qquad \square$$

Axiom 2.7 (large links) *For distinct $A, C \in \mathscr{F}$, the following set is finite:*

$$\{B \in \mathscr{F} \mid A \ne B,\ B \ne C \text{ and } d_B(A, C) > \mathsf{k}\}. \qquad \square$$

We call these the *BBF axioms*. These are called (P0), (P1) and (P2) in [Bestvina et al. 2015].

Suppose that the data $(\mathscr{F}, \pi, \mathsf{k})$ satisfies the BBF axioms. Then, by [Bestvina et al. 2015, Theorem A], for every sufficiently large $\mathsf{K} \ge \mathsf{k}$ there is a metric graph $\mathscr{C}(\mathscr{F}) = \mathscr{C}_{\mathsf{K}}(\mathscr{F})$, called the *quasitree of graphs*. We denote the metric on $\mathscr{C}(\mathscr{F})$ by $d_{\mathscr{C}}$. We now list several properties of $\mathscr{C}(\mathscr{F})$.

Fact 2.8 [Bestvina et al. 2015, Theorem A and Definition 3.6] *By construction, every $A \in \mathcal{F}$ isometrically embeds into $\mathcal{C}(\mathcal{F})$. The image is totally geodesic. For distinct $A, B \in \mathcal{F}$, their images in $\mathcal{C}(\mathcal{F})$ are disjoint. Finally, these images cover the vertices of $\mathcal{C}(\mathcal{F})$.* □

Thus, we may identify the vertices of A with their images in $\mathcal{C}(\mathcal{F})$. We extend the definition of π_B as follows. If $b \in B$ then we take $\pi_B(b) = b$. If $a \in A$, and A is distinct from B, then we take $\pi_B(a) = \pi_B(A)$. One may think of $\pi_B \colon \mathcal{C}(\mathcal{F}) \to B$ as being a "closest-points projection" map. We adopt the shorthand

$$d_B(a, c) = d_B(\pi_B(a), \pi_B(c)).$$

Fact 2.9 [Bestvina et al. 2015, Definition 4.1] *By construction, there is a constant L (depending only on K) with the following property. Suppose that $A, B \in \mathcal{F}$ are graphs such that the set of Axiom 2.7 is empty. Then $d_{\mathcal{C}}(\pi_A(B), \pi_B(A)) \le \mathsf{L}$.* □

We have the following *distance estimate* for $d_{\mathcal{C}}$:

Theorem 2.10 [Bestvina et al. 2015, Theorem 4.13] *Suppose that $\mathcal{F}$ is a family of metric graphs satisfying the BBF axioms. Suppose that K is sufficiently large. Suppose that $\mathcal{C}_{\mathsf{K}}(\mathcal{F})$ is the quasitree of graphs. Then every K' sufficiently larger than K has the following property: for any $a, c \in \mathcal{C}(\mathcal{F})$, we have*

$$\frac{1}{2} \sum [d_B(a, c)]_{\mathsf{K}'} \le d_{\mathcal{C}(\mathcal{F})}(a, c) \le 6\mathsf{K} + 4 \sum [d_B(a, c)]_{\mathsf{K}},$$

where both sums are taken over all $B \in \mathcal{F}$. □

Suppose that G is a group acting on $\mathcal{F}$. We further assume that, for any $g \in G$ and for any $A \in \mathcal{F}$, there is an isometry $g_A \colon A \to g \cdot A$. We suppose that these isometries have the following consistency properties. Suppose that $g, h \in G$ are group elements and $A, B \in \mathcal{F}$ are graphs with $B = g \cdot A$.

- For all $a \in A$, we have $h_B(g_A(a)) = (hg)_A(a)$.
- For any $C \in \mathcal{F}$, we have $g_A(\pi_A(C)) = \pi_B(g \cdot C)$.

From [Bestvina et al. 2015, Section 3.7], we deduce the following: there is an isometric action of G on the quasitree of graphs $\mathcal{C}(\mathcal{F})$ which extends the action of the isometries g_A.

We also have the following control on the asymptotic dimension of $\mathcal{C}(\mathcal{F})$:

Theorem 2.11 [Bestvina et al. 2015, Theorem 4.24] *Suppose that $\mathcal{F}$ is a family of metric graphs satisfying the BBF axioms. Suppose that $\mathcal{F}$ has asymptotic dimension at most D, uniformly. Then $\mathsf{dim}_{\mathsf{asym}}(\mathcal{C}(\mathcal{F})) \le D + 1$.* □

When all of the metric graphs are quasitrees, this can be improved:

Theorem 2.12 [Bestvina et al. 2015, Theorem B(ii)] *Suppose that $\mathcal{F}$ is a family of quasitrees satisfying the BBF axioms. Suppose further that the quasi-isometry constants are uniformly bounded. Then $\mathsf{dim}_{\mathsf{asym}}(\mathcal{C}(\mathcal{F})) \le 1$.* □

The notion of a *quasitree of metric spaces* was introduced and used to prove [Bestvina et al. 2015, Theorem D]: mapping class groups (of connected, compact, oriented surfaces) have finite asymptotic dimension.

2.13 Surfaces, curves and arcs

Let $S = S_{g,b}$ denote the connected, compact, oriented surface of genus g with b boundary components. The *complexity* of S is defined to be $\xi(S) = 3g - 3 + b$. This counts the number of curves in any pants decomposition of S. We will always assume that $\xi(S) \geq 1$. We will also need the *modified complexity* $\xi'(S) = 4g - 3 + b$. If S is closed then we will simply write S_g for $S_{g,0}$.

Suppose that α is an embedded arc or curve in S. We call the embedding *proper* if $\alpha \cap \partial S = \partial \alpha$. A properly embedded arc or curve α in S is *essential* if it does not cut a disk off of S. A properly embedded curve α is *nonperipheral* if it does not cut an annulus off of S.

A *proper isotopy* is an isotopy through proper embeddings. Let $[\alpha]$ denote the proper isotopy class of α. Given α and β, properly embedded arcs or curves, we define their *geometric intersection number*

$$i(\alpha, \beta) = \min\{|\alpha' \cap \beta'| : \alpha' \in [\alpha], \beta' \in [\beta]\}.$$

Note that $i(\alpha, \beta) = 0$ if and only if they have disjoint (proper isotopy) representatives. To lighten the notation, we typically will not distinguish between a curve (or arc) α and its proper isotopy class $[\alpha]$.

A connected, compact subsurface $X \subset S$ is *essential* if every component of ∂X is either a component of ∂S or is essential and nonperipheral in S. If X is essential, we define the *relative boundary* of X to be $\partial_S X = \partial X - \partial S$.

Remark 2.14 If $X \subset Y$ are both essential subsurfaces of S, then $\xi'(X) \leq \xi'(Y)$. Equality holds if and only if X and Y are isotopic. ◁

We say that a properly embedded curve or an arc α *cuts* X if every $\alpha' \in [\alpha]$ intersects X. If α does not cut X, then we say that α *misses* X. Suppose that X and Y are essential, and nonisotopic, subsurfaces of S. We say that X is *nested* in Y if it is (perhaps after an isotopy) contained in Y. We say that X and Y *overlap* if $\partial_S X$ cuts Y and $\partial_S Y$ cuts X.

2.15 Curve and arc graphs

We now define the *curve graph* $\mathscr{C}(S)$. Let $\mathscr{C}^{(0)}(S)$ be the set of proper isotopy classes of essential, nonperipheral curves in S. We have an edge $(\alpha, \beta) \in \mathscr{C}^{(1)}$ exactly when α and β are distinct and $i(\alpha, \beta) = 0$.

We define the *arc graph* $\mathscr{A}(S)$ similarly: $\mathscr{A}^{(0)}(S)$ is the set of proper isotopy classes of essential arcs in S. Again we have an edge $(\alpha, \beta) \in \mathscr{A}^{(1)}$ exactly when α and β are distinct and $i(\alpha, \beta) = 0$. Note that $\mathscr{A}(S)$ is empty when S is closed.

Masur and Schleimer [2013, Definition 7.1] generalise the definition of the arc graph slightly, as follows. Suppose that $\Delta \subset \partial S$ is a nonempty collection of boundary components. We define $\mathscr{A}(S, \Delta)$ to be the subgraph of $\mathscr{A}(S)$ spanned by the arcs having both endpoints in Δ. Note that $\mathscr{A}(S, \partial S) = \mathscr{A}(S)$.

We next define the *arc and curve graph* $\mathscr{AC}(S)$: the zero skeleton is exactly $\mathscr{A}^{(0)}(S) \cup \mathscr{C}^{(0)}(S)$. Edges come from having disjoint representatives, as before. Note that the inclusion of $\mathscr{C}^{(0)}(S)$ into $\mathscr{AC}^{(0)}(S)$ induces a quasi-isometry of graphs.

The definition of the curve complex must be modified when $\xi(S) \le 1$: for $S_{1,1}$ we use $i(\alpha, \beta) = 1$ and for $S_{0,4}$ we use $i(\alpha, \beta) = 2$. For both of these surfaces the graph of curves is a copy of the *Farey graph*. When S is an annulus we define $\mathscr{C}^{(0)}(S)$ to be the set of proper isotopy classes of essential properly embedded arcs, where now isotopies are required to fix boundary points. Two classes span an edge if they have representatives which are disjoint on their interiors.

All of the various curve, arc, and arc and curve graphs are connected when they are nonempty [Masur and Minsky 1999, Lemma 2.1]. We make each of these into a metric graph by decreeing that all edges have length one. It is then a theorem of Masur and Minsky [1999, Theorem 1.1] that, for any surface S with $\xi(S) \ge 1$, the curve complex $\mathscr{C}(S)$ is Gromov hyperbolic. Masur and Schleimer [2013, Theorem 20.3] proved that the same holds for $\mathscr{A}(S, \Delta)$.

2.16 I–bundles

Suppose that F is a connected, compact surface, possibly nonorientable, with nonempty boundary. Let $\rho: T \to F$ be an I–bundle. We call F the *base surface* of the bundle. We define $\partial_v T = \rho^{-1}(\partial F)$ to be the *vertical boundary* of T. We define the closure

$$\partial_h T = \overline{\partial T - \partial_v T}$$

to be the *horizontal boundary* of T. Also, we define the curves

$$\partial(\partial_h T) = \partial(\partial_v T)$$

to be the *corners* of T. Finally, there is an involution $\tau: \partial_h T \to \partial_h T$ associated to ρ obtained by swapping the ends of interval fibres.

We now define $\rho_F: T_F \to F$ to be the *orientation* I–bundle over F. Here the preimage under ρ_F of a simple closed curve α is an annulus or a Möbius band as α is or is not, respectively, orientation-preserving in F. When F is nonorientable, we call T_F *twisted* and so $\partial_h T_F / \tau \cong F$ is nonorientable. If T_F is not twisted then $T_F \cong F \times [-1, 1]$ is a product. In this case, $\tau|_{F \times \{-1\}}$ is a homeomorphism from $F \times \{-1\}$ to $F \times \{1\}$.

2.17 Compression bodies

References on compression bodies, of the type we are interested in here, include [Bonahon 1983, Appendix B; Oertel 2002, Section 1].

Suppose that $S = S_{g,b}$ is a surface. We assume that S is neither a disk nor a sphere. We form $T = S \times I$. We take $\partial^+ T = S \times \{1\}$ and $\partial^- T = S \times \{0\}$ to be the *upper* and *lower* boundaries of T. As before, $\partial_v T = \partial S \times I$ is the *vertical boundary*. We now attach a collection of three-dimensional two- and three-handles to the lower boundary of T to obtain a three-manifold M. We define $\partial^+ M = \partial^+ T$ as well as $\partial_v M = \partial_v T$. Finally, we define

$$\partial^- M = \overline{\partial M - (\partial^+ M \cup \partial_v M)}.$$

Thus, M is a *compression body*. If $\partial^- M$ is not homeomorphic to $\partial^+ M$, then M is *nontrivial*. If $\partial^- M$ has no sphere or disk components, then M is *spotless*. To simplify the notation, we take $S = \partial^+ M$. Note that, if $\partial^- M$ is empty, then M is necessarily a handlebody of positive genus.

We now state the classification of compression bodies.

Theorem 2.18 *Suppose that M and N are compression bodies. Then $(M, \partial_v M)$ is homeomorphic to $(N, \partial_v N)$ if and only if $(\partial^+ M, \partial^- M)$ is homeomorphic to $(\partial^+ N, \partial^- N)$.* □

The proof is similar to that of the classification of surfaces [Farb and Margalit 2012, Theorem 1.1] and of handlebodies [Hempel 1976, Theorem 2.2]. The case of $\partial_v M = \varnothing$ is discussed by Biringer and Vlamis [2017, Corollary 2.3].

2.19 Disk graphs

Suppose that (M, S) is a nontrivial, spotless compression body. Suppose that $(D, \partial D) \subset (M, S)$ is a properly embedded disk. We call D *essential* if ∂D is essential in S. We now define the *disk graph* $\mathcal{D}(M, S)$. The vertices of $\mathcal{D}(M, S)$ are proper isotopy classes of essential disks in (M, S). A pair of distinct vertices D and E give an edge $(D, E) \in \mathcal{D}(M, S)$ if they have disjoint representatives.

2.20 Subsurface projection

We give one of the standard definitions of subsurface projection:

Definition 2.21 [Masur and Schleimer 2013, Definition 4.4] Suppose that X is an essential subsurface, but not a pair of pants, in S. The relation of *subsurface projection*, $\pi_X : \mathcal{AC}(S) \to \mathcal{C}(X)$, is defined as follows. Let $\rho_X : S^X \to S$ be the covering map corresponding to the subgroup $\pi_1(X) < \pi_1(S)$. Note that the Gromov compactification of S^X is homeomorphic to X. This gives an identification of the graphs $\mathcal{C}(S^X)$ and $\mathcal{C}(X)$. For any $\alpha \in \mathcal{AC}(S)$, we define $\alpha^X = \rho_X^{-1}(\alpha)$ to be the full preimage. We define $\alpha|_X$ to be the essential arcs, and essential nonperipheral curves, of α^X. If X is an annulus, then we set $\pi_X(\alpha) = \alpha|_X$. Otherwise, for every $\beta \in \alpha|_X$, we form $N = N(\beta \cup \partial X)$ and we place the essential isotopy classes of $\partial_X N$ into $\pi_X(\alpha)$. ◁

Note that, if α misses X, then $\pi_X(\alpha)$ is empty. Suppose instead that α cuts X. If X is an annulus, then the diameter of $\pi_X(\alpha)$ is at most one. If X is not an annulus, then the diameter of $\pi_X(\alpha)$ is at most two [Masur and Minsky 2000, Lemma 2.3]. If X and Y are essential subsurfaces of S, then we define

$\pi_Y(X) = \pi_Y(\partial_S X)$. If X and Y are disjoint, or if Y is nested in X, then this is empty. We record the following for later use:

Lemma 2.22 *Suppose that X and Y are overlapping essential subsurfaces of S. Then $\pi_Y(X)$ is nonempty and has diameter at most two.* □

We will adopt the following useful shorthand notation. Suppose that α and β are curves or arcs, both cutting an essential subsurface $X \subset S$. Then

$$d_X(\alpha, \beta) = d_{\mathscr{C}(X)}(\pi_X(\alpha), \pi_X(\beta))$$

is the *subsurface projection distance* between α and β in X.

3 Bound for the arc graph

Let $S = S_{g,b}$, where $\xi'(S) > 0$ and $b > 0$. Take $\Delta \subset \partial S$ to be a nonempty union of components. Let $\mathscr{A}(S, \Delta)$ be the graph of essential arcs with endpoints in Δ.

3.1 Witnesses for the arc graph

Definition 3.2 An essential subsurface $X \subset S$ is a *witness* for $\mathscr{A}(S, \Delta)$ if every arc $\alpha \in \mathscr{A}(S, \Delta)$ cuts X. ◁

We repackage a few results [Masur and Schleimer 2013, Lemmas 5.9 and 7.2]:

Lemma 3.3 *Suppose that $X \subset S$ is an essential subsurface, but not an annulus or a pair of pants. The following are equivalent*:

- *X is a witness for $\mathscr{A}(S, \Delta)$.*
- *X contains Δ.*
- *For all arcs $\alpha \in \mathscr{A}(S, \Delta)$, the projection $\pi_X(\alpha)$ is nonempty.*
- *The projection $\pi_X : \mathscr{A}(S, \Delta) \to \mathscr{C}(X)$ is coarsely Lipschitz with a constant of 2.* □

We have a useful corollary:

Corollary 3.4 *Suppose that X and Y are distinct witnesses with $\xi'(X) = \xi'(Y)$. Then X and Y overlap.*

Proof By Lemma 3.3, both X and Y contain Δ; thus, they intersect. Since they have the same modified complexity, by Remark 2.14 they cannot be nested. Thus, they overlap. □

We let $\mathrm{MCG}(S, \Delta)$ be the mapping class group for the pair (S, Δ): the group of mapping classes that preserve Δ setwise. We say that a pair of arcs $\alpha, \beta \in \mathscr{A}(S, \Delta)$ have the same *topological type* (or, more simply, the same *type*) if there is a mapping class $f \in \mathrm{MCG}(S, \Delta)$ such that $f(\alpha) = \beta$.

Lemma 3.5 *The quotient* $\mathcal{A}(S,\Delta)/\mathrm{MCG}(S,\Delta)$ *has diameter at most two.*

Proof Suppose that $\alpha \in \mathcal{A}(S,\Delta)$ is an arc. We break the proof into cases, depending on the number of components of ∂S and of Δ.

Suppose that Δ has at least two components. Then α is disjoint from some $\gamma \in \mathcal{A}(S,\Delta)$ meeting two components of Δ. Since we made no assumptions on α, we find that, in the quotient, all vertices are at most distance one from $[\gamma]$. Thus, the diameter is at most two.

Suppose that Δ has only one component, but ∂S has at least two. Let δ be some component of $\partial S - \Delta$. Then α is disjoint from some arc $\gamma \in \mathcal{A}(S,\Delta)$ such that γ separates δ from the rest of S. We again obtain a diameter bound of two.

Suppose that ∂S has a single component, which necessarily equals Δ. Then α is disjoint from some arc $\gamma \in \mathcal{A}(S,\Delta)$ such that γ is nonseparating. This gives the diameter bound and finishes the proof. □

3.6 Families of witnesses

Fix a number $c \le \xi'(S)$. The collection

$$\mathcal{F}_c = \{X \subset S \mid X \text{ is a witness for } \mathcal{A}(S,\Delta) \text{ and } \xi'(X) = c\}$$

is called a *family*. We only consider nonempty families.

Suppose that $X, Y, Z \in \mathcal{F}_c$ are witnesses, with Y distinct from both X and Z. Then we define

$$d_Y(X,Z) = d_{\mathcal{C}(Y)}(\pi_Y(X), \pi_Y(Z)).$$

This is well defined by Corollary 3.4. We note that there is an abuse of notation here: the family $\mathcal{F}_c$ consists of witnesses X — that is, surfaces — not metric graphs. However, each witness X gives a metric graph, namely $\mathcal{C}(X)$. We trust this will not cause confusion.

Lemma 3.7 *For every* c, *the family* $\mathcal{F}_c$ *satisfies the three BBF axioms given in Section 2.4. Since there are only finitely many of these families, there is a common constant* k *that works for all of them simultaneously.*

Proof Axiom 2.5 follows from Corollary 3.4 and Lemma 2.22.

Axiom 2.6 follows from Corollary 3.4 and the usual Behrstock inequality [2006, Theorem 4.3]. See [Mangahas 2010, Lemma 2.5] for an elementary proof following ideas of Leininger.

Axiom 2.7 appears as [Masur and Minsky 2000, Lemma 6.2]. See [Bestvina et al. 2015, Lemma 5.3] for a proof giving a concrete bound and avoiding the machinery of hierarchies. □

We now apply the BBF construction, outlined in Section 2.4, to each family $\mathscr{F}_c$. This gives us a quasitree of curve graphs $\mathscr{C}(\mathscr{F}_c)$. We deduce that $\mathscr{C}(\mathscr{F}_c)$ is a hyperbolic metric graph where each of the curve complexes $\mathscr{C}(X)$ for $X \in \mathscr{F}_c$ embeds as a totally geodesic subgraph. From [Bestvina and Bromberg 2019, Corollary 1.1] and Theorem 2.11, we deduce that

$$\mathsf{dim}_{\mathsf{asym}}\, \mathscr{C}(\mathscr{F}_c) \le c + 1.$$

If $c = 1$ then Theorem 2.12 allows us to sharpen the bound:

$$\mathsf{dim}_{\mathsf{asym}}\, \mathscr{C}(\mathscr{F}_1) \le 1.$$

On the other hand, if $c = 4g + b - 3$ then $\mathscr{F}_c = \{S\}$ and we have

$$\mathsf{dim}_{\mathsf{asym}}\, \mathscr{C}(\mathscr{F}_c) \le 4g + b - 3.$$

We now define $\mathscr{P}(S, \Delta)$ to be the product, equipped with the ℓ^1 metric, of the quasitrees of curve graphs $\mathscr{C}(\mathscr{F}_c)$ as c ranges from one to $\xi'(S)$. From the above and from Fact 2.2, we deduce the following:

Corollary 3.8 $$\mathsf{dim}_{\mathsf{asym}}\, \mathscr{P}(S, \Delta) \le \tfrac{1}{2}(4g + b)(4g + b - 3) - 2.$$ □

3.9 Embedding the arc graph

In this section we fix the constants k, K and L. We then state and prove Theorem 3.10.

The constant k is the larger of 13 (as explained in the proof of Lemma 3.13) and the constant given by Lemma 3.7. The constant K is now given by [Bestvina et al. 2015, Theorem A]. Finally, the constant L is provided by Fact 2.9. We take $\mathscr{P}(S, \Delta)$ to be the product of the resulting quasitrees. Here is the statement:

Theorem 3.10 *There is a quasi-isometric embedding ϕ of the arc graph $\mathscr{A}(S, \Delta)$ into the product $\mathscr{P}(S, \Delta)$ of quasitrees of curve graphs. Moreover, ϕ is equivariant with respect to the action of the mapping class group* $\mathrm{MCG}(S, \Delta)$.

From this, and from Fact 2.3, we deduce the following:

Corollary 3.11 *Suppose that $S = S_{g,b}$ has nonempty boundary. Suppose that $\Delta \subset \partial S$ is a nonempty union of components. Finally, suppose that $\xi'(S) \ge 1$. Then*

$$\mathsf{dim}_{\mathsf{asym}}\, \mathscr{A}(S, \Delta) \le \tfrac{1}{2}(4g + b)(4g + b - 3) - 2.$$ □

We now turn to the proof of Theorem 3.10. Fix a modified complexity c.

Definition 3.12 Suppose that $\beta \in \mathscr{A}(S, \Delta)$ is an arc. Suppose that $Y \in \mathscr{F}_c$ is a witness, and β has a representative contained in Y. Then we say that Y *carries* β. ◁

Note that $\pi_Y(\beta) \subset \mathscr{C}(Y)$ is one or two essential, nonperipheral curves in Y. Recall, by Fact 2.8, that $\mathscr{C}(Y)$ embeds into $\mathscr{C}(\mathscr{F}_c)$. We now define a relation $\phi_c \colon \mathscr{A}(S, \Delta) \to \mathscr{C}(\mathscr{F}_c)$ as follows:

$$\phi_c(\alpha) = \{\pi_Y(\beta) \mid d_{\mathscr{A}}(\alpha, \beta) \le 2 \text{ and } Y \in \mathscr{F}_c \text{ carries } \beta\}.$$

Lemma 3.13 *The relation ϕ_c is an equivariant coarse Lipschitz map.*

Proof Equivariance follows from the definition.

The set $\phi_c(\alpha)$ is nonempty by Lemma 3.5. Suppose that Y, Y' and Z lie in $\mathcal{F}_c$. Suppose that β and β' are carried by Y and Y', respectively, and are distance at most two from α. Thus, $d_{\mathcal{A}}(\beta, \beta') \le 4$. We deduce that $d_Z(Y, Y')$ is at most 12.

We now recall our choices (made above) of k, K and L. In particular, we have $\mathsf{k} > 12$. Thus, by Fact 2.9, $\pi_Y(\partial Y')$ and $\pi_{Y'}(\partial Y)$ are distance at most L in $\mathcal{C}(\mathcal{F}_c)$. Thus, $d_{\mathcal{C}}(\beta, \beta') \le \mathsf{L} + 20$, bounding the diameter of $\phi_c(\alpha)$. Thus, ϕ_c is a coarse map.

Furthermore, by [Masur and Minsky 2000, Lemma 2.3], if $d_{\mathcal{A}}(\alpha, \alpha') = 1$ then the distance between $\phi_c(\alpha)$ and $\phi_c(\alpha')$ is also bounded in terms of L. Applying the triangle inequality gives the result. □

Lemma 3.14 *Suppose that $\alpha, \gamma \in \mathcal{A}(S, \Delta)$ are arcs and X is a witness with $\xi'(X) = c$. Then*

$$|d_X(\alpha, \gamma) - d_X(\phi_c(\alpha), \phi_c(\gamma))| \le 12$$

Proof Suppose that $\beta \in \mathcal{A}(S, \Delta)$ has $d_{\mathcal{A}}(\alpha, \beta) \le 2$ and β is carried by some witness Y with $\xi'(Y) = c$. Note that α and $\pi_Y(\beta)$ are distance at most three in $\mathcal{AC}(S)$, the arc and curve complex for S.

Now, if $X = Y$, then

$$d_X(\alpha, \pi_Y(\beta)) = d_X(\alpha, \pi_X(\beta)) = d_X(\alpha, \beta) \le 4$$

by [Masur and Minsky 2000, Lemma 2.3]. If $X \neq Y$, then instead we have

$$d_X(\alpha, \pi_Y(\beta)) = d_{\mathcal{C}(X)}(\pi_X(\alpha), \pi_X(Y)) \le 6.$$

This holds for all β arising in the definition of $\phi_c(\alpha)$. The lemma now follows by applying the triangle inequality twice. □

We now define $\phi\colon \mathcal{A}(S, \Delta) \to \mathcal{P}(S, \Delta)$ by taking

$$\phi(\alpha) = (\phi_c(\alpha))_c.$$

All that remains is to prove that ϕ is a quasi-isometric embedding. Suppose that α and γ are arcs in $\mathcal{A}(S, \Delta)$. We must show that $d_{\mathcal{A}}(\alpha, \gamma)$ and $d_{\mathcal{P}}(\phi(\alpha), \phi(\gamma))$ are coarsely equal.

We first bound $d_{\mathcal{P}}(\phi(\alpha), \phi(\gamma))$ from above. Recall that $\mathcal{P}(S, \Delta)$ is equipped with the ℓ^1 metric and so

$$d_{\mathcal{P}}(\phi(\alpha), \phi(\gamma)) = \sum_c d_{\mathcal{C}(\mathcal{F}_c)}(\phi_c(\alpha), \phi_c(\gamma)).$$

Each of the terms on the right-hand side is bounded in terms of $d_{\mathcal{A}}(\alpha, \gamma)$ by Lemma 3.13 and we are done.

We now bound $d_{\mathcal{A}}(\alpha, \gamma)$ from above. Since $\mathcal{A}(S, \Delta)$ is a *combinatorial complex* in the sense of [Masur and Schleimer 2013, Section 5], we have a corollary of [Masur and Schleimer 2013, Theorems 5.14 and 13.1]:

Theorem 3.15 *Suppose that S and Δ are as above. There is a constant L such that, for any $\mathsf{L}' \geq \mathsf{L}$, there is a constant C with the following property. For any arcs α and γ, we have*

$$d_{\mathscr{A}}(\alpha, \gamma) =_{\mathsf{C}} \sum [d_X(\alpha, \gamma)]_{\mathsf{L}'},$$

where the sum is taken over all witnesses X for $\mathscr{A}(S, \Delta)$. □

Take $\mathsf{K}' > 12$ sufficiently larger than the constants K and L appearing in Theorems 2.10 and 3.15, respectively. Set $\mathsf{L}' = \mathsf{K}' + 12$. Fix a witness X and set $c = \xi'(X)$. If a term $d_X(\alpha, \gamma)$ appears in the upper bound of Theorem 3.15, then, by Lemma 3.14, the term $d_X(\phi_c(\alpha), \phi_c(\gamma))$ appears in the lower bound provided by Theorem 2.10 for the family $\mathscr{F}_c$. Also, $d_X(\alpha, \gamma)$ is at most twice $d_X(\phi_c(\alpha), \phi_c(\gamma))$ (by Lemma 3.14 and because $\mathsf{K}' > 12$). Thus, $d_{\mathscr{A}}(\alpha, \gamma)$ is coarsely bounded above by $d_{\mathscr{P}}(\phi(\alpha), \phi(\gamma))$, as desired. This finishes the proof of Theorem 3.10. □

4 Bound for the disk complex

Suppose that M is a spotless compression body, as defined in Section 2.17. Suppose that $S = S_{g,b} = \partial^+ M$ is the upper boundary. We assume that $\xi'(S) > 0$. Let $\mathscr{D}(M, S)$ be the graph of essential disks with boundary in S.

4.1 Witnesses for the disk complex

Definition 4.2 An essential subsurface $X \subset S$ is a *witness* for $\mathscr{D}(M, S)$ if every disk $D \in \mathscr{D}(M, S)$ cuts X. ◁

Some authors call such an X *disk-busting* [Masur and Schleimer 2013]. We call a witness X *large* if it satisfies

$$\operatorname{diam}_X\big(\pi_X(\mathscr{D}(M, S))\big) > 60.$$

We now record the classification of large witnesses [Masur and Schleimer 2013, Theorems 10.1, 11.10 and 12.1]:

Theorem 4.3 *Suppose that (M, S) is a nontrivial spotless compression body. Suppose that $X \subset S$ is a large witness for $\mathscr{D}(M, S)$. Then we have the following:*

- *X is not an annulus.*
- *If X compresses in M, then there are disks D and E with boundary contained in, and filling, X.*
- *If X is incompressible in M, then there is an orientation I–bundle $\rho_F : T_F \to F$ with $(T_F, \partial_h T_F) \subset (M, S)$ and with X being a component of $\partial_h T_F$. Also, $\partial_v T_F$ is properly embedded in (M, S) and at least one component of $\partial_v T_F$ is isotopic into S. Also, F admits a pseudo-Anosov map.* □

Remark 4.4 Suppose that X is a large incompressible witness, as in the third case. Let F be the base of the associated I–bundle T_F. Let P_F be the collection of annuli, embedded in S, which are isotopic, rel boundary, to components of $\partial_v T_F$. We call these annuli the *paring locus* for T_F. The paring locus P_F is nonempty (Theorem 4.3) and is disk-busting [Masur and Schleimer 2013, Remark 12.17]. We call an essential disk $(D, \partial D) \subset (M, S)$ *vertical* for T_F exactly when $D \cap T_F$ is vertical in T_F. Such disks exist by Theorem 4.3. If D is vertical for T_F, then D meets the paring locus P_F in exactly two essential arcs. Finally, the union $P_F \cup \partial_h T_F$ is a compressible witness for $\mathcal{D}(M, S)$. ◁

Definition 4.5 If T_F is a product, then $\partial_h T_F = X \sqcup X'$, where $X' \subset S$ is again a large incompressible witness. In this case, we call X and X' *twins*. ◁

Similar to the case of the arc complex (Corollary 3.4), all large witnesses for the disk complex *interfere*, as follows:

Corollary 4.6 *Suppose that X and Y are disjoint large witnesses with $\xi'(X) = \xi'(Y)$. Then either*

(1) *X and Y are twins, or*

(2) *X and Y have twins, X' and Y', respectively, such that X' overlaps with Y and Y' overlaps with X.*

Proof Suppose that X and Y are not twins. Thus, we may apply [Masur and Schleimer 2013, Lemma 12.21] to find that X has a twin X' and X' intersects Y. Since $\xi'(X') = \xi'(Y)$, neither X' nor Y is nested in the other. Thus, X' overlaps with Y. The remainder of the proof is similar. □

As usual, we take $\mathrm{MCG}(M, S)$ to be the mapping class group for the pair (M, S): that is, the group of mapping classes of M that preserve S setwise. We say that a pair of disks $D, E \in \mathcal{D}(M, S)$ have the same *topological type* (or, more simply, the same *type*) if there is a mapping class $f \in \mathrm{MCG}(M, S)$ such that $f(D) = E$.

Lemma 4.7 *The quotient $\mathcal{D}(M, S)/\mathrm{MCG}(M, S)$ has diameter at most two.*

Proof There are two cases. Suppose first that $\mathcal{D}(M, S)$ contains a nonseparating disk, that is, a disk $(D, \partial D) \subset (M, S)$ such that $M - D$ is connected. By Theorem 2.18 (the classification of compression bodies), all nonseparating disks have the same topological type. Suppose that $E \in \mathcal{D}(M, S)$ is a separating disk. The classification of compression bodies implies that E is disjoint from some nonseparating disk D'. This gives the desired diameter bound.

Suppose instead that all essential disks in (M, S) are separating. Thus, the lower boundary of M is nonempty. Suppose that F is a component of the lower boundary of M. Let D be a separating disk which cuts a copy of $F \times I$ off of M. Now suppose that E is any separating disk. The classification of compression bodies implies that E is disjoint from some disk D' which is a homeomorphic image of D. □

4.8 Families of witnesses

Fix a modified complexity $c \le \xi'(S)$. The collection

$$\mathscr{F}_c = \{X \subset S \mid X \text{ is a large witness for } \mathscr{D}(M,S) \text{ and } \xi'(X) = c\}$$

is called a *complete family*. We now define a *reduced family* as follows. Suppose that X and X' in $\mathscr{F}_c$ are twins. Thus, there is an I–bundle $T = F_T \times I$, where $\partial_h T = X \sqcup X'$. We remove both X and X' from $\mathscr{F}_c$ and replace them by the base surface F_T. We abuse notation and again use $\mathscr{F}_c$ to denote the reduced family. When $A = F_T \in \mathscr{F}_c$ is the base surface associated to T, we abuse notation and define $\partial_S A = \partial_S \partial_h T$.

Lemma 4.9 *Fix (M,S) and c. Suppose that $A, B \in \mathscr{F}_c$ are distinct. Suppose that B is a base surface replacing the twinned surfaces Y and Y'. Then $\partial_S A$ cuts both Y and Y'.*

Proof Let T_B be the product I–bundle associated to B. Let P_B be the paring locus of T_B. Recall that P_B is disk-busting.

Suppose that A is a compressible witness. Thus, P_B cuts A. Suppose that $\partial_S A$ does not cut Y. Thus, Y is (after an isotopy) contained in A. From Remark 2.14, we deduce that Y is isotopic to A. Thus, P_B does not cut A, a contradiction. Thus, $\partial_S A$ cuts Y; a similar argument proves that $\partial_S A$ cuts Y'.

Suppose that A is an incompressible witness, with I–bundle T_A. Let P_A be the paring locus of T_A. There are two subcases as T_A is twisted or a product.

Suppose that T_A is twisted. Thus, $A \cup P_A$ is a compressible witness (Remark 4.4). Again, since P_B is disk-busting, it cuts $A \cup P_A$. If P_B cuts $\partial_S A$, we are done because P_B is parallel into both Y and Y'. If not, then P_B is (after an isotopy) contained in A or contained in P_A. In either case, Y and Y' must cut A. Since $\xi'(Y) = \xi'(Y') = \xi'(A)$, we cannot have Y or Y' contained in A (Remark 2.14). Thus, A overlaps *both* Y and Y', and we are done.

Suppose that T_A is a product. Let X and X' be the twin components of $\partial_h T_A$. Appealing to Corollary 4.6, we may assume that X overlaps with Y. If X overlaps with Y', then we are done. If it does not, then, by Corollary 4.6, we deduce that X' overlaps Y'. Thus, in either case, we are done. □

Definition 4.10 Fix a modified complexity c. Suppose that $A, B \in \mathscr{F}_c$. We now define $\pi_B(A)$. (Note that we are overloading the notation π_B. When the argument is a collection of curves, we use Definition 2.21. When the argument is a witness, we use Definition 4.10.) There are two cases:

(1) Suppose that B is not a base surface. Then we define $\pi_B(A) = \pi_B(\partial_S A)$.

(2) Suppose that B is a base surface. Suppose that $\rho_B \colon T_B \to B$ is the I–bundle associated to B. Then we isotope $\partial_S A$ to meet $\partial_h T_B$ minimally and we define $\pi_B(A) = \pi_B(\rho_B(T_B \cap \partial_S A))$. ◁

In the above definition, we are considering $\rho_B(T_B \cap \partial_S A)$ as a set of arcs and curves in B. We surger them one at a time to obtain a set of curves in B. Also, in both parts of the definition, Corollary 4.6 implies that $\pi_B(A)$ is nonempty.

Suppose that $A, B, C \in \mathcal{F}_c$. Suppose further that B is distinct from both A and C. Then we define

$$d_B(A, C) = d_{\mathcal{C}(B)}(\pi_B(A), \pi_B(C)).$$

We will now abuse notation: the reduced family $\mathcal{F}_c$ contains surfaces A which index metric spaces, namely the curve graphs $\mathcal{C}(A)$, instead of being metric spaces themselves.

We now verify the three BBF axioms, as stated in Section 2.4. Recall that (M, S) is a spotless compression body, together with its upper boundary. Also, $c \in \mathbb{Z}$ is an integer. Here is the proof of Axiom 2.5:

Lemma 4.11 *There is a constant $k > 0$ such that, for every (M, S), for every c and for every $A, B \in \mathcal{F}_c$, we have that* $\operatorname{diam}_B(\pi_B(A)) \le k$.

Proof Suppose that B is not a base surface. Note that $\partial_S A$ is a disjoint collection of curves. By Corollary 4.6, $\partial_S A$ cuts B. By [Masur and Minsky 2000, Lemma 2.3], the diameter of $\pi_B(A)$ in $\mathcal{C}(B)$ is at most two.

Suppose that B is a base surface. Let D be either a compressing disk for A or a vertical disk for T_A, as provided by Theorem 4.3 or Remark 4.4, respectively. If D is a compressing disk, then ∂D is disjoint from $\partial_S A$. If D is a vertical disk, then ∂D meets $\partial_S A = \partial_S \partial_h T_A$ in exactly four points. We now isotope ∂D to have minimal intersection with $\partial_S B$. Applying [Masur and Schleimer 2013, Lemma 12.20], we deduce that $\pi_B(\rho_B(T_B \cap \partial D))$ has bounded diameter. Thus, by the triangle inequality, $\pi_B(\rho_B(T_B \cap \partial_S A))$ also has bounded diameter, and we are done. □

We adopt the notation $d_Y(A, C) = d_Y(\partial_S A, \partial_S C)$.

Lemma 4.12 *Let k be the constant of Lemma 4.11. Fix (M, S) and c. Suppose that $A, B, C \in \mathcal{F}_c$, where B is a base surface replacing the twinned surfaces Y, Y'. Then*

$$d_Y(A, C) \le d_B(A, C) \le d_Y(A, C) + 2k$$

and the same holds for Y'.

Proof By Lemma 4.9, $d_Y(A, C)$ is defined. The first inequality follows from the definitions. The second inequality follows from two applications of Lemma 4.11 and the triangle inequality. □

Here is the proof of Axiom 2.6:

Lemma 4.13 *Let k be the constant of Lemma 4.11. Fix (M, S) and c. For every $A, B, C \in \mathcal{F}_c$, at most one of the following is greater than $12 + 2k$:*

$$d_A(B, C), \quad d_B(A, C), \quad d_C(A, B).$$

Proof It suffices to assume that $d_B(A, C) > 12 + 2k$ and bound $d_A(B, C)$ from above.

Suppose that neither B nor A is a base surface. Then we may apply the usual Behrstock inequality and deduce that $d_A(B, C) < 10$; see [Mangahas 2010, Lemma 2.5].

Suppose instead B is not a base surface but A is. Let X and X' be the twins over A. By Corollary 4.6, both X and X' overlap B. Applying [Masur and Minsky 2000, Lemma 2.3], $d_B(X, C) > 10 + 2k$. The usual Behrstock inequality gives $d_X(B, C) < 10$. Applying Lemma 4.12, we deduce that $d_A(B, C) < 10 + 2k$.

Suppose instead that B is a base surface but A is not. Let Y and Y' be the twins over B. By Lemma 4.12, $d_Y(A, C) \geq 12$ and also $d_{Y'}(A, C) \geq 12$. By Lemma 4.9, both $\partial_S B$ and $\partial_S C$ cut A. Suppose that $\partial_S Y$ cuts A. The usual Behrstock inequality gives $d_A(Y, C) < 10$. We deduce that $d_A(B, C) < 12$, as desired.

Finally, suppose that B and A are base surfaces. Let Y and Y', and X and X', be the twins over B and A, respectively. By Corollary 4.6, we may assume that X and Y overlap. Thus, $d_Y(X, C) > 10$, so $d_X(Y, C) < 10$, and we are done as above. □

Here is the proof of Axiom 2.7:

Lemma 4.14 *Let k be the constant of Lemma 4.11. Fix (M, S) and c. For every $A, C \in \mathcal{F}_c$, the following set is finite*:

$$\{B \in \mathcal{F}_c \mid A \neq B,\ B \neq C \text{ and } d_B(A, C) > 7 + 2k\}.$$

Proof If A and C are not base surfaces, then this follows from [Bestvina et al. 2015, Lemma 5.3] and Lemma 4.12. If A is a base surface but C is not, then suppose that X and X' are the twins over A. We may repeat the previous argument for the pairs (X, C) as well as (X', C), paying at most an additional two [Masur and Minsky 2000, Lemma 2.3] in each case. When both A and C are base surfaces, there are four such pairs and the cost is at most an additional four in each case. □

Since the axioms hold, as in Section 3.6 we may build the product of quasitrees of spaces $\mathcal{P}(M, S)$ for the disk graph. We obtain the following:

Corollary 4.15 *Suppose that (M, S) is a nontrivial spotless compression body with $S = S_{g,b}$. Suppose that $\xi'(S) \geq 1$. Then*

$$\operatorname{dim}_{\mathsf{asym}} \mathcal{P}(M, S) \leq \tfrac{1}{2}(4g + b)(4g + b - 3) - 2.$$ □

4.16 Embedding the disk complex

In this section, we prove the following:

Theorem 4.17 *There is a quasi-isometric embedding ϕ of the disk graph $\mathcal{D}(M, S)$ into the product $\mathcal{P}(M, S)$ of quasitrees of curve graphs. Moreover, ϕ is equivariant with respect to the action of the mapping class group* $\mathrm{MCG}(M, S)$.

We deduce from this, and from Fact 2.3, the following:

Corollary 4.18 *Suppose that (M, S) is a nontrivial spotless compression body with $S = S_{g,b}$. Suppose that $\xi'(S) \geq 1$. Then*

$$\operatorname{dim}_{\mathsf{asym}} \mathcal{D}(M, S) \leq \tfrac{1}{2}(4g + b)(4g + b - 3) - 2.$$ □

Remark 4.19 When $g > 1$ and $b = 0$, the upper bound is smaller by one. See [Bestvina and Bromberg 2019, Corollary 1.1]. ◁

Remark 4.20 Hamenstädt [2019, Theorem 3.6] previously showed, in the special case of a handlebody H_g, that $\operatorname{dim}_{\mathsf{asym}} \mathscr{D}(H_g, \partial H_g) \leq (3g-3)(6g-2)$. Her proof technique is quite different from ours. ◁

The proof of Theorem 4.17 is the same as that of Theorem 3.10, with three changes. First, we replace the diameter bound (Lemma 3.5, for arcs) by Lemma 4.7. Second, we replace the definition of *carries* (Definition 3.12, for arcs) with the following:

Definition 4.21 Suppose that $D \in \mathscr{D}(M, S)$ is a disk and $A \in \mathscr{F}_c$. If $A \subset S$ is compressible (so not a base surface), then A *carries* D exactly when D is a compressing disk for A. If A is a twisted witness, or a base surface, then A *carries* D exactly when D is isotopic to a vertical disk in T_A. ◁

Third and lastly, we replace the distance estimate of Theorem 3.15 with the following:

Theorem 4.22 *Suppose that M and S are as above. There is a constant K such that, for any $\mathsf{K}' \geq \mathsf{K}$, there is a constant C with the following property: for any disks D and E, we have*

$$d_{\mathscr{D}}(D, E) =_{\mathsf{C}} \sum [d_X(D, E)]_{\mathsf{K}'},$$

where the sum is taken over all X in all reduced families for $\mathscr{D}(M, S)$.

Proof The distance estimate [Masur and Schleimer 2013, Theorem 19.9] bounds $d_{\mathscr{D}}(D, E)$ above and below using the sum of projection distances to all witnesses. That is, the sum there is taken over all X in all complete families for $\mathscr{D}(M, S)$. Suppose that B is a base surface for the twins Y and Y'. By [Masur and Schleimer 2013, Theorem 12.20], all three of

$$d_B(D, E), \quad d_Y(D, E) \quad \text{and} \quad d_{Y'}(D, E)$$

are coarsely equal. Here we define $d_B(D, E) = d_B(\pi_B(\partial D), \pi_B(\partial E))$ and $\pi_B(\partial D) = \pi_B(\rho_B(T_B \cap \partial D))$ as in Definition 4.10. This and [Masur and Schleimer 2013, Theorem 19.9] gives the lower bound. The upper bound is proved in the same way, after weakening the constant C by a factor of two. □

Acknowledgement Fujiwara is supported by Grants-in-Aid for Scientific Research from JSPS (20H00114, 15H05739).

References

[Behrstock 2006] **J A Behrstock**, *Asymptotic geometry of the mapping class group and Teichmüller space*, Geom. Topol. 10 (2006) 1523–1578 MR Zbl

[Behrstock et al. 2017] **J Behrstock**, **M F Hagen**, **A Sisto**, *Asymptotic dimension and small-cancellation for hierarchically hyperbolic spaces and groups*, Proc. Lond. Math. Soc. 114 (2017) 890–926 MR Zbl

[Bell and Dranishnikov 2008] **G Bell**, **A Dranishnikov**, *Asymptotic dimension*, Topology Appl. 155 (2008) 1265–1296 MR Zbl

[Bell and Fujiwara 2008] **G C Bell**, **K Fujiwara**, *The asymptotic dimension of a curve graph is finite*, J. Lond. Math. Soc. 77 (2008) 33–50 MR Zbl

[Bestvina and Bromberg 2019] **M Bestvina**, **K Bromberg**, *On the asymptotic dimension of the curve complex*, Geom. Topol. 23 (2019) 2227–2276 MR Zbl

[Bestvina et al. 2015] **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 Zbl

[Biringer and Vlamis 2017] **I Biringer**, **N G Vlamis**, *Automorphisms of the compression body graph*, J. Lond. Math. Soc. 95 (2017) 94–114 MR Zbl

[Bonahon 1983] **F Bonahon**, *Cobordism of automorphisms of surfaces*, Ann. Sci. École Norm. Sup. 16 (1983) 237–270 MR Zbl

[Farb and Margalit 2012] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR Zbl

[Gromov 1993] **M Gromov**, *Asymptotic invariants of infinite groups*, from "Geometric group theory, II" (G A Niblo, M A Roller, editors), London Math. Soc. Lecture Note Ser. 182, Cambridge Univ. Press (1993) 1–295 MR Zbl

[Hamenstädt 2019] **U Hamenstädt**, *Asymptotic dimension and the disk graph, II*, J. Topol. 12 (2019) 675–684 MR Zbl

[Harvey 1981] **W J Harvey**, *Boundary structure of the modular group*, from "Riemann surfaces and related topics" (I Kra, B Maskit, editors), Ann. of Math. Stud. 97, Princeton Univ. Press (1981) 245–251 MR Zbl

[Hempel 1976] **J Hempel**, 3*–Manifolds*, Annals of Math. Studies 86, Princeton Univ. Press (1976) MR Zbl

[Mangahas 2010] **J Mangahas**, *Uniform uniform exponential growth of subgroups of the mapping class group*, Geom. Funct. Anal. 19 (2010) 1468–1480 MR Zbl

[Masur and Minsky 1999] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. 138 (1999) 103–149 MR Zbl

[Masur and Minsky 2000] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR Zbl

[Masur and Schleimer 2013] **H Masur**, **S Schleimer**, *The geometry of the disk complex*, J. Amer. Math. Soc. 26 (2013) 1–62 MR Zbl

[Oertel 2002] **U Oertel**, *Automorphisms of three-dimensional handlebodies*, Topology 41 (2002) 363–410 MR Zbl

[Vokes 2022] **K M Vokes**, *Hierarchical hyperbolicity of graphs of multicurves*, Algebr. Geom. Topol. 22 (2022) 113–151 MR Zbl

Department of Mathematics, Kyoto University
Kyoto, Japan
Mathematics Institute, University of Warwick
Coventry, United Kingdom

kfujiwara@math.kyoto-u.ac.jp, s.schleimer@warwick.ac.uk

Received: 18 May 2022 Revised: 26 October 2022

Algebraic & Geometric Topology **24**:5 (2024) 2673–2705
DOI: 10.2140/agt.2024.24.2673
Published: 19 August 2024

Representation stability for homotopy automorphisms

Erik Lindell
Bashar Saleh

We consider in parallel pointed homotopy automorphisms of iterated wedge sums of finite CW–complexes and boundary-relative homotopy automorphisms of iterated connected sums of manifolds minus a disk. Under certain conditions on the spaces and manifolds, we prove that the rational homotopy groups of these homotopy automorphisms form finitely generated FI–modules, and thus satisfy representation stability for symmetric groups in the sense of Church and Farb. We also calculate explicit bounds on the weights and stability degrees of these FI–modules.

55P62

1 Introduction

Pointed homotopy automorphisms of iterated wedge sums of spaces and boundary-relative homotopy automorphisms of iterated connected sums of manifolds minus a disk, come with stabilization maps that yield questions of whether the homology groups or the homotopy groups of these homotopy automorphisms stabilize in any sense. Previously Berglund and Madsen [2020] have proven rational homological stability for homotopy automorphisms of iterated connected sums of higher-dimensional tori $S^n \times S^n$ for $n \geq 3$, and these results were later expanded by Grey [2019] and Stoll [2024] for homotopy automorphisms of iterated connected sums of certain manifolds of the form $S^n \times S^m$ for $n, m \geq 3$.

We instead study the rational *homotopy groups* of the homotopy automorphisms in question, which we consider as based spaces with the identity map as the basepoint. These homotopy groups do not stabilize in the traditional sense. Instead, we show that they satisfy a different kind of stability, known as *representation stability*. In the two cases we study here, we consider sequences of rational homotopy groups, which in step n are representations of the symmetric group Σ_n. For such representations there is a consistent way to name the irreducible representations for arbitrary n, and representation stability essentially means that as n tends to infinity, the decomposition into irreducible representations eventually becomes constant.

Representation stability was introduced by Church and Farb [2013] and later further developed by Church, Ellenberg and Farb [Church et al. 2015], who showed that for representations of symmetric groups this notion can be encoded by so-called FI–modules, which are functors from the category of finite sets and injections to the category of vector spaces. The stable range of representation stability corresponds to *stability degree* and *weight* of the corresponding FI–module.

We review FI–modules and representation stability in more detail in Section 2. Our first main result is the following:

Theorem A *Let $(X,*)$ be a pointed simply connected space with the homotopy type of a finite CW–complex and let $X_S := \bigvee^S X$ for any finite set S. For each $k \geq 1$, the functor*

$$S \mapsto \pi_k^{\mathbb{Q}}(\mathrm{aut}_*(X_S))$$

is an FI*–module. If $H_n(X,\mathbb{Q}) = 0$ for $n \geq d$, this* FI*–module is of weight $\leq k+d-1$ and stability degree $\leq k+d$.*

For the analogous theorem for connected sums, we need the notion of a boundary-relative homotopy automorphism of a manifold N (with boundary). A boundary-relative homotopy automorphism of N is a homotopy automorphism of N that preserves the boundary $\partial := \partial N$ pointwise. The boundary-relative homotopy automorphisms of N form a topological monoid, with respect to composition, which we will denote by $\mathrm{aut}_\partial(N)$.

Let $M = M^d$ be a closed oriented d–dimensional manifold. For any finite set S, we let M_S denote the S–fold connected sum of M with itself, with an open d–disk removed: $M_S = \#^S M \setminus \mathring{D}^d$. For $\boldsymbol{n} = \{1,2,\dots,n\}$, we denote $M_{\boldsymbol{n}}$ simply by M_n. A homotopy automorphism of M_n does not extend to a homotopy automorphism of M_{n+1} in any canonical way in general. However, boundary-relative homotopy automorphisms of M_n extend by the identity to a boundary-relative homotopy automorphism of M_{n+1}. In particular, there is a stabilization map

$$s_n \colon \mathrm{aut}_\partial(M_n) \to \mathrm{aut}_\partial(M_{n+1}).$$

By picking some basepoint in the boundary of M_1, there is a deformation retract $M_S \xrightarrow{\simeq} \bigvee^S M_1$ (see eg [Félix et al. 2008, Section 3.1.2]), where the wedge sum is taken along this basepoint. It follows by Theorem A that there is an FI–module given on objects by $S \mapsto \pi_k(\mathrm{aut}_*(M_S)) \cong \pi_k(\mathrm{aut}_*(\bigvee^S M_1))$. For any finite set S we have an obvious inclusion map $\mathrm{aut}_\partial(M_S) \hookrightarrow \mathrm{aut}_*(M_S)$, so we may ask whether we can find an FI–module given by $S \mapsto \pi_k(\mathrm{aut}_\partial(M_S))$ that make these maps into a morphism of FI–modules, ie a natural transformation of functors. We will refer to this as "lifting" the FI–module structure. In our second main theorem, we address this problem:

Theorem B *Let $M = M^d$ be a closed simply connected oriented d–dimensional manifold. With M_S defined as above, we have the following:*

(a) *For each $k \geq 1$, the* FI*–module*
$$S \mapsto \pi_k\left(\operatorname{aut}_*\left(\bigvee^S M_1\right)\right) \cong \pi_k(\operatorname{aut}_*(M_S))$$
lifts to an FI*–module*
$$S \mapsto \pi_k(\operatorname{aut}_\partial(M_S))$$
sending the standard inclusion $\boldsymbol{n} \to \boldsymbol{n}+\boldsymbol{1}$ to the map $\pi_k(\operatorname{aut}_\partial(M_n)) \to \pi_k(\operatorname{aut}_\partial(M_{n+1}))$ induced by the stabilization map s_n.

(b) *The rationalization of this* FI*–module is of weight $\leq k+d-2$ and stability degree $\leq k+d-1$.*

Remark 1.1 Theorems A and B are somewhat analogous to those for unordered configuration spaces of manifolds. Rational homological stability for unordered configuration spaces of arbitrary connected manifolds was proven by Church [2012], following integral results for open[1] manifolds by Arnold [1969], McDuff [1975] and Segal [1979]. It was later proven by Kupers and Miller [2018] that the rational *homotopy groups* of unordered configuration spaces on connected, simply connected manifolds of dimension at least 3 satisfy representation stability.

Homotopy automorphisms of iterated wedge sums of *spheres* have been studied by Miller, Patzt and Petersen [Miller et al. 2019]. Using representation stability, they prove that for $d \geq 2$ the sequence $\{B\operatorname{aut}(\bigvee_{i=1}^n S^d)\}_{n\geq 1}$ satisfies homological stability with $\mathbb{Z}[\frac{1}{2}]$–coefficients, which proves homological stability with the same coefficients for $\{B\operatorname{GL}_n(\mathbb{S})\}_{n\geq 1}$, where $\mathbb{S}$ is the sphere spectrum. These results are neither weaker nor stronger than Theorem A, since on one hand they work with $\mathbb{Z}[\frac{1}{2}]$–coefficients and on the other hand we work with wedge sums of more general CW–complexes than spheres.

For a simply connected d–dimensional manifold M, with boundary $\partial M \cong S^{d-1}$, the rational homotopy theory of $\operatorname{aut}_\partial(M)$ has been thoroughly studied by Berglund and Madsen [2020], whose results we will use.

As a byproduct of the techniques used for proving Theorem B(a) we get the following:

Theorem *Let M be a closed oriented simply connected d–dimensional manifold such that the reduced homology of $M \setminus \mathring{D}^d$ is nontrivial. Given a subspace $A \subseteq \partial M_n$, possibly empty, such that $A \subset M_n$ is a cofibration, then the groups $\pi_0(\operatorname{aut}_A(M_n))$, $\pi_0(\operatorname{Diff}_A(M_n))$ and $\pi_0(\operatorname{Homeo}_A(M_n))$ contain a subgroup isomorphic to Σ_n.*

Structure In Section 2 we review the necessary background on FI–modules. The reader familiar with FI–modules may skip directly to Section 2.8, where we introduce the notion of FI–*Lie models* of pointed FI–spaces, which is of key importance for proving the main theorems. In Section 3 we review rational

[1] Integral homological stability is known not to hold for closed manifolds. A simple counterexample is given already by the 2–sphere S^2, where $H_1(B_n(S^2), \mathbb{Z}) \cong \mathbb{Z}/(2n-2)\mathbb{Z}$; see for example [Birman 1974, Theorem 1.11].

homotopy theory for homotopy automorphisms needed for proving the main theorems. In Section 4 we study homotopy automorphisms of wedge sums and prove Theorem A. In Section 5 we study homotopy automorphisms of connected sums and prove Theorem B.

Acknowledgements The authors would like to thank Dan Petersen and Alexander Berglund for many useful comments and suggestions. We would also like to thank the referees, whose careful reading and useful comments have improved the paper greatly from its first version.

Section 5.1, which treats integral homotopy theory for relative homotopy automorphisms, has developed greatly since the first preprint version of this paper, thanks to many other people. Our decision to consider the integral homotopy groups of the homotopy automorphisms of iterated connected sums is inspired by an answer by Ryan Budney to a question by Saleh at MathOverflow. The method used to prove Theorem B(a) was suggested by Manuel Krannich, who has also provided several other very helpful comments. In addition, he was in the committee for the PhD defense of Lindell, where he, together with Fabian Hebestreit, pointed out several minor errors in the paper and had some very helpful suggestions for improvements.

Saleh was supported by the Knut and Alice Wallenberg Foundation through grant 2019.0521.

2 Representation stability, FI–modules and Lie models of FI–spaces

2.1 Conventions

Throughout the paper, we will use R to denote a commutative ring, which we will assume to be Noetherian for convenience. We will mainly work over the field $\mathbb{Q}$, so unless otherwise specifically stated, all vector spaces are over $\mathbb{Q}$. We will use "dg" to abbreviate the term *differential graded*. FI denotes the category of finite sets with injective maps as morphisms.

If S is a finite set, we will use $|S|$ to denote its cardinality, and we will write $\Sigma(S) := \mathrm{Aut}_{\mathrm{FI}}(S)$ for the symmetric group on S. If $S = \boldsymbol{n} := \{1, 2, \dots, n\}$, we will simply write $\Sigma(S) = \Sigma_n$ for brevity.

Recall that the irreducible $\mathbb{Q}$–representations of Σ_n are indexed by partitions of weight n, ie sequences of nonnegative integers $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l \geq 0 \geq \cdots)$ such that $|\lambda| = \lambda_1 + \lambda_2 + \cdots = n$. We will denote the corresponding $\mathbb{Q}$–representation by V_λ. For any $k \geq n + \lambda_1$, we also define the *padded* partition $\lambda[k] := (k - n \geq \lambda_1 \geq \lambda_2 \geq \cdots)$ and write $V(\lambda)_k := V_{\lambda[k]}$.

2.2 Representation stability

Before we introduce the language of FI–modules, let us recall the original notion of representation stability, which is formulated in terms of consistent sequences of Σ_n–representations.

Definition 2.1 Let R be a commutative ring. A *consistent sequence* of Σ_n–representations over R is a sequence $\{V^n, \phi^n\}$, where V^n is an $R[\Sigma_n]$–module and $\phi^n \colon V^n \to V^{n+1}$ is a Σ_n–equivariant map (where V^{n+1} is considered an $R[\Sigma_n]$–module through the standard inclusion $\Sigma_n \hookrightarrow \Sigma_{n+1}$).

If $R = \mathbb{Q}$, we may define (uniform) *representation stability* for such a sequence as follows:

Definition 2.2 A consistent sequence of rational Σ_n–representations $\{V^n, \phi^n\}$ is said to be uniformly representation stable with stable range $n \geq N$ if, for all $n \geq N$,

(i) the map ϕ^n is injective,

(ii) the image of ϕ^n generates V^{n+1} as a Σ_{n+1}–representation,

(iii) for each partition λ the multiplicity of $V(\lambda)_n$ in V^n is independent of n.

Next, we will introduce FI–modules, and recall how representation stability is encoded in that language.

2.3 FI–modules

We first introduce the notion of an FI–object in an arbitrary category.

Definition 2.3 Let $\mathcal{C}$ be a category. A functor $\mathrm{FI} \to \mathcal{C}$ is called an FI–object in $\mathcal{C}$.

Let us review the kinds of FI–objects that will be of interest to us:

- An FI–object in (gr)Mod_R, the category of ($\mathbb{Z}$–graded) R–modules, is called a (graded) FI–*R–module*. An FI–object in dgMod_R, the category of differential graded R–modules, is called a *dg* FI–R–module. For a dg FI–R–module $\mathcal{V}$, we will write $H_*(\mathcal{V})$ for the composition with the homology functor and refer to it as the homology of $\mathcal{V}$.

- An FI–object in dgLie_R, the category of dg Lie algebras, over R, will be called a dg FI–R–Lie algebra.

- An FI–object in Top_*, the category of pointed topological spaces, will be called a *pointed* FI–*space*. If $\mathcal{P}$ is a property of pointed topological spaces, such as being simply connected, we will say that a pointed FI–space $\mathcal{X}$ has property $\mathcal{P}$ if $\mathcal{X}(S)$ has property $\mathcal{P}$, for every finite set S. If $\mathcal{X}$ is a pointed FI–space with $\pi_1(\mathcal{X}(S))$ being abelian for every finite set S, composing with the (rational) homotopy groups functor π_* (resp. $\pi_*^{\mathbb{Q}}$) naturally gives us a graded FI–$\mathbb{Z}$–module (resp. graded FI–$\mathbb{Q}$–module). We will simply write $\pi_*(\mathcal{X})$ (resp. $\pi_*^{\mathbb{Q}}(\mathcal{X})$) for this composite functor and refer to it as the (rational) homotopy groups of $\mathcal{X}$.

We will generally consider the first two examples for $R = \mathbb{Q}$ and $R = \mathbb{Z}$. If the ring is clear from context, or if the choice of R is not important, we will generally drop it from the notation.

Now let us recall some basics from the theory of FI–modules. Since the category of (graded) R–modules is abelian, the category of (graded) FI–R–modules inherits this structure, which means that there are natural notions of (graded) FI–R–submodules as well as quotients, direct sums and tensor products of (graded) FI–R–modules, all defined pointwise; see [Church et al. 2015, Remark 2.1.2].

Remark 2.4 Any FI–R–module $\mathcal{V}$ gives rise to a consistent sequence $\{V^n := \mathcal{V}(\boldsymbol{n}), \phi^n := \mathcal{V}(\boldsymbol{n} \hookrightarrow \boldsymbol{n}+\mathbf{1})\}$ of $R[\Sigma_n]$–modules, where $\boldsymbol{n} \hookrightarrow \boldsymbol{n}+\mathbf{1}$ is the standard inclusion.

Remark 2.5 Not every consistent sequence arises from an FI–module; see [loc. cit., Remark 3.3.1].

Sometimes it will be more convenient to work with consistent sequences than with FI–modules. For this purpose, the following lemma is important:

Lemma 2.6 [loc. cit., Remark 3.3.1] *A consistent sequence $\{V^n, \phi^n\}$ is induced by some FI–module if and only if every $\sigma \in \Sigma_{n+k}$ with $\sigma|_{\boldsymbol{n}} = \mathrm{id}$ acts trivially on*

$$\mathrm{im}(\phi^{n+k-1} \circ \cdots \circ \phi^n \colon V^n \to V^{n+k}).$$

If two FI–modules give rise to isomorphic consistent sequences, then the two FI–modules are isomorphic.

The main property of FI–modules that will be of interest to us is *finite generation*, since this is what encodes representation stability:

Definition 2.7 Let $\boldsymbol{n} := \{1, 2, \dots, n\}$. A (graded) FI–$R$–module $\mathcal{V}$ is said to be *finitely generated* if there exists a finite set $S \subset \bigsqcup_{n \geq 1} \mathcal{V}(\boldsymbol{n})$ such that there is no proper (graded) FI–R–submodule $\mathcal{W}$ of $\mathcal{V}$ such that $S \subset \bigsqcup_{n \geq 1} \mathcal{W}(\boldsymbol{n})$.

Now we can describe how representation stability relates to FI–modules:

Theorem 2.8 [loc. cit., Theorem 1.13] *An FI–$\mathbb{Q}$–module $\mathcal{V}$ is finitely generated if and only if the consistent sequence $\{V^n := \mathcal{V}(\boldsymbol{n})\}$ is uniformly representation stable and each V^n is finite-dimensional.*

What makes working with the category of FI–R–modules for any Noetherian ring R particularly useful is that it is *Noetherian*, ie an FI–R–submodule of such a finitely generated FI–R–module is itself finitely generated; see [Church et al. 2015, Theorem 1.3; 2014, Theorem A]. Finite generation is also preserved by tensor products and quotients. This means that to prove that an FI–R–module is finitely generated, it suffices to show that it is a subquotient of a tensor product of some FI–R–modules that are more obviously finitely generated.

Since we want to use rational homotopy theory to prove our results, we need to consider graded FI–modules in our proofs. For this reason we will need the following definition:

Definition 2.9 If $\mathcal{V}$ is a graded FI–R–module and $m \in \mathbb{Z}$, let $\mathcal{V}_m$ be the degree-m part of $\mathcal{V}$, ie the postcomposition with the functor $\mathrm{grVect}_{\mathbb{Q}} \to \mathrm{Vect}_{\mathbb{Q}}$ given by sending a graded vector space to its degree-m part. If $\mathcal{V}_m = 0$ for $m \leq m'$ (resp. $m \geq m'$), we say that $\mathcal{V}$ is concentrated in degrees above (resp. below) m'. Such a graded FI–module is called bounded from below (resp. above).

2.4 Weight and stability degree

For the rest of Section 2 we will assume that $R = \mathbb{Q}$. We have seen how finite generation of FI–$\mathbb{Q}$–modules corresponds to representation stability of the corresponding consistent sequence of rational Σ_n–representations, but in order to make quantitative statements about stability ranges we need to introduce the *weight* and *stability degree* of such FI–modules.

Recall that if V is a Σ_n–representation, $(V)_{\Sigma_n}$ denotes the quotient of *coinvariants* of V. For an FI–module $\mathcal{V}$, this allows us to define a sequence $\{\phi_a(\mathcal{V})^n\}$ of vector spaces and maps between them, for each $a \geq 0$, by $\phi_a(\mathcal{V})^n := (\mathcal{V}(\boldsymbol{a} \sqcup \boldsymbol{n}))_{\Sigma_n}$. Any inclusion $\iota\colon \boldsymbol{n} \hookrightarrow \boldsymbol{n+1}$ gives us an inclusion $\mathrm{id} \sqcup \iota\colon \boldsymbol{a} \sqcup \boldsymbol{n} \hookrightarrow \boldsymbol{a} \sqcup (\boldsymbol{n+1})$, inducing a map $\phi_a(\mathcal{V})^n \to \phi_a(\mathcal{V})^{n+1}$. Since we quotient by Σ_{n+1}, the choice of inclusion ι does not matter.

With this, we can define the *stability degree* of an FI–module:

Definition 2.10 [Church et al. 2015, Definition 3.1.3] The *injectivity degree* inj-deg$(\mathcal{V})$ (resp. *surjectivity degree* surj-deg$(\mathcal{V})$) of an FI–module $\mathcal{V}$ is the smallest $s \geq 0$ such that for all $a \geq 0$, the map $\phi_a(\mathcal{V})^n \to \phi_a(\mathcal{V})^{n+1}$, defined as above, is injective (resp. surjective) for all $n \geq s$ (and if no such s exists we set the degree to ∞). We define the *stability degree* stab-deg(V) of $\mathcal{V}$ to be the maximum of the injectivity and surjectivity degrees.

Definition 2.11 The weight of an FI–module $\mathcal{V}$, which we denote by weight$(\mathcal{V})$, is the maximum weight $|\lambda|$ over all $V(\lambda)_n$ appearing in the Σ_n–representation $\mathcal{V}(\boldsymbol{n})$, if such a maximum exists. If no maximum exists, we set weight$(\mathcal{V}) = \infty$ and if the FI–module is zero we set it to zero.

These definitions are relevant because of their relation to representation stability, which may now be stated as follows:

Proposition 2.12 [loc. cit., Proposition 3.3.3] *Let $\mathcal{V}$ be an* FI*–module. The consistent sequence $\{V^n, \phi^n\}$ determined by $\mathcal{V}$ is uniformly representation stable with stable range $n \geq$* weight$(\mathcal{V})$ + stab-deg$(\mathcal{V})$.

Remark 2.13 This implies that if an FI–module has finite weight and stability degree, it is finitely generated. For this reason we will only be working with weight and stability degree going forward. However, due to the Noetherian property of FI–modules, it is possible to freely take submodules and quotients and preserve finite generation. This is not the case for stability degree, as we will see, making it much easier to prove finite generation than to obtain an explicit bound on stability degree.

Let us recall some useful properties of weight and stability degree. First, the following is immediate from the definitions:

Proposition 2.14 *Let $\mathcal{V}^1$ and $\mathcal{V}^2$ be* FI*–modules. Then* weight$(\mathcal{V}^1 \oplus \mathcal{V}^2) \leq \max($weight$(\mathcal{V}^1)$, weight$(\mathcal{V}^2))$ *and* stab-deg$(\mathcal{V}^1 \oplus \mathcal{V}^2) \leq \max($stab-deg$(\mathcal{V}^1)$, stab-deg$(\mathcal{V}^2))$.

Next, we will recall how weight and stability degree behave under taking tensor products:

Proposition 2.15 *Suppose that $\mathcal{V}^1, \mathcal{V}^2, \dots, \mathcal{V}^k$ are* FI*–modules with stability degrees $\leq r_1, r_2, \dots, r_k$ and weights $\leq s_1, s_2, \dots, s_k$, respectively. Then*

$$\mathrm{weight}(\mathcal{V}^1 \otimes \cdots \otimes \mathcal{V}^k) \leq s_1 + \cdots + s_k$$

and

$$\text{stab-deg}(\mathcal{V}^1 \otimes \cdots \otimes \mathcal{V}^k) \leq \max(r_1 + s_1, \dots, r_k + s_k, s_1 + \cdots + s_k).$$

Proof The first part is [Church et al. 2015, Proposition 3.2.2], while the second part is [Kupers and Miller 2018, Proposition 2.23]. □

We also need to know how stability degree and weight behave when taking submodules and quotients:

Proposition 2.16 *Let $\mathcal{V}$ be an* FI–*module and $\mathcal{W}$ be an* FI–*submodule of $\mathcal{V}$. Then* $\text{weight}(\mathcal{W}) \le \text{weight}(\mathcal{V})$ *and* $\text{weight}(\mathcal{V}/\mathcal{W}) \le \text{weight}(\mathcal{V})$. *If in addition $\mathcal{V}$ is such that $\mathcal{V}(S)$ is finite-dimensional for every finite set S, we have the following*:

(i) $\text{inj-deg}(\mathcal{W}) \le \text{inj-deg}(\mathcal{V})$.

(ii) $\text{surj-deg}(\mathcal{V}/\mathcal{W}) \le \text{surj-deg}\,(V)$.

(iii) *If* $\text{inj-deg}(\mathcal{V}) \le r$ *and* $\text{surj-deg}(\mathcal{W}) \le r$, *then* $\text{inj-deg}(\mathcal{V}/\mathcal{W}) \le r$.

(iv) *If* $\text{surj-deg}(\mathcal{V}) \le r$ *and* $\text{inj-deg}(\mathcal{V}/\mathcal{W}) \le r$, *then* $\text{surj-deg}(\mathcal{W}) \le r$.

Proof The first part follows directly by the definition of weight, and (i) and (ii) are [Church et al. 2015, Lemma 3.1.6].

To prove (iii) and (iv), note that for each $a \ge 0$, ϕ_a defines a functor from the category of FI–modules to the category of sequences of vector spaces and linear maps, and this functor is exact. Thus the following respective propositions from linear algebra suffice to prove (iii) and (iv): if $f : V \to W$ is a linear map of finite-dimensional vector spaces, $V' \subseteq V$ and $W' \subseteq W$ are subspaces, $f' \colon V' \to W'$ is a linear map such that $f'(v) = f(v)$ for all $v \in V'$ and $f/f' \colon V/V' \to W/W'$ is the induced map between the quotients, then

(iii′) if f is injective and f' is surjective, f/f' is injective,

(iv′) if f is surjective and f/f' is injective, f' is surjective.

These are both simple exercises in linear algebra and therefore left to the reader. □

Note however that given *only* the stability degree of an FI–module we can in general not say anything about the stability degree of its FI–submodules or quotients. However, if an FI–module $\mathcal{V}$ is isomorphic to *both* an FI–submodule of an FI–module and a quotient of an FI–module, for both of which we have bounds on the stability degree, we can use the proposition above to determine a bound on $\text{stab-deg}(\mathcal{V})$. This will be the case for an important class of FI–modules that we consider in Section 2.6. Note that in particular, we get the following corollary:

Corollary 2.17 *Suppose $\mathcal{W}$ is an* FI–*module which is a direct summand of another* FI–*module $\mathcal{V}$, ie that there exists a third* FI–*module $\mathcal{U}$ such that $\mathcal{V} \cong \mathcal{W} \oplus \mathcal{U}$. Then* $\text{stab-deg}(\mathcal{W}) \le \text{stab-deg}(\mathcal{V})$.

Finally, we need a way to determine the weight and stability degree in each degree when taking the homology of a differential graded FI–module. We will prove the following more general statement (see [Kupers and Miller 2018, Proposition 2.19]):

Proposition 2.18 *Let $\mathcal{U} \xrightarrow{f} \mathcal{V} \xrightarrow{g} \mathcal{W}$ be a sequence of* FI*–modules and morphisms of* FI*–modules such that $\mathcal{U}(S), \mathcal{W}(S)$ and $\mathcal{V}(S)$ are finite-dimensional for all $S \in \mathrm{FI}$ and $g \circ f = 0$. Then*

$$\mathrm{weight}(\ker(g)/\mathrm{im}(f)) \le \mathrm{weight}(\mathcal{W}),$$

and if all three FI*–modules have stability degree $\le r$, then* $\text{stab-deg}(\ker(g)/\mathrm{im}(f)) \le r$*. In particular, if $\mathcal{V}$ is a dg* FI*–module such that $\mathcal{V}_m(S)$ is finite-dimensional for each m and finite set S then* $\mathrm{weight}(H_m(\mathcal{V})) \le \mathrm{weight}(\mathcal{V}_m)$*, and if* $\text{stab-deg}(\mathcal{V}_i) \le r$ *for $i \in \{m-1, m, m+1\}$, we have* $\text{stab-deg}(H_m(\mathcal{V})) \le r$*.*

Proof The first part follows directly from the first part of Proposition 2.16. We prove the second part by showing that the homology has injectivity and surjectivity degree $\le r$. For injectivity degree, note that $\ker(g)$ has injectivity degree $\le r$ by Proposition 2.16(i), since it is an FI–submodule of $\mathcal{V}$. Furthermore, since the category of FI–modules is abelian, $\mathrm{im}(f) \cong \mathcal{U}/\ker(f)$, which has surjectivity degree $\le r$ by Proposition 2.16(ii). Thus it follows from Proposition 2.16(iii) that $\ker(g)/\mathrm{im}(f)$ has injectivity degree $\le r$.

For surjectivity degree, we argue similarly as follows: The injectivity degree of $\mathrm{im}(g)$ is at most r by Proposition 2.16(i), and since $\mathrm{im}(g) \cong \mathcal{V}/\ker(g)$, we thus get by Proposition 2.16(iv) that $\ker(g)$ has surjectivity degree $\le r$. Thus the quotient $\ker(g)/\mathrm{im}(f)$ does as well, by Proposition 2.16(ii). □

2.5 FI$^\#$–modules

Many FI–modules appearing “naturally” actually have additional structure, which may be encoded using the notion of an FI$^\#$–module. The category FI$^\#$ has the same objects as FI, but the morphisms $S \to T$ are given by a pair of subsets $A \subset S$ and $B \subset T$ and a bijection $A \to B$. We call these *partial injections*. An FI$^\#$–object in a category $\mathcal{C}$ is simply a functor $\mathrm{FI}^\# \to \mathcal{C}$. Since FI is a subcategory of FI$^\#$, any FI$^\#$–object has an underlying FI–object, so all the notions defined in the previous sections can be defined for (graded) FI$^\#$–modules by simply considering the underlying (graded) FI–module.

We consider FI$^\#$–modules because there is a natural way to define *duals* in this category. Note that the category FI$^\#$ is naturally isomorphic to its opposite category simply by taking the inverse of the bijection (see the end of [Church et al. 2015, Remark 4.1.3]). This allows us to make the following definition:

Definition 2.19 If $\mathcal{V}\colon \mathrm{FI}^\# \to \mathrm{Vect}_{\mathbb{Q}}$, we define the *dual* FI$^\#$–module $\mathcal{V}^*$ as the composite functor

$$\mathrm{FI}^\# \xrightarrow{\cong} (\mathrm{FI}^\#)^{\mathrm{op}} \xrightarrow{\mathcal{V}^{\mathrm{op}}} \mathrm{Vect}_{\mathbb{Q}}^{\mathrm{op}} \xrightarrow{\mathrm{Hom}_{\mathbb{Q}}(-,\mathbb{Q})} \mathrm{Vect}_{\mathbb{Q}}.$$

2.6 Schur functors

The graded FI–modules that we will study will be constructed by composing *Schur functors* with simpler graded FI–modules, which is why they are of finite type. In this section we will review what we mean by Schur functors in this context and their properties when composed with graded FI–modules.

If λ is a partition of $k \geq 0$, we define the Schur functor $\mathbb{S}_\lambda \colon \mathrm{grVect}_\mathbb{Q} \to \mathrm{grVect}_\mathbb{Q}$ on objects by

$$V \mapsto S^\lambda \otimes_{\Sigma_k} V^{\otimes k},$$

considering $V^{\otimes k}$ with the standard Σ_k–action and considering S^λ as a graded vector space concentrated in degree 0. Another definition, which gives an isomorphic functor, is that $\mathbb{S}_\lambda(V)$ is given by the composition of the k^{th} tensor power functor with the action of a certain idempotent operator $c_\lambda \in \mathbb{Q}[\Sigma_k]$, known as a *Young symmetrizer*, acting on $V^{\otimes k}$ (see [Fulton and Harris 1991] for a definition). This characterizes $\mathbb{S}_\lambda(V)$ as a subrepresentation of $V^{\otimes k}$.

If W is a finite-dimensional graded Σ_k–representation, we more generally define its associated Schur functor by

$$V \mapsto W \otimes_{\Sigma_k} V^{\otimes k},$$

and denote it by $\mathbb{S}_W$. Note that since W is finite-dimensional, this functor decomposes as a direct sum of Schur functors $\mathbb{S}_\lambda$ (possibly shifted in degree).

Even more generally, given a symmetric sequence $W = (W(1), W(2), \dots)$ of (graded) vector spaces, ie a sequence in which $W(k)$ is a graded Σ_k–representation, we can associate to it the endofunctor $\bigoplus_{k \geq 1} \mathbb{S}_{W(k)} \circ \mathcal{V}$ of $\mathrm{grVect}_\mathbb{Q}$, which we will denote by $\mathbb{S}_W$ and call the Schur functor associated to W.

Schur functors are of interest to us, since they preserve stability degree and weight in the following way:

Proposition 2.20 *Let $W = (W(1), W(2), \dots)$ be a symmetric sequence of graded vector spaces, where each $W(k)$ is finite-dimensional and concentrated in nonnegative degree, and let $\mathcal{V} \colon \mathrm{FI} \to \mathrm{grVect}_\mathbb{Q}$ be a graded* FI*–module such that $\mathcal{V}(S)$ is concentrated in strictly positive degrees for every $S \in \mathrm{FI}$. Suppose that $\mathcal{V}(S)$ is finite-dimensional in each degree and that* $\mathrm{weight}(\mathcal{V}_i) \leq s$ *and* $\text{stab-deg}(\mathcal{V}_i) \leq r$ *for all $i \leq m$. Then* $\mathrm{weight}((\mathbb{S}_W \circ \mathcal{V})_m) \leq ms$ *and* $\text{stab-deg}((\mathbb{S}_W \circ \mathcal{V})_m) \leq \max(r+s, ms)$.

Proof By definition $\mathbb{S}_W \circ \mathcal{V}$ decomposes as the direct sum

$$\bigoplus_{k \geq 1} \mathbb{S}_{W(k)} \circ \mathcal{V},$$

and we may decompose each summand further as

$$\mathbb{S}_{W(k)} \circ \mathcal{V} = \bigoplus_{j \geq 0} \bigoplus_{i \geq 1} W(k)_j \otimes_{\Sigma_k} (\mathcal{V}^{\otimes k})_i.$$

Since $W(k)$ is concentrated in nonnegative degree and $\mathcal{V}$ is concentrated in positive degree, it follows that $\mathbb{S}_{W(k)} \circ \mathcal{V}$ is concentrated in degrees $\geq k$. We thus have

$$(\mathbb{S}_W \circ \mathcal{V})_m = \bigoplus_{k=1}^{m} \bigoplus_{i+j=m} W(k)_j \otimes_{\Sigma_k} (\mathcal{V}^{\otimes k})_i.$$

By Corollary 2.17, it thus suffices to find bounds on the weight and stability degree of $W(k)_j \otimes_{\Sigma_k} (\mathcal{V}^{\otimes k})_i$ for all $k \leq m$ and all i and j such that $i + j = m$. By definition, this is a quotient of the FI–module

$W(k)_j \otimes (\mathcal{V}^{\otimes k})_i$. Since $W(k)_j$ is a constant FI–module and $(\mathcal{V}^{\otimes k})_i$ decomposes as a direct sum of summands of the form $\mathcal{V}_{l_1} \otimes \cdots \otimes \mathcal{V}_{l_k}$ such that $l_1 + \cdots + l_k = i$, it follows by Propositions 2.15 and 2.16 that $\operatorname{weight}(W(k)_j \otimes (\mathcal{V}^{\otimes k})_i) \le is$ and $\operatorname{surj-deg}(W^j \otimes (\mathcal{V}^{\otimes k})_i) \le \max(r+s, is)$.

By the discussion above, we also have that $W(k)_j \otimes_{\Sigma_k} (\mathcal{V}^{\otimes k})_i$ is isomorphic to a direct sum of graded FI–submodules of $(\mathcal{V}^{\otimes k}[j])_i$ (by decomposing $W(k)_j$ into irreducible Σ_k–representations and applying the corresponding Young symmetrizer for each summand), where $[j]$ denotes a shift of j degrees upwards. Thus we get the same bound on injectivity degree, finishing the proof, since $i \le m$. □

2.7 Derivation Lie algebras as FI–Lie algebras

Now let us introduce more specific examples of FI–modules that will be of interest to us. Here it will be useful to work with FI$^\#$–modules. We make the following definition:

Definition 2.21 Let H be a graded vector space. We define a graded FI$^\#$–module $\mathcal{H}$ by letting $\mathcal{H}(S) := H^{\oplus S}$ for any $S \in \mathrm{FI}$, and for any $A \subset S$, $B \subset T$ and bijection $f\colon A \to B$ we define a linear map $\mathcal{H}(f)\colon \mathcal{H}(S) \to \mathcal{H}(T)$ as the composition

$$\mathcal{H}(S) \twoheadrightarrow \mathcal{H}(A) \to \mathcal{H}(B) \hookrightarrow \mathcal{H}(T),$$

where the first map is the natural projection, the second is the map induced by f and the last is the natural injection.

In the following sections H will be the desuspension of the reduced homology of a simply connected finite CW–complex, so that its homology is finite-dimensional. We then have $\operatorname{weight}(\mathcal{H}) = 1$, since $H^{\oplus S}$ decomposes into a direct sum of trivial and standard representations of $\Sigma(S)$, which correspond to the padded partitions $\lambda[|S|]$ of $\lambda = (1)$ and $\lambda = (0)$, respectively. It is also easily verified that $\operatorname{stab-deg}(\mathcal{H}) = 1$.

Composing with the free graded Lie algebra functor $\mathbb{L}$, we get a new graded FI$^\#$–module, which we denote by $\mathbb{L}\mathcal{H}$.

Since $\mathcal{H}$ is an FI$^\#$–module, we may consider its dual FI$^\#$–module $\mathcal{H}^*$. Let us describe it in some more detail. For a finite set S we simply have $\mathcal{H}^*(S) = \mathcal{H}(S)^* = (H^*)^{\oplus S}$, and if $S \supseteq A \xrightarrow{f} B \subseteq T$ is a partial injection then $\mathcal{H}^*(S \supseteq A \xrightarrow{f} B \subseteq T)$ is the composition

$$\mathcal{H}^*(S) \twoheadrightarrow \mathcal{H}^*(A) \xrightarrow{\circ H(f^{-1})} \mathcal{H}^*(B) \hookrightarrow \mathcal{H}^*(T).$$

Remark 2.22 If we restrict this FI$^\#$–module to FI and $i\colon S \hookrightarrow T$ is an injection, we can describe the map $\mathcal{H}^*(i)$ as follows: Let $\phi \in \mathcal{H}^*(S)$ and x_α be in the summand of $H^{\oplus T}$ corresponding to $\alpha \in T$. Then

$$(\mathcal{H}^*(i)(\phi))(x_\alpha) = \begin{cases} 0 & \text{if } \alpha \in T \setminus i(S), \\ (\phi \circ \mathcal{H}(i)^{-1})(x_\alpha) & \text{if } \alpha \in i(S). \end{cases} \tag{1}$$

Just as for $\mathcal{H}$, the following proposition is easily verified:

Proposition 2.23 *If H is a finite-dimensional graded vector space, then the graded FI$^\#$–module $\mathcal{H}^*$ has weight ≤ 1 and stability degree ≤ 1.*

Next, we will define the graded $\mathrm{FI}^{\#}$–Lie algebra of derivations on the graded $\mathrm{FI}^{\#}$–Lie algebra $\mathbb{L}\mathcal{H}$. Recall that if L is a graded Lie algebra, we define a derivation on L as a (graded) linear map $D\colon L \to L$ which satisfies

$$D[x,y] = [Dx,y] + (-1)^{|x||D|}[x,Dy]$$

for all $x, y \in L$. We denote the graded vector space of all derivations by $\mathrm{Der}(L)$.

Definition 2.24 We define the graded FI–module $\mathrm{Der}(\mathbb{L}\mathcal{H})\colon \mathrm{FI} \to \mathrm{grVect}_{\mathbb{Q}}$ by letting $\mathrm{Der}(\mathbb{L}\mathcal{H})(S) = \mathrm{Der}(\mathbb{L}\mathcal{H}(S))$ for $S \in \mathrm{FI}$, and for $i\colon S \hookrightarrow T$ an injection we define $\mathrm{Der}(\mathbb{L}\mathcal{H})(i)$ as follows: Recall that a derivation on $\mathbb{L}\mathcal{H}(T)$ is uniquely determined by its restriction to $\mathcal{H}(T)$. Suppose therefore that $x_\alpha \in \mathcal{H}(T)$ lies in the direct summand of $\mathcal{H}(T)$ corresponding to $\alpha \in T$ and let $D \in \mathrm{Der}(\mathbb{L}(H^{\oplus S}))$. Then $\mathrm{Der}(\mathbb{L}\mathcal{H})(i)D$ is determined by

$$(\mathrm{Der}(\mathbb{L}\mathcal{H})(i)D)(x_\alpha) = \begin{cases} 0 & \text{if } \alpha \in T \setminus i(S), \\ (\mathbb{L}\mathcal{H}(i) \circ D \circ \mathcal{H}(i)^{-1})(x_\alpha) & \text{if } \alpha \in i(S). \end{cases} \tag{2}$$

Remark 2.25 The functor $\mathrm{Der}(\mathbb{L}\mathcal{H})$ may in fact be extended to all of $\mathrm{FI}^{\#}$ using a similar definition, but since we will not be using this we only consider the simpler functor from $\mathrm{FI}^{\#}$.

For a graded Lie algebra L, the commutator Lie bracket

$$[D,D'] = D \circ D' - (-1)^{|D||D'|} D' \circ D$$

makes $\mathrm{Der}(L)$ into a graded Lie algebra. A straightforward computation using (2) shows that

$$\mathrm{Der}(\mathbb{L}\mathcal{H})(i)[D,D'] = [\mathrm{Der}(\mathbb{L}\mathcal{H})(i)(D), \mathrm{Der}(\mathbb{L}\mathcal{H})(i)(D')],$$

giving us the following result:

Proposition 2.26 *The functor* $\mathrm{Der}(\mathbb{L}\mathcal{H})\colon \mathrm{FI} \to \mathrm{grVect}_{\mathbb{Q}}$ *of Definition 2.24 factors through the forgetful functor* $\mathrm{grLie}_{\mathbb{Q}} \to \mathrm{grVect}_{\mathbb{Q}}$, *where* $\mathrm{grLie}_{\mathbb{Q}}$ *is the category of graded Lie algebras over* $\mathbb{Q}$.

Further, we can determine explicit weights and stability degrees in each degree of this graded FI–module:

Proposition 2.27 *Let H be a finite-dimensional graded vector space concentrated in strictly positive degrees. If the degree of H is bounded strictly below d, for some $d \geq 1$, we have* $\mathrm{weight}(\mathrm{Der}(\mathbb{L}\mathcal{H})_m) \leq m + d$ *and* $\mathrm{stab\text{-}deg}(\mathrm{Der}(\mathbb{L}\mathcal{H})_m) \leq m + d$.

Proof For every $S \in \mathrm{FI}$, we have an isomorphism of graded vector spaces

$$\Psi_S\colon \mathcal{H}^*(S) \otimes \mathbb{L}\mathcal{H}(S) \xrightarrow{\cong} \mathrm{Der}(\mathbb{L}\mathcal{H})(S),$$

given by sending $\phi \otimes A \in \mathcal{H}^*(S) \otimes \mathbb{L}\mathcal{H}(S)$ to the derivation in $\mathrm{Der}(\mathbb{L}\mathcal{H})(S)$ defined by

$$x \mapsto \phi(x)A$$

on $x \in \mathcal{H}(S)$. We want to prove that this defines a map of graded FI–modules, ie that for every morphism $i\colon S \hookrightarrow T$, the diagram

$$\begin{array}{ccc} \mathcal{H}^*(S) \otimes \mathbb{L}\mathcal{H}(S) & \xrightarrow{\Psi_S} & \mathrm{Der}(\mathbb{L}\mathcal{H})(S) \\ \Big\downarrow{\scriptstyle \mathcal{H}^*(i)\otimes\mathbb{L}\mathcal{H}(i)} & & \Big\downarrow{\scriptstyle \mathrm{Der}(\mathbb{L}\mathcal{H})(i)} \\ \mathcal{H}^*(T) \otimes \mathbb{L}\mathcal{H}(T) & \xrightarrow{\Psi_T} & \mathrm{Der}(\mathbb{L}\mathcal{H})(T) \end{array} \tag{3}$$

is commutative. This can be verified by applying the definitions of Ψ_S and Ψ_T, together with the description of $\mathcal{H}^*(i)$ given by (1) and the description of $\mathrm{Der}(\mathbb{L}\mathcal{H})(i)$ given by (2).

Thus $\mathrm{Der}(\mathbb{L}\mathcal{H}) \cong \mathcal{H}^* \otimes \mathbb{L}\mathcal{H}$, as graded FI–modules. Note that $\mathcal{H}^*$ is concentrated in *negative* degrees, which are bounded from below, by the assumption on H. We thus have

$$(\mathcal{H}^* \otimes \mathbb{L}\mathcal{H})_m = \bigoplus_{i=1}^{d-1} (\mathcal{H}^*)_{-i} \otimes (\mathbb{L}\mathcal{H})_{m+i}.$$

Since both $\mathcal{H}$ and $\mathcal{H}^*$ are of weight and stability degree 1, the same argument as in the proof of Proposition 2.20 shows that $(\mathcal{H}^*)_{-i} \otimes (\mathbb{L}\mathcal{H})_{m+i}$ is simultaneously a quotient of an FI–module of weight and stability degree $\le m+i+1$, so $(\mathcal{H}^* \otimes \mathbb{L}\mathcal{H})_m$ thus has both weight and stability degree $m+d$, due to Propositions 2.16 and 2.17. □

The FI–modules that we consider in Theorem A are the homology groups of graded FI–modules of the type $\mathrm{Der}(\mathbb{L}\mathcal{H})$, with H as above, so it will follow immediately from Proposition 2.18 that we get the claimed bounds on weight and stability degree. In the case of Theorem B, it turns out that we can use Proposition 2.20 more directly, due to results from [Berglund and Madsen 2020].

2.8 FI–Lie models

Now, let us introduce the notion of an FI–*Lie model*, which will be one of our main tools. For the basic theory of Lie models in rational homotopy theory, see for example [Félix et al. 2001].

Definition 2.28 Let $\mathcal{X}$ be a simply connected based FI–space and let $\mathcal{L}$ be a dg FI–Lie algebra. We say that $\mathcal{L}$ is an FI–Lie model for $\mathcal{X}$ if

(i) for every $S \in \mathrm{FI}$, $\mathcal{L}(S)$ is a dg Lie model for the space $\mathcal{X}(S)$, and

(ii) for every morphism $S \hookrightarrow T$ in FI, the dgl map

$$\mathcal{L}(S) \to \mathcal{L}(T)$$

is a model for the map $\mathcal{X}(S) \to \mathcal{X}(T)$.

Remark 2.29 If $\mathcal{L}$ is an FI–Lie model for $\mathcal{X}$, then $H_*(\mathcal{L}) \cong \pi_*^{\mathbb{Q}}(\mathcal{X})$ is an isomorphism of FI–modules.

Remark 2.30 A reader may feel that Definition 2.28 is somewhat unnatural. Indeed, it is not the "philosophically" correct definition of FI–Lie model, seen from a modern homotopy-theoretic perspective. There is an equivalence of ∞–categories

$$(\mathrm{dgLie}_{\mathbb{Q}})_{\geq 1} \cong \mathrm{Top}^{\mathbb{Q}}_{\geq 2},$$

between the ∞–categories of connected dg Lie algebras, localized at the quasi-isomorphisms, and simply connected spaces, localized at the rational homotopy equivalences. The usual definition of dg Lie models in rational homotopy theory is that a connected dg Lie algebra (L, d) is a dg Lie model for a simply connected space X if they are isomorphic under this equivalence. Equivalently, it suffices to require that they are isomorphic under the equivalence between the homotopy categories $h(\mathrm{dgLie}_{\mathbb{Q}})_{\geq 1} \cong h\,\mathrm{Top}^{\mathbb{Q}}_{\geq 2}$. The correct definition of FI–Lie model should therefore be that a dg FI–Lie algebra $\mathcal{L}$ is an FI–Lie model of a simply connected pointed FI–space $\mathcal{X}$ if they are isomorphic under the equivalence of the homotopy categories

$$h\,\mathrm{Fun}(\mathrm{FI}, (\mathrm{dgLie}_{\mathbb{Q}})_{\geq 1}) \cong h\,\mathrm{Fun}(\mathrm{FI}, \mathrm{Top}^{\mathbb{Q}}_{\geq 2}).$$

In contrast, our definition is requiring isomorphism under the equivalence of "ordinary" functor categories

$$\mathrm{Fun}(\mathrm{FI}, h(\mathrm{dgLie}_{\mathbb{Q}})_{\geq 1}) \cong \mathrm{Fun}(\mathrm{FI}, h\,\mathrm{Top}^{\mathbb{Q}}_{\geq 2}).$$

Nevertheless, the naive Definition 2.28 is simpler and sufficient for our purposes here.

3 Rational homotopy theory for homotopy automorphisms

In this section we will review some rational homotopy theory for homotopy automorphisms we will need.

Let X be a simply connected topological space homotopy equivalent to a CW–complex. A homotopy automorphism of X is a self-map $\varphi\colon X \to X$ that is a homotopy equivalence. We denote the topological monoid of unpointed and pointed homotopy automorphisms of X by $\mathrm{aut}(X)$ and $\mathrm{aut}_*(X)$, respectively. Given a subspace $A \subset X$, we denote the topological monoid of A–relative homotopy automorphisms of X, ie the homotopy automorphisms that preserve A pointwise, by $\mathrm{aut}_A(X)$. When A is a point or empty we simply write $\mathrm{aut}_*(X)$ and $\mathrm{aut}(X)$, respectively, and when $X = N$ is a manifold with boundary $A = \partial N$, the monoid of boundary-relative homotopy automorphisms of N is denoted by $\mathrm{aut}_\partial(N)$.

If X is well pointed and $A \subset X$ is a cofibration of cofibrant spaces in the Hurewicz model structure, then all of $\mathrm{aut}(X)$, $\mathrm{aut}_*(X)$ and $\mathrm{aut}_A(X)$ are group-like monoids, and thus equivalent to topological groups. We take the basepoint of a topological monoid G to be the identity element and $\pi_k(G, \mathrm{id})$ is abbreviated by $\pi_k(G)$. We denote the classifying space of G by BG and its universal cover by $\widetilde{BG}$. Moreover, if a topological monoid G is group-like, then G and ΩBG are weakly equivalent as topological monoids. Let $G_\circ \subset G$ denote the connected component of the identity. Then $BG_\circ \simeq \widetilde{BG}$. We observe that

$$\pi_k(G) \otimes \mathbb{Q} \cong \pi_{k+1}(\widetilde{BG}) \otimes \mathbb{Q} \cong \pi_{k+1}(BG_\circ) \otimes \mathbb{Q} \cong H_k(\mathfrak{g}_{BG_\circ})$$

for all $k \geq 1$ and where $\mathfrak{g}_{BG_\circ}$ is any dg Lie algebra model for $BG_\circ$.

The identity component of $\mathrm{aut}_A(X)$ is denoted by $\mathrm{aut}_{A,\circ}(X)$.

Remark 3.1 By [Farjoun 1996], there are functorial and continuous rationalization functors that preserve cofibrations. In particular, given a cofibration $A \subset X$, there is a rationalization functor that induces a group homomorphism $r\colon \pi_0(\mathrm{aut}_A(X)) \to \pi_0(\mathrm{aut}_{A_{\mathbb{Q}}}(X_{\mathbb{Q}}))$.

For $k \geq 1$ we have that

$$\pi_k(\mathrm{aut}_A(X)) \otimes \mathbb{Q} \cong \pi_k(\mathrm{aut}_{A_{\mathbb{Q}}}(X_{\mathbb{Q}})),$$

since $B\,\mathrm{aut}_{A,\circ}(X)_{\mathbb{Q}} \simeq B\,\mathrm{aut}_{A_{\mathbb{Q}},\circ}(X_{\mathbb{Q}})$; see [Berglund and Saleh 2020, Proposition 2.4].

A model for $B\,\mathrm{aut}_{A,\circ}(X)$ is given in terms of dg Lie algebras of derivations.

Definition 3.2 Given a dg Lie algebra L, let $\mathrm{Der}(L)$ denote the dg Lie algebra of derivations of L, where the graded Lie bracket is given by

$$[\theta, \eta] = \theta \circ \eta - (-1)^{|\theta||\eta|}\eta \circ \theta$$

and the differential is given by $\partial = [d_L, -]$ where d_L is the differential of L.

Definition 3.3 Given a chain complex $C = C_*$, the positive truncation of C, denoted by C^+, is given by

$$C_i^+ = \begin{cases} C_i & \text{if } i > 1, \\ \ker(C_1 \xrightarrow{d} C_0) & \text{if } i = 1, \\ 0 & \text{if } i < 1. \end{cases}$$

Definition 3.4 A dg Lie algebra $(\mathbb{L}(V), d)$ is called *quasifree* if its underlying graded Lie algebra structure is a free graded Lie algebra on the graded vector space V.

Definition 3.5 We say that a dg Lie algebra map between two quasifree dg Lie algebras $\phi\colon \mathbb{L}(V) \to \mathbb{L}(U)$ is free if ϕ is injective and $\phi(V) \subseteq U$. In particular U has a decomposition $U \cong V \oplus W$.

Remark 3.6 One can show that the free maps between the quasifree dg Lie algebras are exactly the cofibrant maps between them; see the remark after [Quillen 1969, Proposition 5.5].

Proposition 3.7 (a) *Let X be a simply connected space of the homotopy type of a finite CW–complex with a quasifree dg Lie algebra model $\mathbb{L}_X$. A dg Lie model for $B\,\mathrm{aut}_{*,\circ}(X)$ is given by $\mathrm{Der}^+(\mathbb{L}_X)$.*

(b) *Let $A \subset X$ be a cofibration of simply connected spaces of the homotopy type of finite CW–complexes, and let $\mathbb{L}_A \to \mathbb{L}_X$ be a cofibration (ie a free map) of quasifree dg Lie algebras that models the inclusion $A \subset X$. A dg Lie model for $B\,\mathrm{aut}_{A,\circ}(X)$ is given by the positive truncation of the dg Lie algebra of derivations on $\mathbb{L}_X$ that vanish on $\mathbb{L}_A$, denoted by $\mathrm{Der}^+(\mathbb{L}_X \,\|\, \mathbb{L}_A)$.*

(c) *The inclusion $\mathrm{Der}^+(\mathbb{L}_X \,\|\, \mathbb{L}_A) \to \mathrm{Der}^+(\mathbb{L}_X)$ is a model for $B\,\mathrm{aut}_{A,\circ}(X) \to B\,\mathrm{aut}_{*,\circ}(X)$ induced by the inclusion $\mathrm{aut}_{A,\circ}(X) \hookrightarrow \mathrm{aut}_{*,\circ}(X)$.*

Proof For (a), see [Tanré 1983, corollarie VII.4(4)]. For (b), see [Berglund and Saleh 2020, Theorem 1.1]. Statement (c) follows by [loc. cit., Proposition 4.6] and the theory established in [Berglund 2020, Sections 3.4 and 3.5]. □

We recall the notion of geometric realizations of dg Lie algebras. For a detailed account on the subject we refer the reader to [Hinich 1997; Getzler 2009; Berglund 2015; 2020].

Definition 3.8 [Hinich 1997, Definition 2.1.1] Let $\Omega_\bullet = \Omega_\bullet^*$ denote the simplicial commutative dg algebra in which Ω_n^* is the Sullivan–de Rham algebra of polynomial differential forms on the n–simplex. The geometric realization of a positively graded dg Lie algebra L is defined to be the simplicial set $\mathrm{MC}(L \otimes \Omega_\bullet)$ of Maurer–Cartan elements of the simplicial dg Lie algebra $L \otimes \Omega_\bullet$, denoted by $\mathrm{MC}_\bullet(L)$. We recall that the tensor product $L \otimes \Omega$ of a dg Lie algebra L with a commutative dg algebra Ω is again a dg Lie algebra, where $[\ell_1 \otimes c_1, \ell_2 \otimes c_2] = (-1)^{|c_1||\ell_2|}[\ell_1, \ell_2] \otimes c_1 c_2$. A positively graded dg Lie algebra L is a Lie model for a simply connected space X if and only if there exists a zigzag of rational homotopy equivalences between the geometric realization $\mathrm{MC}_\bullet(L)$ and X.

The functor $\mathrm{MC}_\bullet$ takes surjections to Kan fibrations [Getzler 2009, Proposition 4.7] and takes injections to cofibrations (in the classical model structure on simplicial sets). In particular, if $\mathbb{L}_A \to \mathbb{L}_X$ is a free map of dg Lie algebras that models a cofibration $A \subset X$, then the cofibration $\mathrm{MC}_\bullet(\mathbb{L}_A) \hookrightarrow \mathrm{MC}_\bullet(\mathbb{L}_X)$ is a simplicial model for the cofibration $A_\mathbb{Q} \subset X_\mathbb{Q}$. Thus $\mathrm{aut}_{A_\mathbb{Q}}(X_\mathbb{Q})$ and $\mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}_A)}(\mathrm{MC}_\bullet(\mathbb{L}_X))$ are weakly equivalent as topological monoids.

Definition 3.9 The exponential $\exp(\mathfrak{h})$ of a nilpotent Lie algebra $\mathfrak{h}$ concentrated in degree zero is the nilpotent group with underlying set given by $\mathfrak{h}$ and multiplication given by the Baker–Campbell–Hausdorff formula. The exponential of a positively graded dg Lie algebra L, denoted by $\exp_\bullet(L)$, is the simplicial group given by the exponential $\exp(Z_0(L \otimes \Omega_\bullet))$ of the zero cycles in $L \otimes \Omega_\bullet$; see [Berglund 2020].

Proposition 3.10 [loc. cit., Corollary 3.10] *For a positively graded dg Lie algebra L there is an equivalence of topological monoids between $\exp_\bullet(L)$ and the loop space $\Omega\,\mathrm{MC}_\bullet(L)$.*

Definition 3.11 Let $\mathbb{L}(V) \subset \mathbb{L}(V \oplus W)$ be a cofibration of free positively graded dg Lie algebras and let $\mathrm{Der}(\mathbb{L}(V \oplus W) \,\|\, \mathbb{L}(V))$ denote the dg Lie algebra of derivations on $\mathbb{L}(V \oplus W)$ that vanish on $\mathbb{L}(V)$; the differential is $[d_{\mathbb{L}(V\oplus W)}, -]$. There is a left action of $\exp_\bullet\big(\mathrm{Der}^+(\mathbb{L}(V \oplus W) \,\|\, \mathbb{L}(V))\big)$ on $\mathrm{MC}_\bullet(\mathbb{L}(V \oplus W))$ given by

$$\Theta.x = \sum_{i\geq 0} \frac{\Theta^i(x)}{i!}. \tag{4}$$

See [Berglund and Saleh 2020, Section 3.2].

Proposition 3.12 [Berglund 2022, Proposition 3.7] *Let $A \subset X$ be a cofibration of simply connected spaces with homotopy types of finite CW–complexes, and let $\iota\colon \mathbb{L}(V) \to \mathbb{L}(V \oplus W)$ be a free map of quasifree dg Lie algebras that models the inclusion $A \subset X$. Then the topological monoid map*

$$F\colon \exp_\bullet\big(\mathrm{Der}^+(\mathbb{L}(V \oplus W) \,\|\, \mathbb{L}(V))\big) \to \mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V)),\circ}\big(\mathrm{MC}_\bullet(\mathbb{L}(V \oplus W))\big) \simeq \mathrm{aut}_{A_\mathbb{Q},\circ}(X_\mathbb{Q}),$$
$$F(\Theta)(x) = \Theta.x,$$

is a weak equivalence.

Proof Note that the action of $\exp_\bullet\big(\mathrm{Der}^+(\mathbb{L}(V\oplus W)\,\|\,\mathbb{L}(V))\big)$ on $\mathrm{MC}_\bullet(\mathbb{L}(V\oplus W))$ fixes $\mathrm{MC}_\bullet(\mathbb{L}(V))\subset \mathrm{MC}_\bullet(\mathbb{L}(V\oplus W))$ pointwise. In particular, the group action yields a map

$$\exp_\bullet\big(\mathrm{Der}^+(\mathbb{L}(V\oplus W)\,\|\,\mathbb{L}(V))\big)\to \mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V))}\big(\mathrm{MC}_\bullet(\mathbb{L}(V\oplus W))\big).$$

Moreover, since $\exp_\bullet\big(\mathrm{Der}^+(\mathbb{L}(V\oplus W)\,\|\,\mathbb{L}(V))\big)$ is connected and F preserves the identity element, we may replace the codomain by

$$\mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V)),\circ}\big(\mathrm{MC}_\bullet(\mathbb{L}(V\oplus W))\big),$$

ie the component of the identity. We proceed by adapting the proof of [loc. cit., Proposition 3.7] to our situation. Given a positively graded dg Lie algebra $\mathfrak{h}$, there is an isomorphism of abelian groups

$$G\colon H_k(\mathfrak{h})\to \pi_k(\exp_\bullet(\mathfrak{h}))$$

where a homology class of a cycle $z\in Z_k(\mathfrak{h})$ is sent to the homotopy class of the k–simplex $z\otimes\nu_k\in Z_0(\mathfrak{h}\otimes\Omega_k^*)$, where ν_k is the class $k!dt_1\cdots dt_k$. That G defines an isomorphism is motivated in the proof of [loc. cit., Proposition 3.7].

We have that $\nu_k^2=0$, and consequently

$$F(\theta\otimes\nu_k)=\mathrm{id}+\theta\otimes\nu_k.$$

Let us now analyze $\pi_k\big(\mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V)),\circ}\big(\mathrm{MC}_\bullet(\mathbb{L}(V\oplus W))\big)\big)$ for $k\ge 1$, as in the proof of [Berglund and Madsen 2020, Theorem 3.6]. In order to simplify notation, $\mathrm{MC}_\bullet(\mathbb{L}(V))$ is denoted by $A_{\mathbb{Q}}$ and $\mathrm{MC}_\bullet(\mathbb{L}(V\oplus W))$ by $X_{\mathbb{Q}}$. We have that an element $f\in\pi_k(\mathrm{aut}_{A_{\mathbb{Q}},\circ}(X_{\mathbb{Q}}))$ is represented by a map

$$f\colon (S^k\sqcup *)\wedge X_{\mathbb{Q}}\to X_{\mathbb{Q}},$$

where $f(*,x)=x$ for every $x\in X_{\mathbb{Q}}$ and $f(s,a)=a$ for every $a\in A_{\mathbb{Q}}$ and $s\in S^k$.

A dg Lie algebra model for $(S^k\sqcup *)\wedge X_{\mathbb{Q}}$ is given by $(\mathbb{L}(U\oplus s^kU),\partial)$ where $U=V\oplus W$ and with a differential determined by the following: Let d be the differential on $\mathbb{L}(U)$. Then $\partial(u)=d(u)$ for every $u\in U$ and $\partial(s^ku)=(-1)^ks^kd(u)$ for every $s^ku\in s^kU$.

Now, $f\colon (S^k\sqcup *)\wedge X_{\mathbb{Q}}\to X_{\mathbb{Q}}$ is modeled by some map $\varphi_f\colon \mathbb{L}(U\oplus s^kU)\to\mathbb{L}(U)$ that satisfies $\varphi_f(u)=u$ for every $u\in U$ and $\varphi_f(s^kv)=0$ for every $v\in V\subset U$. Let θ_f be the unique derivation on $\mathbb{L}(U)$ that satisfies $\theta_f(u)=\varphi_f(s^ku)$ for every $u\in U$. Note that θ_f is a cycle and that it vanishes on $\mathbb{L}(V)$, ie $\theta_f\in Z_k\big(\mathrm{Der}^+(\mathbb{L}(U)\,\|\,\mathbb{L}(V))\big)$. Also note that if $f=\pi_k(F)[\theta\otimes\nu_k]$ then $\theta_f=\theta$. Let $K\colon \pi_k(\mathrm{aut}_{A_{\mathbb{Q}},\circ}(X_{\mathbb{Q}}))\to H_k\big(\mathrm{Der}^+(\mathbb{L}(U)\,\|\,\mathbb{L}(V))\big)$ be given by $K(f)=\theta_f$. It follows from [Berglund and Madsen 2020; Lupton and Smith 2007] that this map is well defined and is an isomorphism.

Set $\mathfrak{h}=\mathrm{Der}^+(\mathbb{L}(U)\,\|\,\mathbb{L}(V))$. The composition

$$H_k(\mathfrak{h})\xrightarrow{G}\pi_k(\exp_\bullet(\mathfrak{h}))\xrightarrow{\pi_k(F)}\pi_k(\mathrm{aut}_{A_{\mathbb{Q}},\circ}(X_{\mathbb{Q}}))\xrightarrow{K}H_k(\mathfrak{h})$$

is the identity map, which forces $\pi_k(F)$ to be an isomorphism. This proves that F is a weak equivalence. □

Recall that a topological group G' acts on itself by conjugation $G' \to \mathrm{Aut}(G')$ via $g \mapsto \kappa_g$ where $\kappa_g(h) = ghg^{-1}$. If g and g' belong to the same connected component of G', then κ_g and $\kappa_{g'}$ are homotopic and this induce equal maps on the homotopy groups of G'. This group action restricts to a group action on the identity component $G'_\circ$ of G', which in turn induces an action of G' on $BG'_\circ$. This gives that $\pi_0(G')$ acts on $\pi_*(BG'_\circ)$. Since a group-like monoid G is equivalent to a topological group G', we have that $\pi_0(G)$ acts on $\pi_*(G_\circ)$.

In the rest of this section we discuss the action of $\pi_0(\mathrm{aut}_A(X))$ on $\pi_k(\mathrm{aut}_A(X))$ from a rational homotopy point of view. To do so we recall some theory.

Proposition 3.13 [Espic and Saleh 2020, Theorem 1.3] *Given a map $f\colon \mathbb{L}(V) \to \mathfrak{g}$ of positively graded dg Lie algebras, there exists a minimal relative model $q\colon \mathbb{L}(V \oplus W) \xrightarrow{\sim} \mathfrak{g}$ for f in the following sense:*

(a) *$\mathbb{L}(V)$ is a dg subalgebra of $\mathbb{L}(V \oplus W)$ and $f = q \circ \iota$, where $\iota\colon \mathbb{L}(V) \to \mathbb{L}(V \oplus W)$ is the inclusion.*

(b) *Given a quasi-isomorphism $g\colon \mathbb{L}(V \oplus W) \to \mathbb{L}(V \oplus W)$, where g restricts to an automorphism of $\mathbb{L}(V)$, g is an automorphism.*

Definition 3.14 Let $\iota\colon \mathbb{L}(V) \to \mathbb{L}(V \oplus W)$ be a free map of quasifree dg Lie algebras. We say that an endomorphism $\varphi\colon \mathbb{L}(V \oplus W) \to \mathbb{L}(V \oplus W)$ is ι–relative if $\varphi|_{\mathbb{L}(V)} = \mathrm{id}$, and two ι–relative endomorphisms φ and ψ are ι–equivalent if there exists a homotopy $h\colon \mathbb{L}(V \oplus W) \to \mathbb{L}(V \oplus W) \otimes \Lambda(t, dt)$ from φ to ψ that preserves $\mathbb{L}(V)$ in the sense that $h(v) = v \otimes 1$ for every $v \in \mathbb{L}(V)$; see [Félix et al. 2001, Section 14(a)].

We denote the group of ι–relative automorphisms of $\mathbb{L}(V \oplus W)$ by $\mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$.

Lemma 3.15 [Espic and Saleh 2020, Corollary 4.6] *Let $\iota\colon \mathbb{L}(V) \to \mathbb{L}(V \oplus W)$ be a minimal relative dg Lie model for a cofibration $A \subset X$ of simply connected spaces. Then there are group isomorphisms*

$$\mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))/\iota\text{–equivalence} \cong \pi_0\big(\mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V))}\big(\mathrm{MC}_\bullet(\mathbb{L}(V \oplus W))\big)\big) \cong \pi_0(\mathrm{aut}_{A_\mathbb{Q}}(X_\mathbb{Q})).$$

Remark 3.16 By this lemma, it makes sense to refer to an ι–relative automorphisms of $\mathbb{L}(V \oplus W)$ as an algebraic model for an $A_\mathbb{Q}$–relative homotopy automorphism of $X_\mathbb{Q}$.

Definition 3.17 Consider the group action of $\mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$ on $\mathrm{Der}(\mathbb{L}(V \oplus W) \,\|\, \mathbb{L}(V))$ given by the following: for $\varphi \in \mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$ and $\theta \in \mathrm{Der}(\mathbb{L}(V \oplus W) \,\|\, \mathbb{L}(V))$, let

$$\varphi.\theta = \varphi \circ \theta \circ \varphi^{-1}.$$

This induces an action of $\mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$ on $\exp_\bullet\big(\mathrm{Der}(\mathbb{L}(V \oplus W) \,\|\, \mathbb{L}(V))\big)$.

There is also an action of $\mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$ on $\mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V)),\circ}\big(\mathrm{MC}_\bullet(\mathbb{L}(V \oplus W))\big)$; for $\varphi \in \mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$ and $f \in \mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V)),\circ}\big(\mathrm{MC}_\bullet(\mathbb{L}(V \oplus W))\big)$, let

$$\varphi.f = \mathrm{MC}_\bullet(\varphi) \circ f \circ \mathrm{MC}_\bullet(\varphi^{-1}).$$

Proposition 3.18 *The equivalence*

$$F : \exp_\bullet\big(\mathrm{Der}^+(\mathbb{L}(V \oplus W) \parallel \mathbb{L}(V))\big) \to \mathrm{aut}_{\mathrm{MC}_\bullet(\mathbb{L}(V)),\circ}\big(\mathrm{MC}_\bullet(\mathbb{L}(V \oplus W))\big) \simeq \mathrm{aut}_{A_{\mathbb{Q}},\circ}(X_{\mathbb{Q}})$$

of Proposition 3.12 is $\mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$*–equivariant with respect to the actions in Definition 3.17.*

Proof This is a straightforward verification left to the reader. □

Corollary 3.19 *Let* $f \in \mathrm{aut}_{A_{\mathbb{Q}}}(X_Q)$ *and let* $\varphi \in \mathrm{Aut}_\iota(\mathbb{L}(V \oplus W))$ *be a relative model for* f*. The automorphism*

$$\alpha_\varphi : \mathrm{Der}(\mathbb{L}(V \oplus W) \parallel \mathbb{L}(V)) \to \mathrm{Der}(\mathbb{L}(V \oplus W) \parallel \mathbb{L}(V)), \quad \alpha_\varphi(\theta) = \varphi \circ \theta \circ \varphi^{-1},$$

is a model for the delooping of the homotopy automorphism

$$\mathrm{Ad}_f : \mathrm{aut}_{A_{\mathbb{Q}}}(X_{\mathbb{Q}}) \to \mathrm{aut}_{A_{\mathbb{Q}}}(X_{\mathbb{Q}}), \quad \mathrm{Ad}_f(g) = f \circ g \circ f^{-1},$$

where f^{-1} *is an* $A_{\mathbb{Q}}$*–relative homotopy inverse to* f*.*

4 Homotopy automorphisms of wedge sums

We fix some notation for this section. Let $(X, *)$ be a fixed simply connected space, homotopy equivalent to a finite CW–complex. For any finite set S, let $X_S := \bigvee^S X$. For any morphism $S \hookrightarrow T$ in FI, there is an obvious induced basepoint-preserving map $X_S \hookrightarrow X_T$ given by inclusion of wedge summands in the order specified by the injection $S \hookrightarrow T$. Thus the functor $S \mapsto X_S$ is a pointed FI–space, which we will denote by $\mathcal{X}$.

We fix a quasifree dg Lie algebra model $\mathbb{L}(H) = (\mathbb{L}(H), d_{\mathbb{L}(H)})$ for X. A dg Lie model for X_S is given by the S–fold free product of dg Lie algebras

$$\mathbb{L}(H)^{*S} := \mathbb{L}(H) * \cdots * \mathbb{L}(H) \cong \mathbb{L}(H^{\oplus S}).$$

See [Félix et al. 2001, Section 24(f)]. The association $S \mapsto \mathbb{L}(H^{\oplus S})$ defines a dg FI–Lie algebra $\mathbb{L}\mathcal{H}$. Given a morphism $i : S \hookrightarrow T$ in FI, we get an induced inclusion $H^{\oplus S} \hookrightarrow H^{\oplus T}$, which induces an inclusion $\mathbb{L}(H^{\oplus S}) \hookrightarrow \mathbb{L}(H^{\oplus T})$ that models the map $\mathcal{X}(i) : \mathcal{X}(S) \to \mathcal{X}(T)$; this follows from eg Example 1 in Section 12(c) and Example 2 in Section 24(f) of [loc. cit.]. Thus $S \mapsto \mathbb{L}(H^{\oplus S})$ defines a dg FI–Lie model for the pointed FI–space $S \mapsto \mathcal{X}(S)$.

We proceed and define another pointed FI–space $\mathrm{aut}_*(\mathcal{X})$ (the basepoint is always the identity) as follows: For $S \in \mathrm{FI}$, we let $\mathrm{aut}_*(\mathcal{X})(S) := \mathrm{aut}_*(X_S)$. For $i : S \hookrightarrow T$ in FI we get a map $\mathrm{aut}_*(X_S) \hookrightarrow \mathrm{aut}_*(X_T)$, defined, for $x_\alpha \in X_T$ in the wedge summand of X_T corresponding to $\alpha \in T$ and $f \in \mathrm{aut}_*(X_S)$, by

$$(\mathrm{aut}_*(\mathcal{X})(i) f)(x_\alpha) = \begin{cases} x_\alpha & \text{if } \alpha \in T \setminus i(S), \\ (\mathcal{X}(i) \circ f \circ \mathcal{X}(i)^{-1})(x_\alpha) & \text{if } \alpha \in i(S). \end{cases}$$

Note that $\mathrm{aut}_*(\mathcal{X})(i) f$ is in some sense an extension by the identity of f. For instance, if $i_s : \boldsymbol{n} \to \boldsymbol{n+1}$ is the standard inclusion, then $\mathrm{aut}_*(\mathcal{X})(i_s) f$ is the homotopy automorphism of $X_{\boldsymbol{n+1}}$ that coincides with f on the first n wedge summands, and is the identity on the last summand.

Restricting to the identity component gives a pointed sub-FI–space $\mathrm{aut}_{*,\circ}(\mathcal{X})$. We are interested in the rational homotopy groups of this FI–space.

Remark 4.1 It is tempting to say that we will construct an FI–Lie model for $\mathrm{aut}_*(\mathcal{X})$. However, this pointed FI–space is generally not simply connected. Instead we take a functorial classifying space construction $B\colon \mathrm{TopMon} \to \mathrm{Top}_*$ from the category of topological monoids to the category of pointed topological spaces and consider the pointed FI–space $B\,\mathrm{aut}_{*,\circ}(\mathcal{X})$, where $\mathrm{aut}_{*,\circ}(X_S)$ is the identity component of $\mathrm{aut}_*(X_S)$ for every $S \in \mathrm{FI}$. For every S we have $B\,\mathrm{aut}_{*,\circ}(X_S) \simeq \widetilde{B\,\mathrm{aut}_*(X_S)}$, and so this is a simply connected pointed FI–space, which enables us to apply our tools from rational homotopy theory. Furthermore, for every $k \geq 1$

$$\pi_k^{\mathbb{Q}}(\mathrm{aut}_*(X_S)) \cong \pi_{k+1}^{\mathbb{Q}}(B\,\mathrm{aut}_{*,\circ}(X_S)),$$

so

$$\pi_k^{\mathbb{Q}}(\mathrm{aut}_*(\mathcal{X})) \cong \pi_{k+1}^{\mathbb{Q}}(B\,\mathrm{aut}_{*,\circ}(\mathcal{X})),$$

as FI–modules.

We have by Proposition 3.7(a) that a model for $B\,\mathrm{aut}_{*,\circ}(X_S)$ is given by $\mathrm{Der}^+(\mathbb{L}(H^{\oplus S}))$, with differential given by $[d_{\mathbb{L}(H^{\oplus S})}, -]$. The inclusion $\mathbb{L}\mathcal{H}(i)\colon \mathbb{L}(H^{\oplus S}) \hookrightarrow \mathbb{L}(H^{\oplus T})$ induces a graded Lie algebra map $\mathrm{Der}^+(\mathbb{L}(H^{\oplus S})) \hookrightarrow \mathrm{Der}^+(\mathbb{L}(H^{\oplus T}))$, as discussed in Proposition 2.26. Moreover, this map commutes with the differential, ie it is a dg Lie algebra map. This, together with Proposition 2.26, yields that we have a dg FI–Lie algebra $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$.

We will show that $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$ defines an FI–Lie model for the pointed FI–space $B\,\mathrm{aut}_{*,\circ}(\mathcal{X})$.

Proposition 4.2 (a) *Let $i_s\colon \boldsymbol{n} \to \boldsymbol{n+1}$ denote the standard inclusion. Then a dg Lie algebra model for*

$$B\,\mathrm{aut}_{*,\circ}(\mathcal{X})(i_s)\colon B\,\mathrm{aut}_{*,\circ}(X_n) \to B\,\mathrm{aut}_{*,\circ}(X_{n+1})$$

is given by

$$\varphi_n := \mathrm{Der}^+(\mathbb{L}\mathcal{H})(i_s)\colon \mathrm{Der}^+(\mathbb{L}(H^{\oplus n})) \to \mathrm{Der}^+(\mathbb{L}(H^{\oplus n+1})).$$

(b) *The Σ_n–action on $B\,\mathrm{aut}_{*,\circ}(X_n)$ is modeled by the Σ_n–action on* $\mathrm{Der}^+(\mathbb{L}(H^{\oplus n}))$.

Proof (a) To simplify notation, let $\mathbb{L}_k$ denote $\mathbb{L}(H^{\oplus k})$, let $H_l \cong H$ denote the last summand of $H^{\oplus n+1}$ and let $\mathbb{L}_l$ denote $\mathbb{L}(H_l)$. In particular, $\mathbb{L}_{n+1} = \mathbb{L}_n * \mathbb{L}_l$. Let $c_n\colon \mathrm{MC}_\bullet(\mathbb{L}_n) \to \mathrm{MC}_\bullet(\mathbb{L}_{n+1})$ and $c_l\colon \mathrm{MC}_\bullet(\mathbb{L}_l) \to \mathrm{MC}_\bullet(\mathbb{L}_{n+1})$ denote the cofibrations induced by the standard inclusions $\mathbb{L}_n \to \mathbb{L}_n * \mathbb{L}_l$ and $\mathbb{L}_l \to \mathbb{L}_n * \mathbb{L}_l$, respectively.

From Proposition 3.12 we get topological monoid equivalences

$$F_n\colon \exp_\bullet(\mathrm{Der}^+(\mathbb{L}_n)) \to \mathrm{aut}_*(\mathrm{MC}_\bullet(\mathbb{L}_n)).$$

Those maps have adjoints

$$\widetilde{F}_n\colon \exp_\bullet(\mathrm{Der}^+(\mathbb{L}_n)) \times \mathrm{MC}_\bullet(\mathbb{L}_n) \to \mathrm{MC}_\bullet(\mathbb{L}_n).$$

By the explicit formulas for $\{F_n\}$,

$$\widetilde{F}_{n+1} \circ (\exp_\bullet(\varphi_n) \times c_n) = c_n \circ \widetilde{F}_n.$$

In particular, $F_{n+1} \circ \exp_\bullet(\varphi_n)(\Theta)$ is an extension of $F_n(\Theta)$ for $\Theta \in \exp_\bullet(\mathrm{Der}^+(\mathbb{L}_n))$.

We also have that

$$\widetilde{F}_{n+1} \circ (\exp_\bullet(\varphi_n) \times c_l)(g,x) = c_l(x) \quad \text{for all } (g,x) \in \exp_\bullet(\mathrm{Der}^+(\mathbb{L}_n)) \times \mathrm{MC}_\bullet(\mathbb{L}_l).$$

In particular $F_{n+1} \circ \exp_\bullet(\varphi_n)(\Theta)$ restricts to the identity on $\mathrm{MC}_\bullet(\mathbb{L}_l) \subset \mathrm{MC}_\bullet(\mathbb{L}_{n+1})$. That means that $\exp_\bullet(\varphi_n)$ is a simplicial model for $\mathrm{aut}_{*,\circ}(\mathcal{X})(i_s)\colon \mathrm{aut}_{*,\circ}(X_n) \to \mathrm{aut}_{*,\circ}(X_{n+1})$. This gives (a).

(b) This is a direct consequence of Proposition 3.18 and Corollary 3.19. □

Theorem 4.3 *The dg* FI*–Lie algebra* $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$ *is an* FI*–Lie model for the pointed* FI*–space* $B\,\mathrm{aut}_{*,\circ}(\mathcal{X})$.

Proof By the second part of Lemma 2.6, an FI–module is completely determined its underlying consistent sequence. By Proposition 4.2, the stabilization maps and the Σ_n–actions defining the consistent sequence for the dg FI–Lie algebra $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$ models the stabilization maps and the Σ_n–actions defining the consistent sequence for the FI–space $B\,\mathrm{aut}_{*,\circ}(\mathcal{X})$. From this we conclude that $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$ is an FI–Lie model for the pointed FI–space $B\,\mathrm{aut}_{*,\circ}(\mathcal{X})$. □

We now have all the ingredients needed for proving Theorem A.

Theorem A *Let* $(X, *)$ *be a pointed simply connected space with the homotopy type of a finite CW–complex and let* $X_S := \bigvee^S X$ *for any finite set* S*. For each* $k \geq 1$*, the functor*

$$S \mapsto \pi_k^{\mathbb{Q}}(\mathrm{aut}_*(X_S))$$

is an FI*–module. If* $H_n(X, \mathbb{Q}) = 0$ *for* $n \geq d$*, this* FI*–module is of weight* $\leq k + d - 1$ *and stability degree* $\leq k + d$*.*

Proof We will use the established terminology in this section. We have already seen in Theorem 4.3 that $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$ is an FI–Lie model for $B\,\mathrm{aut}_{*,\circ}(\mathcal{X})$. Since $H_k(\mathrm{Der}^+(\mathbb{L}\mathcal{H})) \cong \pi_k^{\mathbb{Q}}(\mathrm{aut}_*(\mathcal{X}))$ (see Remark 4.1) it is enough to prove that $H_k(\mathrm{Der}^+(\mathbb{L}\mathcal{H}))$ has the stated bounds on weight and stability degree.

Since $(\mathrm{Der}^+(\mathbb{L}\mathcal{H}), [d_{\mathbb{L}\mathcal{H}}, -])$ defines a dg FI–Lie algebra, it follows that $H_*(\mathrm{Der}^+(\mathbb{L}\mathcal{H}))$ is a graded FI–module. The truncation is defined precisely so that $H_k(\mathrm{Der}^+(\mathbb{L}\mathcal{H})) \cong H_k(\mathrm{Der}(\mathbb{L}\mathcal{H}))$ for all $k \geq 0$. Since $H = s^{-1}\widetilde{H}_*(X)$ and X is assumed to be simply connected, H is finite-dimensional and concentrated in positive degree, and since we have assumed that the homology of X vanishes in degree at least d, H is concentrated in degrees strictly below $d - 1$. The given bounds on stability degree and weight now follow from Propositions 2.27 and 2.18. □

5 Homotopy automorphisms of connected sums

Let M be a closed oriented d–dimensional manifold. For a nonempty finite set S, let $M_S = (\#^S M) \setminus \mathring{D}^d$ be the space obtained by removing an open d–dimensional disk from the S–fold connected sum of M. If $S = \boldsymbol{n} = \{1, \dots, n\}$ we simply write M_n. We then have a deformation retraction $M_S \xrightarrow{\simeq} \bigvee^S M_1$. Hence there is an FI–module given on objects by $S \mapsto \pi_k(\mathrm{aut}_*(M_S))$, as defined in the previous section.

If we choose a basepoint in the boundary of M_S, there is an inclusion map $\mathrm{aut}_\partial(M_S) \to \mathrm{aut}_*(M_S)$ for every $S \in \mathrm{FI}$. In Section 5.1 we prove that the FI–module $S \mapsto \pi_k(\mathrm{aut}_*(M_S))$ lifts to an FI–module given on objects by $S \mapsto \pi_k(\mathrm{aut}_\partial(M_S))$. In Section 5.2 we prove that $S \mapsto \pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_S))$ is a finitely generated FI–module using certain rational models.

5.1 The integral FI–module structure on the homotopy automorphisms of iterated connected sums

For the purposes of this section, we give an explicit construction of M_n by removing the interiors of n embedded little disks in D^d, which we fix as in Figure 1, left, and gluing n copies of $M \setminus \mathring{D}^d$ along the new boundary components. Note that with this definition, we still have $M_1 = M \setminus \mathring{D}^d$. In Figure 1, right, we see how we can embed M_n into M_{n+1}, and by extending a boundary-relative homotopy automorphism of M_n by the identity thus define a stabilization map

$$s_n \colon \mathrm{aut}_\partial(M_n) \to \mathrm{aut}_\partial(M_{n+1}).$$

In this section we will define a Σ_n–action on the homotopy groups of $\mathrm{aut}_\partial(M_n)$ and, combining this with the stabilization induced by s_n, we obtain our FI–module structure. Before we do this, we need to introduce some notation:

Definition 5.1 For any pointed space X and any finite set S, let us write $Q_{S,X} \colon \Sigma(S) \to \pi_0(\mathrm{aut}_*(\bigvee^S X))$ for the group homomorphism given by sending $\sigma \in \Sigma(S)$ to the homotopy class of the automorphism $\mathcal{X}_X(\sigma) \colon \bigvee^S X \to \bigvee^S X$ described in the beginning of Section 4.

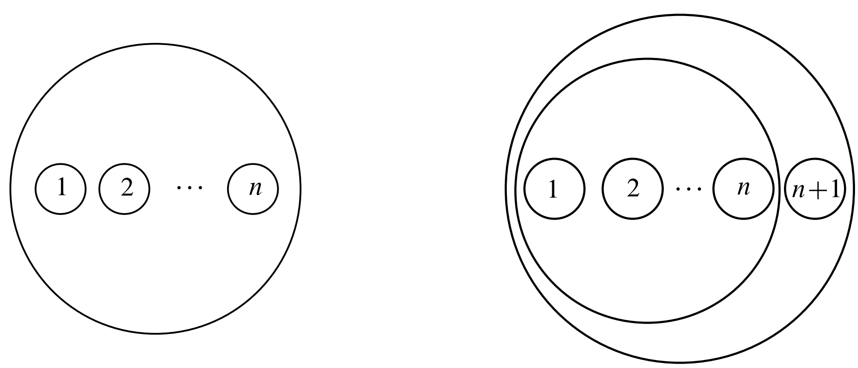


Figure 1: We can define M_n by gluing copies of M_1 into the disks on the left, and we show how to define an embedding $M_n \hookrightarrow M_{n+1}$ on the right.

Remark 5.2 Since $\pi_0\big(\mathrm{aut}_*\big(\bigvee^S X\big)\big)$ acts on $\pi_k\big(\mathrm{aut}_*\big(\bigvee^S X\big)\big)$, we get an induced $\Sigma(S)$–action on $\pi_k\big(\mathrm{aut}_*\big(\bigvee^S X\big)\big)$ by the above. This action coincides with the $\Sigma(S)$–action coming from the FI–module structure discussed in Section 4.

Definition 5.3 The deformation retraction $M_S \to \bigvee^S M_1$ induces an equivalence

$$\mathrm{aut}_*(M_S) \xrightarrow{\sim} \mathrm{aut}_*\Big(\bigvee^S M_1\Big).$$

Composing this map with the inclusion $\mathrm{aut}_\partial(M_S) \hookrightarrow \mathrm{aut}_*(M_S)$ yields a map

$$u\colon \mathrm{aut}_\partial(M_S) \to \mathrm{aut}_*\Big(\bigvee^S M_1\Big)$$

that induces a group homomorphism $\pi_0(u)\colon \pi_0(\mathrm{aut}_\partial(M_S)) \to \pi_0\big(\mathrm{aut}_*\big(\bigvee^S M_1\big)\big)$.

The first thing we will show to construct our FI–module is the following:

Proposition 5.4 *Assuming $d \geq 3$, there is a group homomorphism $\varepsilon_n\colon \Sigma_n \to \pi_0(\mathrm{aut}_\partial(M_n))$ such that Q_{n,M_1} factors as $Q_{n,M_1} = \pi_0(u) \circ \varepsilon_n$.*

Remark 5.5 Since the group $\pi_0(\mathrm{aut}_\partial(M_n))$ acts on the higher homotopy groups $\pi_k(\mathrm{aut}_\partial(M_n))$, this means that there is a Σ_n–action on the higher homotopy groups of $\mathrm{aut}_\partial(M_n)$ which is nontrivial whenever ε_n is nontrivial. This action, together with the stabilization maps, will define our FI–module structure.

We will prove this in a number of steps, so let us first describe the idea: Writing $D := D^d$, we consider the subgroup $G_n \subseteq \mathrm{Diff}_\partial(D)$ consisting of diffeomorphisms which fix the embedded little disks in D from Figure 1, left, up to permutation. There is then a group homomorphism $\pi\colon G_n \to \Sigma_n$, given by sending a diffeomorphism to the permutation it induces on the little disks. We also get a group homomorphism $G_n \to \mathrm{Diff}_\partial(M_n)$, given by constructing M_n as above, and mapping $f \in G_n$ to the boundary-relative diffeomorphism of M_n which is given by f outside the n glued-in copies of M_1, and on $\bigsqcup^n M_1$ is given by $\pi(f)$. We will construct a group homomorphism $\Sigma_n \hookrightarrow \pi_0(G_n)$, which postcomposed with the maps

$$\pi_0(G_n) \to \pi_0(\mathrm{Diff}_\partial(M_n)) \to \pi_0(\mathrm{aut}_\partial(M_n))$$

is the map ε_n described in Proposition 5.4. Let us now give the proof in more detail:

Proof Our choice of embedded disks in D defines an element $e \in \mathrm{Emb}\big(\bigsqcup^n D, D\big)$. Let $\bar{e}$ denote its image in the quotient $\mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n$, where we take the quotient of the action permuting the embedded disks. Restricting to the image of e defines a map $\mathrm{Diff}_\partial(D) \to \mathrm{Emb}\big(\bigsqcup^n D, D\big)$, which is a Serre fibration. The quotient map $\mathrm{Emb}\big(\bigsqcup^n D, D\big) \to \mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n$ is a covering map, so the composition

$$p\colon \mathrm{Diff}_\partial(D) \to \mathrm{Emb}\Big(\bigsqcup^n D, D\Big)\Big/\Sigma_n \tag{5}$$

is also a Serre fibration. The fiber over $\bar{e}$ consists of the diffeomorphisms which restricted to the image of e are permutations, ie $\mathrm{fib}_p(\bar{e}) = G_n$. We thus get a connecting homomorphism

$$\delta\colon \pi_1(\mathrm{Emb}\Big(\bigsqcup^n D, D\Big)\Big/\Sigma_n) \to \pi_0(G_n)$$

in the long exact sequence of homotopy groups. We will therefore first show that there is an injective group homomorphism $\Sigma_n \hookrightarrow \pi_1\big(\mathrm{Emb}\big(\bigsqcup^n D, D\big)\big)$.

Note that if we let $C_n(\mathring{D})$ denote the ordered configuration space of n points in $\mathring{D}$, there is a map $\hat{\rho}\colon \mathrm{Emb}\big(\bigsqcup^n D, D\big) \to C_n(\mathring{D})$ given by restricting to the center of each embedded disk. This map also has a section $\hat{s}$, given by sending a configuration to an embedding of n little disks, centered at the respective points and with radii all equal to the minimum distance between the points and between the points and the boundary of D, divided by three. We also get an induced map

$$\rho\colon \mathrm{Emb}\Big(\bigsqcup^n D, D\Big)\Big/\Sigma_n \to C_n(\mathring{D})/\Sigma_n \tag{6}$$

on orbits, which has a section s defined in the corresponding way. We define $U_n(\mathring{D}) := C_n(\mathring{D})/\Sigma_n$ for brevity. Since we have assumed that $d \geq 3$, we have that $\pi_1(U_n(\mathring{D})) \cong \Sigma_n$ and thus we get a homomorphism $\pi_1(s)\colon \Sigma_n \cong \pi_1(U_n(\mathring{D})) \to \pi_1\big(\mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n\big)$. Furthermore, note that since s is a section, $\pi_1(\rho) \circ \pi_1(s)$ is the identity on $\pi_1(U_n(\mathring{D}))$ and so $\pi_1(s)$ is injective.

By composing with the connecting homomorphism in the long exact sequence associated to p, we thus get a homomorphism $\Sigma_n \to \pi_0(G_n)$. In order to understand this map better, we describe the connecting homomorphism δ in more detail. If γ is a loop in $\mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n$ based at $\bar{e}$, representing an element of $\pi_1\big(\mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n\big)$, it lifts to a path $\tilde{\gamma}$ in $\mathrm{Diff}_\partial(D)$ starting at id_D, since p is a Serre fibration. The connecting homomorphism sends the class of γ to the connected component of G_n containing $\tilde{\gamma}(1)$. If we consider the restriction of δ to the image of inclusion $\pi_1(s)$ above, we see that a permutation σ is sent to the isotopy class of some diffeomorphism in G_n which, restricted to the little disks, is precisely σ. If we finally consider the composite map

$$\Sigma_n \to \pi_0(G_n) \to \pi_0(\mathrm{Diff}_\partial(M_n)) \to \pi_0(\mathrm{aut}_\partial(M_n)) \to \pi_0(\mathrm{aut}_*(M_n)) \cong \pi_0\Big(\mathrm{aut}_*\Big(\bigvee^n M_1\Big)\Big),$$

it follows by the definition of the map $G_n \to \mathrm{Diff}_\partial(M_n)$ that this takes a permutation to the homotopy class of the homotopy automorphism of $\bigvee^n M_1$ given by permuting the wedge summands in the corresponding way. In other words, the composition is equal to Q_{n,M_1}, so we can simply define ε_n as the composition of the first three maps. □

Remark 5.6 If we assume that M_1 has nontrivial homology, then for any nontrivial permutation σ we have that $\mathcal{X}_{M_1}(\sigma)\colon \bigvee^n M_1 \to \bigvee^n M_1$ is not homotopic to the identity, since it induces a nontrivial permutation of the reduced homology $\widetilde{H}_*\big(\bigvee^n M_1\big) = \bigoplus_n \widetilde{H}_*(M_1)$, which is different from the identity map whenever $\widetilde{H}_*(M_1)$ is nontrivial. If that is the case, the homomorphism Q_{n,M_1} is injective, so it follows that ε_n is injective as well, and thus both $\pi_0(\mathrm{aut}_\partial(M_n))$ and $\pi_0\big(\mathrm{aut}_*\big(\bigvee^n M_1\big)\big)$ contain a subgroup isomorphic to Σ_n.

Corollary 5.7 *Under the assumptions of Remark 5.6, fix a subspace $A \subseteq \partial M_n$, possibly empty, such that $A \subset M_n$ is a cofibration. Then all of the groups $\pi_0(\mathrm{aut}_A(M_n))$, $\pi_0(\mathrm{Diff}_A(M_n))$ and $\pi_0(\mathrm{Homeo}_A(M_n))$ contain a subgroup isomorphic to Σ_n.*

Proof Suppose that $A \neq \varnothing$ and let us first consider the case of homotopy automorphisms. The map $u\colon \mathrm{aut}_\partial(M_n) \to \mathrm{aut}_*\big(\bigvee^n M_1\big)$ factors as

$$\mathrm{aut}_\partial(M_n) \to \mathrm{aut}_A(M_n) \to \mathrm{aut}_*\Big(\bigvee^n M_1\Big),$$

proving this case. To get the cases with diffeomorphisms or homeomorphisms, consider the factorization

$$\mathrm{Diff}_\partial(M_n) \to \mathrm{Diff}_A(M_n) \to \mathrm{Homeo}_A(M_n) \to \mathrm{aut}_A(M_n) \to \mathrm{aut}_*\Big(\bigvee^n M_1\Big).$$

For the case where A is empty, we instead postcompose with the map $\mathrm{aut}_*\big(\bigvee^n M_1\big) \to \mathrm{aut}\big(\bigvee^n M_1\big)$, and the resulting map factors as

$$\mathrm{aut}_\partial(M_n) \to \mathrm{aut}(M_n) \to \mathrm{aut}\Big(\bigvee^n M_1\Big).$$

The composition of Q_{n,M_1} with the map induced on π_0 by the rightmost map above will still be injective, and from this the case follows. To get the statement for diffeomorphisms and homeomorphisms, we instead use the factorization

$$\mathrm{Diff}_\partial(M_n) \to \mathrm{Diff}(M_n) \to \mathrm{Homeo}(M_n) \to \mathrm{aut}(M_n) \to \mathrm{aut}\Big(\bigvee^n M_1\Big). \qquad \square$$

Remark 5.8 A referee pointed out that the existence of the homomorphism $\Sigma_n \to \pi_0(\mathrm{aut}_\partial(M_n))$ is likely a consequence of a higher structure. More specifically, it is reasonable to expect that the space $\bigsqcup_{n\geq 1} B\,\mathrm{aut}_\partial(M_n)$ can be endowed with the structure of an E_d–algebra, ie an algebra over the little d–disks operad, in a similar way as, for example, the space $\bigsqcup_{n\geq 1} B\,\mathrm{Diff}_\partial(M_n)$. If this is the case, the E_d–algebra structure maps in particular give us a map

$$E_d(n)/\Sigma_n \to B\,\mathrm{aut}_\partial(M_n),$$

and since $E_d(n)/\Sigma_n \simeq U_n(\mathring{D})$, taking fundamental groups gives us a map $\Sigma_n \to \pi_0(\mathrm{aut}_\partial(M_n))$, which should be precisely ε_n. We expect this to be true, but have elected to use a more hands-on approach, since rigorously constructing the E_d–algebra structure is nontrivial and seems to require using methods from higher homotopy theory that go quite far beyond the scope of this paper. For comparison, what makes this easier in the case of diffeomorphisms is that we have a good model for $B\,\mathrm{Diff}_\partial(M_n)$ as a topological space, in terms of embeddings of M_n into $\mathbb{R}^\infty$ (with certain boundary conditions), modulo the action of $\mathrm{Diff}_\partial(M_n)$. In contrast, it is not clear how to do a similar construction for homotopy automorphisms.

We have now defined the Σ_n–action on the homotopy groups of $\mathrm{aut}_\partial(M_n)$. Next we show that this action is compatible with the stabilization maps s_n.

Proposition 5.9 *There is a commutative diagram*

$$\begin{array}{ccc} \Sigma_n & \longrightarrow & \Sigma_{n+1} \\ {\scriptstyle\varepsilon_n}\downarrow & & \downarrow{\scriptstyle\varepsilon_{n+1}} \\ \pi_0(\mathrm{aut}_\partial(M_n)) & \xrightarrow{\pi_0(s_n)} & \pi_0(\mathrm{aut}_\partial(M_{n+1})) \end{array}$$

where the upper horizontal map is the standard inclusion.

Proof Construct stabilization maps $G_n \to G_{n+1}$, $\mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n \to \mathrm{Emb}\big(\bigsqcup^{n+1} D, D\big)/\Sigma_{n+1}$ and $C_n(\mathring{D})/\Sigma_n \to C_{n+1}(\mathring{D})/\Sigma_{n+1}$ in the same way as $s_n \colon \mathrm{aut}_\partial(M_n) \to \mathrm{aut}_\partial(M_{n+1})$, using Figure 1, right. This gives us a diagram

$$\begin{array}{ccc} \Sigma_n & \longrightarrow & \Sigma_{n+1} \\ \cong\downarrow & & \downarrow\cong \\ \pi_1(U_n(\mathring{D})) & \longrightarrow & \pi_1(U_{n+1}(\mathring{D})) \\ \downarrow & & \downarrow \\ \pi_1\big(\mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n\big) & \longrightarrow & \pi_1\big(\mathrm{Emb}\big(\bigsqcup^{n+1} D, D\big)/\Sigma_{n+1}\big) \\ \downarrow & & \downarrow \\ \pi_0(G_n) & \longrightarrow & \pi_0(G_{n+1}) \\ \downarrow & & \downarrow \\ \pi_0(\mathrm{aut}_\partial(M_n)) & \longrightarrow & \pi_0(\mathrm{aut}_\partial(M_{n+1})) \end{array}$$

where the top horizontal arrow is the standard inclusion. The two upper squares, as well as the bottom square, are all commutative by the definition of the stabilization maps. The second square from the bottom can be shown to be commutative simply by once again considering the definition of the connecting homomorphism in detail as above, but we can also reason as follows: Define a map $\mathrm{Diff}_\partial(D) \to \mathrm{Diff}_\partial(D)$ in the same was as we defined the stabilization maps, using Figure 1, right, and extending by the identity (note however that this map is homotopic to the identity), giving us a commutative diagram

$$\begin{array}{ccc} \mathrm{Diff}_\partial(D) & \longrightarrow & \mathrm{Diff}_\partial(D) \\ \downarrow & & \downarrow \\ \mathrm{Emb}\big(\bigsqcup^n D, D\big)/\Sigma_n & \longrightarrow & \mathrm{Emb}\big(\bigsqcup^{n+1} D, D\big)/\Sigma_{n+1} \end{array}$$

which is a map of Serre fibrations. By functoriality, this induces a map between the long exact sequences of homotopy groups, in which the square we consider appears. □

Corollary 5.10 *For $k \geq 1$, the sequence $\{\pi_k(\mathrm{aut}_\partial(M_n)), \pi_k(s_n)\}$ is a consistent sequence of $\mathbb{Z}[\Sigma_n]$–modules.*

Proof Recall that $\pi_0(\mathrm{aut}_\partial(M_n))$ acts on $\pi_k(\mathrm{aut}_\partial(M_n))$ and thus, through the stabilization map

$$\pi_0(s_n)\colon \pi_0(\mathrm{aut}_\partial(M_n)) \to \pi_0(\mathrm{aut}_\partial(M_{n+1})),$$

$\pi_0(\mathrm{aut}_\partial(M_n))$ acts on $\pi_k(\mathrm{aut}_\partial(M_{n+1}))$ as well. By definition of the stabilization map, $\pi_k(s_n)$ is $\pi_0(\mathrm{aut}_\partial(M_n))$–equivariant.

By considering $\pi_k(\mathrm{aut}_\partial(M_n))$ as a $\mathbb{Z}[\Sigma_n]$–module via the homomorphism $\varepsilon_n\colon \Sigma_n \to \pi_0(\mathrm{aut}_\partial(M_n))$, it follows from Proposition 5.9 and the equivariance discussed above that $\{\pi_k(\mathrm{aut}_\partial(M_n)), \pi_k(s_n)\}$ is a consistent sequence of $\mathbb{Z}[\Sigma_n]$–modules. $\square$

Theorem 5.11 *For each $k \geq 1$, the* FI*–module $S \mapsto \pi_k(\mathrm{aut}_*(M_S)) \cong \pi_k\big(\mathrm{aut}_*\big(\bigvee^S M_1\big)\big)$ lifts to an* FI*–module*

$$S \mapsto \pi_k(\mathrm{aut}_\partial(M_S)),$$

where the standard inclusion $\boldsymbol{n} \hookrightarrow \boldsymbol{n+1}$ gives the induced stabilization map

$$\pi_k(s)\colon \pi_k(\mathrm{aut}_\partial(M_n)) \to \pi_k(\mathrm{aut}_\partial(M_{n+1})).$$

Proof We have shown in Corollary 5.10 that the homotopy groups $\{\pi_k(\mathrm{aut}_\partial(M_n))\}_{n\geq 1}$ form a consistent sequence of $\mathbb{Z}[\Sigma_n]$–modules, and from the previous discussion it is clear that the maps $\mathrm{aut}_\partial(M_n) \to \mathrm{aut}_*\big(\bigvee^n M_1\big)$ induce a map of consistent sequences to $\big\{\pi_k\big(\mathrm{aut}_*\big(\bigvee^n M_1\big)\big)\big\}_{n\geq 1}$, which we know comes from an FI–module. Thus, it is sufficient to show that $\{\pi_k(\mathrm{aut}_\partial(M_n))\}_{n\geq 1}$ also comes from an FI–module.

From Lemma 2.6, it suffices to show that if $\sigma \in \Sigma_{n+m}$ is such that $\sigma|_{\boldsymbol{n}} = \mathrm{id}$, it acts trivially on the image of the stabilization map $\pi_k(\mathrm{aut}_\partial(M_n)) \to \pi_k(\mathrm{aut}_\partial(M_{n+m}))$. Embedding M_n in M_{n+m} according to the composition of the embeddings $M_n \hookrightarrow \cdots \hookrightarrow M_{n+m}$ defined by Figure 1, right, we may represent σ by an automorphism $f_\sigma \in \mathrm{aut}_\partial(M_{n+m})$ which is supported completely on $M_m \subset M_{m+n}$ and is thus the identity on $M_n \subset M_{m+n}$. Any homotopy automorphism $g \in \mathrm{im}(s_{n+m-1}\cdots s_n\colon \mathrm{aut}_\partial(M_n) \to \mathrm{aut}_\partial(M_{m+n}))$ is supported on M_n, so $f_\sigma g f_\sigma^{-1} = g$. Hence σ on acts trivially on the image of the stabilization map $\pi_k(\mathrm{aut}_\partial(M_n)) \to \pi_k(\mathrm{aut}_\partial(M_{n+m}))$. $\square$

5.2 Rational representation stability via algebraic models for relative homotopy automorphisms

We will study a certain dg Lie model for $B\,\mathrm{aut}_{\partial,\circ}(M_n)$ constructed in [Berglund and Madsen 2020], and use it to prove that the FI–module $S \mapsto \pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_S)) = \pi_k(\mathrm{aut}_\partial(M_S)) \otimes \mathbb{Q}$ is finitely generated.

We recall that a quasifree dg Lie algebra $(\mathbb{L}(V), d)$ is said to be minimal if $d(V) \subset [\mathbb{L}(V), \mathbb{L}(V)]$. If two minimal dg Lie algebras are quasi-isomorphic then they are isomorphic. Moreover, if $\mathbb{L}(V)$ is a minimal dg Lie algebra model for a nilpotent space X of finite type, then one can show that V is isomorphic to the desuspension of the reduced rational homology of X, which we will denote by $s^{-1}\widetilde{H}_*(X;\mathbb{Q})$.

In this subsection we fix a d–dimensional simply connected oriented closed manifold M, where $M_1 = M \setminus \mathring{D}$ has a nontrivial rational homology. The intersection form on $H_*(M)$ induces a graded symmetric inner product of degree d on the reduced homology $\widetilde{H}_*(M_1)$. This in turn induces a graded antisymmetric inner product of degree $d-2$ on $H = s^{-1}\widetilde{H}^*(M_1)$.

Definition 5.12 Let H be a graded antisymmetric inner product space of degree $d-2$ (eg $s^{-1}\widetilde{H}_*(M_1)$) with a basis $\{\alpha_1, \dots, \alpha_m\}$. The dual basis $\{\alpha_1^\#, \dots, \alpha_m^\#\}$ is characterized by the following property:

$$\langle \alpha_i, \alpha_j^\# \rangle = \delta_{ij}.$$

Let $\omega_H \in \mathbb{L}^2(H)$ be given by

$$\omega_H = \frac{1}{2}\sum_{i=1}^{m} [\alpha_i^\#, \alpha_i].$$

It turns out that ω_H is independent of choice of basis $\{\alpha_1, \dots, \alpha_m\}$; see [Berglund and Madsen 2020] for details.

Remark 5.13 By the same arguments as above, the graded vector space $s^{-1}\widetilde{H}_*(M_n)$ also has a structure of a graded antisymmetric inner product space of degree $d-2$ which coincides with the one given by the direct sum $(s^{-1}\widetilde{H}_*(M_1))^{\oplus n}$.

The next proposition is due to Stasheff [1983, Theorem 2], and is discussed in [Berglund and Madsen 2020, Theorem 3.11].

Proposition 5.14 *Let $M = M^d$ be a closed oriented d–dimensional manifold, let $M_1 = M \setminus \mathring{D}$ and let $H = s^{-1}\widetilde{H}_*(M_1)$. Then the inclusion $S^{d-1} \cong \partial M_1 \hookrightarrow M_1$ is modeled by a dg Lie algebra map*

$$\iota\colon \mathbb{L}(x) \hookrightarrow \mathbb{L}(H), \quad \iota(x) = (-1)^d \omega_H,$$

where $\mathbb{L}(H)$ and $\mathbb{L}(x)$ denote the minimal dgl models for M_1 and S^{d-1}, respectively.

Given a fixed basis $\{\alpha_1, \dots, \alpha_m\}$ for $H = s^{-1}\widetilde{H}_*(M_1)$ we get a basis for $s^{-1}\widetilde{H}_*(M_n) \cong H^{\oplus n}$ which is of the form

$$\{\alpha_i^j \mid 1 \le i \le m,\ 1 \le j \le n\}.$$

We denote $\omega_{H^{\oplus n}} = \frac{1}{2}\sum_{i,j}[(\alpha_i^j)^\#, \alpha_i^j] \in \mathbb{L}(H^{\oplus n})$ by ω_n. We have that ω_n is invariant under the Σ_n–action on $\mathbb{L}(H^{\oplus n})$ that permutes the summands of $H^{\oplus n}$.

Note that $\iota\colon \mathbb{L}(x) \to \mathbb{L}(H^{\oplus n})$ is not a cofibration. In order to model the inclusion $\partial M_n \subset M_n$ by a cofibration in the model category of dg Lie algebras we need a new model for M_n.

Lemma 5.15 *Let $\mathbb{L}(H^{\oplus n}, x, y)$ be the dg Lie algebra that contains $\mathbb{L}(H^{\oplus n})$ as a dg Lie subalgebra where $|x| = d-2$ and $|y| = d-1$, and where*

$$dx = 0 \quad \text{and} \quad dy = x - (-1)^d \omega_n.$$

Then

$$\hat{\iota}\colon \mathbb{L}(x) \to \mathbb{L}(H^{\oplus n}, x, y), \quad \hat{\iota}(x) = x,$$

is a cofibration that models the inclusion of $\partial M_n \cong S^{d-1}$ into M_n. Moreover this model is a relative minimal model in the sense of [Espic and Saleh 2020].

Proof The dg Lie algebra map $\rho\colon \mathbb{L}(H^{\oplus n}, x, y) \to \mathbb{L}(H^{\oplus n})$ where $\rho|_{H^{\oplus n}} = \mathrm{id}_{H^{\oplus n}}$, $\rho(x) = (-1)^d \omega_n$ and $\rho(y) = 0$ is a quasi-isomorphism. Straightforward computation shows that $\rho \circ \hat{\iota} = \iota$, proving that $\hat{\iota}$ is a model for ι (which is a model for the inclusion of the boundary). Minimality is straightforward verification; see [loc. cit., Section 3]. □

By Proposition 3.7(b), a dg Lie algebra model for $B\,\mathrm{aut}_{\partial,\circ}(M_n)$ is given by $\mathrm{Der}^+(\mathbb{L}(H^{\oplus n}, x, y) \,\|\, \mathbb{L}(x))$. However, we will use another model thanks to the following result:

Proposition 5.16 [Berglund and Madsen 2020, Theorem 3.12] *Let $\mathrm{Der}(\mathbb{L}(H^{\oplus n}) \| \omega_n)$ denote the dg Lie algebra of derivations on $\mathbb{L}(H^{\oplus n})$ that vanish on ω_n and where the differential is given by $[d_{\mathbb{L}(H^{\oplus n})}, -]$. Then there is an equivalence of dg Lie algebras*

$$\mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \,\|\, \omega_n) \to \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}, x, y) \,\|\, \mathbb{L}(x)), \quad \theta \mapsto \hat{\theta},$$

where $\hat{\theta}|_{\mathbb{L}(H^{\oplus n})} = \theta$ and $\theta(x) = \theta(y) = 0$.

Remark 5.17 It follows that $\mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \,\|\, \omega_n)$ is a dg Lie algebra model for $B\,\mathrm{aut}_{\partial,\circ}(M_n)$ and the inclusion $\mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \| \omega_n) \to \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}))$ is a model for the map $B\,\mathrm{aut}_{\partial,\circ}(M_n) \to B\,\mathrm{aut}_{*,\circ}(M_n)$, induced by the inclusion $\mathrm{aut}_{\partial,\circ}(M_n) \hookrightarrow \mathrm{aut}_{*,\circ}(M_n)$.

Definition 5.18 With the terminology of Section 2, we define a dg FI–Lie algebra $\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})$ as follows: For $S \in \mathrm{FI}$, we let $\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})(S) := \mathrm{Der}(\mathbb{L}\mathcal{H}(S) \,\|\, \omega_S)$ be the dg Lie algebra of derivations on $\mathbb{L}\mathcal{H}(S) = \mathbb{L}(H^{\oplus S})$ that vanish on ω_S. For $i\colon S \hookrightarrow T$ in FI, we get a map

$$\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})(i)\colon \mathrm{Der}(\mathbb{L}(H^{\oplus S}) \,\|\, \omega_S) \hookrightarrow \mathrm{Der}(\mathbb{L}(H^{\oplus T}) \,\|\, \omega_T),$$

defined as follows: Suppose $x_\alpha \in \mathcal{H}(T)$ lies in the direct summand of $\mathcal{H}(T)$ corresponding to $\alpha \in T$ and let $D \in \mathrm{Der}(\mathbb{L}(H^{\oplus S}) \,\|\, \omega_S)$. Then $\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})(i)D$ is determined by

$$(\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})(i)D)(x_\alpha) = \begin{cases} 0 & \text{if } \alpha \in T \setminus i(S), \\ (\mathbb{L}\mathcal{H}(i) \circ D \circ \mathcal{H}(i)^{-1})(x_\alpha) & \text{if } \alpha \in i(S). \end{cases}$$

We conclude from having such a dg FI–Lie algebra the following:

Remark 5.19 The above dg FI–Lie algebra structure induces an FI–module structure on the homology. For $k \geq 1$, we have that $H_k(\mathrm{Der}(\mathbb{L}(H^{\oplus S}) \| \omega_S)) \cong \pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_S))$, which gives an FI–module structure on $\{\pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_S))\}_{S \in \mathrm{FI}}$. We will show that this FI–module structure coincides with the one obtained by rationalizing the FI–module structure on $\{\pi_k(\mathrm{aut}_\partial(M_S))\}_{S \in \mathrm{FI}}$ defined in Section 5.1.

Proposition 5.20 *A dg Lie algebra model for the stabilization map $B\,\mathrm{aut}_{\partial,\circ}(M_n) \to B\,\mathrm{aut}_{\partial,\circ}(M_{n+1})$ is given by*

$$\varphi_n\colon \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \,\|\, \omega_n) \to \mathrm{Der}^+(\mathbb{L}(H^{\oplus n+1}) \,\|\, \omega_{n+1}),$$

where $\varphi_n(\theta)$ coincides with θ on the first n summands of $H^{\oplus n+1}$ and vanishes on the last summand.

Proof The proof is omitted since it is very similar to the proof of Proposition 4.2. □

Proposition 5.21 *The Σ_n–action on $\pi_*^{\mathbb{Q}}(\mathrm{aut}_\partial(M_n))$ induced by $\varepsilon_n\colon \Sigma_n \to \pi_0(\mathrm{aut}_\partial(M_n))$ is modeled by the Σ_n–action on $H_k(\mathrm{Der}(\mathbb{L}(H^{\oplus n})) \parallel \omega_n)$ induced by the* FI*–module structure from Definition 5.18.*

Proof For every $\sigma \in \Sigma_n$, let $\eta_\sigma \in \mathrm{aut}_\partial(M_n))$ denote a representative for $\varepsilon_n(\sigma) \in \pi_0(\mathrm{aut}_\partial(M_n))$, and define a self-equivalence

$$\mathrm{Ad}_\sigma\colon \mathrm{aut}_\partial(M_n) \to \mathrm{aut}_\partial(M_n), \quad \mathrm{Ad}_\sigma(f) = \eta_\sigma \circ f \circ \eta_{\sigma^{-1}}.$$

This induces a Σ_n–action on $\pi_k(\mathrm{aut}_\partial(M_n))$ given by $\sigma.a = \pi_k(\mathrm{Ad}_\sigma)(a)$ which is precisely the Σ_n–action given by the FI–module structure.

As we saw in Lemma 5.15, $\hat{\imath}\colon \mathbb{L}(x) \to \mathbb{L}(H^{\oplus n}, x, y)$ is a minimal relative model for the inclusion $\partial M_n \hookrightarrow M_n$.

By Lemma 3.15, η_σ is modeled by an $\hat{\imath}$–relative automorphism $\zeta_\sigma \in \mathrm{Aut}_{\hat{\imath}}(\mathbb{L}(H^{\oplus n}, x, y))$, and hence, by Corollary 3.19, the automorphism

$$\alpha_{\zeta_\sigma}\colon \mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x)) \to \mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x)), \quad \alpha_{\zeta_\sigma}(\theta) = \zeta_\sigma \circ \theta \circ \zeta_\sigma^{-1},$$

is a model for the delooping of Ad_σ. In particular, $H_k(\alpha_{\zeta_\sigma})$ is a model for $\pi_k(\mathrm{Ad}_\sigma)$. Moreover, this defines a Σ_n–action on $H_k\big(\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x))\big)$ given by $\sigma.b = H_k(\alpha_{\zeta_\sigma})(b)$ that models the Σ_n–action on $\pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_n))$ described above.

Since the isomorphism of Lemma 3.15 is not explicit, we do not know what ζ_σ is. However, viewing ζ_σ as a nonrelative automorphism that models pointed homotopy automorphisms, we know that it models the permutation of the summands of $\bigvee_{i=1}^n M_1$ corresponding to $\sigma \in \Sigma_n$. A model for this pointed map is given by $\psi_\sigma\colon \mathbb{L}(H^{\oplus n}, x, y) \to \mathbb{L}(H^{\oplus n}, x, y)$, where $\psi_\sigma(\alpha_i^j) = \alpha_i^{\sigma(j)}$, $\psi_\sigma(x) = x$ and $\psi_\sigma(y) = y$. Since ψ_σ and ζ_σ model the same pointed homotopy class of pointed maps they have to be homotopic as dg Lie algebra maps, and thus α_{ζ_σ} and α_{ψ_σ} induce the same map on the homology of $\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y))$. In particular, for every cycle $\theta \in Z\big(\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y))\big)$, the difference $\alpha_{\zeta_\sigma}(\theta) - \alpha_{\psi_\sigma}(\theta)$ is a boundary $\partial\nu$ for some $\nu \in \mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y))$.

Note that ψ_σ is also $\hat{\imath}$–relative, but not necessarily $\hat{\imath}$–equivalent, to ζ_σ. Since ζ_σ and ψ_σ are $\hat{\imath}$–relative, α_{ζ_σ} and α_{ψ_σ} define automorphisms of $\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x))$. We will show that these automorphisms induce the same map on homology. Given a cycle $\theta \in Z\big(\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x))\big)$, we have that θ is also a cycle in $\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y))$, and thus by the above there is some $\nu \in \mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y))$ such that $\alpha_{\zeta_\sigma}(\theta) - \alpha_{\psi_\sigma}(\theta) = \partial\nu$. By this equality $\partial\nu(x) = 0$.

Let $\tilde{\nu} \in \mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x))$ be given by $\tilde{\nu}|_{\mathrm{span}(H^{\oplus n}, y)} = \nu|_{\mathrm{span}(H^{\oplus n}, y)}$ and $\tilde{\nu}(x) = 0$. Now it is straightforward to see that

$$\alpha_{\zeta_\sigma}(\theta) - \alpha_{\psi_\sigma}(\theta) = \partial\nu = \partial\tilde{\nu}.$$

Hence α_{ζ_σ} and α_{ψ_σ} induce the same morphisms on $H_*\big(\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) | \mathbb{L}(x))\big)$. From this we conclude that the Σ_n–action on $H_k\big(\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y) | \mathbb{L}(x))\big)$ given by $\sigma.b = H_k(\alpha_{\psi_\sigma})(b)$ is a model for the Σ_n–action on $\pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_n))$.

Now consider the ω_n–preserving automorphism $\phi_\sigma\colon \mathbb{L}(H^{\oplus n}) \to \mathbb{L}(H^{\oplus n})$ given by $\phi_\sigma = \psi_\sigma|_{\mathbb{L}(H^{\oplus n})}$. This yields an automorphism

$$\alpha_{\phi_\sigma}\colon \mathrm{Der}(\mathbb{L}(H^{\oplus n}) \parallel \omega_n) \to \mathrm{Der}(\mathbb{L}(H^{\oplus n}) \parallel \omega_n), \quad \alpha_{\phi_\sigma}(\theta) = \phi_\sigma \circ \theta \circ \phi_\sigma^{-1}.$$

The Σ_n–action on $\mathrm{Der}(\mathbb{L}(H^{\oplus n}) \parallel \omega_n)$ given by $\sigma.b = \alpha_{\phi_\sigma}(b)$ is the same Σ_n–action coming from the FI–module structure described in Definition 5.18.

The diagram

$$\begin{array}{ccc} \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \parallel \omega_n) & \xrightarrow{\alpha_{\phi_\sigma}} & \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \parallel \omega_n) \\ \sim\downarrow & & \downarrow\sim \\ \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x)) & \xrightarrow{\alpha_{\psi_\sigma}} & \mathrm{Der}^+(\mathbb{L}(H^{\oplus n}, x, y) \parallel \mathbb{L}(x)) \end{array}$$

where the vertical maps are the quasi-isomorphisms of dg Lie algebras described in Proposition 5.16, is commutative, which gives that the induced Σ_n–action on $H_k(\mathrm{Der}^+(\mathbb{L}(H^{\oplus n}) \parallel \omega_n))$ is a model for the Σ_n–action on $H_k\big(\mathrm{Der}(\mathbb{L}(H^{\oplus n}, x, y)|\mathbb{L}(x))\big)$ — which, in turn, is a model for the Σ_n–action on $\pi_k^{\mathbb{Q}}(\mathrm{aut}_\partial(M_n))$. □

We recall that the Lie operad $\mathscr{L}ie$ is a cyclic operad, ie that the Σ_n–action on $\mathscr{L}ie(n)$ extends to a Σ_{n+1}–action. Let $\mathscr{L}ie_c(n+1)$ denote $\mathscr{L}ie(n)$ viewed as a Σ_{n+1}–representation.

Proposition 5.22 [Berglund and Madsen 2020, Proposition 6.6] *There is an isomorphism of* FI*–modules*

$$\mathrm{Der}(\mathbb{L}\mathcal{H} \parallel \omega_{\mathcal{H}}) \cong s^{2-d}\mathbb{S}_{\mathscr{L}ie_c}(\mathcal{H}).$$

Proof We will prove that this isomorphism is a special case of the more general isomorphism of Berglund and Madsen, where the authors consider the category of graded antisymmetric inner product spaces of degree $2-d$, with morphisms being linear maps of degree 0 that preserve the inner product. They call this category Sp^{2-d}. An Sp^{2-d}–module is a functor from Sp^{2-d} to the category of graded vector spaces. By [loc. cit., Proposition 6.6], $V \mapsto \mathrm{Der}(\mathbb{L}(V) \parallel \omega_V)$ defines an Sp^{2-d}–module that is isomorphic to the Sp^{2-d}–module given by $V \mapsto s^{2-d}\mathbb{S}_{\mathscr{L}ie_c}(V)$.

For any morphism $i\colon S \to T$ in FI, the associated map $\mathcal{H}(i)\colon \mathcal{H}(S) = H^{\oplus S} \to \mathcal{H}(T) = H^{\oplus T}$ is a morphism of Sp^{2-d}–modules. Thus the isomorphism above follows. □

Theorem B *Let $M = M^d$ be a closed simply connected oriented d–dimensional manifold. With M_S defined as above, we have the following*:

(a) *For each $k \geq 1$, the* FI*–module*

$$S \mapsto \pi_k\bigg(\mathrm{aut}_*\Big(\bigvee^S M_1\Big)\bigg) \cong \pi_k(\mathrm{aut}_*(M_S))$$

lifts to an FI*–module*

$$S \mapsto \pi_k(\mathrm{aut}_\partial(M_S))$$

sending the standard inclusion $\boldsymbol{n} \to \boldsymbol{n}+\mathbf{1}$ *to the map* $\pi_k(\mathrm{aut}_\partial(M_n)) \to \pi_k(\mathrm{aut}_\partial(M_{n+1}))$ *induced by the stabilization map* s_n*.*

(b) *The rationalization of this* FI–*module is of weight* $\le k+d-2$ *and stability degree* $\le k+d-1$*.*

Proof (a) This is Theorem 5.11.

(b) By the isomorphism in Proposition 5.22,

$$\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})_k \cong \mathbb{S}_{\mathscr{L}ie_c}(\mathcal{H})_{k+d-2}.$$

By Proposition 2.20,

$$\mathrm{weight}(\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})_k) \le k+d-2$$

and

$$\mathrm{stab\text{-}deg}(\mathrm{Der}(\mathbb{L}\mathcal{H} \,\|\, \omega_{\mathcal{H}})_k) \le k+d-2.$$

The weight and the stability degree for the homology follow from Proposition 2.18. □

References

[Arnold 1969] **V I Arnold**, *The cohomology ring of the group of dyed braids*, Mat. Zametki 5 (1969) 227–231 MR Zbl In Russian; translated in Math. Notes 5 (1969) 138–140

[Berglund 2015] **A Berglund**, *Rational homotopy theory of mapping spaces via Lie theory for L_∞–algebras*, Homology Homotopy Appl. 17 (2015) 343–369 MR Zbl

[Berglund 2020] **A Berglund**, *Rational models for automorphisms of fiber bundles*, Doc. Math. 25 (2020) 239–265 MR Zbl

[Berglund 2022] **A Berglund**, *Characteristic classes for families of bundles*, Selecta Math. 28 (2022) art. id. 51 MR Zbl

[Berglund and Madsen 2020] **A Berglund**, **I Madsen**, *Rational homotopy theory of automorphisms of manifolds*, Acta Math. 224 (2020) 67–185 MR Zbl

[Berglund and Saleh 2020] **A Berglund**, **B Saleh**, *A dg Lie model for relative homotopy automorphisms*, Homology Homotopy Appl. 22 (2020) 105–121 MR Zbl

[Birman 1974] **J S Birman**, *Braids, links, and mapping class groups*, Annals of Math. Studies 82, Princeton Univ. Press (1974) MR Zbl

[Church 2012] **T Church**, *Homological stability for configuration spaces of manifolds*, Invent. Math. 188 (2012) 465–504 MR Zbl

[Church and Farb 2013] **T Church**, **B Farb**, *Representation theory and homological stability*, Adv. Math. 245 (2013) 250–314 MR Zbl

[Church et al. 2014] **T Church**, **J S Ellenberg**, **B Farb**, **R Nagpal**, *FI-modules over Noetherian rings*, Geom. Topol. 18 (2014) 2951–2984 MR Zbl

[Church et al. 2015] **T Church**, **J S Ellenberg**, **B Farb**, *FI-modules and stability for representations of symmetric groups*, Duke Math. J. 164 (2015) 1833–1910 MR Zbl

[Espic and Saleh 2020] **H Espic**, **B Saleh**, *On the group of homotopy classes of relative homotopy automorphisms*, preprint (2020) arXiv 2002.12083

[Farjoun 1996] **E D Farjoun**, *Cellular spaces, null spaces and homotopy localization*, Lecture Notes in Math. 1622, Springer (1996) MR Zbl

[Félix et al. 2001] **Y Félix**, **S Halperin**, **J-C Thomas**, *Rational homotopy theory*, Graduate Texts in Math. 205, Springer (2001) MR Zbl

[Félix et al. 2008] **Y Félix**, **J Oprea**, **D Tanré**, *Algebraic models in geometry*, Oxford Graduate Texts in Math. 17, Oxford Univ. Press (2008) MR Zbl

[Fulton and Harris 1991] **W Fulton**, **J Harris**, *Representation theory*, Graduate Texts in Math. 129, Springer (1991) MR Zbl

[Getzler 2009] **E Getzler**, *Lie theory for nilpotent L_∞–algebras*, Ann. of Math. 170 (2009) 271–301 MR Zbl

[Grey 2019] **M Grey**, *On rational homological stability for block automorphisms of connected sums of products of spheres*, Algebr. Geom. Topol. 19 (2019) 3359–3407 MR Zbl

[Hinich 1997] **V Hinich**, *Descent of Deligne groupoids*, Int. Math. Res. Not. (1997) 223–239 MR Zbl

[Kupers and Miller 2018] **A Kupers**, **J Miller**, *Representation stability for homotopy groups of configuration spaces*, J. Reine Angew. Math. 737 (2018) 217–253 MR Zbl

[Lupton and Smith 2007] **G Lupton**, **S B Smith**, *Rationalized evaluation subgroups of a map, II: Quillen models and adjoint maps*, J. Pure Appl. Algebra 209 (2007) 173–188 MR Zbl

[McDuff 1975] **D McDuff**, *Configuration spaces of positive and negative particles*, Topology 14 (1975) 91–107 MR Zbl

[Miller et al. 2019] **J Miller**, **P Patzt**, **D Petersen**, *Representation stability, secondary stability, and polynomial functors*, preprint (2019) arXiv 1910.05574

[Quillen 1969] **D Quillen**, *Rational homotopy theory*, Ann. of Math. 90 (1969) 205–295 MR Zbl

[Segal 1979] **G Segal**, *The topology of spaces of rational functions*, Acta Math. 143 (1979) 39–72 MR Zbl

[Stasheff 1983] **J Stasheff**, *Rational Poincaré duality spaces*, Illinois J. Math. 27 (1983) 104–109 MR Zbl

[Stoll 2024] **R Stoll**, *The stable cohomology of self-equivalences of connected sums of products of spheres*, Forum Math. Sigma 12 (2024) art. id. e1 MR Zbl

[Tanré 1983] **D Tanré**, *Homotopie rationnelle: modèles de Chen, Quillen, Sullivan*, Lecture Notes in Math. 1025, Springer (1983) MR Zbl

Department of Mathematical Sciences, University of Copenhagen
Copenhagen, Denmark

Department of Mathematics, Stockholm University
Stockholm, Sweden

erikjlindell@gmail.com, basharsaleh1@gmail.com

Received: 7 June 2022 Revised: 17 July 2023

Algebraic & Geometric Topology **24**:5 (2024) 2707–2719
DOI: 10.2140/agt.2024.24.2707
Published: 19 August 2024

The strong Haken theorem via sphere complexes

SEBASTIAN HENSEL
JENNIFER SCHULTENS

We give a short proof of Scharlemann's strong Haken theorem for closed 3–manifolds (and manifolds with spherical boundary). As an application, we also show that given a decomposing sphere R for a 3–manifold M that splits off an $S^2 \times S^1$ summand, any Heegaard splitting of M restricts to the standard Heegaard splitting on the summand.

57K30

1 Introduction

Any (closed oriented connected) 3–dimensional manifold M admits a *Heegaard splitting*, that is, it can be decomposed into two 3–dimensional handlebodies V and V' of the same genus g along an embedded surface $S \subset M$:

$$M = V \cup_S V'.$$

In theory, all information about the 3–manifold is encoded in the identification of the two handlebodies. However, in practice, interpreting topological properties of M using a Heegaard splitting is often nontrivial.

A basic example of this occurs when studying spheres in M. If $\alpha \subset S$ is a curve which bounds disks D and D' in both V and V', then gluing these disks yields a 2–sphere $D \cup D' \subset M$ which intersects S in the single curve α. When essential, such a sphere is called a *Haken sphere* — but a priori it is completely unclear what kind of spheres in M are of this form.

A classical theorem of Haken [6] shows that if M admits any essential sphere σ, then it also admits a Haken sphere σ'. In fact, Scharlemann [16] recently proved a *strong Haken theorem*, showing that σ' can in fact be chosen to be isotopic to σ:

Theorem 1.1 (strong Haken theorem) *Let $M = V \cup_S V'$ be a Heegaard splitting. Every essential 2–sphere in M is isotopic to a Haken sphere for $M = V \cup_S V'$.*

Our purpose here is to give an independent short proof of Theorem 1.1 for any M which is closed or has spherical boundary. We want to mention that Scharlemann's version of the strong Haken theorem is in fact more general, allowing for manifolds with arbitrary boundary, and also showing that any properly embedded disk is isotopic to a Haken disk. This more general case could also be obtained from our methods; for clarity we focus on the closed case.

To prove Theorem 1.1, we make crucial use of the (*surviving*) *sphere complex*, which is a combinatorial complex encoding the intersection pattern of essential (surviving) spheres in M. Such complexes have already been used successfully in the study of outer automorphism groups of free groups (via mapping class groups of connected sums of $S^2 \times S^1$). Here, we show that this perspective can also be useful in streamlining arguments in low-dimensional topology. The other crucial ingredient is the classical Waldhausen theorem on Heegaard splittings of the 3–sphere [18]. Together, these allow an inductive approach to Theorem 1.1.

Our methods and results also allow control over Heegaard splittings of reducible manifolds. As a motivating example, we prove:

Proposition 1.2 *Every Heegaard splitting of $W_n = n(S^2 \times S^1)$ is isotopic to a stabilization of the standard Heegaard splitting.*

Combining the uniqueness for W_1 with the strong Haken theorem, we obtain the following structural result on Heegaard splittings of arbitrary reducible 3–manifolds:

Corollary 1.3 *For a reducible 3–manifold M with a Heegaard splitting $M = V \cup_S V'$, any decomposing sphere that splits off an $S^2 \times S^1$ summand can be isotoped so that S is standard in this summand.*

Acknowledgements We would like to thank Martin Scharlemann for finding a mistake in an earlier draft, and many helpful comments.

2 Heegaard splittings of closed 3–manifolds

In this section, we recall some preliminaries on closed 3–manifolds, their Heegaard splittings, and spheres in such manifolds. The results presented here are classical.

2.1 Heegaard splittings

Definition 2.1 (Heegaard splitting) A *handlebody* is a 3–manifold that is homeomorphic to a regular neighborhood of a graph in S^3. A *Heegaard splitting* of a 3–manifold M is a decomposition $M = V \cup_S V'$, where V and V' are handlebodies and $S = \partial V = \partial V' = V \cap V'$.

The surface S is called the *splitting surface* or *Heegaard surface*. Heegaard splittings are considered *equivalent* if their splitting surfaces are isotopic.

Remark 2.2 The Heegaard splitting $M = V \cup_S V'$ is completely specified by the pair (M, S), so we will sometimes write (M, S) instead of $M = V \cup_S V'$.

Remark 2.3 The connected sum of two 3–manifolds M_1 and M_2 with Heegaard splittings (M_1, S_1) and (M_2, S_2) inherits a Heegaard splitting $(M_1 \# M_2,\ S_1 \# S_2)$. This Heegaard splitting is unique in the sense that it is completely determined by the construction. Later, we will briefly consider a refined notion of equivalence for Heegaard splittings where we distinguish between the two sides of the splitting surface. With respect to this refined notion of equivalence, the Heegaard splitting of a connected sum is then not (a priori) unique, as the construction allows two different choices, namely which side of S_1 is identified to which side of S_2.

Definition 2.4 Given a Heegaard splitting (M, S), the Heegaard splitting obtained from the pairwise connected sum $(M, S) \# (S^3, T)$, where T is the standard unknotted torus in S^3, is called a *stabilization* of (M, S). A Heegaard splitting is *stabilized* if it is the stabilization of another Heegaard splitting and *unstabilized* otherwise.

A sphere that separates $(M, S) \# (S^3, T)$, ie a sphere that splits off a punctured 3–ball containing an unknotted punctured torus, is called a *stabilizing* sphere.

A *stabilizing pair* of disks is a pair (D, D') of disks such that D is properly embedded in V, D' is properly embedded in V' and $\partial D \cap \partial D'$ is exactly one point.

Remark 2.5 A Heegaard splitting is stabilized if and only if it admits a stabilizing pair of disks. Indeed, consider the standard unknotted torus in the 3–sphere and observe that it separates S^3 into two solid tori. The boundaries of the meridian disks of these solid tori intersect in exactly one point.

A crucial theorem of Waldhausen's characterizes all Heegaard splittings of the 3–sphere; see [18].

Theorem 2.6 (Waldhausen's theorem) *Every Heegaard splitting of the* 3*–sphere is a stabilization of the unique standard genus*-0 *Heegaard splitting.*

2.2 Sphere complexes

A core tool in our argument is the following simplicial complex, which encodes the intersection patterns of spheres in M:

Definition 2.7 (sphere complex) A sphere S in a 3–manifold is *compressible* if it bounds a 3–ball. Otherwise it is *incompressible*. We say that a sphere is *peripheral* if it is isotopic into the boundary of the manifold.

The *sphere complex* of a 3–manifold M is the simplicial complex $\mathcal{S}(M)$ determined by the following three conditions:

(1) Vertices of $\mathcal{S}(M)$ correspond to isotopy classes of incompressible nonperipheral embedded 2–spheres.
(2) Edges of $\mathcal{S}(M)$ correspond to pairs of vertices with disjoint representatives.
(3) The complex $\mathcal{S}(M)$ is flag.

It is not hard to see that a simplex in the sphere complex corresponds to a collection of nonisotopic spheres that can be realized disjointly. See Figure 1 for an example of a simplex.

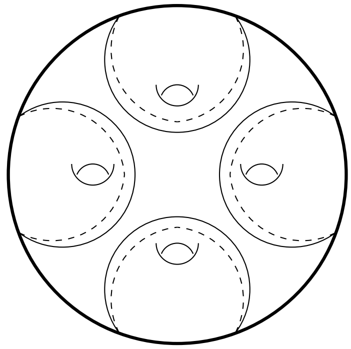


Figure 1: A 3–simplex in the sphere complex of W_4, the double of a genus-4 handlebody (alternatively, the connected sum of four copies of $S^2 \times S^1$). Here, only one of the handlebodies is pictured; the spheres comprising the simplex intersect it in the pictured disks.

Furthermore, a standard argument involving surgery at innermost intersection circles shows that the sphere complex of any closed 3–manifold is connected (if it is nonempty). See eg [7] for a proof in the case of doubled handlebodies, which also works in general.

2.3 Haken spheres

Our central aim will be to understand how essential spheres in M interact with Heegaard splittings of M. The following notion is crucial:

Definition 2.8 Let $M = V \cup_S V'$ be a Heegaard splitting. An essential sphere in M that meets the Heegaard surface S in a single simple closed curve is called a *Haken sphere*. A (not necessarily essential) sphere that intersects S in a single simple closed curve essential in S is called a *reducing sphere*.

The following theorem was originally proved by Haken in [6]. Proofs can be found in the standard references on 3–manifolds; see [8; 9; 17].

Theorem 2.9 (Haken's lemma) *If a closed 3–manifold M contains an essential sphere and $M = V \cup_S V'$ is a Heegaard splitting, then M admits a Haken sphere.*

In general, the Haken sphere is obtained by modifying the given essential sphere by surgery, and so cannot be guaranteed to be related to the sphere given at the outset.

3 3–Manifolds with spherical boundary

In this section, we present versions of the results and notions of the previous section for 3–manifolds with spherical boundary. These appear naturally in our inductive proof of the strong Haken theorem (even if one is just interested in proving it in the closed case). For ease of notation, if M is a 3–manifold with spherical boundary, then we call each component of ∂M a *puncture*. Similarly, we call such a manifold a *punctured manifold*.

3.1 Heegaard splittings

To define Heegaard splittings of punctured manifolds, we use spotted handlebodies.

Definition 3.1 A *spotted* handlebody is a handlebody with a specified set of disks $D_1 \sqcup \cdots \sqcup D_k$ in its boundary. Each disk is called a *spot*. A Heegaard splitting of a 3–manifold M with spherical boundary is a decomposition $M = V \cup_S V'$, where V and V' are spotted handlebodies with spots $D_1 \sqcup \cdots \sqcup D_k$ and $D'_1 \sqcup \cdots \sqcup D'_k$, respectively, and $S = \partial V - (D_1 \sqcup \cdots \sqcup D_k) = \partial V' - (D'_1 \sqcup \cdots \sqcup D'_k)$.

Remark 3.2 In a Heegaard splitting of a 3–manifold with spherical boundary, each puncture meets the splitting surface in a single simple closed curve. This simple closed curve is the boundary of a spot on each of the handlebodies.

Suppose M_1 and M_2 are two punctured manifolds with boundary components $\partial_i \subset M_i$, and $M = M_1 \cup_{\partial_1 = \partial_2} M_2$ is the manifold obtained by gluing the boundary components. Given Heegaard splittings of M_1 and M_2, the manifold M inherits a Heegaard splitting which is obtained by gluing the handlebodies at the corresponding spots.

The glued boundary components yield an essential 2–sphere σ in M, which intersects the induced Heegaard splitting in a single circle (ie it becomes a Haken sphere). Conversely, given a Haken sphere σ for any manifold M, one can cut the manifold and the splitting at σ.

We need a version of Waldhausen's theorem in the context of punctured 3–spheres (which is a fairly straightforward consequence of Waldhausen's theorem for S^3).

Theorem 3.3 (Waldhausen's theorem for punctured 3–spheres) *Every Heegaard splitting of a punctured 3–sphere is a stabilization of a unique standard genus-*0 *Heegaard splitting.*

Proof Let M be a punctured 3–sphere and $M = V \cup_S V'$ a Heegaard splitting. Construct $\widehat{M} = S^3$ from M by attaching a 3–ball to each puncture. By Alexander's theorem, the result does not depend on the attaching map. Moreover, the attaching maps can be chosen so that a meridional disk of each 3–ball caps off a component of ∂S. We thus obtain a closed surface $\widehat{S}$ that defines a Heegaard splitting $S^3 = \widehat{V} \cup_{\widehat{S}} \widehat{V}'$.

Each 3–ball that has been attached to a puncture is a regular neighborhood of a point and, as such, arbitrarily small. By Waldhausen's theorem, $S^3 = \widehat{V} \cup_{\widehat{S}} \widehat{V}'$ is a stabilization of the standard genus-0 Heegaard splitting of S^3. The stabilizing pairs of disks can be chosen to be disjoint from the attached 3–balls. Thus, after destabilizing, if necessary, we may assume that S is genus 0.

Hence, to prove the theorem, it suffices to show that any genus-0 splitting of a punctured S^3 is standard. To this end, observe that the spotted genus-0 handlebody $V \subset S^3$ can be isotoped to be a regular neighborhood of a graph $\Gamma \subset V$. The graph Γ can be chosen to have the following form: It has one vertex v_0 in the interior of M, and one vertex on each boundary component. Each vertex on a boundary component is joined to v_0 by an edge. We are done, once we show that any two such graphs Γ and Γ' are isotopic.

We may assume that Γ and Γ' have the same vertex set, and edges are disjoint or equal. Pick an edge $e \subset \Gamma$ and the edge $e' \subset \Gamma'$ joining the same vertices. Then e and e' are isotopic rel endpoints by an isotopy fixing $\Gamma \setminus e$. This follows by the light bulb trick, noting that a small regular neighborhood of $\Gamma \setminus e$ and all boundary spheres it touches is a sphere. Inductively, it follows that Γ and Γ' are isotopic, proving the theorem. □

At this point, we briefly want to address the ambiguity appearing in the previous proof when filling the boundary and drilling it out again — namely, one can isotope a pair of stabilizing disks across a puncture. This leads to a homeomorphism of the manifold preserving the Heegaard surface. Given a Heegaard splitting of a 3–manifold, the *Goeritz group* of the splitting is the group of isotopy classes of orientation-preserving diffeomorphisms of the manifold that preserve the splitting. Loosely speaking, the Goeritz group will be small if the surface automorphism that defines the Heegaard splitting is complicated relative to the handlebodies. Conversely, the Goeritz group will be as large as possible in the case of W_n, the manifold for which this surface automorphism is the identity, and the Goeritz group is equal to the handlebody group. Scharlemann finds a system of $4g+1$ generators for the Goeritz group of a handlebody; see [14]. On the other hand, the Goeritz group of the 3–sphere is still largely mysterious. We refer the interested reader to the recent [15].

3.2 Sphere complexes

We now want to define a useful sphere complex for punctured manifolds. One obvious change is that for the vertices one should also exclude *peripheral* spheres, ie spheres which are homotopic into the boundary (otherwise, such spheres are adjacent to any other vertex, rendering the resulting complex useless). However, even with this modification, the resulting sphere complex will be somewhat problematic for our purposes, as it may often be disconnected. Namely, suppose that M_0 is an aspherical 3–manifold with infinite fundamental group. Let M be the manifold obtained from M_0 by removing two open balls. The manifold M admits many essential nonperipheral spheres obtained by joining the two punctures by a nontrivial tube. In fact, by asphericity of M_0, any essential nonperipheral sphere in M is of this form. In particular, no two such are disjoint.

To sidestep this issue, we use the following variant of the sphere complex:

Definition 3.4 (surviving sphere complex) We call a sphere S in a punctured 3–manifold M *almost peripheral* if a component of $M \setminus S$ is a punctured 3–ball. Equivalently, S is almost peripheral if S is inessential in the manifold obtained by filling the punctures of M.

If S is not almost peripheral, then it is *surviving*.

The *surviving sphere complex* of a 3–manifold M is the simplicial complex $\mathcal{S}^s(M)$ determined by the following three conditions:

(1) Vertices of $\mathcal{S}^s(M)$ correspond to isotopy classes of incompressible embedded surviving 2–spheres.

(2) Edges of $\mathcal{S}^s(M)$ correspond to pairs of vertices with disjoint representatives.

(3) The complex $\mathcal{S}^s(M)$ is flag.

The terminology stems from the fact that the spheres "survive filling in the punctures" and is in analogy to the surviving curve complex used in the study of mapping class groups of surfaces; see eg [1; 5]. It turns out that these complexes are much better behaved in our setting.

Lemma 3.5 *Let M be a* 3*–manifold. Then the surviving sphere complex $\mathcal{S}^s(M)$ is connected* (*if it is nonempty*).

Proof Let σ and σ' be two incompressible embedded surviving 2–spheres in M. Up to isotopy, we may assume that σ and σ' intersect transversely. Further, we may assume that up to isotopy, the number of intersection components $\sigma \cap \sigma'$ is minimal.

Let $C \subset \sigma \cap \sigma'$ be an innermost intersection circle, ie suppose that it bounds a disk $D \subset \sigma$ with $D \cap \sigma' = \partial D = C$. Denote by $S^+, S^- \subset \sigma'$ the two disks bounded by C, and denote by $\sigma^\pm = S^\pm \cup D$ the two 2–spheres obtained by disk-swapping. Observe that up to isotopy, both of these are disjoint from σ', and intersect σ in at least one fewer circle than σ'. If either σ^+ or σ^- were compressible, then we could reduce the number of components in $\sigma \cap \sigma'$ by sliding σ over the ball bounded by the compressible sphere, which is impossible by our choice.

Assume that σ^- is almost peripheral. Then, after filling in the punctures of M, the spheres σ and σ^+ are isotopic (by sliding D over the now unpunctured ball bounded by σ^-; see Figure 2). In particular, σ^+ is surviving, as the same is true for σ.

Hence, at least one of $\sigma^\pm$ is surviving, and we are done. Indeed, repeating this process produces a sequence of spheres corresponding to vertices in a path, in $\mathcal{S}^s(M)$, between $[\sigma]$ and $[\sigma']$. □

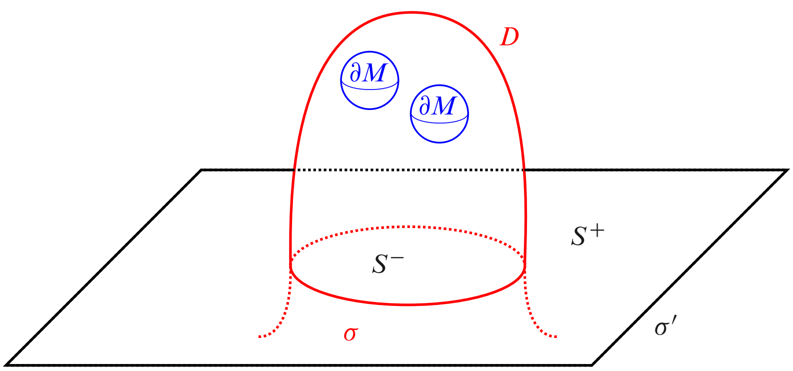


Figure 2: In the proof of Lemma 3.5: The innermost intersection circle of σ and σ' cuts σ' into two disks S^+ and S^-. If σ^- is almost peripheral then σ^+ is isotopic to σ after filling the punctures.

3.3 Haken spheres

Just as in the closed case, we call an essential sphere which intersects a Heegaard splitting of a punctured manifold in a single curve a *Haken sphere*. For punctured manifolds, almost peripheral and surviving Haken spheres behave slightly differently.

On the one hand, using the same strategy as in the proof of Theorem 3.3, we obtain the following corollary of Theorem 2.9:

Theorem 3.6 (surviving Haken's lemma) *If M contains a surviving sphere and $M = V \cup_S V'$ is a Heegaard splitting, then there is a surviving Haken sphere.*

Proof Denote by σ an essential surviving sphere in M. Let M' be the 3–manifold obtained from M by gluing a ball to each boundary component. Denote by $B \subset M'$ the disjoint union of the resulting balls. By definition of almost peripheral, the image of σ in M' is still essential. Thus, Haken's lemma (Theorem 2.9) applies, and yields a Haken sphere $\sigma' \subset M'$. By an isotopy preserving the Heegaard surface, we may assume that σ' is disjoint from B. We can thus interpret σ' as a sphere in $M \subset M'$, where it is the desired Haken sphere. □

On the other hand, almost peripheral spheres are also Haken spheres:

Lemma 3.7 (almost peripheral strong Haken theorem) *Let M be a 3–manifold with at least two punctures, and $M = V \cup_S V'$ be a Heegaard splitting. Then any almost peripheral sphere σ in M is isotopic to a Haken sphere.*

Proof We begin with the case where σ cuts off exactly two punctures δ_1 and δ_2. The almost peripheral sphere σ is then isotopic to the boundary of a regular neighborhood of $\delta_1 \cup \alpha \cup \delta_2$, where $\alpha \subset M$ is a properly embedded arc. We may homotope α to lie in S, as any arc in a handlebody is homotopic into the boundary. However, the arc may now not be embedded anymore. We can remove the self-intersections by "popping subarcs over $\delta_1 \cap S$". To be more precise, parametrize $\alpha\colon [0,1] \to S$ so that it starts on $\delta_1 \cap S$, and homotope so that all self-intersections are transverse. Consider the first self-intersection point $\alpha(t) = \alpha(s)$ for $t < s$. In particular, this implies that $\alpha|_{[0,t]}$ is an embedded arc.

Now homotope a small subarc $\alpha|_{[s-\epsilon,s+\epsilon]}$ to instead be the arc obtained by following $\alpha|_{[0,t]}$ backwards to $\delta_1 \cap S$, following around $\delta_1 \cap S$, and returning along $\alpha|_{[0,t]}$ (see Figure 3). This homotopy is possible in V, and the resulting arc has at least one fewer self-intersection.

After a finite number of modifications of this type, the boundary of a regular neighborhood of $\delta_1 \cup \alpha \cup \delta_2$ (which is homotopic to σ) intersects S in a single curve, and thus is a Haken sphere. By a theorem of Laudenbach [10], homotopy and isotopy are the same for spheres in 3–manifolds; hence the claim follows.[1]

[1]One could also avoid citing this theorem by isotoping α into a regular neighborhood of S and resolving crossings of the projection to S by isotopies which slide strands over the puncture similar to Figure 3.

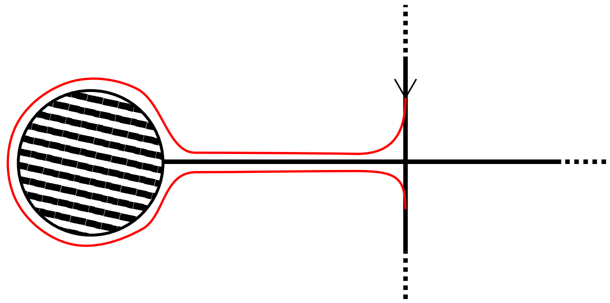


Figure 3: Removing self-intersections of an arc joining two spots.

Now we suppose σ is a sphere cutting off $k > 2$ punctures. Then there is a sphere σ', disjoint from σ, which cuts off two punctures, and which is contained in the punctured S^3 bounded by σ. By the initial case, σ' may be assumed to be Haken. Let M' be the manifold obtained by cutting M at σ', with the induced Heegaard splitting; observe that $\sigma \subset M'$ is still almost peripheral, but now cuts off at most $k-1$ spheres. By induction, σ is a Haken sphere. □

Since any essential nonperipheral sphere in a punctured S^3 is almost peripheral, this implies the following:

Corollary 3.8 (strong Haken theorem for punctured 3–spheres) *Any essential sphere in a punctured 3–sphere is isotopic to a Haken sphere.*

4 Heegaard splittings of $n(S^1 \times S^2)$

In this section, we study Heegaard splittings of a specific manifold, namely:

Definition 4.1 We denote the double of the genus-n handlebody by W_n. It is the connected sum of n copies of $S^2 \times S^1$.

A reader only interested in the strong Haken theorem may safely skip ahead to the next section. Our goal here will be to prove that, similar to Waldhausen's theorem for the 3–sphere, all Heegaard splittings of W_n are "standard" in the following sense:

Definition 4.2 A Heegaard splitting of W_n is *standard* if it is the double of a genus-n handlebody. A *standard* Heegaard splitting of W_n is a Heegaard splitting that is the connected sum of n copies of W_1 with the standard Heegaard splitting.

Waldhausen seems to claim in [18] that all Heegaard splittings of W_n are standard (although it is not entirely clear up to which equivalence relation, and the proof sketch is incomplete). In the unpublished preprint [3], Oertel and Navarro Carvalho prove the result, using results on the homeomorphism groups of handlebodies and W_n (in a very similar way to the argument we will use below). In this section, we show that these techniques could also be used to prove a strong Haken theorem (and obtain the uniqueness of splittings as a corollary). We want to emphasize that this of course follows from the general strong Haken theorem (Theorem 1.1), but consider the argument using homeomorphisms of W_n interesting enough to warrant this alternative proof.

Proposition 4.3 *Every unstabilized Heegaard splitting of W_1 is standard.*

Proof Suppose that $W_1 = V \cup_S V'$ is a Heegaard splitting. We wish to show that $W_1 = V \cup_S V'$ is standard. Since W_1 is reducible, Haken's lemma tells us that there is a Haken sphere R for $W_1 = V \cup_S V'$. Denote $V \cap R$ by D and $V' \cap R$ by D'. Note that all essential spheres in W_1, in particular R, are isotopic to $S^2 \times \text{(point)}$.

We may assume that S intersects a bicollar of R in an annulus $(S \cap R) \times [-1, 1]$. Removing this bicollar leaves a copy of $S^2 \times [-1, 1]$, ie a twice-punctured 3–sphere that inherits a Heegaard splitting. By Theorem 3.3, this Heegaard splitting is either of genus 0 or stabilized.

Since $W_1 = V \cup_S V'$ is unstabilized, the Heegaard splitting obtained on the complement of $S^2 \times [-1, 1]$ must be of genus 0. Specifically, the splitting surface is a twice-punctured 2–sphere, ie an annulus. Hence we can reconstruct $W_1 = V \cup_S V'$: Indeed, say, V, is composed of a 3–ball attached to the two copies $D \times \{\pm 1\}$ of D. It follows that V is a solid torus. The same is true of V', and hence $W_1 = V \cup_S V'$ is the standard Heegaard splitting. □

First, we have the following classical result due to Griffiths [4]:

Theorem 4.4 *The action of the mapping class group of a handlebody V_n on its fundamental group $\pi_1(V_n) = F_n$ induces a surjection*

$$\mathrm{Mcg}(V_n) \to \mathrm{Out}(F_n) \to 1.$$

We remark that the kernel of this map is quite complicated, and generated by twists about disk-bounding curves [12]. Next, we need a theorem of François Laudenbach [10] (see also [2; 11] for a modern proof):

Theorem 4.5 *The action of the mapping class group of a doubled handlebody W_n on its fundamental group $\pi_1(W_n) = F_n$ induces a short exact sequence*

$$1 \to K \to \mathrm{Mcg}(W_n) \to \mathrm{Out}(F_n) \to 1.$$

The kernel K is finite, generated by Dehn twists about nonseparating spheres and acts trivially on the isotopy class of every embedded sphere or loop.

Corollary 4.6 *For the standard Heegaard splitting of W_n, every essential sphere in W_n is isotopic to a Haken sphere.*

Proof First, any two nonseparating spheres in W_n can be mapped to each other by a homeomorphism. Namely, the complement of such a sphere is homeomorphic to W_{n-1} with two punctures. Similarly, separating spheres can be mapped to each other if and only if the fundamental groups of the complements are free groups of the same rank (as the complement is a disjoint union of once-punctured W_k and W_{n-k}).

Next, observe that there are Haken spheres of all such possible types, obtained by doubling a suitable disk in the handlebody.

Let $i: V_n \to W_n$ be the inclusion induced by doubling. Observe that on the one hand, the boundary of V_n maps under i to the standard Heegaard splitting of W_n, and on the other hand i induces an isomorphism i_* of fundamental groups. For any outer automorphism $\varphi \in \mathrm{Out}(\pi_1(V_n))$ of the fundamental group of V_n, by Theorem 4.4 there is a homeomorphism $f: V_n \to V_n$ inducing it. Let $F: W_n \to W_n$ be the homeomorphism of W_n obtained by doubling f. Observe that F preserves the standard Heegaard splitting of W_n by construction, and F induces φ via the isomorphism $i_*: \pi_1(V_n) \to \pi_1(W_n)$. Since φ was arbitrary, this shows that any outer automorphism of $\pi_1(W_n)$ can in fact be realized by a homeomorphism of W_n preserving the standard Heegaard splitting.

Together with Laudenbach's Theorem 4.5 this shows that any sphere is isotopic to the image of a Haken sphere under a homeomorphism preserving the standard Heegaard splitting — hence, it is isotopic to a Haken sphere. □

Lemma 4.7 *There is a unique Heegaard splitting of W_n of genus n.*

Proof Connected sum decompositions of W_n are not unique. However, let $W_n = V \cup_S V'$ be the standard Heegaard splitting and let $W_n = X \cup_Y X'$ be any Heegaard splitting of genus n. By repeated application of Theorem 2.9, there are Haken spheres $R_1 \cup \dots \cup R_{n-1}$ for W_n that cut $W_n = X \cup_Y X'$ into standard Heegaard splittings of W_1. By Corollary 4.6, $R_1, \dots, R_{n-1}$ are also Haken spheres for $W_n = V \cup_S V'$. By an Euler characteristic argument, these cut $W_n = V \cup_S V'$ into genus-1 Heegaard splittings of the summands. By Proposition 4.3 these are standard. In particular, Y is isotopic to S. □

Proof of Proposition 1.2 For an unstabilized Heegaard splitting, this is Lemma 4.7. Furthermore, if $n = 1$, then this follows from Proposition 4.3. So suppose that $n > 1$ and $W_n = V \cup_S V'$ is stabilized. By Corollary 4.6, there is a Haken sphere R that decomposes W_n into $W_1 \# W_{n-1}$. Moreover, by [13], one of the Heegaard splittings inherited by the summands is stabilized. By induction, $W_n = V \cup_S V'$ is a stabilization of a connected sum of standard Heegaard splittings, ie a stabilization of the standard Heegaard splitting of W_n. □

5 Strong Haken theorem

Combining the uniqueness of Heegaard splittings for W_n (Proposition 1.2) with Corollary 4.6 yields a *strong Haken theorem* for W_n: any sphere in W_n is isotopic to a Haken sphere. This statement was recently proved by Scharlemann [16] for all 3–manifolds. In this section, we provide a short independent proof of this theorem for closed manifolds and manifold with spherical boundary.

The following proof proceeds by two nested inductions. It naturally involves 3–manifolds with spherical boundary, even if we just want to prove the theorem in the closed case. Recall that for such 3–manifolds, we decree that each boundary sphere (puncture) meets the splitting surface in a single simple closed curve.

Proof of Theorem 1.1 We prove the theorem by considering all punctured 3–manifolds and all Heegaard splittings, ordered according to a suitable complexity. Namely, if $M = V \cup_S V'$ is a Heegaard splitting, we define the *complexity* as the pair $(g(S), n(S))$ of genus and number of spots of the handlebodies (ordered lexicographically). We perform a nested induction on the genus g and the number of boundary components n. The argument for the inductive step is in fact the same in both cases, and so we describe both inductions simultaneously.

Induction start ($g = 0$ and $n \geq 0$) Observe that the only punctured 3–manifold that can be obtained from a Heegaard splitting of genus 0 is the 3–sphere. Thus, the strong Haken theorem in this case is simply Corollary 3.8.

Induction steps Now suppose that the strong Haken theorem is known for all manifolds of complexity at most (g, n) and that M is a manifold of complexity $(g, n+1)$, or suppose that the strong Haken theorem is known for all manifolds of complexity $(g, k), k \geq 0$ and that M is a manifold of complexity $(g+1, 0)$.

First observe that by Lemma 3.7 any almost peripheral sphere in M is isotopic to a Haken sphere. We thus have to show that surviving spheres in M are also isotopic to Haken spheres.

Claim 5.1 *Suppose that R is a surviving Haken sphere in M, and suppose that R' is a surviving sphere disjoint from R and not isotopic to R. Then R' is isotopic to a Haken sphere.*

Proof Denote by $M - R$ the punctured 3–manifold obtained by cutting at R. $M - R$ has two punctures more than M, corresponding to the two sides of R. $M - R$ has one or two components, depending on whether R is separating or not.

Let M' be the component of $M - R$ containing R'. This manifold inherits a Heegaard splitting from $V \cup_S V'$ with splitting surface a component of $S' = S - (R \cap S)$. If $R \cap S$ is nonseparating, then $g(S') < g(S)$. If $R \cap S$ is separating, then either the genus or the number of boundary components is smaller for S'. In either case, $(g(S'), n(S')) < (g(S), n(S))$ lexicographically.

The sphere R' defines an essential sphere in M': if it would bound a ball in M', the same would be true in M (violating incompressibility of R' in M), and if it were isotopic to a boundary component of M', then R' would be peripheral in M or isotopic to R (both of which we exclude).

If R' is almost peripheral in M', then by Theorem 3.6 it is isotopic to a Haken sphere in M'. Otherwise, since the complexity of the splitting of M' is smaller than the original one, we can use the inductive hypothesis on M' to conclude that R' is isotopic to a Haken sphere in M'. Interpreting M' as a submanifold of M, and using that the Heegaard splitting of M' is inherited from M, this shows that R' is isotopic to a Haken sphere in M as well. □

If M contains any surviving spheres, then the surviving Haken lemma (Theorem 3.6) implies that there is a surviving Haken sphere σ_0. Connectivity of the surviving sphere complex (Lemma 3.5), together with Claim 5.1, then inductively implies that any surviving sphere is isotopic to a Haken sphere. □

References

[1] **J Bowden**, **S W Hensel**, **R Webb**, *Quasi-morphisms on surface diffeomorphism groups*, J. Amer. Math. Soc. 35 (2022) 211–231 MR Zbl

[2] **T Brendle**, **N Broaddus**, **A Putman**, *The mapping class group of connect sums of* $S^2 \times S^1$, Trans. Amer. Math. Soc. 376 (2023) 2557–2572 MR Zbl

[3] **L N Carvalho**, **U Oertel**, *A classification of automorphisms of compact* 3*–manifolds*, preprint (2005) arXiv math/0510610

[4] **H B Griffiths**, *Automorphisms of a* 3*–dimensional handlebody*, Abh. Math. Sem. Univ. Hamburg 26 (1964) 191–210 MR Zbl

[5] **F Gültepe**, **C J Leininger**, **W Pho-On**, *A universal Cannon–Thurston map and the surviving curve complex*, Trans. Amer. Math. Soc. Ser. B 9 (2022) 99–143 MR Zbl

[6] **W Haken**, *Some results on surfaces in* 3*–manifolds*, from "Studies in modern topology" (P J Hilton, editor), Studies in Math. 5, Math. Assoc. America, Washington, DC (1968) 39–98 MR Zbl

[7] **A Hatcher**, *Homological stability for automorphism groups of free groups*, Comment. Math. Helv. 70 (1995) 39–62 MR Zbl

[8] **J Hempel**, 3*–Manifolds*, Ann. of Math. Stud. 86, Princeton Univ. Press (1976) MR Zbl

[9] **W Jaco**, *Lectures on three-manifold topology*, CBMS Regional Conf. Ser. Math. 43, Amer. Math. Soc., Providence, RI (1980) MR Zbl

[10] **F Laudenbach**, *Sur les* 2*–sphères d'une variété de dimension* 3, Ann. of Math. 97 (1973) 57–81 MR Zbl

[11] **F Laudenbach**, *Topologie de la dimension trois: homotopie et isotopie*, Astérisque 12, Soc. Math. France, Paris (1974) MR Zbl

[12] **E Luft**, *Actions of the homeotopy group of an orientable* 3*–dimensional handlebody*, Math. Ann. 234 (1978) 279–292 MR Zbl

[13] **R Qiu**, **M Scharlemann**, *A proof of the Gordon conjecture*, Adv. Math. 222 (2009) 2085–2106 MR Zbl

[14] **M Scharlemann**, *Generating the genus* $g+1$ *Goeritz group of a genus* g *handlebody*, from "Geometry and topology down under" (C D Hodgson, W H Jaco, M G Scharlemann, S Tillmann, editors), Contemp. Math. 597, Amer. Math. Soc., Providence, RI (2013) 347–369 MR Zbl

[15] **M Scharlemann**, *Generating the Goeritz group of* S^3 (2020) arXiv 2011.10613 To appear in J. Assoc. Math. Res.

[16] **M Scharlemann**, *A strong Haken theorem*, Algebr. Geom. Topol. 24 (2024) 717–753 MR Zbl

[17] **J Schultens**, *Introduction to* 3*–manifolds*, Graduate Studies in Math. 151, Amer. Math. Soc., Providence, RI (2014) MR Zbl

[18] **F Waldhausen**, *Heegaard–Zerlegungen der* 3*–Sphäre*, Topology 7 (1968) 195–203 MR Zbl

Department Mathematisches Institut, LMU München
Munich, Germany
Department of Mathematics, University of California Davis
Davis, CA, United States

hensel@math.lmu.de, jcs@math.ucdavis.edu

Received: 6 July 2022 Revised: 6 February 2023

Algebraic & Geometric Topology **24**:5 (2024) 2721–2777
DOI: 10.2140/agt.2024.24.2721
Published: 19 August 2024

What are GT–shadows?

Vasily A Dolgushev
Khanh Q Le
Aidan A Lorenz

Let B_4 (resp. PB_4) be the braid group (resp. the pure braid group) on 4 strands and $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ be the poset whose elements are finite-index normal subgroups N of B_4 that are contained in PB_4. We introduce GT–shadows, which may be thought of as "approximations" to elements of the profinite version $\widehat{\mathsf{GT}}$ of the Grothendieck–Teichmüller group (Drinfeld 1991). We prove that GT–shadows form a groupoid whose objects are elements of the underlying set $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. GT–shadows coming from elements of $\widehat{\mathsf{GT}}$ satisfy various additional properties and we investigate these properties. We establish an explicit link between GT–shadows and the group $\widehat{\mathsf{GT}}$. Selected results of computer experiments on GT–shadows are presented. In the appendix we give a complete description of GT–shadows in the abelian setting. We also prove that, in the abelian setting, every GT–shadow comes from an element of $\widehat{\mathsf{GT}}$. Objects very similar to GT–shadows were introduced by D Harbater and L Schneps (1997). A variation of the concept of GT–shadows for the gentle version of $\widehat{\mathsf{GT}}$ was studied by P Guillot (2016 and 2018).

14F35, 14H30, 18M60; 14H57

1 Introduction

The absolute Galois group $G_{\mathbb{Q}}$ of the field $\mathbb{Q}$ of rational numbers and the Grothendieck–Teichmüller group $\widehat{\mathsf{GT}}$ introduced by V Drinfeld in [7] are among the most mysterious objects in mathematics. A far from complete list of references includes Ellenberg [8], Fresse [9], Harbater and Schneps [14], Ihara [15], Lochak and Schneps [19], Nakamura and Schneps [22], Pop [23] and Schneps [25].

Using the outer action of $G_{\mathbb{Q}}$ on the algebraic fundamental group of $\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, 1, \infty\}$, one can produce a natural group homomorphism

$$G_{\mathbb{Q}} \to \widehat{\mathsf{GT}} \tag{1-1}$$

and, due to Belyi's theorem [3], this homomorphism is injective. Although both $G_{\mathbb{Q}}$ and $\widehat{\mathsf{GT}}$ are uncountable, it is very hard to produce explicit examples of elements in $G_{\mathbb{Q}}$ and in $\widehat{\mathsf{GT}}$. In particular, the famous question on surjectivity of (1-1) posed by Ihara at his ICM address [15] is still open.

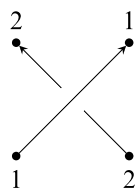
The group $G_{\mathbb{Q}}$ can be studied by investigating finite-degree field extensions of $\mathbb{Q}$. In fact $G_{\mathbb{Q}}$ coincides with the limit of the functor that sends a finite-degree Galois extension K of $\mathbb{Q}$ to the Galois group $\mathrm{Gal}(K/\mathbb{Q})$. The goal of this paper is to propose a loose analog of such a functor for $\widehat{\mathsf{GT}}$.

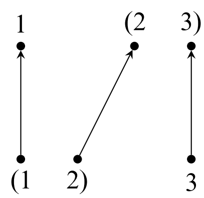
The most elegant definition of the group $\widehat{\mathsf{GT}}$ involves (the profinite completion $\widehat{\mathsf{PaB}}$ of) the operad PaB of parenthesized braids; see Bar-Natan [1], Fresse [9, Chapter 6] and Tamarkin [28]. PaB is an operad in the category of groupoids that is "assembled from" braid groups B_n for all $n \geq 1$. The objects of $\mathsf{PaB}(n)$ are words of the free magma generated by symbols $1, 2, \dots, n$ in which each generator appears exactly once. For example, $\mathsf{PaB}(3)$ has exactly 12 objects: (12)3, (21)3, (23)1, (32)1, (31)2, (13)2, 1(23), 2(13), 2(31), 3(21), 3(12), 1(32). For every $n \geq 2$ and every object τ of PaB, we have

$$\mathrm{Aut}_{\mathsf{PaB}(n)}(\tau) = \mathrm{PB}_n,$$

where PB_n is the pure braid group on n strands.

As an operad in the category of groupoids, PaB is generated by the two morphisms

(1-2) $\beta :=$ 

and $\alpha :=$ 

.

Moreover, any relation on β and α in PaB is a consequence of the pentagon relation and the two hexagon relations; see (A-13), (A-14) and (A-15) in Section A.3. The hexagon relations come from two ways of connecting (12)3 to 3(12) and two ways of connecting 1(23) to (23)1 in $\mathsf{PaB}(3)$. Similarly, the pentagon relation comes from two ways of connecting ((12)3)4 to 1(2(34)) in $\mathsf{PaB}(4)$. For more details about the operad PaB and its profinite completion $\widehat{\mathsf{PaB}}$, see Appendix A.

By definition, $\widehat{\mathsf{GT}}$ is the group $\mathrm{Aut}(\widehat{\mathsf{PaB}})$ of (continuous) automorphisms[1] of the profinite completion $\widehat{\mathsf{PaB}}$ of PaB.

Since the morphisms β and α from (1-2) are topological generators of $\widehat{\mathsf{PaB}}$, every $\hat{T} \in \widehat{\mathsf{GT}}$ is uniquely determined by its values

$$\hat{T}(\beta) \in \mathrm{Hom}_{\widehat{\mathsf{PaB}}}((1,2),(2,1)) \quad \text{and} \quad \hat{T}(\alpha) \in \mathrm{Hom}_{\widehat{\mathsf{PaB}}}((1,2)3, 1(2,3)). \tag{1-3}$$

[1] We tacitly assume that our automorphisms act as identity on objects.

Moreover, since $\mathrm{Aut}_{\widehat{\mathsf{PaB}}}((1,2)3) = \widehat{\mathrm{PB}}_3$, $\mathrm{Aut}_{\widehat{\mathsf{PaB}}}((1,2)) = \widehat{\mathrm{PB}}_2$ and $\widehat{\mathrm{PB}}_2 \cong \widehat{\mathbb{Z}}$, the underlying set of $\widehat{\mathsf{GT}}$ can be identified with the subset of pairs $(\hat{m}, \hat{f}) \in \widehat{\mathbb{Z}} \times \widehat{\mathrm{PB}}_3$ satisfying some relations and technical conditions.

Recall that PB_3 is isomorphic to the direct product $\mathsf{F}_2 \times \mathbb{Z}$ of the free group F_2 on two generators and the infinite cyclic group. The F_2–factor is generated by the two standard generators x_{12} and x_{23}, and the $\mathbb{Z}$–factor is generated by the element $c := x_{23}x_{12}x_{13}$. In this paper, we tacitly identify F_2 (resp. its profinite completion $\widehat{\mathsf{F}}_2$) with the subgroup $\langle x_{12}, x_{23}\rangle \leq \mathrm{PB}_3$ (resp. the topological closure of $\langle x_{12}, x_{23}\rangle$ in $\widehat{\mathrm{PB}}_3$). Occasionally, we denote the standard generators of F_2 by x and y.

One can show[2] (see, for example, Corollary 2.22 in Section 2 of this paper) that, for every $\hat{T} \in \widehat{\mathsf{GT}}$, the corresponding element $\hat{f} \in \widehat{\mathrm{PB}}_3$ belongs to the topological closure $([\widehat{\mathsf{F}}_2, \widehat{\mathsf{F}}_2])^{\mathrm{cl}}$ of the commutator subgroup of $\widehat{\mathsf{F}}_2$.

Remark 1.1 Due to Proposition 2.19, the restriction of every (continuous) automorphism $\hat{T} \in \mathrm{Aut}(\widehat{\mathsf{PaB}})$ to $\widehat{\mathsf{F}}_2 \leq \widehat{\mathrm{PB}}_3 = \mathrm{Aut}_{\widehat{\mathsf{PaB}}}((1,2)3)$ gives us an automorphism of $\widehat{\mathsf{F}}_2$. In fact, many authors introduce $\widehat{\mathsf{GT}}$ as the subgroup of (continuous) automorphisms of $\widehat{\mathsf{F}}_2$ of the form

$$x \mapsto x^{\hat{\lambda}}, \quad y \mapsto \hat{f}^{-1} y^{\hat{\lambda}} \hat{f},$$

where the pair $(\hat{\lambda}, \hat{f}) \in \widehat{\mathbb{Z}}^{\times} \times ([\widehat{\mathsf{F}}_2, \widehat{\mathsf{F}}_2])^{\mathrm{cl}}$ satisfies certain cocycle relations and the "invertibility condition". Another equivalent definition of $\widehat{\mathsf{GT}}$ is based on the use of the outer automorphisms of the profinite completions of the pure mapping class groups. For more details about this definition, we refer the reader to [14, Main Theorem].

Remark 1.2 It is known [18, Theorem 2] that, for every $(\hat{m}, \hat{f}) \in \widehat{\mathsf{GT}}$, the element $\hat{f}$ satisfies further, rather subtle, properties. It would be interesting to investigate whether GT–shadows satisfy consequences of these properties.

1.1 The link between $G_{\mathbb{Q}}$ and $\widehat{\mathsf{GT}}$

For completeness, we briefly recall here the link between the absolute Galois group $G_{\mathbb{Q}}$ of rationals and the Grothendieck–Teichmüller group $\widehat{\mathsf{GT}}$.

Applying the basic theory of the algebraic fundamental group (see for instance Grothendieck [11] and Szamuely [27, Section 5.6]) to

$$\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, 1, \infty\},$$

we get an outer action of the absolute Galois group $G_{\mathbb{Q}}$ on $\widehat{\mathsf{F}}_2$. Using the fact that this action preserves the inertia subgroups, we can lift this outer action to an honest action of the form

$$g(x) = x^{\chi(g)}, \quad g(y) = \hat{f}_g(x,y)^{-1} y^{\chi(g)} \hat{f}_g(x,y) \quad \text{for } g \in G_{\mathbb{Q}}, \tag{1-4}$$

[2] This statement can also be found in many introductory papers on $\widehat{\mathsf{GT}}$.

where $\chi\colon G_{\mathbb{Q}} \to \widehat{\mathbb{Z}}^{\times}$ is the cyclotomic character and $\hat{f}_g(x, y)$ is an element of $([\hat{\mathsf{F}}_2, \hat{\mathsf{F}}_2])^{\mathrm{cl}}$ that depends only on g.

It is known — see Drinfeld [7, Section 4], Ihara [15, Section 3] and Szamuely [27, Theorem 4.7.7 and Fact 4.7.8] — that:

- For all $g \in G_{\mathbb{Q}}$, the pair $\left(\frac{1}{2}(\chi(g) - 1), \hat{f}_g(x, y)\right) \in \widehat{\mathbb{Z}} \times \hat{\mathsf{F}}_2$ defines an element of $\widehat{\mathsf{GT}}$.
- The assignment $g \in G_{\mathbb{Q}} \mapsto \left(\frac{1}{2}(\chi(g)-1), \hat{f}_g(x, y)\right) \in \widehat{\mathbb{Z}} \times \widehat{F}_2$ defines the group homomorphism (1-1).
- Finally, using Belyi's theorem [3], one can prove that the homomorphism (1-1) is injective.

For more details, we refer the reader to Ihara [16].

1.2 The groupoid GTSh of GT–shadows and its link to $\widehat{\mathsf{GT}}$

Let us denote by $\mathsf{PaB}^{\leq 4}$ the truncation of the operad PaB up to arity 4, ie

$$\mathsf{PaB}^{\leq 4} := \mathsf{PaB}(1) \sqcup \mathsf{PaB}(2) \sqcup \mathsf{PaB}(3) \sqcup \mathsf{PaB}(4).$$

Moreover, let $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ be the poset of finite-index normal subgroups $\mathsf{N} \lhd \mathrm{B}_4$ such that $\mathsf{N} \leq \mathrm{PB}_4$.

To every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, we assign an equivalence relation $\sim_{\mathsf{N}}$ on $\mathsf{PaB}^{\leq 4}$ that is compatible with the structure of the truncated operad and the composition of morphisms. For every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the quotient

$$\mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}}$$

is a truncated operad in the category of *finite* groupoids.

In this paper, we introduce a groupoid GTSh whose objects are elements of the underlying set $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. Morphisms from $\widetilde{\mathsf{N}}$ to N are isomorphisms of truncated operads

(1-5) $$\mathsf{PaB}^{\leq 4}/\sim_{\widetilde{\mathsf{N}}} \xrightarrow{\cong} \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}}.$$

We call these isomorphisms GT–*shadows*.

Just like PaB, the truncated operad $\mathsf{PaB}^{\leq 4}$ is generated by the braiding $\beta \in \mathsf{PaB}(2)$ and the associator $\alpha \in \mathsf{PaB}(3)$. Hence morphisms of GTSh with the target $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ are in bijection with pairs

(1-6) $$(m + N_{\mathrm{ord}}\mathbb{Z}, f\mathsf{N}_{\mathrm{PB}_3}) \in \mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z} \times \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$$

that satisfy appropriate versions of the hexagon relations, the pentagon relation and some technical conditions. Here, the integer N^{ord} and the (finite-index) normal subgroup $\mathsf{N}_{\mathrm{PB}_3} \trianglelefteq \mathrm{PB}_3$ are obtained from N via a precise procedure described in Section 2.2.

We denote by $\mathsf{GT}(\mathsf{N})$ the set of such pairs (1-6) and identify them with GT–shadows whose target is N. From now on, we denote by $[(m, f)]$ the GT–shadow represented by a pair $(m, f) \in \mathbb{Z} \times \mathrm{PB}_3$.

A GT–shadow $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ is called *genuine* if there exists an element $\hat{T} \in \widehat{\mathsf{GT}}$ such that the diagram

$$\begin{array}{ccc} \widehat{\mathsf{PaB}}^{\le 4} & \xrightarrow{\hat{T}} & \widehat{\mathsf{PaB}}^{\le 4} \\ \downarrow & & \downarrow \\ \mathsf{PaB}^{\le 4}/\sim_{\tilde{\mathsf{N}}} & \xrightarrow{\cong} & \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}} \end{array} \tag{1-7}$$

commutes. In (1-7), the lower-horizontal arrow is the isomorphism corresponding to $[(m, f)]$ and the vertical arrows are the canonical projections. If such $\hat{T}$ does not exist, we say that the GT–shadow $[(m, f)]$ is *fake*.[3]

In this paper, we show that genuine GT–shadows satisfy additional conditions. For example, every genuine GT–shadow in $\mathsf{GT}(\mathsf{N})$ can be represented by a pair (m, f) with[4]

$$f \in [\mathsf{F}_2, \mathsf{F}_2], \tag{1-8}$$

where $[\mathsf{F}_2, \mathsf{F}_2]$ is the commutator subgroup of $\mathsf{F}_2 \le \mathrm{PB}_3$.

A GT–shadow $[(m, f)]$ satisfying all these additional conditions (see Definition 2.20) is called *charming*. In this paper, we show that charming GT–shadows form a subgroupoid of GTSh and we denote this subgroupoid by $\mathsf{GTSh}^{\heartsuit}$.

The groupoid $\mathsf{GTSh}^{\heartsuit}$ is highly disconnected. However, it is easy to see that, for every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the connected component $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ is a finite groupoid; see Proposition 3.1. In all examples we have considered so far (see Dolgushev [4] and Section 4 of this paper), $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ has at most two objects and, for many of examples of $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the groupoid $\mathsf{GTSh}_{\mathrm{conn}}(\mathsf{N})$ has exactly one object, ie $\mathsf{GT}(\mathsf{N})$ is a group. Such elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ play a special role and we call them *isolated*. We denote by $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ the subposet of isolated elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

In this paper, we show that the subposet $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ is cofinal, ie for every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, there exists $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ such that $\mathsf{K} \le \mathsf{N}$. We show that the assignment $\mathsf{N} \mapsto \mathsf{GT}(\mathsf{N})$ upgrades to a functor $\mathcal{ML}$ from the poset $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the category of finite groups and we prove that the limit of this functor is precisely the Grothendieck–Teichmüller group $\widehat{\mathsf{GT}}$; see Theorem 3.8.

Remark 1.3 Recall from Harbater and Schneps [14] that, omitting the pentagon relation from the definition of $\widehat{\mathsf{GT}}$, we get the gentle version $\widehat{\mathsf{GT}}_0$ of the Grothendieck–Teichmüller group. It is not hard to show that $\widehat{\mathsf{GT}}_0$ is the group of continuous automorphisms of the truncated operad $\widehat{\mathsf{PaB}}^{\le 3}$ and $\widehat{\mathsf{GT}}$ is a subgroup of $\widehat{\mathsf{GT}}_0$. P Guillot [12; 13] studies a variant of GT–shadows for this gentle version $\widehat{\mathsf{GT}}_0$ of the Grothendieck–Teichmüller group.

[3] This name was suggested to the authors by David Harbater.

[4] It should be mentioned that, in the computer implementation [4], we only considered GT–shadows of the form $[(m, f)]$ with $f \in \mathsf{F}_2 \le \mathrm{PB}_3$.

1.3 Organization of the paper

In Section 2, we introduce the poset of compatible equivalence relations on the truncated operad $\mathsf{PaB}^{\le 4}$, and we show that $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ can be identified with the subposet of this poset. We introduce the concept of GT–pair and show that GT–pairs coming from elements of $\widehat{\mathsf{GT}}$ satisfy certain conditions. This consideration motivates the concept of GT–shadow; see Definition 2.9. We prove that GT–shadows form a groupoid GTSh: objects of this groupoid are elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and morphisms are GT–shadows.

In Section 2, we also investigate further conditions on GT–shadows coming from elements of $\widehat{\mathsf{GT}}$, introduce charming GT–shadows and prove that charming GT–shadows form a subgroupoid of GTSh. In this section, we introduce a natural functor $\mathsf{Ch}_{\mathrm{cyclot}}$ from GTSh to the category of finite cyclic groups. We call this functor *the virtual cyclotomic character*.

In Section 3, we introduce an important subposet $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and construct a functor $\mathcal{ML}$ from $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the category of finite groups. In this section, we prove that the limit of the functor $\mathcal{ML}$ is precisely the Grothendieck–Teichmüller group $\widehat{\mathsf{GT}}$.

In Section 4, we present selected results of computer experiments. We outline the basic information about 35 selected elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and the corresponding connected components of the groupoid GTSh. We say a few words about selected remarkable examples. Finally, we discuss two versions of the Furusho property (see Properties 4.2 and 4.3) and list selected open questions.

In Appendix A, we give a brief reminder of (pure) braid groups, the operad PaB and its completion.

In Appendix B, we give a complete description of charming GT–shadows in the abelian setting and we prove that, in this setting, every charming GT–shadow is genuine; see Theorem B.2.

1.4 Notational conventions

For a set X with an equivalence relation and $a \in X$ we will denote by $[a]$ the equivalence class that contains the element a. For a category $\mathscr{C}$, $\mathrm{Ob}(\mathscr{C})$ denotes the set of objects of $\mathscr{C}$. Every poset J is naturally a category: $\mathrm{Ob}(J) := J$, for $a, b \in J$, $J(a,b)$ is a singleton if $a \le b$ and $J(a,b) := \varnothing$ if $a \not\le b$. For a groupoid $\mathscr{G}$, the notation $\gamma \in \mathscr{G}$ means that γ is a *morphism* of this groupoid.

Every finite group is tacitly considered with the discrete topology. For a group G, $\widehat{G}$ denotes the profinite completion (see Ribes and Zalesskii [24]) of G. The notation $[G,G]$ is reserved for the commutator subgroup of G. For a normal subgroup $H \trianglelefteq G$ of finite index, we denote by $\mathsf{NFI}_H(G)$ the poset of finite-index normal subgroups N in G such that $N \le H$. Moreover, $\mathsf{NFI}(G) := \mathsf{NFI}_G(G)$, ie $\mathsf{NFI}(G)$ is the poset of normal finite-index subgroups of a group G.

For a group G and elements $\mathsf{K} \le \mathsf{N}$ of the poset $\mathsf{NFI}(G)$, the notation $\mathscr{P}_{\mathsf{N}}$ (resp. $\mathscr{P}_{\mathsf{K},\mathsf{N}}$) is reserved for the reduction homomorphism $G \to G/\mathsf{N}$ (resp. $G/\mathsf{K} \to G/\mathsf{N}$). The notation $\widehat{\mathscr{P}}_{\mathsf{N}}$ is reserved for the canonical (continuous) homomorphism from $\widehat{G}$ to G/N. Similar notation is used for the canonical functors to finite quotients of a groupoid.

The notation B_n (resp. PB_n) is reserved for the Artin braid group on n strands (resp. the pure braid group on n strands). S_n denotes the symmetric group on n letters. The standard generators of B_n are denoted by $\sigma_1, \dots, \sigma_{n-1}$ and the standard generators of PB_n are denoted by x_{ij} for $1 \le i < j \le n$. We will tacitly identify the free group F_2 on two generators with the subgroup $\langle x_{12}, x_{23} \rangle$ of PB_3.

We will freely use the language of operads as in Dolgushev and Rogers [6, Section 3], Fresse [9, Chapter 1], Loday and Vallette [20], Markl, Shnider and Stasheff [21] and Stasheff [26]. In this paper, we work with operads in the category of sets and in the category of (topological) groupoids. The category of topological groupoids is understood in the "strict sense". For example, the associativity axioms for the *elementary insertions*[5] $\circ_i$ (for operads in the category of groupoids) are satisfied "on the nose".

For an integer $q \ge 1$, a *q–truncated operad* in the category of groupoids is a collection of groupoids $\{\mathcal{G}(n)\}_{1 \le n \le q}$ such that

- for every $1 \le n \le q$, the groupoid $\mathcal{G}(n)$ is equipped with an action of S_n,
- for every triple of integers i, n, m such that $1 \le i \le n$, $n, m, n+m-1 \le q$ we have functors

$$\circ_i : \mathcal{G}(n) \times \mathcal{G}(m) \to \mathcal{G}(n+m-1), \tag{1-9}$$

- the axioms of the operad for $\{\mathcal{G}(n)\}_{1 \le n \le q}$ are satisfied in the cases where all the arities are $\le q$.

For every operad $\mathbb{O}$ and every integer $q \ge 1$, the disjoint union $\mathbb{O}^{\le q} := \bigsqcup_{n=0}^{q} \mathbb{O}(n)$ is clearly a q–truncated operad. In this paper, we only consider 4–truncated operads. So we will simply call them *truncated operads*.

The operad PaB of parenthesized braids, its truncation $\mathsf{PaB}^{\le 4}$ and its completion $\widehat{\mathsf{PaB}}^{\le 4}$ play the central role in this paper. See Appendix A for more details.

Acknowledgements We are thankful to Benjamin Collas, David Harbater, Julia Hartmann, Florian Pop, Leila Schneps, Dmitry Vaintrob, John Voight and Jingfeng Xia for useful discussions. Dolgushev is thankful to Pavol Severa for showing him the works of Pierre Guillot. Dolgushev discussed vague ideas of the construction presented in this paper during a walk near Zürich with Thomas Willwacher in October of 2016, and is thankful to Thomas for leaving him with the question "What are GT–shadows?" and for giving the title to this paper! Dolgushev is especially thankful to Leila Schneps for her unbounded enthusiasm about this project and its possible continuations, and is also thankful to Leila Schneps for her suggestion to modify the original definition of charming GT–shadows, her comments about the introduction, and her input concerning hypothetical versions of Furusho's property. Lorenz acknowledges both the Temple University Honors Program and the Undergraduate Research Program for their active support of undergraduate researchers. Dolgushev and Le acknowledge partial support from NSF grant DMS-1501001.

[5] In the literature, elementary insertions are sometimes called *partial compositions.*

2 GT–pairs and GT–shadows

2.1 The poset of compatible equivalence relations on $\mathsf{PaB}^{\le 4}$

An equivalence relation $\sim$ on the disjoint union of groupoids[6]

$$\mathsf{PaB}^{\le 4} = \mathsf{PaB}(1) \sqcup \mathsf{PaB}(2) \sqcup \mathsf{PaB}(3) \sqcup \mathsf{PaB}(4)$$

is an equivalence relation on the set of morphisms of $\mathsf{PaB}^{\le 4}$ such that, if $\gamma \sim \tilde{\gamma}$, then the source (resp. the target) of γ coincides with the source (resp. the target) of $\tilde{\gamma}$. In particular, $\gamma \sim \tilde{\gamma}$ implies that γ and $\tilde{\gamma}$ have the same arity.

Definition 2.1 An equivalence relation $\sim$ on $\mathsf{PaB}^{\le 4}$ is called *compatible* if

- for every pair of composable morphisms $\gamma, \tilde{\gamma} \in \mathsf{PaB}(n)$, the equivalence class of the composition $\gamma \cdot \tilde{\gamma}$ depends only on the equivalence classes of γ and $\tilde{\gamma}$;
- for every $\gamma, \tilde{\gamma} \in \mathsf{PaB}(n)$ and every $\theta \in S_n$,
$$\gamma \sim \tilde{\gamma} \iff \theta(\gamma) \sim \theta(\tilde{\gamma});$$
- for every tuple of integers i, n, m, $1 \le i \le n$, $n, m, n+m-1 \le 4$, and every $\gamma_1, \tilde{\gamma}_1 \in \mathsf{PaB}(n)$, $\gamma_2, \tilde{\gamma}_2 \in \mathsf{PaB}(m)$, we have
$$\gamma_1 \sim \tilde{\gamma}_1 \implies \gamma_1 \circ_i \gamma_2 \sim \tilde{\gamma}_1 \circ_i \gamma_2 \qquad \text{and} \qquad \gamma_2 \sim \tilde{\gamma}_2 \implies \gamma_1 \circ_i \gamma_2 \sim \gamma_1 \circ_i \tilde{\gamma}_2.$$

It is clear that, for every compatible equivalence relation $\sim$ on $\mathsf{PaB}^{\le 4}$, the set

$$\mathsf{PaB}^{\le 4}/\sim \tag{2-1}$$

of equivalence classes of morphisms in $\mathsf{PaB}^{\le 4}$ is a truncated operad in the category of groupoids. The set of objects of (2-1) coincides with the set of objects of $\mathsf{PaB}^{\le 4}$. The action of symmetric groups and the elementary insertions are defined by the formulas

$$\theta([\gamma]) := [\theta(\gamma)] \qquad \text{for } \theta \in S_n \text{ and } \gamma \in \mathsf{PaB}(n),$$
$$[\gamma_1] \circ_i [\gamma_2] := [\gamma_1 \circ_i \gamma_2] \quad \text{for } \gamma_1 \in \mathsf{PaB}(n) \text{ and } \gamma_2 \in \mathsf{PaB}(m).$$

The conditions of Definition 2.1 guarantee that the composition of morphisms, the action of the symmetric groups on $\mathsf{PaB}(n)/\sim$ and the elementary operadic insertions are well defined. The axioms of the (truncated) operad follow directly from their counterparts for $\mathsf{PaB}^{\le 4}$.

Compatible equivalence relations on $\mathsf{PaB}^{\le 4}$ form a poset with the following obvious partial order: we say that $\sim_1 \le \sim_2$ if $\sim_1$ is finer than $\sim_2$, ie

$$\gamma \sim_1 \tilde{\gamma} \implies \gamma \sim_2 \tilde{\gamma}.$$

[6] Recall that $\mathsf{PaB}(0)$ is the empty groupoid.

It is clear that, for every pair of compatible equivalence relations $\sim_1, \sim_2$ on $\mathsf{PaB}^{\le 4}$ such that $\sim_1 \le \sim_2$, we have a natural onto morphism of truncated operads

$$\mathcal{P}_{\sim_1,\sim_2} \colon \mathsf{PaB}^{\le 4}/\sim_1 \to \mathsf{PaB}^{\le 4}/\sim_2. \tag{2-2}$$

Moreover, the assignment $\sim \mapsto \mathsf{PaB}^{\le 4}/\sim$ upgrades to a functor from the poset of compatible equivalence relations to the category of truncated operads.

For every compatible equivalence relation $\sim$ on $\mathsf{PaB}^{\le 4}$, we denote by $\mathcal{P}_\sim$ the natural (onto) morphism of truncated operads:

$$\mathcal{P}_\sim \colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim. \tag{2-3}$$

2.2 From $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the poset of compatible equivalence relations

In this paper, we mostly consider compatible equivalence relations for which the set of morphisms of (2-1) is finite and a large supply of such compatible equivalence relations come from elements of the poset $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

For $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, we set

$$\mathsf{N}_{\mathrm{PB}_3} := \varphi_{123}^{-1}(\mathsf{N}) \cap \varphi_{12,3,4}^{-1}(\mathsf{N}) \cap \varphi_{1,23,4}^{-1}(\mathsf{N}) \cap \varphi_{1,2,34}^{-1}(\mathsf{N}) \cap \varphi_{234}^{-1}(\mathsf{N}), \tag{2-4}$$

$$\mathsf{N}_{\mathrm{PB}_2} := \varphi_{12}^{-1}(\mathsf{N}_{\mathrm{PB}_3}) \cap \varphi_{12,3}^{-1}(\mathsf{N}_{\mathrm{PB}_3}) \cap \varphi_{1,23}^{-1}(\mathsf{N}_{\mathrm{PB}_3}) \cap \varphi_{23}^{-1}(\mathsf{N}_{\mathrm{PB}_3}), \tag{2-5}$$

where φ_{123}, $\varphi_{12,3,4}$, $\varphi_{1,23,4}$, $\varphi_{1,2,34}$ and φ_{234} are the homomorphisms defined in (A-16), and φ_{12}, $\varphi_{12,3}$, $\varphi_{1,23}$ and φ_{23} are the homomorphisms defined in (A-17); see also the explicit formulas in (A-18) and (A-19).

Proposition 2.2 *For every* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, *the subgroup* $\mathsf{N}_{\mathrm{PB}_3}$ (*resp.* $\mathsf{N}_{\mathrm{PB}_2}$) *is an element of the poset* $\mathsf{NFI}_{\mathrm{PB}_3}(\mathrm{B}_3)$ (*resp. the poset* $\mathsf{NFI}_{\mathrm{PB}_2}(\mathrm{B}_2)$).

Proof Since every subgroup of B_2 is normal and $\mathsf{N}_{\mathrm{PB}_2}$ has a finite index in PB_2, the statement about $\mathsf{N}_{\mathrm{PB}_2}$ is obvious.

It is also easy to see that $\mathsf{N}_{\mathrm{PB}_3}$ is a subgroup of finite index in PB_3. So it suffices to prove that

$$g \mathsf{N}_{\mathrm{PB}_3} g^{-1} \le \mathsf{N}_{\mathrm{PB}_3} \quad \text{for all } g \in B_3.$$

Let $h \in \mathsf{N}_{\mathrm{PB}_3}$ and $g \in \mathrm{B}_3$. Then

$$\varphi_{1,23,4}(g \cdot h \cdot g^{-1}) = \mathfrak{ou}(\mathfrak{m}(g \cdot h \cdot g^{-1}) \circ_2 \mathrm{id}_{12}), \tag{2-6}$$

where the map $\mathfrak{ou} \colon \mathsf{PaB}(n) \to \mathrm{B}_n$ and its right inverse $\mathfrak{m} \colon \mathrm{B}_n \to \mathsf{PaB}(n)$ are defined in Section A.2.

Using identity (A-11), we get

$$\mathfrak{m}(g \cdot h \cdot g^{-1}) = \theta(\mathfrak{m}(g)) \cdot \theta(\chi) \cdot \mathfrak{m}(g^{-1}),$$

where $\theta = \rho(g)$ and $\chi := \mathfrak{m}(h)$.

Therefore

$$\mathfrak{m}(g\cdot h\cdot g^{-1})\circ_2 \mathrm{id}_{12} = \big(\theta(\mathfrak{m}(g))\circ_2 \mathrm{id}_{12}\big)\cdot(\theta(\chi)\circ_2 \mathrm{id}_{12})\cdot(\mathfrak{m}(g^{-1})\circ_2 \mathrm{id}_{12}). \tag{2-7}$$

Combining (2-6) with (2-7), we get

$$\varphi_{1,23,4}(g\cdot h\cdot g^{-1}) = \mathfrak{ou}\big(\theta(\mathfrak{m}(g))\circ_2 \mathrm{id}_{12}\big)\cdot\mathfrak{ou}(\theta(\chi)\circ_2 \mathrm{id}_{12})\cdot\mathfrak{ou}(\mathfrak{m}(g^{-1})\circ_2 \mathrm{id}_{12}). \tag{2-8}$$

Since

$$\begin{aligned}\mathfrak{ou}\big(\theta(\mathfrak{m}(g))\circ_2 \mathrm{id}_{12}\big)\cdot\mathfrak{ou}(\mathfrak{m}(g^{-1})\circ_2 \mathrm{id}_{12}) &= \mathfrak{ou}\big(\theta(\mathfrak{m}(g))\circ_2 \mathrm{id}_{12}\cdot\mathfrak{m}(g^{-1})\circ_2 \mathrm{id}_{12}\big)\\ &= \mathfrak{ou}\big(\big(\theta(\mathfrak{m}(g))\cdot\mathfrak{m}(g^{-1})\big)\circ_2 \mathrm{id}_{12}\big)\\ &= \mathfrak{ou}\big(\mathfrak{m}(g\cdot g^{-1})\circ_2 \mathrm{id}_{12}\big) = \mathfrak{ou}(\mathrm{id}_{(1(23))4}) = 1_{\mathrm{B}_4},\end{aligned}$$

the element $\varphi_{1,23,4}(g\cdot h\cdot g^{-1})\in \mathrm{B}_4$ can be rewritten as

$$\varphi_{1,23,4}(g\cdot h\cdot g^{-1}) = \widetilde{g}\cdot\mathfrak{ou}(\theta(\chi)\circ_2 \mathrm{id}_{12})\cdot\widetilde{g}^{-1}, \quad \text{where } \widetilde{g} := \mathfrak{ou}\big(\theta(\mathfrak{m}(g))\circ_2 \mathrm{id}_{12}\big).$$

Thus it remains to prove that

$$\mathfrak{ou}(\theta(\chi)\circ_2 \mathrm{id}_{12})\in \mathsf{N}. \tag{2-9}$$

For this purpose, we consider the three possible cases: $\theta(1)=2$, $\theta(2)=2$ and $\theta(3)=2$.

- If $\theta(1)=2$, then $\mathfrak{ou}(\theta(\chi)\circ_2 \mathrm{id}_{12}) = \varphi_{12,3,4}(h)$ and (2-9) is a consequence of $h\in\varphi^{-1}_{12,3,4}(\mathsf{N})$.
- If $\theta(2)=2$, then $\mathfrak{ou}(\theta(\chi)\circ_2 \mathrm{id}_{12}) = \varphi_{1,23,4}(h)$ and (2-9) is a consequence of $h\in\varphi^{-1}_{1,23,4}(\mathsf{N})$.
- If $\theta(3)=2$, then $\mathfrak{ou}(\theta(\chi)\circ_2 \mathrm{id}_{12}) = \varphi_{1,2,34}(h)$ and (2-9) is a consequence of $h\in\varphi^{-1}_{1,2,34}(\mathsf{N})$.

We proved that the element ghg^{-1} belongs to $\varphi^{-1}_{1,23,4}(\mathsf{N})\subset \mathrm{PB}_3$. The proofs for the remaining four homomorphisms φ_{123}, $\varphi_{12,3,4}$, $\varphi_{1,2,34}$ and φ_{234} are similar and we omit them. □

It is clear that $\mathsf{N}_{\mathrm{PB}_2} = \langle x_{12}^{N_{\mathrm{ord}}}\rangle$, where N_{ord} is the index of $\mathsf{N}_{\mathrm{PB}_2}$ in PB_2, ie N_{ord} is the least common multiple of orders of $x_{12}\mathsf{N}_{\mathrm{PB}_3}$, $x_{23}\mathsf{N}_{\mathrm{PB}_3}$, $x_{12}x_{13}\mathsf{N}_{\mathrm{PB}_3}$ and $x_{13}x_{23}\mathsf{N}_{\mathrm{PB}_3}$ in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$.

Using the identities $x_{12}x_{13} = x_{23}^{-1}c$ and $x_{13}x_{23} = x_{12}^{-1}c$ involving the generator c (see (A-5)) of the center of PB_3, it is easy to prove the following statement:

Proposition 2.3 *Let $\mathsf{N}_{\mathrm{PB}_2} = \langle x_{12}^{N_{\mathrm{ord}}}\rangle$ be the subgroup of PB_2 defined in (2-5). Then N_{ord} coincides with*

(1) *the least common multiple of orders of elements $x_{12}\mathsf{N}_{\mathrm{PB}_3}$, $x_{23}\mathsf{N}_{\mathrm{PB}_3}$ and $x_{12}x_{13}\mathsf{N}_{\mathrm{PB}_3}$;*

(2) *the least common multiple of orders of elements $x_{12}\mathsf{N}_{\mathrm{PB}_3}$, $x_{23}\mathsf{N}_{\mathrm{PB}_3}$ and $x_{13}x_{23}\mathsf{N}_{\mathrm{PB}_3}$; and*

(3) *the least common multiple of orders of elements $x_{12}\mathsf{N}_{\mathrm{PB}_3}$, $x_{23}\mathsf{N}_{\mathrm{PB}_3}$ and $c\mathsf{N}_{\mathrm{PB}_3}$.* □

Given $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and the corresponding normal subgroups $\mathsf{N}_{\mathrm{PB}_3}$ and $\mathsf{N}_{\mathrm{PB}_2}$, we will now define an equivalence relation $\sim_{\mathsf{N}}$ on the set of morphisms in $\mathsf{PaB}^{\le 4}$.

The groupoid $\mathsf{PaB}(1)$ has exactly one object and exactly one (identity) morphism. So $\mathsf{PaB}(1)$ has only one equivalence relation.

Let $n \in \{2,3,4\}$ and

$$\gamma, \widetilde{\gamma} \in \mathrm{Hom}_{\mathsf{PaB}(n)}(\tau_1, \tau_2).$$

We declare that $\gamma \sim_{\mathsf{N}} \widetilde{\gamma}$ if and only if

$$\mathfrak{ou}(\gamma^{-1} \cdot \widetilde{\gamma}) \in \mathsf{N}_{\mathrm{PB}_n}, \tag{2-10}$$

where $\mathsf{N}_{\mathrm{PB}_4} := \mathsf{N}$. In other words, $\gamma \sim_{\mathsf{N}} \widetilde{\gamma}$ if and only if

$$\widetilde{\gamma} = \gamma \cdot \eta,$$

where $\mathfrak{ou}(\eta) \in \mathsf{N}_{\mathrm{PB}_n}$ and the source of γ coincides with the target of η.

Proposition 2.4 *For every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $\sim_{\mathsf{N}}$ is a compatible equivalence relation on $\mathsf{PaB}^{\le 4}$ in the sense of Definition 2.1. Moreover, the assignment*

$$\mathsf{N} \mapsto \sim_{\mathsf{N}}$$

upgrades to a functor from the poset $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the poset of compatible equivalence relations on $\mathsf{PaB}^{\le 4}$.

Proof The first property of $\sim_{\mathsf{N}}$ follows from the fact that $\mathsf{N}_{\mathrm{PB}_4} := \mathsf{N}$ (resp. $\mathsf{N}_{\mathrm{PB}_3}$, $\mathsf{N}_{\mathrm{PB}_2}$) is normal in B_4 (resp. B_3, B_2).

The second property of $\sim_{\mathsf{N}}$ follows from the obvious identity

$$\mathfrak{ou}(\gamma) = \mathfrak{ou}(\theta(\gamma)) \quad \text{for all } \gamma \in \mathsf{PaB}(n) \text{ and } \theta \in S_n.$$

The proof of the last property is based on the observation that elementary operadic insertions for PaB can be expressed in terms of the operations $? \circ_i \mathrm{id}_\tau$, $\mathrm{id}_\tau \circ_i ?$, and the composition of morphisms in the groupoids $\mathsf{PaB}(3)$ and $\mathsf{PaB}(4)$.

Let $n \in \{2,3\}$ and η be a morphism in $\mathsf{PaB}(n)$ whose target coincides with its source. In particular, $\mathfrak{ou}(\eta) \in \mathrm{PB}_n$. Let us prove that, if $\mathfrak{ou}(\eta) \in \mathsf{N}_{\mathrm{PB}_n}$, then, for every $\tau \in \mathrm{Ob}(\mathsf{PaB}(m))$ with $n+m-1 \le 4$, we have

$$\mathfrak{ou}(\eta \circ_i \mathrm{id}_\tau) \in \mathrm{PB}_{n+m-1} \quad \text{for all } 1 \le i \le n, \tag{2-11}$$

$$\mathfrak{ou}(\mathrm{id}_\tau \circ_i \eta) \in \mathrm{PB}_{n+m-1} \quad \text{for all } 1 \le i \le m. \tag{2-12}$$

Let $h = \mathfrak{ou}(\eta)$. If $m = 2$, then there exists $1 \le j \le n$ (resp. $j \in \{1,2\}$) such that $\mathfrak{ou}(\mathfrak{m}(h) \circ_j \mathrm{id}_{12}) = \mathfrak{ou}(\eta \circ_i \mathrm{id}_\tau)$ (resp. $\mathfrak{ou}(\mathrm{id}_{12}) \circ_j \mathfrak{m}(h) = \mathfrak{ou}(\mathrm{id}_\tau \circ_i \eta)$).

Thus, if $m = 2$, (2-11) and (2-12) follow directly from the definitions of $\mathsf{N}_{\mathrm{PB}_3}$, $\mathsf{N}_{\mathrm{PB}_2}$, (2-4), (2-5) and the definitions of the homomorphisms $\varphi_{123}, \dots, \varphi_{12}, \dots$; see (A-16) and (A-17).

If $m = 3$, then there exist $j, k \in \{1, 2\}$ such that

$$\mathfrak{ou}(\eta \circ_i \mathrm{id}_\tau) = \mathfrak{ou}((\eta \circ_j \mathrm{id}_{12}) \circ_k \mathrm{id}_{12}).$$

For example, if $\eta \in \mathrm{Hom}_{\mathsf{PaB}}((2,1),(2,1))$, then $\mathfrak{ou}(\eta \circ_2 \mathrm{id}_{2(1,3)}) = \mathfrak{ou}((\eta \circ_2 \mathrm{id}_{12}) \circ_3 \mathrm{id}_{12})$.

Thus (2-11) for $m = 3$ follows from (2-11) for $m = 2$. Similarly, (2-12) for $m = 3$ follows from (2-12) for $m = 2$.

We will now use (2-11) and (2-12) to prove the last property of $\sim_{\mathsf{N}}$.

Consider $\gamma_1, \widetilde{\gamma}_1 \in \mathrm{Hom}_{\mathsf{PaB}(n)}(\tau_1, \tau_2)$ and $\gamma_2, \widetilde{\gamma}_2 \in \mathrm{Hom}_{\mathsf{PaB}(m)}(\omega_1, \omega_2)$. First suppose $\gamma_1 \sim_{\mathsf{N}} \widetilde{\gamma}_1$, so $\widetilde{\gamma}_1 = \gamma_1 \cdot \eta$ for some $\eta \in \mathrm{Hom}_{\mathsf{PaB}(n)}(\tau_1, \tau_1)$ such that $\mathfrak{ou}(\eta) \in \mathsf{N}_{\mathrm{PB}_n}$. It follows that

$$\widetilde{\gamma}_1 \circ_i \gamma_2 = (\gamma_1 \cdot \eta) \circ_i (\gamma_2 \cdot \mathrm{id}_{\omega_1}) = (\gamma_1 \circ_i \gamma_2) \cdot (\eta \circ_i \mathrm{id}_{\omega_1}).$$

Due to (2-11), $\mathfrak{ou}(\eta \circ_i \mathrm{id}_{\omega_1}) \in \mathsf{N}_{\mathrm{PB}_{n+m-1}}$ and hence $\widetilde{\gamma}_1 \circ_i \gamma_2 \sim \gamma_1 \circ_i \gamma_2$.

Now suppose $\gamma_2 \sim_{\mathsf{N}} \widetilde{\gamma}_2$, so $\widetilde{\gamma}_2 = \gamma_2 \cdot \eta'$ for some $\eta' \in \mathrm{Hom}_{\mathsf{PaB}(m)}(\omega_1, \omega_1)$ such that $\mathfrak{ou}(\eta') \in \mathsf{N}_{\mathrm{PB}_m}$. It follows that

$$\gamma_1 \circ_i \widetilde{\gamma}_2 = (\gamma_1 \cdot \mathrm{id}_{\tau_1}) \circ_i (\gamma_2 \cdot \eta') = (\gamma_1 \circ_i \gamma_2) \cdot (\mathrm{id}_{\tau_1} \circ_i \eta').$$

Due to (2-12), $\mathfrak{ou}(\mathrm{id}_{\tau_1} \circ_i \eta') \in \mathsf{N}_{\mathrm{PB}_{n+m-1}}$ and hence $\gamma_1 \circ_i \widetilde{\gamma}_2 \sim \gamma_1 \circ_i \gamma_2$.

This completes the proof of the fact that $\sim_{\mathsf{N}}$ is indeed a compatible equivalence relation on $\mathsf{PaB}^{\leq 4}$.

It is clear that, if $\widetilde{\mathsf{N}}, \mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and $\widetilde{\mathsf{N}} \leq \mathsf{N}$, then $\widetilde{\mathsf{N}}_{\mathrm{PB}_3} \leq \mathsf{N}_{\mathrm{PB}_3}$ and $\widetilde{\mathsf{N}}_{\mathrm{PB}_2} \leq \mathsf{N}_{\mathrm{PB}_2}$.

Thus the assignment $\mathsf{N} \mapsto \sim_{\mathsf{N}}$ upgrades to a functor from the poset $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the poset of compatible equivalence relations. □

Later, we will need the following technical statement about $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$:

Proposition 2.5 (A) *For every* $\mathsf{N} \in \mathsf{NFI}(\mathrm{PB}_3)$, *there exists* $\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *satisfying the property*

$$\mathsf{K}_{\mathrm{PB}_3} \leq \mathsf{N}.$$

(B) *For every* $\mathsf{N} \in \mathsf{NFI}(\mathrm{PB}_2)$, *there exists* $\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *such that* $\mathsf{K}_{\mathrm{PB}_2} \leq \mathsf{N}$.

Proof Stronger versions of these statements are proved in Section 3.1; see Proposition 3.9. So we omit the proof of this proposition. □

2.3 The set of GT–pairs $\mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$

Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and $\sim_{\mathsf{N}}$ be the corresponding compatible equivalence relation on $\mathsf{PaB}^{\leq 4}$. Let $\mathsf{N}_{\mathrm{PB}_3}$ (resp. $\mathsf{N}_{\mathrm{PB}_2}$) be the corresponding normal subgroup of PB_3 (resp. PB_2) and N_{ord} be the index of $\mathsf{N}_{\mathrm{PB}_2}$ in PB_2.

Since the groupoid $\mathsf{PaB}(0)$ is empty, Theorem A.1 implies that the truncated operad $\mathsf{PaB}^{\leq 4}$ is generated by morphisms α and β shown in equation (1-2).

Moreover any relation on α and β in $\mathsf{PaB}^{\leq 4}$ is a consequence of the pentagon relation

(2-13)
$$\begin{array}{ccccc} & (1(23))4 & \xrightarrow{\alpha \circ_2 \mathrm{id}_{12}} & 1((23)4) & \\ \nearrow^{\mathrm{id}_{12} \circ_1 \alpha} & & & & \searrow^{\mathrm{id}_{12} \circ_2 \alpha} \\ ((12)3)4 & & & & 1(2(34)) \\ & \searrow_{\alpha \circ_1 \mathrm{id}_{12}} & (12)(34) & \nearrow_{\alpha \circ_3 \mathrm{id}_{12}} & \end{array}$$

and the hexagon relations

(2-14)
$$\begin{array}{ccccc} (12)3 & \xrightarrow{\beta \circ_1 \mathrm{id}_{12}} & 3(12) & \xleftarrow{(1,3,2)\,\alpha} & (31)2 \\ \alpha \big\downarrow & & & & \big\uparrow (2,3)\,(\mathrm{id}_{12} \circ_1 \beta) \\ 1(23) & \xrightarrow{\mathrm{id}_{12} \circ_2 \beta} & 1(32) & \xrightarrow{(2,3)\,\alpha^{-1}} & (13)2 \end{array}$$

(2-15)
$$\begin{array}{ccccc} 1(23) & \xrightarrow{\beta \circ_2 \mathrm{id}_{12}} & (23)1 & \xrightarrow{(1,2,3)\,\alpha} & 2(31) \\ \alpha^{-1} \big\downarrow & & & & \big\uparrow (1,2)\,(\mathrm{id}_{12} \circ_2 \beta) \\ (12)3 & \xrightarrow{\mathrm{id}_{12} \circ_1 \beta} & (21)3 & \xrightarrow{(1,2)\,\alpha} & 2(13) \end{array}$$

Thus morphisms of truncated operads

(2-16)
$$T : \mathsf{PaB}^{\leq 4} \to \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}}$$

are in bijection with pairs

(2-17)
$$(m + N_{\mathrm{ord}}\mathbb{Z},\ f\mathsf{N}_{\mathrm{PB}_3}) \in \mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z} \times \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$$

satisfying in $\mathrm{B}_3/\mathsf{N}_{\mathrm{PB}_3}$ the relations

(2-18)
$$\sigma_1 x_{12}^m f^{-1} \sigma_2 x_{23}^m f \mathsf{N}_{\mathrm{PB}_3} = f^{-1} \sigma_1 \sigma_2 (x_{13} x_{23})^m \mathsf{N}_{\mathrm{PB}_3},$$

(2-19)
$$f^{-1} \sigma_2 x_{23}^m f \sigma_1 x_{12}^m \mathsf{N}_{\mathrm{PB}_3} = \sigma_2 \sigma_1 (x_{12} x_{13})^m f \mathsf{N}_{\mathrm{PB}_3},$$

and in PB_4/N the relation

(2-20)
$$\varphi_{234}(f) \varphi_{1,23,4}(f) \varphi_{123}(f) \mathsf{N} = \varphi_{1,2,34}(f) \varphi_{12,3,4}(f) \mathsf{N}.$$

More precisely, this bijection sends a pair (2-17) to the morphism $T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$ of truncated operads defined by the formulas

$$T_{m,f}(\alpha) := [\alpha \cdot \mathfrak{m}(f)] \quad \text{and} \quad T_{m,f}(\beta) := [\beta \cdot \mathfrak{m}(x_{12}^m)],$$

where $\mathfrak{m}$ is the map from B_n to $\mathsf{PaB}(n)$ defined in Section A.2.

This observation motivates our definition of a GT–pair:

Definition 2.6 For $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the set $\mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$ consists of pairs

$$(m + N_{\mathrm{ord}}\mathbb{Z},\, f\mathsf{N}_{\mathrm{PB}_3}) \in \mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z} \times \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$$

satisfying (2-18), (2-19) and (2-20). Elements of $\mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$ are called GT–*pairs*.

We will represent GT–pairs by tuples $(m, f) \in \mathbb{Z} \times \mathrm{PB}_3$. It is straightforward to see that, if relations (2-18), (2-19) and (2-20) are satisfied for a tuple (m, f), then they are also satisfied for $(m + qN_{\mathrm{ord}}, fh)$, where q is an arbitrary integer and h is an arbitrary element in $\mathsf{N}_{\mathrm{PB}_3}$. A GT–pair in $\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z} \times \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ represented by a tuple $(m, f) \in \mathbb{Z} \times \mathrm{PB}_3$ will be often denoted by

$$[(m, f)].$$

For $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and a tuple (m, f) representing a GT–pair in $\mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$, we denote by $T_{m,f}$ the corresponding morphism of truncated operads,

$$T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}. \tag{2-21}$$

It is clear that, for every $n \in \{2, 3, 4\}$, the assignment $\mathfrak{ou}$ from Section A.2 induces the obvious map

$$\mathsf{PaB}(n)/\sim_{\mathsf{N}} \to \mathrm{B}_n/\mathsf{N}_{\mathrm{PB}_n}, \quad \text{where } \mathsf{N}_{\mathrm{PB}_4} := \mathsf{N}. \tag{2-22}$$

By abuse of notation, we will use the same symbol $\mathfrak{ou}$ for the map (2-22).

Using (2-22) together with the map $\mathfrak{m}\colon \mathrm{B}_n \to \mathsf{PaB}(n)$ from Section A.2 and morphism (2-21), we define group homomorphisms $\mathrm{B}_2 \to \mathrm{B}_2/\mathsf{N}_{\mathrm{PB}_2}$, $\mathrm{B}_3 \to \mathrm{B}_3/\mathsf{N}_{\mathrm{PB}_3}$ and $\mathrm{B}_4 \to \mathrm{B}_4/\mathsf{N}$. Restricting these homomorphisms to PB_2, PB_3 and PB_4, we get group homomorphisms $\mathrm{PB}_2 \to \mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}$, $\mathrm{PB}_3 \to \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ and $\mathrm{PB}_4 \to \mathrm{PB}_4/\mathsf{N}$, respectively. More precisely, we have the following statement:

Corollary 2.7 *Suppose that* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. *For every pair* $(m + N_{\mathrm{ord}}\mathbb{Z}, f\mathsf{N}_{\mathrm{PB}_3}) \in \mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$, *and every* $n \in \{2, 3, 4\}$, *the assignment*

$$T^{\mathrm{B}_n}_{m,f}(g) := \mathfrak{ou} \circ T_{m,f} \circ \mathfrak{m}(g) \tag{2-23}$$

defines a group homomorphism $\mathrm{B}_n \to \mathrm{B}_n/\mathsf{N}_{\mathrm{PB}_n}$ (*with* $\mathsf{N}_{\mathrm{PB}_4} := \mathsf{N}$). *The restriction of* $T^{\mathrm{B}_n}_{m,f}$ *to* PB_n *gives us a group homomorphism*

$$T^{\mathrm{PB}_n}_{m,f}\colon \mathrm{PB}_n \to \mathrm{PB}_n/\mathsf{N}_{\mathrm{PB}_n}. \tag{2-24}$$

The action of $T^{\mathrm{B}_4}_{m,f}$ *on the generators of* B_4 *is given by the formulas*

$$
\begin{aligned}
T^{\mathrm{B}_4}_{m,f}(\sigma_1) &:= \sigma_1 x_{12}^m \mathsf{N},\\
T^{\mathrm{B}_4}_{m,f}(\sigma_2) &:= \varphi_{123}(f)^{-1}(\sigma_2 x_{23}^m)\varphi_{123}(f)\mathsf{N},\\
T^{\mathrm{B}_4}_{m,f}(\sigma_3) &:= \varphi_{12,3,4}(f)^{-1}(\sigma_3 x_{34}^m)\varphi_{12,3,4}(f)\mathsf{N}.
\end{aligned}
\tag{2-25}
$$

The action of $T^{\mathrm{B}_3}_{m,f}$ *on the generators of* B_3 *are given by the formulas*

$$
T^{\mathrm{B}_3}_{m,f}(\sigma_1) := \sigma_1 x_{12}^m \mathsf{N}_{\mathrm{PB}_3} \quad \text{and} \quad T^{\mathrm{B}_3}_{m,f}(\sigma_2) := f^{-1}(\sigma_2 x_{23}^m) f \mathsf{N}_{\mathrm{PB}_3}.
\tag{2-26}
$$

Finally, $T^{\mathrm{B}_2}_{m,f}$ *sends* σ_1 *to* $\sigma_1 x_{12}^m \mathsf{N}_{\mathrm{PB}_2}$.

Proof It is clear that, for every two composable morphisms $\gamma_1, \gamma_2 \in \mathsf{PaB}(n)/\sim_{\mathsf{N}}$, we have

$$
\mathfrak{ou}(\gamma_1 \cdot \gamma_2) = \mathfrak{ou}(\gamma_1) \cdot \mathfrak{ou}(\gamma_2).
\tag{2-27}
$$

Then using (A-11) from Section A.2 and the compatibility of $T_{m,f}$ with the structure of the truncated operad we get

$$
\begin{aligned}
T^{\mathrm{B}_n}_{m,f}(g_1 \cdot g_2) &= \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(g_1 \cdot g_2))\big) = \mathfrak{ou}\big(T_{m,f}\big(\rho(g_2)^{-1}(\mathfrak{m}(g_1)) \cdot \mathfrak{m}(g_2)\big)\big)\\
&= \mathfrak{ou}\big(T_{m,f}\big(\rho(g_2)^{-1}(\mathfrak{m}(g_1))\big) \cdot T_{m,f}(\mathfrak{m}(g_2))\big) = \mathfrak{ou}\big(\rho(g_2)^{-1} T_{m,f}(\mathfrak{m}(g_1)) \cdot T_{m,f}(\mathfrak{m}(g_2))\big)\\
&= \mathfrak{ou}\big(\rho(g_2)^{-1} T_{m,f}(\mathfrak{m}(g_1))\big) \cdot \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(g_2))\big) = \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(g_1))\big) \cdot \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(g_2))\big)\\
&= T^{\mathrm{B}_n}_{m,f}(g_1) \cdot T^{\mathrm{B}_n}_{m,f}(g_2),
\end{aligned}
$$

whence $T^{\mathrm{B}_n}_{m,f}$ is a group homomorphism.

The second statement of the corollary follows immediately from the fact $\mathfrak{m}$ is a right inverse of $\mathfrak{ou}$ and $T_{m,f}$ acts trivially on objects of PaB.

We will now prove (2-25). The easier cases of $T^{\mathrm{B}_3}_{m,f}$ and $T^{\mathrm{B}_2}_{m,f}$ are left for the reader.

For the generator σ_1, we have

$$
T^{\mathrm{B}_4}_{m,f}(\sigma_1) = \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(\sigma_1))\big) = \mathfrak{ou}(T_{m,f}(\mathrm{id}_{(12)3} \circ_1 \beta)) = \mathfrak{ou}\big(\mathrm{id}_{(12)3} \circ_1 [\beta \cdot \mathfrak{m}(x_{12}^m)]\big) = \sigma_1 x_{12}^m \mathsf{N}.
$$

For the generator σ_2, we have

$$
\begin{aligned}
T^{\mathrm{B}_4}_{m,f}(\sigma_2) &= \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(\sigma_2))\big) = \mathfrak{ou}\big(T_{m,f}((2,3)(\mathrm{id}_{12} \circ_1 \alpha^{-1}) \cdot (\mathrm{id}_{(12)3} \circ_2 \beta) \cdot (\mathrm{id}_{12} \circ_1 \alpha))\big)\\
&= \mathfrak{ou}\big((2,3)\big(\mathrm{id}_{12} \circ_1 [\mathfrak{m}(f^{-1}) \cdot \alpha^{-1}]\big) \cdot \big(\mathrm{id}_{(12)3} \circ_2 [\beta \cdot \mathfrak{m}(x_{12}^m)]\big) \cdot \big(\mathrm{id}_{12} \circ_1 [\alpha \cdot \mathfrak{m}(f)]\big)\big)\\
&= \varphi_{123}(f)^{-1}(\sigma_2 x_{23}^m)\varphi_{123}(f)\mathsf{N}.
\end{aligned}
$$

Finally, for the generator σ_3, we have

$$
\begin{aligned}
T^{\mathrm{B}_4}_{m,f}(\sigma_3) &= \mathfrak{ou}\big(T_{m,f}(\mathfrak{m}(\sigma_3))\big) = \mathfrak{ou}\big(T_{m,f}((3,4)(\alpha^{-1} \circ_1 \mathrm{id}_{12}) \cdot (\mathrm{id}_{(12)3} \circ_3 \beta) \cdot (\alpha \circ_1 \mathrm{id}_{12}))\big)\\
&= \mathfrak{ou}\big((3,4)\big([\mathfrak{m}(f)^{-1} \cdot \alpha^{-1}] \circ_1 \mathrm{id}_{12}\big) \cdot \big(\mathrm{id}_{(12)3} \circ_3 [\beta \cdot \mathfrak{m}(x_{12}^m)]\big) \cdot \big([\alpha \cdot \mathfrak{m}(f)] \circ_1 \mathrm{id}_{12}\big)\big)\\
&= \varphi_{12,3,4}(f)^{-1}(\sigma_3 x_{34}^m)\varphi_{12,3,4}(f)\mathsf{N}.
\end{aligned}
$$

□

Let us now use Corollary 2.7 to prove the following statement.

Corollary 2.8 *Let* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $[(m, f)] \in \mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$ *and* c *be the generator of the center of* PB_3*; see* (A-5). *Then*

$$T^{\mathrm{PB}_3}_{m,f}(x_{12}) = x_{12}^{2m+1}\mathsf{N}_{\mathrm{PB}_3}, \qquad T^{\mathrm{PB}_3}_{m,f}(x_{23}) = f^{-1}x_{23}^{2m+1}f\mathsf{N}_{\mathrm{PB}_3}, \tag{2-28}$$

$$T^{\mathrm{PB}_3}_{m,f}(x_{13}) = x_{12}^{-m}\sigma_1^{-1}f^{-1}x_{23}^{2m+1}f\sigma_1x_{12}^{m}\mathsf{N}_{\mathrm{PB}_3}, \qquad T^{\mathrm{PB}_3}_{m,f}(c) = c^{2m+1}\mathsf{N}_{\mathrm{PB}_3}. \tag{2-29}$$

Proof The first equation in (2-28) is a simple consequence of the first equation in (2-26).

Using the second equation in (2-26), we get

$$T^{\mathrm{PB}_3}_{m,f}(x_{23}) = T^{\mathrm{PB}_3}_{m,f}(\sigma_2^2) = \big(f^{-1}(\sigma_2x_{23}^m)f\big)^2\,\mathsf{N}_{\mathrm{PB}_3} = f^{-1}(\sigma_2^2x_{23}^{2m})f\,\mathsf{N}_{\mathrm{PB}_3} = f^{-1}\,x_{23}^{2m+1}\,f\,\mathsf{N}_{\mathrm{PB}_3}.$$

Thus the second equation in (2-28) is proved.

Using the definition of $x_{13} := \sigma_1^{-1}\sigma_2^2\sigma_1 = \sigma_1^{-1}x_{23}\sigma_1$, the first equation in (2-26) and the second equation in (2-28), we get

$$T^{\mathrm{PB}_3}_{m,f}(x_{13}) = T^{\mathrm{PB}_3}_{m,f}(\sigma_1^{-1}x_{23}\sigma_1) = x_{12}^{-m}\sigma_1^{-1}\,f^{-1}x_{23}^{2m+1}f\,\sigma_1x_{12}^m\,\mathsf{N}_{\mathrm{PB}_3}.$$

Thus the first equation in (2-29) is also satisfied.

To prove the second equation in (2-29), we use the formulas (2-18), (2-19), (2-26), and the identities $x_{13}x_{23} = x_{12}^{-1}c$, $x_{12}x_{13} = x_{23}^{-1}c$ extensively.

$$\begin{aligned}
T^{\mathrm{PB}_3}_{m,f}(c) &= T^{\mathrm{PB}_3}_{m,f}((\sigma_1\sigma_2)^3) = (T^{\mathrm{B}_3}_{m,f}(\sigma_1\sigma_2))^3\\
&= \sigma_1x_{12}^mf^{-1}\sigma_2x_{23}^mf\sigma_1x_{12}^mf^{-1}\sigma_2x_{23}^mf\sigma_1x_{12}^mf^{-1}\sigma_2x_{23}^mf\mathsf{N}_{\mathrm{PB}_3}\\
&= f^{-1}\sigma_1\sigma_2(x_{13}x_{23})^m\sigma_1x_{12}^m\sigma_2\sigma_1(x_{12}x_{13})^mff^{-1}\sigma_2x_{23}^mf\mathsf{N}_{\mathrm{PB}_3}\\
&= f^{-1}\sigma_1\sigma_2(x_{12}^{-1}c)^m\sigma_1x_{12}^m\sigma_2\sigma_1(x_{23}^{-1}c)^m\sigma_2x_{23}^mf\mathsf{N}_{\mathrm{PB}_3}\\
&= f^{-1}\sigma_1\sigma_2x_{12}^{-m}c^m\sigma_1x_{12}^m\sigma_2\sigma_1x_{23}^{-m}c^m\sigma_2x_{23}^mf\mathsf{N}_{\mathrm{PB}_3}\\
&= c^{2m}f^{-1}(\sigma_1\sigma_2\sigma_1\sigma_2\sigma_1\sigma_2)f\mathsf{N}_{\mathrm{PB}_3} = c^{2m+1}\mathsf{N}_{\mathrm{PB}_3}.
\end{aligned}$$

□

2.4 GT–pairs coming from automorphisms of $\widehat{\mathsf{PaB}}$

Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and $\sim_{\mathsf{N}}$ be the corresponding compatible equivalence relation. Since the groupoids $\mathsf{PaB}(n)/{\sim_{\mathsf{N}}}$ for $1 \le n \le 4$ are finite, we have a canonical continuous onto morphism of truncated operads

$$\widehat{\mathcal{P}}_{\mathsf{N}} : \widehat{\mathsf{PaB}}^{\le 4} \to \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{N}}}. \tag{2-30}$$

Thus, given any continuous automorphism $\widehat{T} : \widehat{\mathsf{PaB}} \to \widehat{\mathsf{PaB}}$, we can produce the morphism of truncated operads

$$T_{\mathsf{N}} : \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{N}}}$$

by setting

$$T_{\mathsf{N}} := \widehat{\mathcal{P}}_{\mathsf{N}} \circ \widehat{T} \circ \mathcal{I}, \tag{2-31}$$

where $\mathcal{I}$ is the natural embedding of truncated operads

$$\mathcal{I}\colon \mathsf{PaB}^{\le 4} \to \widehat{\mathsf{PaB}}^{\le 4}. \tag{2-32}$$

In other words, for every continuous automorphism of $\widehat{\mathsf{PaB}}$ and every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, we get a GT–pair $[(m, f)]$ corresponding to T_{N}. In this situation, we say that the GT–pair $[(m, f)]$ *comes from* the automorphism $\widehat{T}$.

GT–pairs coming from automorphisms of $\widehat{\mathsf{PaB}}$ satisfy additional properties. Indeed, since $\mathcal{I}(\mathsf{PaB}^{\le 4})$ is dense in $\widehat{\mathsf{PaB}}^{\le 4}$ and the morphism

$$\mathcal{P}_{\mathsf{N}} \circ \widehat{T}\colon \widehat{\mathsf{PaB}}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$$

is continuous and onto, the morphism T_{N} is also onto.

Thus, if a GT–pair $[(m, f)]$ comes from a (continuous) automorphism of $\widehat{\mathsf{PaB}}$, the group homomorphisms

$$T^{\mathrm{PB}_4}_{m,f}\colon \mathrm{PB}_4 \to \mathrm{PB}_4/\mathsf{N}, \tag{2-33}$$

$$T^{\mathrm{PB}_3}_{m,f}\colon \mathrm{PB}_3 \to \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}, \tag{2-34}$$

$$T^{\mathrm{PB}_2}_{m,f}\colon \mathrm{PB}_2 \to \mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}, \tag{2-35}$$

are onto.

GT–pairs satisfying these properties are called GT–*shadows*. More precisely:

Definition 2.9 Let N be a finite-index normal subgroup of B_4 such that $\mathsf{N} \le \mathrm{PB}_4$. Furthermore, let $\mathsf{N}_{\mathrm{PB}_3}$ and $\mathsf{N}_{\mathrm{PB}_2}$ be the corresponding normal subgroups of B_3 and B_2, respectively, and let N_{ord} be the index of $\mathsf{N}_{\mathrm{PB}_2}$ in PB_2. The set $\mathsf{GT}(\mathsf{N})$ consists of GT–pairs $[(m, f)] \in \mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$ for which group homomorphisms (2-33), (2-34) and (2-35) are onto. Elements of $\mathsf{GT}(\mathsf{N})$ are called GT–*shadows*.

It is easy to see that homomorphism (2-35) is onto if and only if

$$(2m+1) + N_{\mathrm{ord}}\mathbb{Z} \in (\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z})^{\times}. \tag{2-36}$$

We say that a GT–pair $[(m, f)]$ is *friendly* if m satisfies condition (2-36).

Due to the following proposition, only homomorphisms (2-34) and (2-35) matter.

Proposition 2.10 *Let* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $[(m, f)] \in \mathsf{GT}_{\mathrm{pr}}(\mathsf{N})$ *and* $T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$ *be the corresponding map of truncated operads*; *see* (2-21)*. The following statements are equivalent*:

(1) *The pair* $[(m, f)]$ *is a* GT*–shadow.*
(2) *Group homomorphisms* (2-34) *and* (2-35) *are onto.*
(3) *The map* $T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$ *is onto.*

Proof The implication (1) $\Longrightarrow$ (2) is obvious. It is also clear that if $T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$ is onto, then group homomorphisms (2-33), (2-34) and (2-35) are onto. Thus the implication (3) $\Longrightarrow$ (1) is also obvious.

It remains to prove the implication (2) $\Longrightarrow$ (3).

Since group homomorphism (2-35) is onto, there exists $\gamma_2 \in \mathrm{Hom}_{\mathsf{PaB}}(12, 12)$ such that

$$T_{m,f}(\gamma_2) = [\mathfrak{m}(x_{12}^{-m})].$$

Therefore,

$$T_{m,f}(\beta \cdot \gamma_2) = T_{m,f}(\beta) \cdot T_{m,f}(\gamma_2) = [\beta].$$

Since homomorphism (2-35) is onto, there exists $\gamma_3 \in \mathrm{Hom}_{\mathsf{PaB}}((12)3, (12)3)$ such that

$$T_{m,f}(\gamma_3) = [\mathfrak{m}(f^{-1})].$$

Therefore,

$$T_{m,f}(\alpha \cdot \gamma_3) = T_{m,f}(\alpha) \cdot T_{m,f}(\gamma_3) = T_{m,f}(\alpha) \cdot [\mathfrak{m}(f^{-1})] = [\alpha].$$

Since, as a truncated operad in the category of groupoids, $\mathsf{PaB}^{\le 4}$ is generated by β and α, the truncated operad $\mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$ is generated by the equivalence classes $[\beta] \in \mathsf{PaB}(2)/\sim_{\mathsf{N}}$ and $[\alpha] \in \mathsf{PaB}(3)/\sim_{\mathsf{N}}$.

Using the fact that $[\beta]$ and $[\alpha]$ belong to the image of $T_{m,f}$, we conclude that the morphism of truncated operads $T_{m,f}$ is indeed onto.

Since the implication (2) $\Longrightarrow$ (3) is established, the proposition is proved. □

2.5 The groupoid GTSh

Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and $[(m, f)] \in \mathsf{GT}(\mathsf{N})$. The morphism of truncated operads

$$T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$$

gives us the obvious compatible equivalence relation $\sim_{\mathfrak{s}}$:

$$\gamma_1 \sim_{\mathfrak{s}} \gamma_2 \iff T_{m,f}(\gamma_1) = T_{m,f}(\gamma_2). \tag{2-37}$$

Due to the following proposition we have $\sim_{\mathfrak{s}} = \sim_{\mathsf{N}^{\mathfrak{s}}_{m,f}}$, where $\mathsf{N}^{\mathfrak{s}}_{m,f} := \ker(T^{\mathrm{PB}_4}_{m,f})$ and $\mathsf{N}^{\mathfrak{s}}_{m,f} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. This will allow us to construct a groupoid GTSh with $\mathrm{Ob}(\mathsf{GTSh}) := \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. We will see that $\mathsf{GT}(\mathsf{N})$ is the set of morphisms of GTSh with the target N and, for every morphism $[(m, f)] \in \mathsf{GT}(\mathsf{N})$, its source is $\mathsf{N}^{\mathfrak{s}}_{m,f} := \ker(T^{\mathrm{PB}_4}_{m,f})$.

Proposition 2.11 *Let* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ *and*

$$\mathsf{N}^{\mathfrak{s}} = \mathsf{N}^{\mathfrak{s}}_{m,f} := \ker(T^{\mathrm{PB}_4}_{m,f}) \trianglelefteq \mathrm{PB}_4.$$

Then $\mathsf{N}^{\mathfrak{s}} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *and the compatible equivalence relation* $\sim_{\mathfrak{s}}$ *coincides with* $\sim_{\mathsf{N}^{\mathfrak{s}}}$.

Proof To prove the first statement, we observe that, since $\mathsf{N} \in \mathrm{PB}_4$, the standard homomorphism $\rho\colon B_4 \to S_4$ induces a group homomorphism $\widetilde{\rho}\colon \mathrm{B}_4/\mathsf{N} \to S_4$. Furthermore, using equations (2-25), it is easy to see that the composition

$$\psi := \widetilde{\rho} \circ T^{\mathrm{B}_4}_{m,f}\colon \mathrm{B}_4 \to S_4$$

coincides with ρ. Thus $\mathsf{N}^{\mathfrak{s}}$ is the kernel of a group homomorphism $T^{\mathrm{B}_4}_{m,f}$ from B_4 to a finite group B_4/N. Hence $\mathsf{N}^{\mathfrak{s}}$ is a finite-index normal subgroup of B_4. Since we also have $\mathsf{N}^{\mathfrak{s}} \le \mathrm{PB}_4$, we conclude that $\mathsf{N}^{\mathfrak{s}} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

Although the proof of the second statement is rather technical, the main idea is to show that group homomorphisms $T^{\mathrm{PB}_n}_{m,f}$ for $n = 2, 3, 4$ are, in some sense, compatible with the homomorphisms

$$\varphi_{123}, \quad \varphi_{12,3,4}, \quad \varphi_{1,23,4}, \quad \varphi_{1,2,34}, \quad \varphi_{234}, \quad \varphi_{12}, \quad \varphi_{12,3}, \quad \varphi_{1,23}, \quad \varphi_{23}.$$

See equations (A-18) and (A-19). This fact is deduced from the compatibility of $T_{m,f}$ with the structures of truncated operads. Then the desired second statement of Proposition 2.11 is a simple consequence of this compatibility property of homomorphisms $T^{\mathrm{PB}_n}_{m,f}$ for $n = 2, 3, 4$.

Let us consider $h \in \mathrm{PB}_n$ (for $n \in \{2, 3\}$) and denote by $\widetilde{h}$ any representative of the coset $T^{\mathrm{PB}_n}_{m,f}(h)$ in $\mathrm{PB}_n/\mathsf{N}_{\mathrm{PB}_n}$. Our first goal is to prove that, for every

$$\varphi \in \begin{cases} \{\varphi_{123}, \varphi_{12,3,4}, \varphi_{1,23,4}, \varphi_{1,2,34}, \varphi_{234}\} & \text{if } n = 3, \\ \{\varphi_{12}, \varphi_{12,3}, \varphi_{1,23}, \varphi_{23}\} & \text{if } n = 2, \end{cases}$$

there exists $g \in \mathrm{PB}_{n+1}/\mathsf{N}_{\mathrm{PB}_{n+1}}$ such that

$$g^{-1} T^{\mathrm{PB}_{n+1}}_{m,f}(\varphi(h)) g = \varphi(\widetilde{h})\, \mathsf{N}_{\mathrm{PB}_{n+1}}. \tag{2-38}$$

Indeed, let $n = 3$ and $\varphi = \varphi_{1,23,4}$. Setting $\eta := \mathfrak{m}(h)$ and using the compatibility of $T_{m,f}$ with operadic insertions and compositions we get

$$T_{m,f}(\eta) \circ_2 \mathrm{id}_{12} = T_{m,f}(\eta \circ_2 \mathrm{id}_{12}). \tag{2-39}$$

Applying $\mathfrak{ou}$ to the left-hand side of (2-39), we get

$$\mathfrak{ou}(T_{m,f}(\eta) \circ_2 \mathrm{id}_{12}) = \varphi_{1,23,4}(\widetilde{h}) \mathsf{N}_{\mathrm{PB}_4}, \tag{2-40}$$

where $\widetilde{h}$ is an element of the coset $T^{\mathrm{PB}_3}_{m,f}(h)$ in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$.

As for the right-hand side of (2-39), we have

$$\begin{aligned} T_{m,f}(\eta \circ_2 \mathrm{id}_{12}) &= T_{m,f}\big(\alpha^{(1(2,3))4}_{((1,2)3)4} \cdot \mathfrak{m}(\varphi_{1,23,4}(h)) \cdot \alpha^{((1,2)3)4}_{(1(2,3))4}\big) \\ &= T_{m,f}(\alpha^{(1(2,3))4}_{((1,2)3)4}) \cdot T_{m,f}\big(\mathfrak{m}(\varphi_{1,23,4}(h))\big) \cdot T_{m,f}(\alpha^{((1,2)3)4}_{(1(2,3))4}). \end{aligned}$$

Thus

$$\mathfrak{ou}(T_{m,f}(\eta \circ_2 \mathrm{id}_{12})) = g^{-1} T^{\mathrm{PB}_4}_{m,f}(\varphi_{1,23,4}(h)) g, \tag{2-41}$$

where $g = \mathfrak{ou}(T_{m,f}(\alpha^{((1,2)3)4}_{(1(2,3))4}))$.

Combining (2-40) with (2-41), we conclude that (2-38) holds for $n = 3$ and $\varphi = \varphi_{1,23,4}$.

Let us now consider the case when $n = 2$ and $\varphi = \varphi_{12}$.

As above, setting $\eta := \mathfrak{m}(h)$ and using the compatibility of $T_{m,f}$ with operadic insertions and compositions we get

$$\mathrm{id}_{12} \circ_1 T_{m,f}(\eta) = T_{m,f}(\mathrm{id}_{12} \circ_1 \eta). \tag{2-42}$$

Applying $\mathfrak{ou}$ to the left-hand side of (2-42), we get

$$\mathfrak{ou}(\mathrm{id}_{12} \circ_1 T_{m,f}(\eta)) = \varphi_{12}(\tilde{h})\mathsf{N}_{\mathrm{PB}_3}, \tag{2-43}$$

where $\tilde{h}$ is an element of the coset $T^{\mathrm{PB}_2}_{m,f}(h)$ in $\mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}$.

The right-hand side of (2-42) can be rewritten as

$$T_{m,f}(\mathrm{id}_{12} \circ_1 \eta) = T_{m,f}\big(\mathfrak{m}(\varphi_{12}(h))\big).$$

Hence

$$\mathfrak{ou}(T_{m,f}(\mathrm{id}_{12} \circ_1 \eta)) = \mathfrak{ou}\big(T_{m,f}\big(\mathfrak{m}(\varphi_{12}(h))\big)\big) = T^{\mathrm{PB}_3}_{m,f}(\varphi_{12}(h)). \tag{2-44}$$

Combining (2-43) with (2-44), we conclude that (2-38) holds for $n = 2$ and $\varphi = \varphi_{12}$ with $g = 1_{\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}}$.

The proof of (2-38) for the remaining case proceeds in the similar way.

Let us now prove that, for every $n \in \{2, 3, 4\}$,

$$h \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_n} \implies \mathfrak{m}(h) \sim_{\mathfrak{s}} \mathrm{id}_{((1,2)\dots}, \tag{2-45}$$

where $((1,2)\dots$ denotes 12 (resp. $(1,2)3$, $((1,2)3)4$) if $n = 2$ (resp. $n = 3$, $n = 4$).

For $n = 4$, (2-45) is a straightforward consequence of the definition of $\mathsf{N}^{\mathfrak{s}}$. So let $n = 3$ and $\tilde{h}$ be an element of the coset $T^{\mathrm{PB}_3}_{m,f}(h)$ in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$.

Since $T^{\mathrm{PB}_4}_{m,f}(\varphi(h)) = 1$ in PB_4/N for every $\varphi \in \{\varphi_{123}, \varphi_{12,3,4}, \varphi_{1,23,4}, \varphi_{1,2,34}, \varphi_{234}\}$, equation (2-38) implies that

$$\tilde{h} \in \varphi^{-1}_{123}(\mathsf{N}) \cap \varphi^{-1}_{12,3,4}(\mathsf{N}) \cap \varphi^{-1}_{1,23,4}(\mathsf{N}) \cap \varphi^{-1}_{1,2,34}(\mathsf{N}) \cap \varphi^{-1}_{234}(\mathsf{N}).$$

In other words, $\tilde{h} \in \mathsf{N}_{\mathrm{PB}_3}$ and hence $T^{\mathrm{PB}_3}_{m,f}(h) = 1$ in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ Thus (2-45) holds for $n = 3$.

Let us now consider the case $n = 2$ and denote by $\tilde{h}$ an element of the coset $T^{\mathrm{PB}_2}_{m,f}(h)$ in $\mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}$.

Since $\varphi(h) \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_3}$ for every $\varphi \in \{\varphi_{12}, \varphi_{12,3}, \varphi_{1,23}, \varphi_{23}\}$ and implication (2-45) is proved for $n = 3$, we conclude that

$$T^{\mathrm{PB}_3}_{m,f}(\varphi(h)) = 1 \quad \text{for all } \varphi \in \{\varphi_{12}, \varphi_{12,3}, \varphi_{1,23}, \varphi_{23}\}.$$

Therefore, equation (2-38) implies that

$$\tilde{h} \in \varphi^{-1}_{12}(\mathsf{N}_{\mathrm{PB}_3}) \cap \varphi^{-1}_{12,3}(\mathsf{N}_{\mathrm{PB}_3}) \cap \varphi^{-1}_{1,23}(\mathsf{N}_{\mathrm{PB}_3}) \cap \varphi^{-1}_{23}(\mathsf{N}_{\mathrm{PB}_3}).$$

In other words, $\tilde{h} \in \mathsf{N}_{\mathrm{PB}_2}$ and hence $T^{\mathrm{PB}_2}_{m,f}(h) = 1$ in $\mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}$. Thus implication (2-45) holds for $n = 2$ as well.

Let us now prove that, for every $n \in \{2,3,4\}$ and $h \in \mathrm{PB}_n$,

$$\mathfrak{m}(h) \sim_{\mathfrak{s}} \mathrm{id}_{((1,2)\dots} \implies h \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_n}. \tag{2-46}$$

Again, for $n=4$, (2-46) is a straightforward consequence of the definition of $\mathsf{N}^{\mathfrak{s}}$. So let $h \in \mathrm{PB}_3$. Since $\mathfrak{m}(h) \sim_{\mathfrak{s}} \mathrm{id}_{(1,2)3}$, $T^{\mathrm{PB}_3}_{m,f}(h)$ is the identity element of $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$. Hence, equation (2-38) implies that $\varphi(h) \in \mathsf{N}^{\mathfrak{s}}$ for every $\varphi \in \{\varphi_{123}, \varphi_{12,3,4}, \varphi_{1,23,4}, \varphi_{1,2,34}, \varphi_{234}\}$, or equivalently $h \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_3}$.

Similarly, if $h \in \mathrm{PB}_2$ and $\mathfrak{m}(h) \sim_{\mathfrak{s}} \mathrm{id}_{12}$, then $T^{\mathrm{PB}_2}_{m,f}(h)$ is the identity element of $\mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}$. Hence, equation (2-38) implies that $T^{\mathrm{PB}_3}_{m,f}(\varphi(h)) = 1$ in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ for every

$$\varphi \in \{\varphi_{12}, \varphi_{12,3}, \varphi_{1,23}, \varphi_{23}\},$$

or, equivalently,

$$\mathfrak{m}(\varphi(h)) \sim_{\mathfrak{s}} \mathrm{id}_{((1,2)3} \quad \text{for all } \varphi \in \{\varphi_{12}, \varphi_{12,3}, \varphi_{1,23}, \varphi_{23}\}.$$

Since implication (2-46) is already proved for $n=3$, we conclude that

$$\varphi(h) \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_3} \quad \text{for all } \varphi \in \{\varphi_{12}, \varphi_{12,3}, \varphi_{1,23}, \varphi_{23}\}.$$

Thus $h \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_2}$ and (2-46) is proved for $n=2$.

Let $n \in \{2,3,4\}$, $\tau \in \mathrm{Ob}(\mathsf{PaB}(n))$, $\eta \in \mathrm{Aut}_{\mathsf{PaB}}(\tau)$ and $h := \mathfrak{ou}(\eta) \in \mathrm{PB}_n$. Our next goal is to prove that

$$h \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_n} \iff \eta \sim_{\mathfrak{s}} \mathrm{id}_{\tau}. \tag{2-47}$$

Since $T_{m,f}$ is compatible with the action of the symmetric groups, we may assume, without loss of generality, that the underlying permutation of τ is the identity permutation in S_n.

Therefore

$$\eta = \alpha^{\tau}_{((1,2)\dots}\mathfrak{m}(h)\alpha^{((1,2)\dots}_{\tau} \tag{2-48}$$

and hence $T_{m,f}(\eta) = \mathrm{id}_{\tau}$ if and only if $T_{m,f}(\mathfrak{m}(h)) = \mathrm{id}_{((1,2)\dots}$; the latter is equivalent to $\mathfrak{m}(h) \sim_{\mathfrak{s}} \mathrm{id}_{((1,2)\dots}$.

Thus (2-47) is a consequence of implications (2-45) and (2-46).

Finally, let us use (2-47) to prove the statement of the proposition.

Let $\gamma, \tilde{\gamma} \in \mathsf{PaB}(n)$ (with $n \in \{2,3,4\}$) and τ be the source of both morphisms. Clearly, $\gamma \sim_{\mathfrak{s}} \tilde{\gamma}$ if and only if $\eta \sim_{\mathfrak{s}} \mathrm{id}_{\tau}$, where $\eta = \gamma^{-1} \cdot \tilde{\gamma}$.

Thus, due to (2-47), $\gamma \sim_{\mathfrak{s}} \tilde{\gamma}$ if and only if $\mathfrak{ou}(\gamma^{-1} \cdot \tilde{\gamma}) \in \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_n}$. □

Proposition 2.11 has the following important consequences:

Corollary 2.12 *For every* GT*–shadow* $[(m,f)] \in \mathsf{GT}(\mathsf{N})$,

- $|\mathrm{PB}_4 : \mathsf{N}^{\mathfrak{s}}| = |\mathrm{PB}_4 : \mathsf{N}|$,
- $|\mathrm{PB}_3 : \mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_3}| = |\mathrm{PB}_3 : \mathsf{N}_{\mathrm{PB}_3}|$, *and*
- $N^{\mathfrak{s}}_{\mathrm{ord}} = N_{\mathrm{ord}}$, *or equivalently*, $\mathsf{N}^{\mathfrak{s}}_{\mathrm{PB}_2} = \mathsf{N}_{\mathrm{PB}_2}$. □

Corollary 2.13 *For every* GT*–shadow* $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ *the morphism* $T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}}$ *of truncated operads factors as*

$$\begin{array}{ccc} \mathsf{PaB}^{\le 4} & \xrightarrow{T_{m,f}} & \\ \downarrow \mathcal{P}_{\mathsf{N}^{\mathfrak{s}}} & \searrow & \\ \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}^{\mathfrak{s}}} & \xrightarrow{T^{\mathrm{isom}}_{m,f}} & \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}} \end{array}$$

where $\mathcal{P}_{\mathsf{N}^{\mathfrak{s}}}$ *is the canonical projection and* $T^{\mathrm{isom}}_{m,f}$ *is an isomorphism of truncated operads.*

The assignment $[(m, f)] \mapsto T^{\mathrm{isom}}_{m,f}$ *gives us a bijection from the set*

$$\{[(m, f)] \in \mathsf{GT}(\mathsf{N}) \mid \mathsf{N}^{\mathfrak{s}} = \ker(T^{\mathrm{PB}_4}_{m,f})\}$$

to the set $\mathrm{Isom}(\mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}^{\mathfrak{s}}}, \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}})$ *of isomorphisms of truncated operads (in the category of groupoids).*

Proof Due to Proposition 2.10 and the definition of the equivalence relation $\sim_{\mathfrak{s}}$, we have the commutative diagram of morphisms of truncated operads

$$\begin{array}{ccc} \mathsf{PaB}^{\le 4} & \xrightarrow{T_{m,f}} & \\ \downarrow & \searrow & \\ \mathsf{PaB}^{\le 4}/\!\sim_{\mathfrak{s}} & \xrightarrow{T^{\mathrm{isom}}_{m,f}} & \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}} \end{array}$$

with $T^{\mathrm{isom}}_{m,f}$ being a bijection[7] on the level of morphisms.

Thanks to Proposition 2.11, the equivalence relation $\sim_{\mathfrak{s}}$ coincides with $\sim_{\mathsf{N}^{\mathfrak{s}}}$. Hence $T^{\mathrm{isom}}_{m,f}$ is a morphism of truncated operads

$$T^{\mathrm{isom}}_{m,f}\colon \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}^{\mathfrak{s}}} \xrightarrow{\cong} \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}}. \tag{2-49}$$

Let us denote by $S^{\mathrm{isom}}_{m,f}\colon \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}} \to \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}^{\mathfrak{s}}}$ the inverse of $T^{\mathrm{isom}}_{m,f}$ (viewed as a map of morphisms) and show that $S^{\mathrm{isom}}_{m,f}$ is compatible with the composition of morphisms and with the operadic insertions.

As for the compatibility with operadic insertions, we have

$$\begin{aligned} S^{\mathrm{isom}}_{m,f}([\gamma_1] \circ_i [\gamma_2]) &= S^{\mathrm{isom}}_{m,f}(T^{\mathrm{isom}}_{m,f}([\tilde{\gamma}_1]) \circ_i T^{\mathrm{isom}}_{m,f}([\tilde{\gamma}_2])) = S^{\mathrm{isom}}_{m,f}(T^{\mathrm{isom}}_{m,f}([\tilde{\gamma}_1] \circ_i [\tilde{\gamma}_2])) \\ &= [\tilde{\gamma}_1] \circ_i [\tilde{\gamma}_2] = S^{\mathrm{isom}}_{m,f}([\gamma_1]) \circ_i S^{\mathrm{isom}}_{m,f}([\gamma_2]) \end{aligned}$$

for any $[\gamma_1] \in \mathsf{PaB}(n)/\!\sim_{\mathsf{N}}$, $[\gamma_2] \in \mathsf{PaB}(k)/\!\sim_{\mathsf{N}}$ and $k \le 4$, $1 \le i \le n \le 4$.

The compatibility of $S^{\mathrm{isom}}_{m,f}$ with the composition of morphisms is proved in a similar fashion.

Let us now consider an isomorphism of truncated operads

$$T^{\mathrm{isom}}\colon \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}^{\mathfrak{s}}} \xrightarrow{\cong} \mathsf{PaB}^{\le 4}/\!\sim_{\mathsf{N}}.$$

[7]We tacitly assume that $T^{\mathrm{isom}}_{m,f}$ acts as the identity on the level of objects.

Pre-composing T^{isom} with the canonical projection $\mathcal{P}_{\mathsf{N}^\mathfrak{s}} : \mathsf{PaB}^{\leq 4} \to \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}^\mathfrak{s}}$ we get an onto morphism $T := T^{\text{isom}} \circ \mathcal{P}_{\mathsf{N}^\mathfrak{s}}$ of truncated operads. Since T is uniquely determined by a GT–shadow $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ such that $\ker(T^{\mathrm{PB}_4}_{m,f}) = \mathsf{N}^\mathfrak{s}$, we conclude that the assignment $[(m, f)] \mapsto T^{\text{isom}}_{m,f}$ is indeed a bijection

$$\{[(m, f)] \in \mathsf{GT}(\mathsf{N}) \mid \mathsf{N}^\mathfrak{s} = \ker(T^{\mathrm{PB}_4}_{m,f})\} \xrightarrow{\cong} \mathrm{Isom}(\mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}^\mathfrak{s}}, \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}}). \tag{2-50}$$

Corollary 2.13 is proved. □

Let us now observe that the assignment

$$\mathrm{Hom}(\tilde{\mathsf{N}}, \mathsf{N}) := \mathrm{Isom}(\mathsf{PaB}^{\leq 4}/\sim_{\tilde{\mathsf{N}}}, \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}}) \tag{2-51}$$

upgrades the set $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to a groupoid. The set of objects of this groupoid is $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and the set of morphisms from $\tilde{\mathsf{N}}$ to N is the set $\mathrm{Isom}(\mathsf{PaB}^{\leq 4}/\sim_{\tilde{\mathsf{N}}}, \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}})$ of isomorphisms of truncated operads (in the category of groupoids). Morphisms of this groupoid are composed in the standard way.

The second statement of Corollary 2.13 allows us to tacitly identify (2-51) with the set

$$\{[(m, f)] \in \mathsf{GT}(\mathsf{N}) \mid \ker(T^{\mathrm{PB}_4}_{m,f}) = \tilde{\mathsf{N}}\}.$$

We will use the identification in the remainder of this paper and we denote by GTSh the resulting *groupoid of* GT*–shadows.*

The following proposition gives us an explicit formula for the composition of morphisms in GTSh.

Proposition 2.14 *Let* $\mathsf{N}^{(1)}$, $\mathsf{N}^{(2)}$ *and* $\mathsf{N}^{(3)}$ *be elements of* $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *and*

$$[(m_1, f_1)] \in \mathrm{Hom}_{\mathsf{GTSh}}(\mathsf{N}^{(1)}, \mathsf{N}^{(2)}), \qquad [(m_2, f_2)] \in \mathrm{Hom}_{\mathsf{GTSh}}(\mathsf{N}^{(2)}, \mathsf{N}^{(3)}).$$

Then their composition $[(m_2, f_2)] \circ [(m_1, f_1)]$ *is represented by the pair* (m, f) *where*

$$m := 2m_1m_2 + m_1 + m_2 \quad \text{and} \quad f\mathsf{N}^{(3)}_{\mathrm{PB}_3} := f_2\mathsf{N}^{(3)}_{\mathrm{PB}_3} \cdot T^{\mathrm{PB}_3}_{m_2,f_2}(f_1). \tag{2-52}$$

Proof Let $[(m_2, f_2)] \in \mathsf{GT}(\mathsf{N}^{(3)})$ and $[(m_1, f_1)] \in \mathsf{GT}(\mathsf{N}^{(2)})$, where

$$\mathsf{N}^{(2)} := \ker(T^{\mathrm{PB}_4}_{m_2,f_2}) \qquad \text{and} \qquad \mathsf{N}^{(1)} := \ker(T^{\mathrm{PB}_4}_{m_1,f_1}).$$

In other words, the GT–shadow $[(m_1, f_1)]$ (resp. $[(m_2, f_2)]$) is a morphism from $\mathsf{N}^{(1)}$ to $\mathsf{N}^{(2)}$ (resp. a morphism from $\mathsf{N}^{(2)}$ to $\mathsf{N}^{(3)}$) in GTSh.

By Corollary 2.13, we have the diagram of morphisms of truncated operads

$$\begin{array}{ccccc} \mathsf{PaB}^{\leq 4} & & \xrightarrow{\quad T_{m,f} \quad} & & \\ \downarrow{\scriptstyle \mathcal{P}_{\mathsf{N}^{(1)}}} & & & & \searrow \\ \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}^{(1)}} & \xrightarrow{T^{\text{isom}}_{m_1,f_1}} & \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}^{(2)}} & \xrightarrow{T^{\text{isom}}_{m_2,f_2}} & \mathsf{PaB}^{\leq 4}/\sim_{\mathsf{N}^{(3)}} \end{array} \tag{2-53}$$

where the vertical arrow is the canonical projection.

Formula (2-52) is obtained by looking at the image of the associator $[\alpha] \in \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{N}^{(1)}}}$ (resp. the braiding $[\beta] \in \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{N}^{(1)}}}$) under $T^{\mathrm{isom}}_{m_2,f_2} \circ T^{\mathrm{isom}}_{m_1,f_1}$. For $[\alpha]$, we have

$$\begin{aligned} T_{m,f}(\alpha) &= T^{\mathrm{isom}}_{m_2,f_2}(T^{\mathrm{isom}}_{m_1,f_1}[\alpha]) = T^{\mathrm{isom}}_{m_2,f_2}(T_{m_1,f_1}(\alpha)) = T^{\mathrm{isom}}_{m_2,f_2}\big([\alpha\cdot\mathfrak{m}(f_1)]\big) \\ &= T_{m_2,f_2}(\alpha\cdot\mathfrak{m}(f_1)) = T_{m_2,f_2}(\alpha)\cdot T_{m_2,f_2}(\mathfrak{m}(f_1)) \\ &= [\alpha\cdot\mathfrak{m}(f_2)]\cdot\mathfrak{m}(T^{\mathrm{PB}_3}_{m_2,f_2}(f_1)) = [\alpha\cdot\mathfrak{m}(f)], \end{aligned}$$

where f is any representative of the coset $f_2\mathsf{N}^{(3)}_{\mathrm{PB}_3}\cdot T^{\mathrm{PB}_3}_{m_2,f_2}(f_1)$ in $\mathrm{PB}_3/\mathsf{N}^{(3)}_{\mathrm{PB}_3}$.

Similarly, computing $T_{m,f}(\beta)$, it is easy to see that $m \equiv 2m_1m_2 + m_1 + m_2 \bmod N^{(3)}_{\mathrm{ord}}$. □

Remark 2.15 Later we will see that it makes sense to focus only on GT–shadows that can be represented by pairs (m, f) with

$$f \in \mathsf{F}_2 \le \mathrm{PB}_3. \tag{2-54}$$

Let us call such GT–shadows *practical*.

Using (2-28) and (2-52), for the composition $[(m, f)] := [(m_2, f_2)]\circ[(m_1, f_1)]$ of practical GT–shadows $[(m_2, f_2)]$ and $[(m_1, f_1)]$ we get the formulas

$$\begin{aligned} m &:= 2m_1m_2 + m_1 + m_2, \\ f(x,y) &:= f_2(x,y)\, f_1(x^{2m_2+1}, f_2(x,y)^{-1} y^{2m_2+1} f_2(x,y)). \end{aligned} \tag{2-55}$$

Due to this observation, practical GT–shadows form a subgroupoid of GTSh.

Remark 2.16 The authors *do not know* whether there exists $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and an onto morphism of truncated operads $\mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{N}}}$ that *cannot* be represented by a pair $(m, f) \in \mathbb{Z}\times\mathsf{F}_2$.

2.5.1 The virtual cyclotomic character Let us observe that to every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ we assign the (finite) cyclic group

$$\mathrm{PB}_2/\langle x_{12}^{N_{\mathrm{ord}}}\rangle \cong \mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z},$$

where N_{ord} is the index of $\mathsf{N}_{\mathrm{PB}_2}$ in PB_2. Moreover, if $[(m, f)]$ is a morphism from K to N in the groupoid GTSh, then both N and K give the same quotient $\mathrm{PB}_2/\langle x_{12}^{N_{\mathrm{ord}}}\rangle$ of PB_2, ie $K_{\mathrm{ord}} = N_{\mathrm{ord}}$.

Proposition 2.14 implies that the assignment $\mathsf{N} \mapsto \mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$ upgrades to a functor $\mathsf{Ch}_{\mathrm{cyclot}}$ from GTSh to the category of finite cyclic groups. More precisely:

Corollary 2.17 *Let $[(m, f)]$ be a morphism from $\mathsf{N}^{(1)}$ to $\mathsf{N}^{(2)}$ in the groupoid GTSh. The assignments*

$$\begin{gathered} \mathsf{N} \mapsto \mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}, \quad [(m, f)] \mapsto \mathsf{Ch}_{\mathrm{cyclot}}(m, f) \in \mathrm{Aut}(\mathrm{PB}_2/\mathsf{N}^{(2)}_{\mathrm{PB}_2}), \\ \mathsf{Ch}_{\mathrm{cyclot}}(m, f)(x_{12}\mathsf{N}^{(2)}_{\mathrm{PB}_2}) := x_{12}^{2m+1}\mathsf{N}^{(2)}_{\mathrm{PB}_2}, \end{gathered} \tag{2-56}$$

define a functor $\mathsf{Ch}_{\mathrm{cyclot}}$ from the groupoid GTSh to the category of finite cyclic groups.

Proof Since, for every GT shadow $[(m, f)]$, $2m+1$ represents an invertible element of the ring $\mathbb{Z}/N^{(2)}_{\mathrm{ord}}\mathbb{Z}$, $\mathsf{Ch}_{\mathrm{cyclot}}(m, f)$ is clearly an automorphism of $\mathrm{PB}_2/\mathsf{N}^{(2)}_{\mathrm{PB}_2} = \mathrm{PB}_2/\mathsf{N}^{(1)}_{\mathrm{PB}_2}$.

Thus it remains to show that $\mathsf{Ch}_{\mathrm{cyclot}}$ is compatible with the composition of GT–shadows.

For this purpose, we consider two composable GT–shadows: $[(m_1, f_1)] \in \mathrm{Hom}_{\mathsf{GTSh}}(\mathsf{N}^{(1)}, \mathsf{N}^{(2)})$ and $[(m_2, f_2)] \in \mathrm{Hom}_{\mathsf{GTSh}}(\mathsf{N}^{(2)}, \mathsf{N}^{(3)})$.

Since $\mathsf{N}^{(1)}$, $\mathsf{N}^{(2)}$ and $\mathsf{N}^{(3)}$ belong to the same connected component of GTSh, $\mathsf{N}^{(1)}_{\mathrm{PB}_2} = \mathsf{N}^{(2)}_{\mathrm{PB}_2} = \mathsf{N}^{(3)}_{\mathrm{PB}_2}$ or equivalently, $N^{(1)}_{\mathrm{ord}} = N^{(2)}_{\mathrm{ord}} = N^{(3)}_{\mathrm{ord}}$. So let us set $\mathsf{N}_{\mathrm{PB}_2} := \mathsf{N}^{(1)}_{\mathrm{PB}_2}$ and $N_{\mathrm{ord}} := N^{(1)}_{\mathrm{ord}}$.

Let $[(m, f)] := [(m_2, f_2)] \circ [(m_1, f_1)]$.

Due to the first equation in (2-52), $m \equiv 2m_1m_2 + m_1 + m_2 \mod N_{\mathrm{ord}}$. Hence

$$\begin{aligned}\mathsf{Ch}_{\mathrm{cyclot}}(m, f)(x_{12}\mathsf{N}_{\mathrm{PB}_2}) &= x_{12}^{2(2m_1m_2+m_1+m_2)+1}\mathsf{N}_{\mathrm{PB}_2} = x_{12}^{4m_1m_2+2m_1+2m_2+1}\mathsf{N}_{\mathrm{PB}_2} \\ &= x_{12}^{(2m_1+1)(2m_2+1)}\mathsf{N}_{\mathrm{PB}_2} = (x_{12}^{(2m_1+1)}\mathsf{N}_{\mathrm{PB}_2})^{2m_2+1} \\ &= \mathsf{Ch}_{\mathrm{cyclot}}(m_2, f_2) \circ \mathsf{Ch}_{\mathrm{cyclot}}(m_1, f_1)\,(x_{12}\mathsf{N}_{\mathrm{PB}_2}).\end{aligned}$$

Thus $\mathsf{Ch}_{\mathrm{cyclot}}$ is indeed a functor from GTSh to the category of finite cyclic groups. □

We call the functor $\mathsf{Ch}_{\mathrm{cyclot}}$ the *virtual cyclotomic character*. This name is justified by the following remark.

Remark 2.18 Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $g \in G_{\mathbb{Q}}$ and $[(m, f)]$ be the GT–shadow in $\mathsf{GT}(\mathsf{N})$ induced by the element in $\widehat{\mathsf{GT}}$ corresponding to g. Then

$$\mathsf{Ch}_{\mathrm{cyclot}}(m, f)(x_{12}\mathsf{N}_{\mathrm{PB}_2}) = x_{12}^{\chi(g)_{N_{\mathrm{ord}}}}\mathsf{N}_{\mathrm{PB}_2}, \tag{2-57}$$

where $\chi\colon G_{\mathbb{Q}} \to \widehat{\mathbb{Z}}^{\times} \cong \mathrm{Aut}(\widehat{\mathbb{Z}})$ is the cyclotomic character and $\chi(g)_{N_{\mathrm{ord}}}$ represents the image of $\chi(g)$ in $\mathrm{Aut}(\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}) \cong (\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z})^{\times}$. Equation (2-57) follows from the discussion in [27, Example 4.7.4 and Remark 4.7.5]. See also [16, Proposition 1.6].

2.6 Charming GT–shadows

Recall that PB_3 is isomorphic to $\mathsf{F}_2 \times \mathbb{Z}$ where the F_2–factor is freely generated by x_{12} and x_{23} and the $\mathbb{Z}$–factor is generated by the central element c given in (A-5). This implies that $\widehat{\mathrm{PB}}_3 \cong \widehat{\mathsf{F}}_2 \times \widehat{\mathbb{Z}}$. Due to the following proposition, the action of $\widehat{\mathsf{GT}}$ on $\widehat{\mathrm{PB}}_3$ (viewed as the automorphism group of (12)3 in $\widehat{\mathsf{PaB}}$) respects this decomposition:

Proposition 2.19 *For every (continuous) automorphism $\widehat{T}$ of $\widehat{\mathsf{PaB}}^{\le 4}$, its restriction to the subgroup $\widehat{\mathsf{F}}_2 \le \widehat{\mathrm{PB}}_3$ gives us an automorphism*[8] *of $\widehat{\mathsf{F}}_2$*

$$\widehat{T}|_{\widehat{\mathsf{F}}_2}\colon \widehat{\mathsf{F}}_2 \to \widehat{\mathsf{F}}_2$$

defined by the formulas

$$\widehat{T}(x) := x^{2\widehat{m}+1} \quad \text{and} \quad \widehat{T}(y) := \widehat{f}^{-1} y^{2\widehat{m}+1} \widehat{f}. \tag{2-58}$$

[8] In fact, some specialists like to define $\widehat{\mathsf{GT}}$ as a certain subgroup of continuous automorphisms of $\widehat{\mathsf{F}}_2$.

The restriction of $\hat{T}$ to the central factor $\hat{\mathbb{Z}}$ of $\widehat{\mathrm{PB}}_3$ gives us the continuous automorphism of $\hat{\mathbb{Z}}$ defined by the formula

$$\hat{T}(c) := c^{2\hat{m}+1}. \tag{2-59}$$

Proof Due to Proposition 2.5, the action of $\hat{T}$ on $\widehat{\mathrm{PB}}_3$ is determined by the group homomorphisms

$$T^{\mathrm{PB}_3}_{m,f} \colon \mathrm{PB}_3 \to \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}, \quad \text{where } \mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4),$$

corresponding to GT–shadows $[(m, f)]$ that come from $\hat{T}$.

Combining this observation with equations (2-28) and the second equation in (2-29) and using the fact that c is a central element of PB_3, we conclude that the restrictions of $\hat{T}$ to $\hat{\mathsf{F}}_2$ and to $\hat{\mathbb{Z}}$ give us group homomorphisms

$$\hat{T}\big|_{\hat{\mathsf{F}}_2} \colon \hat{\mathsf{F}}_2 \to \hat{\mathsf{F}}_2 \quad \text{and} \quad \hat{T}\big|_{\hat{\mathbb{Z}}} \colon \hat{\mathbb{Z}} \to \hat{\mathbb{Z}}, \tag{2-60}$$

respectively.

Since the restrictions of the inverse of $\hat{T}$ to $\hat{\mathsf{F}}_2$ and to $\hat{\mathbb{Z}}$ give us inverses of the two homomorphisms in (2-60), respectively, the homomorphisms in (2-60) are indeed automorphisms.

Explicit formulas (2-58) and (2-59) are consequences of equations (2-28) and the second equation in (2-29). □

If a GT–shadow $[(m, f)]$ comes from an automorphism of $\widehat{\mathsf{PaB}}$ then it satisfies further conditions. The following definition is motivated by these conditions.

Definition 2.20 Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. A GT–shadow $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ is called *genuine* if it comes from an automorphism of $\widehat{\mathsf{PaB}}$. Otherwise, $[(m, f)]$ is called *fake*. Further, a GT–shadow $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ is called *charming* if

- the coset $f\mathsf{N}_{\mathrm{PB}_3}$ can be represented by $f_1 \in [\mathsf{F}_2, \mathsf{F}_2]$, and
- the group homomorphism

$$T^{\mathsf{F}_2}_{m,f} := T^{\mathrm{PB}_3}_{m,f}\big|_{\mathsf{F}_2} \colon \mathsf{F}_2 \to \mathsf{F}_2/(\mathsf{N}_{\mathrm{PB}_3} \cap \mathsf{F}_2) \tag{2-61}$$

  is onto.

Since the intersection $\mathsf{N}_{\mathrm{PB}_3} \cap \mathsf{F}_2$ plays an important role, we will denote it by $\mathsf{N}_{\mathsf{F}_2}$,

$$\mathsf{N}_{\mathsf{F}_2} := \mathsf{N}_{\mathrm{PB}_3} \cap \mathsf{F}_2. \tag{2-62}$$

Clearly, the kernel of the homomorphism $T^{\mathsf{F}_2}_{m,f} \colon \mathsf{F}_2 \to \mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ coincides with $\mathsf{N}^{\mathfrak{s}}_{\mathsf{F}_2}$ and $|\mathsf{F}_2 : \mathsf{N}^{\mathfrak{s}}_{\mathsf{F}_2}| = |\mathsf{F}_2 : \mathsf{N}_{\mathsf{F}_2}|$ for every charming GT–shadow $[(m, f)]$.

Let us prove that:

Proposition 2.21 *Every genuine* GT*–shadow is charming.*

Proof Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and $[(m, f)] \in \mathsf{GT}(\mathsf{N})$ be a genuine GT–shadow. The element $f \in \mathrm{PB}_3$ can be written uniquely as

$$f = g\, c^k,$$

where $g \in \mathsf{F}_2$, $k \in \mathbb{Z}$ and c is defined in (A-5).

Since $\mathsf{N}_{\mathrm{PB}_3}$ is a normal subgroup of finite index in PB_3, the subgroup $\mathsf{N}_{\mathsf{F}_2} := \mathsf{N}_{\mathrm{PB}_3} \cap \mathsf{F}_2$ is normal in F_2 and it has a finite index in F_2. Similarly, the subgroup $\mathsf{N}_{\mathbb{Z}} := \mathsf{N}_{\mathrm{PB}_3} \cap \mathbb{Z}$ has a finite index in $\mathbb{Z}$. Therefore, the subgroup $\mathsf{N}_{\mathsf{F}_2} \times \mathsf{N}_{\mathbb{Z}}$ is normal and it has finite index in PB_3.

Due to Proposition 2.5, there exists $\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ such that $\mathsf{K}_{\mathrm{PB}_3}$ is contained in $\mathsf{N}_{\mathsf{F}_2} \times \mathsf{N}_{\mathbb{Z}}$. Since $[(m, f)]$ is a genuine GT–shadow, there exists $(m_1, f_1) \in \mathbb{Z} \times \mathrm{PB}_3$ such that (m_1, f_1) represents the same GT–shadow $[(m, f)]$ in $\mathsf{GT}(\mathsf{N})$ and $[(m_1, f_1)] \in \mathsf{GT}(\mathsf{K})$.

Thus, without loss of generality, we may assume that $m = m_1$ and $f = f_1$, ie $[(m, f)] \in \mathsf{GT}(\mathsf{K})$.

Using relation (2-18), we have

$$\sigma_1 x_{12}^m f^{-1} \sigma_2 x_{23}^m f \mathsf{K}_{\mathrm{PB}_3} = f^{-1} \sigma_1 \sigma_2 (x_{13} x_{23})^m \mathsf{K}_{\mathrm{PB}_3}.$$

Next, using (A-5) and the fact that c is a central element of B_3, we get that

$$x_{12}^m \sigma_2^{-1} \sigma_1^{-1} g(x_{12}, x_{23}) \sigma_1 x_{12}^m g(x_{12}, x_{23})^{-1} \sigma_2 x_{23}^m g(x_{12}, x_{23}) c^{-m+k} \in \mathsf{K}_{\mathrm{PB}_3}.$$

Using equations in (A-6) from Section A.1, we deduce that

$$x_{12}^m g(x_{23}^{-1} x_{12}^{-1} c, x_{12})(x_{23}^{-1} x_{12}^{-1} c)^m g(x_{23}^{-1} x_{12}^{-1} c, x_{23})^{-1} x_{23}^m g(x_{12}, x_{23}) c^{-m+k} \in \mathsf{K}_{\mathrm{PB}_3}$$

or

$$x_{12}^m g(x_{23}^{-1} x_{12}^{-1}, x_{12})(x_{23}^{-1} x_{12}^{-1})^m g(x_{23}^{-1} x_{12}^{-1}, x_{23})^{-1} x_{23}^m g(x_{12}, x_{23}) c^k \in \mathsf{K}_{\mathrm{PB}_3}.$$

Since $\mathsf{K}_{\mathrm{PB}_3}$ is a subgroup of $\mathsf{N}_{\mathsf{F}_2} \times \mathsf{N}_{\mathbb{Z}}$, we have $c^k \in \mathsf{N}_{\mathbb{Z}} \subset \mathsf{N}_{\mathrm{PB}_3}$. Hence $f c^{-k} \mathsf{N}_{\mathrm{PB}_3} = g(x_{12}, x_{23}) \mathsf{N}_{\mathrm{PB}_3}$, and so the GT–shadow has a representative of the form (m, f) where $f \in \mathsf{F}_2$.

It remains to show that

- $[(m, f)]$ can be represented by a pair (m, f_1) with $f_1 \in [\mathsf{F}_2, \mathsf{F}_2]$, and
- the homomorphism (2-61) is onto.

Since homomorphism (2-61) does not depend on the choice of the representative of the GT–shadow $[(m, f)]$, we first prove that this homomorphism is indeed onto.

Due to Proposition 2.19, we have the following commutative diagram:

$$\begin{array}{ccc}
\widehat{\mathsf{F}}_2 & \xrightarrow{\widehat{T}|_{\widehat{\mathsf{F}}_2}} & \widehat{\mathsf{F}}_2 \\
{\scriptstyle i}\uparrow & & \downarrow{\scriptstyle \widehat{\mathcal{P}}_{\mathsf{N}_{\mathsf{F}_2}}} \\
\mathsf{F}_2 & \xrightarrow{T^{\mathsf{F}_2}_{m,f}} & \mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}
\end{array}$$

Since F_2 is dense in $\widehat{\mathsf{F}}_2$, we get that the composition $\widehat{\mathcal{P}}_{\mathsf{N}_{\mathsf{F}_2}} \circ \widehat{T}|_{\widehat{\mathsf{F}}_2} \circ i$ is surjective, whence we conclude $T^{\mathsf{F}_2}_{m,f}$ is onto.

Let us now prove that $[(m, f)]$ can be represented by a pair $(m, \tilde{f})$ with $\tilde{f} \in [\mathsf{F}_2, \mathsf{F}_2]$.

Let q be the least common multiple of the orders of $x_{12}\mathsf{N}_{\mathsf{F}_2}$ and $x_{23}\mathsf{N}_{\mathsf{F}_2}$ in $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ and $\psi_x \colon \mathrm{PB}_4 \to S_q$, $\psi_y \colon \mathrm{PB}_4 \to S_q$ be the group homomorphisms defined by equations

$$\psi_x(x_{12}) := (1,2,\dots,q), \quad \psi_x(x_{23}) = \psi_x(x_{13}) = \psi_x(x_{14}) = \psi_x(x_{24}) = \psi_x(x_{34}) := \mathrm{id}_{S_q},$$
$$\psi_y(x_{34}) := (1,2,\dots,q), \quad \psi_y(x_{12}) = \psi_y(x_{23}) = \psi_y(x_{13}) = \psi_y(x_{14}) = \psi_y(x_{24}) := \mathrm{id}_{S_q},$$

respectively.

Let K be an element of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ such that

$$\mathsf{K} \le \mathsf{N} \cap \ker(\psi_x) \cap \ker(\psi_y). \tag{2-63}$$

Since $[(m, f)]$ is a genuine GT–shadow, there exists a GT–shadow $[(m_1, f_1)] \in \mathsf{GT}(K)$ such that (m_1, f_1) is also a representative of $[(m, f)]$. We can assume, without loss of generality, that $f_1 \in \mathsf{F}_2$.

Applying equation (2-20) to f_1 we see that

$$f_1^{-1}(x_{13}x_{23}, x_{34}) f_1^{-1}(x_{12}, x_{23}x_{24}) f_1(x_{23}, x_{34}) f_1(x_{12}x_{13}, x_{24}x_{34}) f_1(x_{12}, x_{23}) \in \mathsf{K}. \tag{2-64}$$

Inclusions (2-63) and (2-64) imply that

$$\psi_x\big(f_1^{-1}(x_{13}x_{23}, x_{34}) f_1^{-1}(x_{12}, x_{23}x_{24}) f_1(x_{23}, x_{34}) f_1(x_{12}x_{13}, x_{24}x_{34}) f_1(x_{12}, x_{23})\big) = \mathrm{id}_{S_q},$$
$$\psi_y\big(f_1^{-1}(x_{13}x_{23}, x_{34}) f_1^{-1}(x_{12}, x_{23}x_{24}) f_1(x_{23}, x_{34}) f_1(x_{12}x_{13}, x_{24}x_{34}) f_1(x_{12}, x_{23})\big) = \mathrm{id}_{S_q}.$$

Hence the sum s_x of exponents of x_{12} in f_1 and the sum s_y of exponents of x_{23} in f_1 are multiples of q, ie $x_{12}^{-s_x} \in \mathsf{N}_{\mathsf{F}_2}$ and $x_{23}^{-s_y} \in \mathsf{N}_{\mathsf{F}_2}$.

Thus $(m, f_1 x_{12}^{-s_x} x_{23}^{-s_y})$ is yet another representative of the GT–shadow $[(m, f)]$ in $\mathsf{GT}(\mathsf{N})$ and, by construction, $f_1 x_{12}^{-s_x} x_{23}^{-s_y} \in [\mathsf{F}_2, \mathsf{F}_2]$. □

The following statement can be found in many introductory (and "not so introductory") papers on the Grothendieck–Teichmüller group $\widehat{\mathsf{GT}}$. Here, we deduce it from Proposition 2.21.

Corollary 2.22 *For every $(\hat{m}, \hat{f}) \in \widehat{\mathsf{GT}}$, $\hat{f}$ belongs to the topological closure of commutator subgroup of $\widehat{\mathsf{F}}_2$.*

Proof It suffices to show that, for every $\mathsf{N} \in \mathsf{NFI}(\mathsf{F}_2)$, the element $\widehat{\mathcal{P}}_{\mathsf{N}}(\hat{f}) \in \mathsf{F}_2/\mathsf{N}$ can be represented by $f_1 \in [\mathsf{F}_2, \mathsf{F}_2]$. Let us observe that $\mathsf{N} \times \langle c \rangle \in \mathsf{NFI}(\mathrm{PB}_3)$.

Due to Proposition 2.5, there exists $\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ such that $\mathsf{K}_{\mathrm{PB}_3} \le \mathsf{N} \times \langle c \rangle$. Clearly, $\mathsf{K}_{\mathsf{F}_2} \le \mathsf{N}$.

Since the pair $(\widehat{\mathcal{P}}_{K_{\mathrm{ord}}}(\hat{m}), \widehat{\mathcal{P}}_{\mathsf{K}_{\mathsf{F}_2}}(\hat{f}))$ is a charming GT–shadow in $\mathsf{GT}(\mathsf{K})$, the element $\widehat{\mathcal{P}}_{\mathsf{K}_{\mathsf{F}_2}}(\hat{f}) \in \mathsf{F}_2/\mathsf{K}_{\mathsf{F}_2}$ can be represented by $f_1 \in [\mathsf{F}_2, \mathsf{F}_2]$. Since $\mathsf{K}_{\mathsf{F}_2} \le \mathsf{N}$, the same element $f_1 \in [\mathsf{F}_2, \mathsf{F}_2]$ represents the coset $\widehat{\mathcal{P}}_{\mathsf{N}}(\hat{f}) \in \mathsf{F}_2/\mathsf{N}$. □

Let us denote by $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ the subset of all charming GT–shadows in $\mathsf{GT}(\mathsf{N})$, and prove that $\mathsf{GT}(\mathsf{N})$ can be safely replaced by $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ in all the constructions of Section 2.5. More precisely:

Proposition 2.23 *The assignment*

$$\text{(2-65)} \qquad \mathrm{Hom}_{\mathsf{GTSh}^{\heartsuit}}(\tilde{\mathsf{N}}, \mathsf{N}) := \{[(m,f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N}) \mid \tilde{\mathsf{N}} = \ker(T^{\mathrm{PB}_4}_{m,f})\} \quad \text{for } \tilde{\mathsf{N}}, \mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$$

upgrades the set $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *to a subgroupoid* $\mathsf{GTSh}^{\heartsuit}$ *of* GTSh*.*

Proof Let $[(m_1,f_1)] \in \mathrm{Hom}_{\mathsf{GTSh}^{\heartsuit}}(\mathsf{N}^{(1)}, \mathsf{N}^{(2)})$ and $[(m_2,f_2)] \in \mathrm{Hom}_{\mathsf{GTSh}^{\heartsuit}}(\mathsf{N}^{(2)}, \mathsf{N}^{(3)})$. The GT–shadows $[(m_1,f_1)]$ and $[(m_2,f_2)]$ are charming so we may assume, without loss of generality, that $f_1, f_2 \in [\mathsf{F}_2, \mathsf{F}_2]$.

Due to Remark 2.15, the composition $[(m_2,f_2)] \circ [(m_1,f_1)]$ is represented by a pair (m,f) with

$$f = f_2 f_1(x^{2m_2+1}, f_2^{-1}(x,y)y^{2m_2+1}f_2(x,y)).$$

Since $f_1, f_2 \in [\mathsf{F}_2, \mathsf{F}_2]$, it is clear that f also belongs to $[\mathsf{F}_2, \mathsf{F}_2]$.

Since $T^{\mathsf{F}_2}_{m,f} \colon \mathsf{F}_2 \to \mathsf{F}_2/\mathsf{N}^{(3)}$ is the composition of the onto homomorphism $T^{\mathsf{F}_2}_{m_1,f_1} \colon \mathsf{F}_2 \to \mathsf{F}_2/\mathsf{N}^{(2)}$ and the isomorphism $T^{\mathsf{F}_2,\mathrm{isom}}_{m_2,f_2} \colon \mathsf{F}_2/\mathsf{N}^{(2)} \to \mathsf{F}_2/\mathsf{N}^{(3)}$, the homomorphism $T^{\mathsf{F}_2}_{m,f}$ is also onto.

We proved that the subset of charming GT–shadows is closed under composition.

To prove that the subset of charming GT–shadows is closed under taking inverses, we start with a charming GT–shadow $[(m,f)] \in \mathrm{Hom}_{\mathsf{GTSh}^{\heartsuit}}(\mathsf{N}^{\mathfrak{s}}, \mathsf{N})$ and assume that $f \in [\mathsf{F}_2, \mathsf{F}_2]$. Let $[(\tilde{m}, \tilde{f})] \in \mathrm{Hom}_{\mathsf{GTSh}}(\mathsf{N}, \mathsf{N}^{\mathfrak{s}})$ be the inverse of $[(m,f)]$ in GTSh. In other words,

$$\text{(2-66)} \qquad \begin{aligned} 2m\tilde{m} + m + \tilde{m} &\equiv 0 \bmod N_{\mathrm{ord}}, \\ f\,T^{\mathrm{PB}_3}_{m,f}(\tilde{f}) &= 1_{\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}}. \end{aligned}$$

Our goal is to show that the coset $\tilde{f}\mathsf{N}_{\mathrm{PB}_3}$ can be represented by $g \in [\mathsf{F}_2, \mathsf{F}_2]$.

Since f^{-1} belongs to $[\mathsf{F}_2, \mathsf{F}_2]$, we have

$$f^{-1} = [g_{11}, g_{12}][g_{21}, g_{22}] \cdots [g_{r1}, g_{r2}],$$

where each $g_{ij} \in \mathsf{F}_2$ and $[g_1, g_2] := g_1 g_2 g_1^{-1} g_2^{-1}$.

Since the homomorphism $T^{\mathsf{F}_2}_{m,f} \colon \mathsf{F}_2 \to \mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ is onto, for every g_{ij} there exists $\tilde{g}_{ij} \in \mathsf{F}_2$ such that $T^{\mathsf{F}_2}_{m,f}(\tilde{g}_{ij}) = g_{ij}\mathsf{N}_{\mathsf{F}_2}$. Hence, for

$$g := [\tilde{g}_{11}, \tilde{g}_{12}][\tilde{g}_{21}, \tilde{g}_{22}] \cdots [\tilde{g}_{r1}, \tilde{g}_{r2}] \in [\mathsf{F}_2, \mathsf{F}_2],$$

we have $T^{\mathrm{PB}_3}_{m,f}(g) = f^{-1}\mathsf{N}_{\mathrm{PB}_3}$, or, equivalently,

$$\text{(2-67)} \qquad f\,T^{\mathrm{PB}_3}_{m,f}(g) = 1_{\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}}.$$

Combining (2-66) with (2-67) we conclude that $g^{-1}\tilde{f}$ belongs to the kernel of $T^{\mathrm{PB}_3}_{m,f} \colon \mathrm{PB}_3 \to \mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$. Thus, due to Proposition 2.11, g also represents the coset $\tilde{f}\mathsf{N}_{\mathrm{PB}_3}$.

Since, by construction, $g \in [\mathsf{F}_2, \mathsf{F}_2]$, the desired statement is proved. □

3 The Main Line functor $\mathcal{ML}$ and $\widehat{\mathsf{GT}}$

In this section, we use (charming) GT–shadows to construct a functor $\mathcal{ML}$ from a certain subposet of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the category of finite groups. We prove that the limit of the functor $\mathcal{ML}$ is isomorphic to the Grothendieck–Teichmüller group $\widehat{\mathsf{GT}}$.

3.1 Connected components of $\mathsf{GTSh}^{\heartsuit}$, settled GT–shadows and isolated elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$

Since the set $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ is infinite, so is the groupoid $\mathsf{GTSh}^{\heartsuit}$. Moreover, the groupoid $\mathsf{GTSh}^{\heartsuit}$ is highly disconnected. Indeed, if $\tilde{\mathsf{N}}$ and N are connected by a morphism in $\mathsf{GTSh}^{\heartsuit}$, then they must have the same index in PB_4.

For $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, we denote by

$$\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$$

the connected component of N in the groupoid $\mathsf{GTSh}^{\heartsuit}$. Clearly, an element $\tilde{\mathsf{N}}$ of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ is an object of $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ if and only if there exists $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N})$ such that

$$\tilde{\mathsf{N}} = \ker(T^{\mathrm{PB}_4}_{m,f}).$$

We call objects of the groupoid $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ *conjugates* of N.

Since $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is a finite set for every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, it is easy to show that:

Proposition 3.1 *For every* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, *the (connected) groupoid* $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ *is finite.* □

To establish a more precise link between (charming) GT–shadows and the group $\widehat{\mathsf{GT}}$, we will be interested in a certain subposet of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. Let us start with the following definition:

Definition 3.2 Let $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N})$. A charming GT–shadow $[(m, f)]$ is called *settled* if its source $\mathsf{N}^{\mathfrak{s}}$ coincides with N, ie $\ker(T^{\mathrm{B}_4}_{m,f}) = \mathsf{N}$. An element N of the poset $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ is called *isolated* if every GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is settled.

Clearly, a GT–shadow $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N})$ is settled if and only if $[(m, f)]$ is an automorphism of the object N in the groupoid $\mathsf{GTSh}^{\heartsuit}$. Moreover, an element $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ is isolated if and only if the groupoid $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ has exactly one object. In this case, $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is the group of automorphisms of the object N in the groupoid $\mathsf{GTSh}^{\heartsuit}$.

The following proposition gives us a simple way to produce many examples of isolated elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

Proposition 3.3 *For every* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, *the normal subgroup*

$$\mathsf{N}^\sharp := \bigcap_{\mathsf{K} \in \mathrm{Ob}(\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N}))} \mathsf{K} \tag{3-1}$$

is an isolated element of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

Proof Let $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N}^\sharp)$ and $\mathsf{N}^{\sharp,\mathfrak{s}}$ be the source of the corresponding morphism in $\mathsf{GTSh}^{\heartsuit}$, ie $\mathsf{N}^{\sharp,\mathfrak{s}} := \ker(T^{\mathrm{PB}_4}_{m,f})$.

Since $\mathsf{N}^\sharp \le \mathsf{K}$, the same pair $(m, f) \in \mathbb{Z} \times \mathsf{F}_2$ represents a GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{K})$. Moreover, the homomorphism from PB_4 to PB_4/K corresponding to $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{K})$ is the composition $\mathcal{P}_{\mathsf{N}^\sharp,\mathsf{K}} \circ T^{\mathrm{PB}_4}_{m,f}$ of $T^{\mathrm{PB}_4}_{m,f}$ with the canonical projection

$$\mathcal{P}_{\mathsf{N}^\sharp,\mathsf{K}} \colon \mathrm{PB}_4/\mathsf{N}^\sharp \to \mathrm{PB}_4/\mathsf{K}.$$

Let $h \in \mathsf{N}^\sharp$, let $\tilde{h} \in \mathrm{PB}_4$ be a representative of $T^{\mathrm{PB}_4}_{m,f}(h)$ and let $\mathsf{K}^{\mathfrak{s}}$ be the source of the GT–shadow $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{K})$, ie $\mathsf{K}^{\mathfrak{s}} := \ker(\mathcal{P}_{\mathsf{N}^\sharp,\mathsf{K}} \circ T^{\mathrm{PB}_4}_{m,f})$.

Since $\mathsf{N}^\sharp \le \mathsf{K}^{\mathfrak{s}}$, we have

$$\mathcal{P}_{\mathsf{N}^\sharp,\mathsf{K}}(T^{\mathrm{PB}_4}_{m,f}(h)) = 1_{\mathrm{PB}_4/\mathsf{K}}. \tag{3-2}$$

The identity (3-2) implies that $\tilde{h} \in \mathsf{K}$ for every $\mathsf{K} \in \mathrm{Ob}(\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N}))$ and hence $\tilde{h} \in \mathsf{N}^\sharp$. Therefore, $T^{\mathrm{PB}_4}_{m,f}(h) = 1_{\mathrm{PB}_4/\mathsf{N}^\sharp}$ or equivalently, $h \in \mathsf{N}^{\sharp,\mathfrak{s}}$.

We proved that

$$\mathsf{N}^\sharp \le \mathsf{N}^{\sharp,\mathfrak{s}}. \tag{3-3}$$

Since these subgroups have the same index in PB_4, inclusion (3-3) implies that $\mathsf{N}^{\sharp,\mathfrak{s}} = \mathsf{N}^\sharp$.

Since we started with an arbitrary GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N}^\sharp)$, we proved that $\mathsf{N}^\sharp$ is indeed an isolated element of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. □

Remark 3.4 In all examples we have considered so far (see Section 4 on selected results of computer experiments), $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ has at most two objects. Hence equation (3-1) gives us a practical way to produce examples of isolated elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

Let us denote by

$$\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4) \tag{3-4}$$

the subposet of isolated elements of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

Since $\mathsf{N}^\sharp \le \mathsf{N}$ for every $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, Proposition 3.3 implies that:

Corollary 3.5 *The subposet* $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *of* $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *is* ***cofinal***. *In other words, for every element* N *of* $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, *there exists* $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *such that* $\mathsf{K} \le \mathsf{N}$. □

Although Corollary 3.5 implies that the poset $\mathsf{NFI}^{\text{isolated}}_{\text{PB}_4}(\text{B}_4)$ is directed (it is a cofinal subposet of a directed poset), it is still useful to know that the intersection of two isolated elements of $\mathsf{NFI}_{\text{PB}_4}(\text{B}_4)$ is an isolated element of $\mathsf{NFI}_{\text{PB}_4}(\text{B}_4)$:

Proposition 3.6 *For every* $\mathsf{N}^{(1)}, \mathsf{N}^{(2)} \in \mathsf{NFI}^{\text{isolated}}_{\text{PB}_4}(\text{B}_4)$,

$$\mathsf{N}^{(1)} \cap \mathsf{N}^{(2)}$$

is also an isolated element of $\mathsf{NFI}_{\text{PB}_4}(\text{B}_4)$.

Proof $\mathsf{K} := \mathsf{N}^{(1)} \cap \mathsf{N}^{(2)}$ is clearly an element of $\mathsf{NFI}_{\text{PB}_4}(\text{B}_4)$. So our goal is to prove that K is isolated.

Let $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{K})$, and let $\mathsf{K}^{\mathfrak{s}}$ be the kernel of the homomorphism $T^{\text{PB}_4}_{m,f} \colon \text{PB}_4 \to \text{PB}_4/\mathsf{K}$.

Recall that $\mathcal{P}_{\mathsf{K},\mathsf{N}^{(1)}}$ (resp. $\mathcal{P}_{\mathsf{K},\mathsf{N}^{(2)}}$) is the canonical homomorphism from PB_4/K to $\text{PB}_4/\mathsf{N}^{(1)}$ (resp. to $\text{PB}_4/\mathsf{N}^{(2)}$). Since $\mathsf{K} \le \mathsf{N}^{(1)}$ and $\mathsf{K} \le \mathsf{N}^{(2)}$, the pair (m, f) also represents a GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(1)})$ and a GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(2)})$. Moreover, the compositions $\mathcal{P}_{\mathsf{K},\mathsf{N}^{(1)}} \circ T^{\text{PB}_4}_{m,f}$ and $\mathcal{P}_{\mathsf{K},\mathsf{N}^{(2)}} \circ T^{\text{PB}_4}_{m,f}$ are the homomorphisms $\text{PB}_4 \to \text{PB}_4/\mathsf{N}^{(1)}$ and $\text{PB}_4 \to \text{PB}_4/\mathsf{N}^{(2)}$ corresponding to these GT–shadows in $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(1)})$ and $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(2)})$, respectively.

Let us now consider $h \in \mathsf{K}^{\mathfrak{s}}$. Since $T^{\text{PB}_4}_{m,f}(h) = 1_{\text{PB}_4/\mathsf{K}}$, we have

$$\mathcal{P}_{\mathsf{K},\mathsf{N}^{(1)}} \circ T^{\text{PB}_4}_{m,f}(h) = 1_{\text{PB}_4/\mathsf{N}^{(1)}} \quad \text{and} \quad \mathcal{P}_{\mathsf{K},\mathsf{N}^{(2)}} \circ T^{\text{PB}_4}_{m,f}(h) = 1_{\text{PB}_4/\mathsf{N}^{(2)}}. \tag{3-5}$$

Since $\mathsf{N}^{(1)}$, $\mathsf{N}^{(2)}$ are both isolated, identities (3-5) imply that $h \in \mathsf{N}^{(1)}$ and $h \in \mathsf{N}^{(2)}$. Hence $h \in \mathsf{K}$.

Since we showed that $\mathsf{K}^{\mathfrak{s}} \le \mathsf{K}$ and both subgroups have the same (finite) index in PB_4, we have the desired equality $\mathsf{K}^{\mathfrak{s}} = \mathsf{K}$. □

Recall that, for every isolated element $\mathsf{N} \in \mathsf{NFI}_{\text{PB}_4}(\text{B}_4)$, the set $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is a finite group. More precisely, $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is the (finite) group of automorphisms of N in the groupoid $\mathsf{GTSh}^{\heartsuit}$. Let us denote this finite group by $\mathcal{ML}(\mathsf{N})$, and prove that:

Proposition 3.7 *The assignment*

$$\mathsf{N} \mapsto \mathcal{ML}(\mathsf{N})$$

upgrades to a functor $\mathcal{ML}$ *from the poset* $\mathsf{NFI}^{\text{isolated}}_{\text{PB}_4}(\text{B}_4)$ *to the category of finite groups.*

Proof Let $\mathsf{K} \le \mathsf{N}$ be isolated elements of $\mathsf{NFI}_{\text{PB}_4}(\text{B}_4)$. Our goal is to define a group homomorphism

$$\mathcal{ML}_{\mathsf{K},\mathsf{N}} \colon \mathcal{ML}(\mathsf{K}) \to \mathcal{ML}(\mathsf{N}) \tag{3-6}$$

and show that, for every triple of nested elements $\mathsf{N}^{(1)} \le \mathsf{N}^{(2)} \le \mathsf{N}^{(3)}$ of $\mathsf{NFI}^{\text{isolated}}_{\text{PB}_4}(\text{B}_4)$,

$$\mathcal{ML}_{\mathsf{N}^{(2)},\mathsf{N}^{(3)}} \circ \mathcal{ML}_{\mathsf{N}^{(1)},\mathsf{N}^{(2)}} = \mathcal{ML}_{\mathsf{N}^{(1)},\mathsf{N}^{(3)}}. \tag{3-7}$$

For this proof, it is convenient to identify GT–shadows $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{K})$ with the corresponding onto morphisms $T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$ of truncated operads. So let $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{K})$ and $T_{m,f}$ be the corresponding morphism.

Recall that $\mathcal{P}_{\mathsf{K},\mathsf{N}}$ denotes the canonical onto morphism of truncated operads

$$\mathcal{P}_{\mathsf{K},\mathsf{N}}\colon \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}.$$

Composing $\mathcal{P}_{\mathsf{K},\mathsf{N}}$ with $T_{m,f}$ we get an onto morphism

$$\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_{m,f}\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}},$$

and hence an element of $\mathsf{GT}^{\heartsuit}(\mathsf{N})$.

We set

$$\mathcal{ML}_{\mathsf{K},\mathsf{N}}(T_{m,f}) := \mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_{m,f}. \tag{3-8}$$

To prove that $\mathcal{ML}_{\mathsf{K},\mathsf{N}}$ is a group homomorphism from $\mathcal{ML}(\mathsf{K})$ to $\mathcal{ML}(\mathsf{N})$, we recall that, since K is isolated, every onto morphism of truncated operads $T\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$ factors as

$$\begin{array}{ccc}
\mathsf{PaB}^{\le 4} & & \\
{\scriptstyle \mathcal{P}_{\mathsf{K}}}\downarrow & \overset{T}{\searrow} & \\
\mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}} & \xrightarrow{T^{\mathrm{isom}}} & \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}
\end{array}$$

Let us now show that, for every onto morphism of truncated operads $T\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$, the following diagram commutes:

$$\begin{array}{ccc}
\mathsf{PaB}^{\le 4} & & \\
{\scriptstyle \mathcal{P}_{\mathsf{K}}}\downarrow & \overset{T}{\searrow} & \\
\mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}} & \xrightarrow{T^{\mathrm{isom}}} & \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}} \\
{\scriptstyle \mathcal{P}_{\mathsf{K},\mathsf{N}}}\downarrow & & \downarrow{\scriptstyle \mathcal{P}_{\mathsf{K},\mathsf{N}}} \\
\mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}} & \xrightarrow{(\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T)^{\mathrm{isom}}} & \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}
\end{array} \tag{3-9}$$

Since the top triangle of (3-9) commutes by definition of T^{isom}, we only need to prove the commutativity of the square. Let $\gamma \in \mathsf{PaB}^{\le 4}$ and $[\gamma]_{\mathsf{K}}$ (resp. $[\gamma]_{\mathsf{N}}$) be equivalence classes of γ in $\mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$ (resp. in $\mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$). Since $T^{\mathrm{isom}}([\gamma]_{\mathsf{K}}) = T(\gamma)$, $(\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T)^{\mathrm{isom}}([\gamma]_{\mathsf{N}}) = \mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T(\gamma)$ and $\mathcal{P}_{\mathsf{K},\mathsf{N}}([\gamma]_{\mathsf{K}}) = [\gamma]_{\mathsf{N}}$, we have

$$\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T^{\mathrm{isom}}([\gamma]_{\mathsf{K}}) = \mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T(\gamma) = (\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T)^{\mathrm{isom}}([\gamma]_{\mathsf{N}}) = (\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T)^{\mathrm{isom}} \circ \mathcal{P}_{\mathsf{K},\mathsf{N}}([\gamma]_{\mathsf{K}}).$$

Thus (3-9) indeed commutes.

Now let T_1 and T_2 be onto morphisms (of truncated operads)

$$T_1, T_2\colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}.$$

Since

$$T_1^{\mathrm{isom}} \circ T_2 \colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$$

is the composition of T_1 and T_2 in $\mathsf{GTSh}^{\heartsuit}$ and

$$(\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_1)^{\mathrm{isom}} \circ (\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_2) \colon \mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{N}}$$

is the composition of $\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_1$ and $\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_2$ in $\mathsf{GTSh}^{\heartsuit}$, our goal is to prove that

$$\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ (T_1^{\mathrm{isom}} \circ T_2) = (\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_1)^{\mathrm{isom}} \circ (\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_2). \tag{3-10}$$

Due to commutativity of (3-9), for $T = T_1$ we have

$$\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_1^{\mathrm{isom}} \circ T_2 = (\mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_1)^{\mathrm{isom}} \circ \mathcal{P}_{\mathsf{K},\mathsf{N}} \circ T_2.$$

Thus equation (3-10) indeed holds and we proved that $\mathcal{ML}_{\mathsf{K},\mathsf{N}}$ is a group homomorphism.

Let us now consider isolated elements $\mathsf{N}^{(1)} \le \mathsf{N}^{(2)} \le \mathsf{N}^{(3)}$ of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. Since

$$\mathcal{P}_{\mathsf{N}^{(1)},\mathsf{N}^{(3)}} = \mathcal{P}_{\mathsf{N}^{(2)},\mathsf{N}^{(3)}} \circ \mathcal{P}_{\mathsf{N}^{(1)},\mathsf{N}^{(2)}},$$

we have

$$\mathcal{ML}_{\mathsf{N}^{(2)},\mathsf{N}^{(3)}} \circ \mathcal{ML}_{\mathsf{N}^{(1)},\mathsf{N}^{(2)}}(T_{m,f}) = \mathcal{P}_{\mathsf{N}^{(2)},\mathsf{N}^{(3)}} \circ \mathcal{P}_{\mathsf{N}^{(1)},\mathsf{N}^{(2)}} \circ T_{m,f} = \mathcal{P}_{\mathsf{N}^{(1)},\mathsf{N}^{(3)}} \circ T_{m,f} = \mathcal{ML}_{\mathsf{N}^{(1)},\mathsf{N}^{(3)}}(T_{m,f})$$

for every $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N}^{(1)})$.

Thus the desired identity (3-7) holds and the proposition is proved. □

We call the functor[9] $\mathcal{ML}$ the *Main Line functor*.

In the next section, we will prove the following theorem:

Theorem 3.8 *The (profinite version)* $\widehat{\mathsf{GT}}$ *of the Grothendieck–Teichmüller group is isomorphic to*

$$\lim(\mathcal{ML}).$$

3.2 Proof of Theorem 3.8

We will need the following auxiliary statements:

Proposition 3.9 (A) *For every* $\mathsf{N} \in \mathsf{NFI}(\mathrm{PB}_3)$, *there exists* $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *satisfying the property*

$$\mathsf{K}_{\mathrm{PB}_3} \le \mathsf{N}.$$

(B) *For every* $\mathsf{N} \in \mathsf{NFI}(\mathrm{PB}_2)$ *there exists* $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *such that* $\mathsf{K}_{\mathrm{PB}_2} \le \mathsf{N}$.

[9] One of the authors of this paper is trying to live in the sequence of suburbs of Philadelphia called the Main Line. The functor $\mathcal{ML}$ is named after this beautiful sequence of suburbs.

Proof Let $\mathsf{N} \in \mathsf{NFI}(\mathrm{PB}_3)$ and ψ be a group homomorphism from PB_3 to S_n such that $\ker(\psi) = \mathsf{N}$.

Using relations (A-3) on the generators of PB_4, it is easy to show that the equations

$$\tilde{\psi}(x_{12}) := \psi(x_{12}), \quad \tilde{\psi}(x_{23}) := \psi(x_{23}), \quad \tilde{\psi}(x_{13}) := \psi(x_{13}),$$
$$\tilde{\psi}(x_{14}) = \tilde{\psi}(x_{24}) = \tilde{\psi}(x_{34}) := \mathrm{id}_{S_n},$$

define a group homomorphism $\tilde{\psi}\colon \mathrm{PB}_4 \to S_n$.

Moreover, the kernel of $\tilde{\psi}$ satisfies the property

$$\varphi_{123}^{-1}(\ker(\tilde{\psi})) = \mathsf{N}.$$

Hence

$$\varphi_{123}^{-1}(\ker(\tilde{\psi})) \cap \varphi_{12,3,4}^{-1}(\ker(\tilde{\psi})) \cap \varphi_{1,23,4}^{-1}(\ker(\tilde{\psi})) \cap \varphi_{1,2,34}^{-1}(\ker(\tilde{\psi})) \cap \varphi_{234}^{-1}(\ker(\tilde{\psi})) \le \mathsf{N}. \tag{3-11}$$

Let $\tilde{\mathsf{N}}$ be the normal subgroup of PB_4 obtained by intersecting all normal subgroups of PB_4 of index $|\mathrm{PB}_4 : \ker(\tilde{\psi})|$. Since $\tilde{\mathsf{N}}$ is a characteristic subgroup of PB_4 of finite index (in PB_4), we have

$$\tilde{\mathsf{N}} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4).$$

Furthermore, due to Corollary 3.5, there exists an isolated element K of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ satisfying the property $\mathsf{K} \le \tilde{\mathsf{N}}$. Combining $\mathsf{K} \le \tilde{\mathsf{N}}$ with $\tilde{\mathsf{N}} \le \ker(\tilde{\psi})$ and (3-11), we deduce that

$$\mathsf{K}_{\mathrm{PB}_3} \le \mathsf{N}.$$

Thus the desired statement (A) is proved.

Just as for statement (A), we start with a group homomorphism $\kappa\colon \mathrm{PB}_2 \to S_n$ whose kernel coincides with N.

It is easy to see that the equations

$$\tilde{\kappa}(x_{12}) := \kappa(x_{12}), \quad \tilde{\kappa}(x_{23}) := \kappa(x_{12})^{-1}, \quad \tilde{\kappa}(x_{13}) := \mathrm{id}_{S_n}, \quad \tilde{\kappa}(x_{14}) = \tilde{\kappa}(x_{24}) = \tilde{\kappa}(x_{34}) := \mathrm{id}_{S_n},$$

define a group homomorphism $\tilde{\kappa}\colon \mathrm{PB}_4 \to S_n$.

The kernel of $\tilde{\kappa}$ satisfies the property

$$\varphi_{12}^{-1}\big(\varphi_{123}^{-1}(\ker(\tilde{\kappa}))\big) = \mathsf{N}. \tag{3-12}$$

Let $\tilde{\mathsf{N}}$ be the normal subgroup of PB_4 obtained by intersecting all normal subgroups of PB_4 of index $|\mathrm{PB}_4 : \ker(\tilde{\kappa})|$. Since $\tilde{\mathsf{N}}$ is a characteristic subgroup of PB_4 of finite index (in PB_4), we have

$$\tilde{\mathsf{N}} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4).$$

As above, there exists an isolated element K of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ satisfying the property $\mathsf{K} \le \tilde{\mathsf{N}}$. Combining $\mathsf{K} \le \tilde{\mathsf{N}}$ with $\tilde{\mathsf{N}} \le \ker(\tilde{\kappa})$ and (3-12), we deduce that

$$\mathsf{K}_{\mathrm{PB}_2} \le \mathsf{N}.$$

Thus statement (B) is also proved. □

Proposition 3.9 allows us to produce a more practical description of $\widehat{\mathsf{PaB}}^{\le 4}$. To give this description, we note that the assignment $\mathsf{K} \mapsto \mathsf{PaB}^{\le 4}/\mathsf{K}$ upgrades to a functor from the poset $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ to the category of truncated operads in finite groupoids. Indeed, for every pair $\mathsf{K}_1 \le \mathsf{K}_2$ of elements of $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ we have the obvious morphism of truncated operads

$$\mathcal{P}_{\mathsf{K}_1,\mathsf{K}_2} \colon \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}_1} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}_2}.$$

Moreover, for every triple $\mathsf{K}_1 \le \mathsf{K}_2 \le \mathsf{K}_3$ of elements of $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, we have $\mathcal{P}_{\mathsf{K}_2,\mathsf{K}_3} \circ \mathcal{P}_{\mathsf{K}_1,\mathsf{K}_2} = \mathcal{P}_{\mathsf{K}_1,\mathsf{K}_3}$.

Let us denote by

$$\widetilde{\mathsf{PaB}}^{\le 4} \tag{3-13}$$

the limit of this functor.

More concretely, $\widetilde{\mathsf{PaB}}(n)$ consists of functions

$$\gamma \colon \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4) \to \bigsqcup_{\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)} \mathsf{PaB}(n)/\sim_{\mathsf{K}}$$

satisfying these two conditions:

- For every $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $\gamma(\mathsf{K}) \in \mathsf{PaB}(n)/\sim_{\mathsf{K}}$.
- For every pair $\mathsf{K}_1 \le \mathsf{K}_2$ in $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, $\mathcal{P}_{\mathsf{K}_1,\mathsf{K}_2}(\gamma(\mathsf{K}_1)) = \gamma(\mathsf{K}_2)$.

Since for every pair $\mathsf{K}_1 \le \mathsf{K}_2$ of elements of $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$ we have

$$\mathcal{P}_{\mathsf{K}_1,\mathsf{K}_2} \circ \widehat{\mathcal{P}}_{\mathsf{K}_1} = \widehat{\mathcal{P}}_{\mathsf{K}_2},$$

the assignment

$$\Psi(\hat{\gamma})(\mathsf{K}) := \widehat{\mathcal{P}}_{\mathsf{K}}(\hat{\gamma}) \quad \text{for } \hat{\gamma} \in \widehat{\mathsf{PaB}}(n)$$

defines a morphism of truncated operads

$$\Psi \colon \widehat{\mathsf{PaB}}^{\le 4} \to \widetilde{\mathsf{PaB}}^{\le 4}. \tag{3-14}$$

Let us prove:

Corollary 3.10 *The morphism* Ψ *in* (3-14) *is an isomorphism of truncated operads in the category of topological groupoids.*

Proof Since the compatibility with the structures of truncated operads and the composition of morphisms is obvious, it suffices to prove that Ψ is a homeomorphism of topological spaces.

Let τ, τ' be objects of $\mathsf{PaB}(n)$ and $\hat{\gamma}_1, \hat{\gamma}_2 \in \mathrm{Hom}_{\widehat{\mathsf{PaB}}}(\tau, \tau')$ such that $\Psi(\hat{\gamma}_1) = \Psi(\hat{\gamma}_2)$ or equivalently, for every $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$,

$$\Psi(\hat{\gamma}_2^{-1} \cdot \hat{\gamma}_1)(\mathsf{K})$$

is the identity automorphism of τ in $\mathsf{PaB}(n)/\sim_{\mathsf{K}}$.

Thus, due to Proposition 3.9, the image of $\hat{\gamma}_2^{-1}\cdot\hat{\gamma}_1$ in PB_n/N is the identity element for every $\mathsf{N}\in\mathsf{NFI}(\mathrm{PB}_n)$. Therefore $\hat{\gamma}_2^{-1}\cdot\hat{\gamma}_1$ is the identity element of $\widehat{\mathrm{PB}}_n$ and hence

$$\hat{\gamma}_1=\hat{\gamma}_2.$$

We have proved that Ψ is one-to-one.

Let $\gamma\in\widetilde{\mathsf{PaB}}(n)$, and let τ and τ' be the source and the target of γ, respectively. Let λ be any isomorphism from τ to τ' in $\mathsf{PaB}(n)$. By abuse of notation, we will use symbol λ for its obvious image in $\widehat{\mathsf{PaB}}(n)$ and in $\widetilde{\mathsf{PaB}}(n)$.

Due to Proposition 3.9, there exists an element $\hat{h}\in\widehat{\mathrm{PB}}_n$ such that

$$\hat{\mathcal{P}}_{\mathsf{K}}(\hat{h})=(\lambda^{-1}\cdot\gamma)(\mathsf{K})\quad\text{for all }\ \mathsf{K}\in\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4). \tag{3-15}$$

Equation (3-15) implies that $\Psi(\lambda\cdot\hat{h})=\gamma$. Thus we have proved that Ψ is onto.

Since, for every $\mathsf{K}\in\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the composition of Ψ with the canonical projection

$$\widetilde{\mathsf{PaB}}^{\le 4}\to\mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}}$$

coincides with the continuous map

$$\hat{\mathcal{P}}_{\mathsf{K}}\colon\widehat{\mathsf{PaB}}^{\le 4}\to\mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}},$$

we conclude that Ψ is continuous.

Since Ψ is a continuous bijection from a compact space $\widehat{\mathsf{PaB}}^{\le 4}$ to a Hausdorff space $\widetilde{\mathsf{PaB}}^{\le 4}$, Ψ is indeed a homeomorphism. □

Due to Corollary 3.10, we can safely replace $\widehat{\mathsf{PaB}}^{\le 4}$ by $\widetilde{\mathsf{PaB}}^{\le 4}$ in all further considerations. We will also use the same symbol $\mathcal{I}$ (resp. $\hat{\mathcal{P}}_{\mathsf{K}}$ for $\mathsf{K}\in\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$) for the canonical embedding $\mathcal{I}\colon\mathsf{PaB}^{\le 4}\to\widetilde{\mathsf{PaB}}^{\le 4}$ and the canonical projection $\hat{\mathcal{P}}_{\mathsf{K}}\colon\widetilde{\mathsf{PaB}}^{\le 4}\to\mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}}$.

Recall that, for every $\hat{T}\in\widehat{\mathsf{GT}}$ and $\mathsf{K}\in\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the formula $T_{\mathsf{K}}:=\hat{\mathcal{P}}_{\mathsf{K}}\circ\hat{T}\circ\mathcal{I}$ defines an onto morphism of truncated operads $\mathsf{PaB}^{\le 4}\to\mathsf{PaB}^{\le 4}/\mathsf{K}$. Since K is an isolated element of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, Corollary 2.13 implies that the onto morphism T_{K} factors as

$$T_{\mathsf{K}}=T^{\mathrm{isom}}_{\mathsf{K}}\circ\mathcal{P}_{\mathsf{K}}, \tag{3-16}$$

where $T^{\mathrm{isom}}_{\mathsf{K}}$ is an isomorphism of truncated operads

$$T^{\mathrm{isom}}_{\mathsf{K}}\colon\mathsf{PaB}^{\le 4}/\mathsf{K}\xrightarrow{\cong}\mathsf{PaB}^{\le 4}/\mathsf{K}$$

and $\mathcal{P}_{\mathsf{K}}$ is the canonical projection $\mathsf{PaB}^{\le 4}\to\mathsf{PaB}^{\le 4}/\mathsf{K}$.

Proposition 3.11 *For every $\hat{T}\in\widehat{\mathsf{GT}}$ and for every $\mathsf{K}\in\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the following diagram commutes*:

$$\begin{array}{ccc}\widetilde{\mathsf{PaB}}^{\le 4} & \xrightarrow{\hat{T}} & \widetilde{\mathsf{PaB}}^{\le 4}\\ {\scriptstyle\hat{\mathcal{P}}_{\mathsf{K}}}\downarrow & & \downarrow{\scriptstyle\hat{\mathcal{P}}_{\mathsf{K}}}\\ \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}} & \xrightarrow{T^{\mathrm{isom}}_{\mathsf{K}}} & \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}}\end{array} \tag{3-17}$$

Proof By definition of $T_{\mathsf{K}}^{\mathrm{isom}}$, $\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T} \circ \mathcal{I}(\gamma) = T_{\mathsf{K}}^{\mathrm{isom}} \circ \mathcal{P}_{\mathsf{K}}(\gamma)$ for every $\gamma \in \mathsf{PaB}^{\le 4}$.

Hence

$$\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}(\mathcal{I}(\gamma)) = T_{\mathsf{K}}^{\mathrm{isom}} \circ \widehat{\mathcal{P}}_{\mathsf{K}}(\mathcal{I}(\gamma)) \quad \text{for all } \gamma \in \mathsf{PaB}^{\le 4}. \tag{3-18}$$

Since the image $\mathcal{I}(\mathsf{PaB}^{\le 4})$ of $\mathsf{PaB}^{\le 4}$ in $\widetilde{\mathsf{PaB}}^{\le 4}$ is dense in $\widetilde{\mathsf{PaB}}^{\le 4}$ and the target $\mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$ of the compositions $\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}$ and $T_{\mathsf{K}}^{\mathrm{isom}} \circ \mathcal{P}_{\mathsf{K}}$ is Hausdorff, identity (3-18) implies that diagram (3-17) indeed commutes. □

Proof of Theorem 3.8 Let K and $\widetilde{\mathsf{K}}$ be elements of $\mathsf{NFI}_{\mathrm{PB}_4}^{\mathrm{isolated}}(\mathrm{B}_4)$ such that $\widetilde{\mathsf{K}} \le \mathsf{K}$ and $\mathcal{P}_{\widetilde{\mathsf{K}},\mathsf{K}}$ be the canonical projection from $\mathsf{PaB}^{\le 4}/\sim_{\widetilde{\mathsf{K}}}$ to $\mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$. Furthermore, let T_{K} and $T_{\widetilde{\mathsf{K}}}$ be onto morphisms from $\mathsf{PaB}^{\le 4}$ to $\mathsf{PaB}^{\le 4}/\mathsf{K}$ and $\mathsf{PaB}^{\le 4}/\widetilde{\mathsf{K}}$, respectively, coming from $\widehat{T} \in \widehat{\mathsf{GT}}$.

Since $\widehat{\mathcal{P}}_{\mathsf{K}} = \widehat{\mathcal{P}}_{\widetilde{\mathsf{K}},\mathsf{K}} \circ \mathcal{P}_{\widetilde{\mathsf{K}}}$, the diagram

$$\begin{array}{ccc} \mathsf{PaB}^{\le 4} & \xrightarrow{T_{\widetilde{\mathsf{K}}}} & \mathsf{PaB}^{\le 4}/\sim_{\widetilde{\mathsf{K}}} \\ & \searrow^{T_{\mathsf{K}}} & \downarrow {\scriptstyle \mathcal{P}_{\widetilde{\mathsf{K}},\mathsf{K}}} \\ & & \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}} \end{array}$$

commutes. Hence the assignment $\widehat{T} \mapsto \{T_{\mathsf{K}}\}_{\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}^{\mathrm{isolated}}(\mathrm{B}_4)}$ gives us a map

$$\widehat{\mathsf{GT}} \to \lim(\mathcal{ML}). \tag{3-19}$$

Let us show that the map (3-19) is a group homomorphism.

Indeed, let $\widehat{T}^{(1)}, \widehat{T}^{(2)} \in \widehat{\mathsf{GT}}$, put $\widehat{T} := \widehat{T}^{(1)} \circ \widehat{T}^{(2)}$ and let $\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}^{\mathrm{isolated}}(\mathrm{B}_4)$. Using Proposition 3.11, we get

$$\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T} = \widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}^{(1)} \circ \widehat{T}^{(2)} = T_{\mathsf{K}}^{(1),\,\mathrm{isom}} \circ \widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}^{(2)} = T_{\mathsf{K}}^{(1),\,\mathrm{isom}} \circ T_{\mathsf{K}}^{(2),\,\mathrm{isom}} \circ \widehat{\mathcal{P}}_{\mathsf{K}}.$$

On the other hand, $\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T} = T_{\mathsf{K}}^{\mathrm{isom}} \circ \widehat{\mathcal{P}}_{\mathsf{K}}$ and hence

$$T_{\mathsf{K}}^{\mathrm{isom}} \circ \widehat{\mathcal{P}}_{\mathsf{K}} = T_{\mathsf{K}}^{(1),\,\mathrm{isom}} \circ T_{\mathsf{K}}^{(2),\,\mathrm{isom}} \circ \widehat{\mathcal{P}}_{\mathsf{K}}. \tag{3-20}$$

Since $\widehat{\mathcal{P}}_{\mathsf{K}} \colon \widetilde{\mathsf{PaB}}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}}$ is onto, identity (3-20) implies that

$$T_{\mathsf{K}}^{\mathrm{isom}} = T_{\mathsf{K}}^{(1),\mathrm{isom}} \circ T_{\mathsf{K}}^{(2),\mathrm{isom}}.$$

Thus the map (3-19) is indeed a group homomorphism.

Our next goal is to show that homomorphism (3-19) is one-to-one and onto.

To prove that (3-19) is one-to-one, we consider $\widehat{T} \in \widehat{\mathsf{GT}}$ such that T_{K} coincides with the canonical projection

$$\mathsf{PaB}^{\le 4} \to \mathsf{PaB}^{\le 4}/\sim_{\mathsf{K}} \quad \text{for every } \mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}^{\mathrm{isolated}}(\mathrm{B}_4).$$

Hence, for every $\gamma \in \mathsf{PaB}^{\le 4}$, we have

$$\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}(\mathcal{I}(\gamma)) = \widehat{\mathcal{P}}_{\mathsf{K}} \circ \mathcal{I}(\gamma) \quad \text{for all } \mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}^{\mathrm{isolated}}(\mathrm{B}_4).$$

This means that the restriction of $\widehat{T}$ to the subset $\mathcal{I}(\mathsf{PaB}^{\le 4}) \subset \widetilde{\mathsf{PaB}}^{\le 4}$ coincides with the restriction of the identity map $\mathrm{id}\colon \widetilde{\mathsf{PaB}}^{\le 4} \to \widetilde{\mathsf{PaB}}^{\le 4}$ to the subset $\mathcal{I}(\mathsf{PaB}^{\le 4})$. Since the subset $\mathcal{I}(\mathsf{PaB}^{\le 4})$ is dense in $\widetilde{\mathsf{PaB}}^{\le 4}$ and the space $\widetilde{\mathsf{PaB}}^{\le 4}$ is Hausdorff, we conclude that $\widehat{T}$ is the identity map $\mathrm{id}\colon \widetilde{\mathsf{PaB}}^{\le 4} \to \widetilde{\mathsf{PaB}}^{\le 4}$. Thus the injectivity of (3-19) is established.

Note that an element of $\lim(\mathcal{ML})$ is a family $\{\mathcal{T}^{\mathrm{isom}}_{\mathsf{K}}\}_{\mathsf{K}\in\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)}$ of isomorphisms of truncated operads

$$\mathcal{T}^{\mathrm{isom}}_{\mathsf{K}}\colon \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}} \xrightarrow{\cong} \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}}$$

satisfying the following property: for every pair $\mathsf{K} \le \widetilde{\mathsf{K}}$ in $\mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$, the diagram

$$\begin{array}{ccc} \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}} & \xrightarrow{\mathcal{T}^{\mathrm{isom}}_{\mathsf{K}}} & \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}} \\ {\scriptstyle \mathcal{P}_{\mathsf{K},\widetilde{\mathsf{K}}}}\downarrow & & \downarrow{\scriptstyle \mathcal{P}_{\mathsf{K},\widetilde{\mathsf{K}}}} \\ \mathsf{PaB}^{\le 4}/{\sim_{\widetilde{\mathsf{K}}}} & \xrightarrow{\mathcal{T}^{\mathrm{isom}}_{\widetilde{\mathsf{K}}}} & \mathsf{PaB}^{\le 4}/{\sim_{\widetilde{\mathsf{K}}}} \end{array} \tag{3-21}$$

commutes.

Due to commutativity of (3-21), the formula

$$\widehat{T}(\gamma)(\mathsf{K}) := \mathcal{T}^{\mathrm{isom}}_{\mathsf{K}}(\gamma(\mathsf{K})) \tag{3-22}$$

defines a morphism of truncated operads in groupoids $\widehat{T}\colon \widetilde{\mathsf{PaB}}^{\le 4} \to \widetilde{\mathsf{PaB}}^{\le 4}$.

To prove that $\widehat{T}$ is continuous, we need to show that the composition

$$\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}\colon \widetilde{\mathsf{PaB}}^{\le 4} \to \mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}}$$

is continuous for every $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$.

By definition of $\widehat{T}$ in (3-22),

$$\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T} = \mathcal{T}^{\mathrm{isom}}_{\mathsf{K}} \circ \widehat{\mathcal{P}}_{\mathsf{K}} \quad \text{for every } \mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4). \tag{3-23}$$

Since $\mathcal{T}^{\mathrm{isom}}_{\mathsf{K}}$ is an automorphism of the (finite) groupoid $\mathsf{PaB}^{\le 4}/{\sim_{\mathsf{K}}}$ equipped with the discrete topology and $\widehat{\mathcal{P}}_{\mathsf{K}}$ is continuous, identity (3-23) implies that the composition $\widehat{\mathcal{P}}_{\mathsf{K}} \circ \widehat{T}$ is indeed continuous.

Thus equation (3-22) defines a continuous endomorphism of the operad $\widetilde{\mathsf{PaB}}^{\le 4}$.

To find the inverse of $\widehat{T}$, we denote by $\mathcal{S}^{\mathrm{isom}}_{\mathsf{K}}$ the inverse of $\mathcal{T}^{\mathrm{isom}}_{\mathsf{K}}$ for every $\mathsf{K} \in \mathsf{NFI}^{\mathrm{isolated}}_{\mathrm{PB}_4}(\mathrm{B}_4)$. Then it is easy to see that the formula

$$\widehat{S}(\gamma)(\mathsf{K}) := \mathcal{S}^{\mathrm{isom}}_{\mathsf{K}}(\gamma(\mathsf{K}))$$

defines the inverse of $\widehat{T}$.

The proof of surjectivity of (3-19) is complete. □

Let us consider $\mathsf{K}, \mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ with $\mathsf{K} \le \mathsf{N}$ and a pair $(m, f) \in \mathbb{Z} \times \mathsf{F}_2$ that represents a GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{K})$. Clearly, the same pair (m, f) also represents a GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N})$. In other words, if $\mathsf{K} \le \mathsf{N}$, then we have a natural map

$$\mathsf{GT}^{\heartsuit}(\mathsf{K}) \to \mathsf{GT}^{\heartsuit}(\mathsf{N}). \tag{3-24}$$

It makes sense to consider this map even if neither K nor N are isolated.

Definition 3.12 We say that a GT–shadow $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N})$ *survives into* K if $[(m, f)]$ belongs to the image of the map (3-24). In other words, there exists $(m_1, f_1) \in \mathbb{Z} \times \mathsf{F}_2$ such that $[(m_1, f_1)] \in \mathsf{GT}^{\heartsuit}(\mathsf{K})$, $m_1 \cong m \bmod N_{\mathrm{ord}}$ and $f_1 \mathsf{N}_{\mathsf{F}_2} = f \mathsf{N}_{\mathsf{F}_2}$.

The following statement is a straightforward consequence of Proposition 3.3 and Theorem 3.8.

Corollary 3.13 *Let* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *and* $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N})$. *The* GT*–shadow* $[(m, f)]$ *is genuine if and only if* $[(m, f)]$ *survives into* K *for every* $\mathsf{K} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *such that* $\mathsf{K} \le \mathsf{N}$. □

4 Selected results of computer experiments

In the computer implementation [4], an element N of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ is represented by a group homomorphism ψ from PB_4 to a symmetric group such that $\mathsf{N} = \ker(\psi)$. Each homomorphism $\psi : \mathrm{PB}_4 \to S_d$ is, in turn, represented by a tuple of permutations

$$(g_{12}, g_{23}, g_{13}, g_{14}, g_{24}, g_{34}) \in (S_d)^6 \tag{4-1}$$

satisfying the relations of PB_4; see (A-3).

It should be mentioned that, in [4], we consider only practical GT–shadows; see Remark 2.15. In particular, throughout this section, $\mathsf{GT}(\mathsf{N})$ denotes the set of practical GT–shadows with the target N. Clearly, every charming GT–shadow is practical.

Table 1 presents basic information about 35 selected elements

$$\mathsf{N}^{(0)}, \mathsf{N}^{(1)}, \dots, \mathsf{N}^{(34)} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4). \tag{4-2}$$

For every $\mathsf{N}^{(i)}$ in this list, the quotient $\mathsf{F}_2/\mathsf{N}^{(i)}_{\mathsf{F}_2}$ is nonabelian. Table 1 also shows $N^{(i)}_{\mathrm{ord}} := |\mathrm{PB}_2 : \mathsf{N}^{(i)}_{\mathrm{PB}_2}|$, the size of $\mathsf{GT}(\mathsf{N}^{(i)})$ (ie the total number of practical GT–shadows with the target $\mathsf{N}^{(i)}$) and the size of $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(i)})$. The last column indicates whether $\mathsf{N}^{(i)}$ is isolated or not.

For every nonisolated element N in the list (4-2), the connected component $\mathsf{GTSh}^{\heartsuit}_{\mathrm{conn}}(\mathsf{N})$ has exactly two objects. More precisely,

- $\mathsf{N}^{(4)}$ is a conjugate of $\mathsf{N}^{(3)}$ and $\mathsf{N}^{(3)} \cap \mathsf{N}^{(4)} = \mathsf{N}^{(14)}$,
- $\mathsf{N}^{(11)}$ is a conjugate of $\mathsf{N}^{(10)}$ and $\mathsf{N}^{(10)} \cap \mathsf{N}^{(11)} = \mathsf{N}^{(24)}$,
- $\mathsf{N}^{(17)}$ is a conjugate of $\mathsf{N}^{(16)}$ and $\mathsf{N}^{(16)} \cap \mathsf{N}^{(17)} = \mathsf{N}^{(30)}$,
- $\mathsf{N}^{(27)}$ is a conjugate of $\mathsf{N}^{(26)}$ and $\mathsf{N}^{(26)} \cap \mathsf{N}^{(27)} = \mathsf{N}^{(34)}$.

For $\mathsf{N}^{(31)}$, $\mathsf{GT}(\mathsf{N}^{(31)})$ has 588 elements. To find the size of $\mathsf{GT}(\mathsf{N}^{(31)})$, the computer had to look at $\approx 9 \cdot 10^6$ elements of the group $\mathsf{F}_2/\mathsf{N}^{(31)}_{\mathsf{F}_2}$. For the iMac with the processor 3.4 GHz, Intel Core i5, it took over nine full days to complete this task.

For $\mathsf{N}^{(32)}$, $\mathsf{GT}(\mathsf{N}^{(32)})$ has 800 elements. To find the size of $\mathsf{GT}(\mathsf{N}^{(32)})$, the computer had to look at over $9 \cdot 10^6$ elements of the group $\mathsf{F}_2/\mathsf{N}^{(32)}_{\mathsf{F}_2}$. For the iMac with the processor 3.4 GHz, Intel Core i5, it took almost 10 full days to complete this task.

Remark 4.1 Recall that the definition of an isolated element of $\mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$ (see Definition 3.2) is based on charming GT–shadows. In principle, it is possible that there exists an isolated element $\mathsf{N} \in \mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$ for which $\mathsf{GT}(\mathsf{N})$ has a nonsettled element. We *did not* encounter such examples in our experiments.

4.1 Selected remarkable examples

For the 19th example $\mathsf{N}^{(19)}$ in Table 1, the quotient $\mathsf{F}_2/\mathsf{N}^{(19)}_{\mathsf{F}_2}$ has order $7776 = 2^5 \cdot 3^5$. Due to the similarity between this order and the historic year 1776, we decided to call the subgroup $\mathsf{N}^{(19)}$ the *Philadelphia subgroup* of PB_4. This subgroup is the kernel of the homomorphism from PB_4 to S_9 that sends the standard generators of PB_4 to the permutations

$$
\begin{aligned}
&g_{12} := (1,3,2)(4,6,5), \quad g_{23} := (1,4,9)(2,7,6), \quad g_{13} := (1,7,5)(3,6,9),\\
&g_{14} := (2,6,7)(3,8,5), \quad g_{24} := (1,8,6)(3,4,7), \quad g_{34} := (1,2,3)(7,9,8),
\end{aligned} \tag{4-3}
$$

respectively.

Since $\mathsf{N}^{(19)}$ is isolated, $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(19)})$ is a group. We showed that $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(19)})$ is isomorphic to the dihedral group $D_6 = \langle r, s \mid r^6, s^2, rsrs \rangle$ of order 12. We also showed that the kernel of the restriction of the virtual cyclotomic character to $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(19)})$ coincides with the cyclic subgroup $\langle r \rangle$ of order 6.

The last element $\mathsf{N}^{(34)}$ in (4-2) has the biggest index $762,048 = 2^6 \cdot 3^5 \cdot 7^2$ in PB_4. This subgroup is the kernel of the homomorphism from PB_4 to S_{18} that sends the standard generators of PB_4 to

$$
\begin{aligned}
g_{12} &:= (1,3,5,7,9,2,4,6,8)(10,12,14,16,18,11,13,15,17),\\
g_{23} &:= (1,3,7,8,2,4,9,6,5)(10,15,17,11,12,16,18,14,13),\\
g_{13} &:= (1,3,8,5,4,9,2,6,7)(10,11,15,17,13,12,18,14,16),\\
g_{14} &:= (1,3,7,8,2,4,9,6,5)(10,15,17,11,12,16,18,14,13),\\
g_{24} &:= (1,7,6,2,4,8,9,3,5)(10,15,14,11,16,18,12,13,17),\\
g_{34} &:= (1,3,5,7,9,2,4,6,8)(10,12,14,16,18,11,13,15,17),
\end{aligned} \tag{4-4}
$$

respectively. We call this subgroup the *Mighty Dandy*.

Due to Proposition 3.3, the Mighty Dandy is an isolated element and hence $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(34)})$ is a group. This is what we showed about this group:

- $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(34)})$ has order $486 = 2 \cdot 3^5$.
- The kernel Ker_{34} of the restriction of the virtual cyclotomic character to $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(34)})$ is an abelian subgroup of order $81 = 3^4$; in fact, Ker_{34} is isomorphic to $\mathscr{Z}_9 \times \mathscr{Z}_9$.
- $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(34)})$ is isomorphic to the semidirect product
$$(\mathscr{Z}_2 \times \mathscr{Z}_3) \ltimes (\mathscr{Z}_9 \times \mathscr{Z}_9).$$
- The Sylow 3–subgroup Syl of $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(34)})$ is a nonabelian group of order $3^5 = 243$; Syl is a normal subgroup of $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(34)})$ and it is isomorphic to the semidirect product
$$\mathscr{Z}_3 \ltimes (\mathscr{Z}_9 \times \mathscr{Z}_9).$$

Although every element N in the list (4-2) has the property $|\mathsf{F}_2 : \mathsf{N}_{\mathsf{F}_2}| > |\mathrm{PB}_4 : \mathsf{N}|$, there are examples $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ for which $|\mathrm{PB}_4 : \mathsf{N}|$ is significantly bigger than the index $|\mathsf{F}_2 : \mathsf{N}_{\mathsf{F}_2}|$.

One such example was suggested to us by Leila Schneps. *Leila's subgroup* $\mathsf{N}^{\mathscr{L}}$ of PB_4 is the kernel of a homomorphism from PB_4 to S_{130} and it can be retrieved from one of the storage files in [4]. Here is what we know about $\mathsf{N}^{\mathscr{L}}$:

- The index of $\mathsf{N}^{\mathscr{L}}$ in PB_4 is $2^{29} \cdot 3^{12} = 285\,315\,214\,344\,192$.
- The index of $\mathsf{N}^{\mathscr{L}}_{\mathrm{PB}_3}$ in PB_3 is $2^{12} \cdot 3^6 = 2\,985\,984$.
- The index of $\mathsf{N}^{\mathscr{L}}_{\mathsf{F}_2}$ in F_2 is $2^{10} \cdot 3^5 = 248\,832$.
- $N^{\mathscr{L}}_{\mathrm{ord}} = 12$.
- The order of the commutator subgroup of $\mathsf{F}_2/\mathsf{N}^{\mathscr{L}}_{\mathsf{F}_2}$ is $2^6 \cdot 3^3 = 1728$.
- There are only $48 = 2^4 \cdot 3$ charming GT–shadows for $\mathsf{N}^{\mathscr{L}}$.
- $\mathsf{N}^{\mathscr{L}}$ is an *isolated* element of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ and hence $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{\mathscr{L}})$ is a group.

We found that the group $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{\mathscr{L}})$ is isomorphic to the semidirect product

$$\mathscr{Z}_2 \ltimes (\mathscr{Z}_2 \times \mathscr{Z}_2 \times \mathscr{Z}_2 \times \mathscr{Z}_3), \tag{4-5}$$

where the nontrivial element of $\mathscr{Z}_2$ acts on

$$\mathscr{Z}_2 \times \mathscr{Z}_2 \times \mathscr{Z}_2 \times \mathscr{Z}_3 = \langle a | a^2 \rangle \times \langle b | b^2 \rangle \times \langle c | c^2 \rangle \times \langle d | d^3 \rangle \tag{4-6}$$

by the automorphism

$$a \mapsto b, \quad b \mapsto a, \quad c \mapsto c, \quad d \mapsto d^{-1}.$$

The restriction of the virtual cyclotomic character to $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{\mathscr{L}})$ gives us the group homomorphism

$$\mathsf{GT}^{\heartsuit}(\mathsf{N}^{\mathscr{L}}) \to (\mathbb{Z}/12\mathbb{Z})^{\times},$$

and the kernel of this homomorphism is the subgroup of (4-6) generated by ab, c and d.

i	$\lvert \mathrm{PB}_4 : \mathsf{N}^{(i)} \rvert$	$\lvert \mathsf{F}_2 : \mathsf{N}^{(i)}_{\mathsf{F}_2} \rvert$	$\lvert [\mathsf{F}_2/\mathsf{N}^{(i)}_{\mathsf{F}_2}, \mathsf{F}_2/\mathsf{N}^{(i)}_{\mathsf{F}_2}] \rvert$	$N^{(i)}_{\mathrm{ord}}$	$\lvert \mathsf{GT}(\mathsf{N}^{(i)}) \rvert$	$\lvert \mathsf{GT}^{\heartsuit}(\mathsf{N}^{(i)}) \rvert$	isolated?
0	8	16	2	4	4	4	True
1	8	16	2	4	8	4	True
2	12	36	4	3	18	6	True
3	21	63	7	3	36	12	False
4	21	63	7	3	36	12	False
5	24	288	8	6	72	12	True
6	24	144	4	6	72	12	True
7	48	144	4	6	72	12	True
8	60	1500	60	5	100	20	True
9	60	900	4	15	360	24	True
10	72	144	18	4	16	8	False
11	72	144	18	4	16	8	False
12	108	972	27	6	72	12	True
13	120	6000	60	10	400	40	True
14	147	441	49	3	216	72	True
15	168	8232	168	7	294	42	True
16	168	1344	168	4	64	32	False
17	168	1344	168	4	64	32	False
18	180	13 500	60	15	600	40	True
19	216	7776	216	6	72	12	True
20	240	6000	60	10	400	40	True
21	324	8748	108	9	486	54	True
22	504	40 824	504	9	486	54	True
23	504	24 696	504	7	294	42	True
24	648	1296	162	4	32	16	True
25	720	54 000	240	15	1800	120	True
26	1512	40 824	504	9	486	54	False
27	1512	40 824	504	9	486	54	False
28	2520	63 000	2520	5	200	40	True
29	2520	45 360	2520	6	144	48	True
30	28 224	225 792	28 224	4	512	256	True
31	181 440	8 890 560	181 440	7	**588**	84	True
32	181 440	9 072 000	181 440	10	**800**	160	True
33	181 440	40 824 000	181 440	15	$\geq$ **1800**	120	True
34	762 048	20 575 296	254 016	9	$\geq$ **4374**	486	True

Table 1: The basic information about selected 35 compatible equivalence relations.

4.2 Is there a charming GT–shadow that is also fake?

Table 1 shows that the set $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ of charming GT–shadows corresponding to a given $\mathsf{N} \in \mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$ is typically a *proper subset* of $\mathsf{GT}(\mathsf{N})$. For example, for the Philadelphia subgroup $\mathsf{N}^{(19)}$, we have 72 GT–shadows and only 12 of them are charming.

Due to Proposition 2.21, every noncharming GT–shadow is fake. Thus, for a typical N from our list of 35 elements of $\mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$, we have many examples of a fake GT–shadows. For instance, $\mathsf{GT}(\mathsf{N}^{(19)})$ contains at least 60 fake GT–shadows.

It is more challenging to find examples of charming GT–shadows that are fake. At the time of writing, we did *not* find a single example of a charming GT–shadow that is also fake.

Here is what we did. In the list (4-2), there are exactly 24 pairs $(\mathsf{N}^{(i)}, \mathsf{N}^{(j)})$ with $i \neq j$ such that

$$\mathsf{N}^{(j)} \leq \mathsf{N}^{(i)}.$$

For each such pair, we showed that every GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(i)})$ survives into $\mathsf{N}^{(j)}$, ie the natural map $\mathsf{GT}^{\heartsuit}(\mathsf{N}^{(j)}) \to \mathsf{GT}^{\heartsuit}(\mathsf{N}^{(i)})$ is onto. We also looked at other selected examples of elements $\mathsf{K} \leq \mathsf{N}$ in $\mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$ in which N belongs to the list (4-2) and K is obtained by intersecting N with another element of (4-2). In all examples we have considered so far, the natural map $\mathsf{GT}^{\heartsuit}(\mathsf{K}) \to \mathsf{GT}^{\heartsuit}(\mathsf{N})$ is onto.

4.3 Versions of the Furusho property and selected open questions

Two versions of the Furusho property are motivated by a remarkable theorem which says roughly that, in the prounipotent setting, the pentagon relation implies the hexagon relations. For a precise statement, we refer the reader to [2, Theorem 3.1] and [10, Theorem 1].

We say that an element $\mathsf{N} \in \mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$ satisfies *the strong Furusho property* if:

Property 4.2 For every $f\mathsf{N}_{\mathsf{F}_2} \in \mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ satisfying pentagon relation (2-20) modulo N, there exists $m \in \mathbb{Z}$ such that

- $2m+1$ represents a unit in $\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$, and
- the pair (m, f) satisfies hexagon relations (2-18) and (2-19).

Furthermore, we say that an element $\mathsf{N} \in \mathsf{NFI}_{\mathsf{PB}_4}(\mathsf{B}_4)$ satisfies *the weak Furusho property* if:

Property 4.3 For every $f\mathsf{N}_{\mathsf{F}_2} \in [\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}, \mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}]$ satisfying pentagon relation (2-20) modulo N, there exists $m \in \mathbb{Z}$ such that

- $2m+1$ represents a unit in $\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$, and
- the pair (m, f) satisfies hexagon relations (2-18) and (2-19).

Using [4], we showed that the following 11 elements of the list (4-2)

$$\mathsf{N}^{(1)},\quad \mathsf{N}^{(2)},\quad \mathsf{N}^{(3)},\quad \mathsf{N}^{(4)},\quad \mathsf{N}^{(6)},\quad \mathsf{N}^{(7)},\quad \mathsf{N}^{(9)},\quad \mathsf{N}^{(10)},\quad \mathsf{N}^{(11)},\quad \mathsf{N}^{(14)},\quad \mathsf{N}^{(24)} \tag{4-7}$$

satisfy Property 4.2 and the remaining 24 elements of (4-2) do *not* satisfy Property 4.2.

For instance, for the Philadelphia subgroup $\mathsf{N}^{(19)}$, $N_{\mathrm{ord}}^{(19)} = 6$ and there are 216 elements $f\mathsf{N}_{\mathsf{F}_2}^{(19)}$ in $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}^{(19)}$ that satisfy the pentagon relation modulo $\mathsf{N}^{(19)}$. However, for only 36 of these 216 elements, there exists $m \in \{0, 1, \dots, 5\}$ such that $2m+1$ represents a unit in $\mathbb{Z}/6\mathbb{Z}$ and the pair (m, f) satisfies hexagon relations (2-18) and (2-19) (modulo $\mathsf{N}_{\mathrm{PB}_3}^{(19)}$).

Using [4], we also showed that the 13 elements of the list (4-2)

$$\mathsf{N}^{(0)},\ \mathsf{N}^{(1)},\ \mathsf{N}^{(2)},\ \mathsf{N}^{(3)},\ \mathsf{N}^{(4)},\ \mathsf{N}^{(5)},\ \mathsf{N}^{(6)},\ \mathsf{N}^{(7)},\ \mathsf{N}^{(9)},\ \mathsf{N}^{(10)},\ \mathsf{N}^{(11)},\ \mathsf{N}^{(14)},\ \mathsf{N}^{(24)} \tag{4-8}$$

satisfy Property 4.3, and the remaining 22 elements of (4-2) do *not* satisfy Property 4.3.

For instance, for the Mighty Dandy $\mathsf{N}^{(34)}$, $N_{\mathrm{ord}}^{(34)} = 9$ and there are 4096 elements[10] in $[\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}^{(34)}, \mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}^{(34)}]$ that satisfy the pentagon relation modulo $\mathsf{N}^{(34)}$. However, for only 243 of them does there exist some $m \in \{0, 1, \dots, 8\}$ such that $2m+1$ represents a unit in $\mathbb{Z}/9\mathbb{Z}$ and the pair (m, f) satisfies hexagon relations (2-18) and (2-19) (modulo $\mathsf{N}_{\mathrm{PB}_3}^{(34)}$).

We conclude this section with selected open questions. Most of these questions are motivated by our experiments [4].

Question 4.4 *Let* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *and* $(m, f) \in \mathbb{Z} \times \mathsf{F}_2$ *be a pair satisfying* (2-18), (2-19), (2-20) (*relative to* $\sim_{\mathsf{N}}$). *Recall that, due to Proposition 2.10, if the group homomorphisms* $T_{m,f}^{\mathrm{PB}_2}$ *and* $T_{m,f}^{\mathrm{PB}_3}$ *are onto then so is the group homomorphism*

$$T_{m,f}^{\mathrm{PB}_4} \colon \mathrm{PB}_4 \to \mathrm{PB}_4/\mathsf{N}.$$

Using [4], *the authors could not find an example of a pair* $(m, f) \in \mathbb{Z} \times \mathsf{F}_2$ *for which* $T_{m,f}^{\mathrm{PB}_4}$ *is onto but* $T_{m,f}^{\mathrm{PB}_2}$ *is not onto or* $T_{m,f}^{\mathrm{PB}_3}$ *is not onto. Can one prove that, if* $T_{m,f}^{\mathrm{PB}_4}$ *is onto, then so are the group homomorphisms* $T_{m,f}^{\mathrm{PB}_2}$ *and* $T_{m,f}^{\mathrm{PB}_3}$?

Question 4.5 *Is it possible to find an example of a nonisolated* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *for which the connected component* $\mathsf{GTSh}_{\mathrm{conn}}^{\heartsuit}(\mathsf{N})$ *has more than* 2 *objects? In other words, is it possible to find* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *that has* > 2 *distinct conjugates?*

Question 4.6 *Is it possible to find* $\mathsf{K}, \mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *such that* $\mathsf{K} \le \mathsf{N}$ *and the natural map*

$$\mathsf{GT}^{\heartsuit}(\mathsf{K}) \to \mathsf{GT}^{\heartsuit}(\mathsf{N})$$

is ***not*** *onto? In other words, can one produce an example of a charming* GT*–shadow that is also fake?*

Question 4.7 *Is it possible to find* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ *for which* $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ *is* ***nonabelian*** *and we can identify all genuine* GT*–shadows in the set* $\mathsf{GT}^{\heartsuit}(\mathsf{N})$?

Note that, if $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ is abelian, all charming GT–shadows can be described completely and they are *all genuine*. (See Theorem B.2 in Appendix B.)

[10] For the iMac with the processor 3.4 GHz, Intel Core i5, it took over 52 hours to find all these elements.

Appendix A The operad PaB and its profinite completion

The operad PaB of parenthesized braids is an operad in the category of groupoids and it was introduced[11] by D Tamarkin in [28].

In this appendix, we give a brief reminder of the operad PaB and its profinite completion. For a more detailed exposition, we refer the reader to [9, Chapter 6].

A.1 The groups B_n and PB_n

The Artin braid group B_n on n strands is, by definition, the fundamental group of the orbifold

$$\mathsf{Conf}(n,\mathbb{C})/S_n,$$

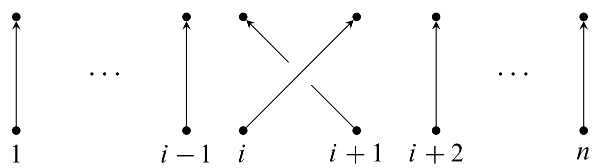
where $\mathsf{Conf}(n,\mathbb{C})$ denotes the configuration space of n (labeled) points on $\mathbb{C}$, ie

$$\mathsf{Conf}(n,\mathbb{C}) := \{(z_1,\dots,z_n)\in\mathbb{C}^n \mid z_i\neq z_j \text{ if } i\neq j\}.$$

It is known [17, Chapter 1] that B_n has the presentation

$$\langle \sigma_1,\sigma_2,\dots,\sigma_{n-1} \mid \sigma_i\sigma_j\sigma_i^{-1}\sigma_j^{-1} \text{ if } |i-j|\geq 2, \sigma_i\sigma_{i+1}\sigma_i\sigma_{i+1}^{-1}\sigma_i^{-1}\sigma_{i+1}^{-1} \text{ for } 1\leq i\leq n-2\rangle, \tag{A-1}$$

where σ_i is the element depicted below:



Recall that the *pure braid group* PB_n on n strands is the kernel of the standard group homomorphism $\rho\colon \mathrm{B}_n\to S_n$. This homomorphism sends the generator σ_i to the transposition $(i,i+1)$.

For $1\leq i<j\leq n$, we denote by x_{ij} the elements of PB_n given by

$$x_{ij} := \sigma_{j-1}\cdots\sigma_{i+1}\sigma_i^2\sigma_{i+1}^{-1}\cdots\sigma_{j-1}^{-1}, \tag{A-2}$$

and recall [17, Section 1.3] that PB_n has the presentation

$$\mathrm{PB}_n \cong \langle \{x_{ij}\}_{1\leq i<j\leq n} \mid \text{relations}\rangle$$

with the relations

$$x_{rs}^{-1}x_{ij}x_{rs} = \begin{cases} x_{ij} & \text{if } s<i \text{ or } i<r<s<j,\\ x_{rj}x_{ij}x_{rj}^{-1} & \text{if } s=i,\\ x_{rj}x_{sj}x_{ij}x_{sj}^{-1}x_{rj}^{-1} & \text{if } r=i<s<j,\\ x_{rj}x_{sj}x_{rj}^{-1}x_{sj}^{-1}x_{ij}x_{sj}x_{rj}x_{sj}^{-1}x_{rj}^{-1} & \text{if } r<i<s<j. \end{cases} \tag{A-3}$$

[11] A very similar construction appeared in the beautiful paper [1] by D Bar-Natan.

For example, the standard generators of PB_3 are

$$x_{12} := \sigma_1^2, \quad x_{23} := \sigma_2^2, \quad x_{13} := \sigma_2\sigma_1^2\sigma_2^{-1}. \tag{A-4}$$

The element

$$c := x_{23}x_{12}x_{13} = x_{12}x_{13}x_{23} = (\sigma_1\sigma_2)^3 = (\sigma_2\sigma_1)^3 \tag{A-5}$$

has infinite order; it generates the center of PB_3 and the center of B_3.

The elements x_{12} and x_{23} generate a free subgroup in PB_3. Thus PB_3 is isomorphic to $\mathsf{F}_2 \times \mathbb{Z}$.

A direction calculation shows that

$$\sigma_1^{-1}x_{23}\sigma_1 = x_{13}, \quad \sigma_2^{-1}x_{12}\sigma_2 = x_{23}^{-1}x_{12}^{-1}c, \quad \sigma_2^{-1}x_{13}\sigma_2 = x_{12}. \tag{A-6}$$

A.2 The groupoid $\mathsf{PaB}(n)$

Objects of $\mathsf{PaB}(n)$ are parenthesizations of sequences $(\tau(1), \tau(2), \dots, \tau(n))$ where τ is a permutation S_n. For example, $\mathsf{PaB}(2)$ has exactly two objects (1 2) and (2 1) and $\mathsf{PaB}(3)$ has 12 objects:

$$(12)3,\quad (21)3,\quad (23)1,\quad (32)1,\quad (31)2,\quad (13)2,\quad 1(23),\quad 2(13),\quad 2(31),\quad 3(21),\quad 3(12),\quad 1(32).$$

To define morphisms in $\mathsf{PaB}(n)$, we denote by $\mathfrak{p}$ the obvious projection from the set of objects of $\mathsf{PaB}(n)$ onto S_n. For example,

$$\mathfrak{p}((23)1) := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

For two objects τ_1 and τ_2 of $\mathsf{PaB}(n)$, we set

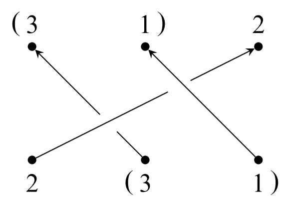
$$\mathrm{Hom}_{\mathsf{PaB}}(\tau_1, \tau_2) := \rho^{-1}(\mathfrak{p}(\tau_2)^{-1} \circ \mathfrak{p}(\tau_1)) \subset \mathrm{B}_n, \tag{A-7}$$

where ρ is the standard homomorphism B_n to S_n.

For instance, $\mathrm{Hom}_{\mathsf{PaB}}(2(31), (31)2)$ consist of elements $g \in \mathrm{B}_n$ such that

$$\rho(g) = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

An example of an isomorphism from $2(31)$ to $(31)2$ in $\mathsf{PaB}(3)$ is shown below:



The composition of morphisms in $\mathsf{PaB}(n)$ comes from the multiplication in B_n. For example, if η is the element of $\mathrm{Hom}_{\mathsf{PaB}}(\tau_1, \tau_2)$ corresponding to $h \in B_n$ and γ is the element of $\mathrm{Hom}_{\mathsf{PaB}}(\tau_2, \tau_3)$ corresponding to $g \in B_n$ then their composition $\gamma \cdot \eta$ is the element of $\mathrm{Hom}_{\mathsf{PaB}}(\tau_1, \tau_3)$ corresponding to $g \cdot h$. We use $\cdot$ for the composition of morphisms in PaB and the multiplication of elements in braid groups.

By definition of morphisms, we have a natural forgetful map

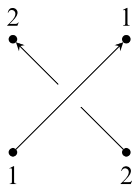
$$\mathfrak{ou}\colon \mathsf{PaB}(n) \to \mathrm{B}_n. \tag{A-8}$$

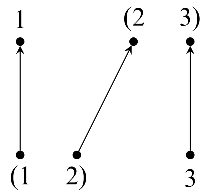
This map assigns to a morphism $\gamma \in \mathsf{PaB}(n)$ the corresponding element of the braid group B_n. Moreover, since the composition of morphisms in $\mathsf{PaB}(n)$ comes from the multiplication in B_n, we have

$$\mathfrak{ou}(\gamma \cdot \eta) = \mathfrak{ou}(\gamma) \cdot \mathfrak{ou}(\eta)$$

for every pair γ, η of composable morphisms.

The isomorphisms $\alpha \in \mathsf{PaB}(3)$ and $\beta \in \mathsf{PaB}(2)$ depicted as

$\beta :=$ 

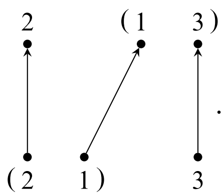
and $\alpha :=$ 

as in (1-2) play a very important role. We call β *the braiding* and α *the associator*. Note that, although α corresponds to the identity element in B_3, it is not an identity morphism in $\mathsf{PaB}(3)$ because $(12)3 \neq 1(23)$.

The symmetric group S_n acts on $\mathrm{Ob}(\mathsf{PaB}(n))$ in the obvious way. Moreover, for every $\theta \in S_n$ and $\gamma \in \mathrm{Hom}_{\mathsf{PaB}(n)}(\tau_1, \tau_2)$, we denote by $\theta(\gamma)$ the morphism from $\theta(\tau_1)$ to $\theta(\tau_2)$ that corresponds to the same element of the braid group B_n, ie

$$\mathfrak{ou}(\theta(\gamma)) = \mathfrak{ou}(\gamma). \tag{A-9}$$

For example, if $\theta = (1,2) \in S_3$ then

$\theta(\alpha) =$ 

.

For our purposes, it is convenient to assign to every element $g \in \mathrm{B}_n$ the corresponding morphism $\mathfrak{m}(g) \in \mathsf{PaB}(n)$ from $(..(1,2)3)\dots n)$ to $(..(i_1,i_2)i_3)\dots i_n)$, where $i_k := \rho(g)^{-1}(k)$. It is easy to see that the map

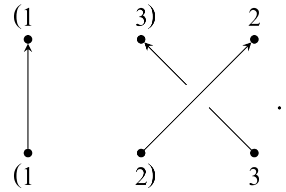
$$\mathfrak{m}\colon \mathrm{B}_n \to \mathsf{PaB}(n) \tag{A-10}$$

defined in this way is a right inverse of $\mathfrak{ou}$; see (A-8).

It is also easy to see that, for every pair $g_1, g_2 \in \mathrm{B}_n$, we have

$$\mathfrak{m}(g_1 \cdot g_2) = \rho(g_2)^{-1}(\mathfrak{m}(g_1)) \cdot \mathfrak{m}(g_2). \tag{A-11}$$

For example, for $\sigma_1, \sigma_2 \in \mathrm{B}_3$, $\mathfrak{m}(\sigma_1) = \mathrm{id}_{12} \circ_1 \beta$ and

$\mathfrak{m}(\sigma_2) :=$ 

.

The composition $\mathfrak{m}(\sigma_2) \cdot \mathfrak{m}(\sigma_1)$ is not defined because the source of $\mathfrak{m}(\sigma_2)$ does not coincide with the target of $\mathfrak{m}(\sigma_1)$. On the other hand, the source of $(1,2)(\mathfrak{m}(\sigma_2))$ coincides with the target of $\mathfrak{m}(\sigma_1)$ and $(1,2)(\mathfrak{m}(\sigma_2)) \cdot \mathfrak{m}(\sigma_1) = \mathfrak{m}(\sigma_2 \cdot \sigma_1)$.

A.3 The operad structure on PaB

We already explained how the symmetric group S_n acts on the groupoid $\mathsf{PaB}(n)$. Furthermore, it is easy to see that $\{\mathsf{Ob}(\mathsf{PaB}(n))\}_{n \geq 1}$ is the underlying collection of the free operad (in the category of sets) generated by the collection T with

$$\mathsf{T}(n) := \begin{cases} \{1\,2, 2\,1\} & \text{if } n = 2, \\ \varnothing & \text{otherwise.} \end{cases}$$

Thus the functors

$$\circ_i \colon \mathsf{PaB}(n) \times \mathsf{PaB}(m) \to \mathsf{PaB}(n+m-1) \tag{A-12}$$

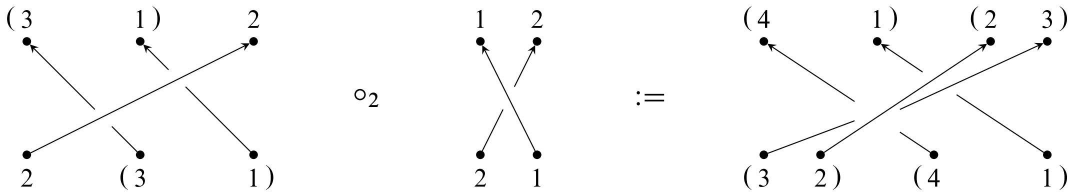
act on the level of objects in the obvious way.

For example,

$$(23)1 \circ_2 12 := ((23)4)1, \quad 21 \circ_1 (23)1 := 4((23)1), \quad 2(3(14)) \circ_3 1(32) := 2((3(54))(16)),$$

where we use the gray color to indicate what happens with the inserted sequence. For instance, in the third example, $1(32) \mapsto (3(54))$.

To define the action of the functor $\circ_i$ on the level of morphisms, we proceed as follows: given $\gamma \in \mathsf{PaB}(n)$, $\tilde{\gamma} \in \mathsf{PaB}(m)$ and $1 \leq i \leq n$, we set $g := \mathfrak{ou}(\gamma)$ and $\tilde{g} := \mathfrak{ou}(\tilde{\gamma})$; we compute the source and the target of $\gamma \circ_i \tilde{\gamma}$ using the rules of operad $\{\mathsf{Ob}(\mathsf{PaB}(k))\}_{k \geq 1}$. Finally, to get the element of B_{n+m-1} corresponding to $\gamma \circ_i \tilde{\gamma}$, we replace the strand of g that originates at the position labeled by i by a "thin" version of $\tilde{g}$. For example,



.

For a more precise definition of operadic multiplications on PaB we refer the reader to [9, Chapter 6].

The (iso)morphisms α and β satisfy the *pentagon relation*

$$
\begin{array}{ccccc}
 & (1(23))4 & \xrightarrow{\alpha \circ_2 \mathrm{id}_{12}} & 1((23)4) & \\
\nearrow^{\mathrm{id}_{12} \circ_1 \alpha} & & & & \searrow^{\mathrm{id}_{12} \circ_2 \alpha} \\
((12)3)4 & & & & 1(2(34)) \\
 & \searrow_{\alpha \circ_1 \mathrm{id}_{12}} & & \nearrow_{\alpha \circ_3 \mathrm{id}_{12}} & \\
 & & (12)(34) & &
\end{array}
\tag{A-13}
$$

and the two *hexagon relations*

$$
\begin{array}{ccccc}
(12)3 & \xrightarrow{\beta \circ_1 \mathrm{id}_{12}} & 3(12) & \xleftarrow{(1,3,2)\,\alpha} & (31)2 \\
{\scriptstyle \alpha}\downarrow & & & & \uparrow{\scriptstyle (2,3)\,(\mathrm{id}_{12} \circ_1 \beta)} \\
1(23) & \xrightarrow{\mathrm{id}_{12} \circ_2 \beta} & 1(32) & \xrightarrow{(2,3)\,\alpha^{-1}} & (13)2
\end{array}
\tag{A-14}
$$

$$
\begin{array}{ccccc}
1(23) & \xrightarrow{\beta \circ_2 \mathrm{id}_{12}} & (23)1 & \xrightarrow{(1,2,3)\,\alpha} & 2(31) \\
{\scriptstyle \alpha^{-1}}\downarrow & & & & \uparrow{\scriptstyle (1,2)\,(\mathrm{id}_{12} \circ_2 \beta)} \\
(12)3 & \xrightarrow{\mathrm{id}_{12} \circ_1 \beta} & (21)3 & \xrightarrow{(1,2)\,\alpha} & 2(13)
\end{array}
\tag{A-15}
$$

It is known [9, Theorem 6.2.4] that:[12]

Theorem A.1 *As the operad in the category of groupoids,* PaB *is generated by morphisms α and β of* (1-2). *Moreover, any relation on α and β in* PaB *is a consequence of* (A-13), (A-14) *and* (A-15).

A.4 The cosimplicial homomorphisms for pure braid groups in arities 2, 3, 4

The collection $\{\mathrm{PB}_n\}_{n \ge 1}$ of pure braid groups can be equipped with the structure of a cosimplicial group. For our purposes we will need the cofaces of this cosimplicial structure only in arities 2, 3 and 4.

Let τ_1 and τ_2 be objects of $\mathsf{PaB}(n)$ which differ only by parenthesizations, ie $\mathfrak{p}(\tau_1) = \mathfrak{p}(\tau_2)$. For such objects, we denote by $\alpha_{\tau_1}^{\tau_2}$ the isomorphism from τ_1 to τ_2 given by the identity element of B_n. For example, the associator α is precisely $\alpha_{(12)3}^{1(23)}$ and α^{-1} is precisely $\alpha_{1(23)}^{(12)3}$.

Using the identity morphism $\mathrm{id}_{12} \in \mathsf{PaB}(2)$, the maps $\mathfrak{ou}$, $\mathfrak{m}$ (see (A-8) and (A-10)) and the operadic insertions, we define the following maps from PB_3 to PB_4 and the maps from PB_2 to PB_3:

$$
\begin{gathered}
\varphi_{123}(h) := \mathfrak{ou}(\mathrm{id}_{12} \circ_1 \mathfrak{m}(h)), \quad \varphi_{12,3,4}(h) := \mathfrak{ou}(\mathfrak{m}(h) \circ_1 \mathrm{id}_{12}), \\
\varphi_{1,2,34}(h) := \mathfrak{ou}(\mathfrak{m}(h) \circ_3 \mathrm{id}_{12}), \qquad \varphi_{234}(h) := \mathfrak{ou}(\mathrm{id}_{12} \circ_2 \mathfrak{m}(h)), \\
\varphi_{1,23,4}(h) := \mathfrak{ou}(\mathfrak{m}(h) \circ_2 \mathrm{id}_{12}),
\end{gathered}
\tag{A-16}
$$

[12] A very similar statement is proved in [1]. See Claim 2.6 in loc. cit. It goes without saying that Theorem A.1 can be thought of as a version of Mac Lane's coherence theorem for braided monoidal categories.

(A-17)
$$\varphi_{12}(h) := \mathfrak{ou}(\mathrm{id}_{12} \circ_1 \mathfrak{m}(h)), \qquad \varphi_{23}(h) := \mathfrak{ou}(\mathrm{id}_{12} \circ_2 \mathfrak{m}(h)),$$
$$\varphi_{12,3}(h) := \mathfrak{ou}(\mathfrak{m}(h) \circ_1 \mathrm{id}_{12}), \quad \varphi_{1,23}(h) := \mathfrak{ou}(\mathfrak{m}(h) \circ_2 \mathrm{id}_{12}).$$

Proposition A.2 *The equations in* (A-16) (*resp. in* (A-17)) *define group homomorphisms from* PB_3 (*resp.* PB_2) *to* PB_4 (*resp.* PB_3).

Proof Let us consider the map $\varphi_{1,23,4} \colon \mathrm{PB}_3 \to \mathrm{PB}_4$. For elements $h, \tilde{h} \in \mathrm{PB}_3$, we set
$$\gamma := \mathfrak{m}(h) \quad \text{and} \quad \tilde{\gamma} := \mathfrak{m}(\tilde{h}).$$
Since PaB is an operad in the category of groupoids, we have
$$(\gamma \cdot \tilde{\gamma}) \circ_2 \mathrm{id}_{12} = (\gamma \circ_2 \mathrm{id}_{12}) \cdot (\tilde{\gamma} \circ_2 \mathrm{id}_{12}).$$
Hence
$$\varphi_{1,23,4}(h) \cdot \varphi_{1,23,4}(\tilde{h}) = \mathfrak{ou}(\gamma \circ_2 \mathrm{id}_{12}) \cdot \mathfrak{ou}(\tilde{\gamma} \circ_2 \mathrm{id}_{12}) = \mathfrak{ou}((\gamma \circ_2 \mathrm{id}_{12}) \cdot (\tilde{\gamma} \circ_2 \mathrm{id}_{12}))$$
$$= \mathfrak{ou}((\gamma \cdot \tilde{\gamma}) \circ_2 \mathrm{id}_{12}) = \varphi_{1,23,4}(h \cdot \tilde{h}),$$
where the last identity is a consequence of $\gamma \cdot \tilde{\gamma} = \mathfrak{m}(h \cdot \tilde{h})$.

The proofs for the remaining eight maps are very similar and we leave it to the reader. □

Since all nine maps in (A-16) and (A-17) are group homomorphisms, they are uniquely determined by their values on generators of PB_3 and PB_4, respectively. It is easy to see that

(A-18)
$$\begin{aligned}
&\varphi_{123}(x_{12}) = x_{12}, && \varphi_{123}(x_{23}) = x_{23}, && \varphi_{123}(x_{13}) = x_{13},\\
&\varphi_{234}(x_{12}) = x_{23}, && \varphi_{234}(x_{23}) = x_{34}, && \varphi_{234}(x_{13}) = x_{24},\\
&\varphi_{12,3,4}(x_{12}) = x_{13}x_{23}, && \varphi_{12,3,4}(x_{23}) = x_{34}, && \varphi_{12,3,4}(x_{13}) = x_{14}x_{24},\\
&\varphi_{1,23,4}(x_{12}) = x_{12}x_{13}, && \varphi_{1,23,4}(x_{23}) = x_{24}x_{34}, && \varphi_{1,23,4}(x_{13}) = x_{14},\\
&\varphi_{1,2,34}(x_{12}) = x_{12}, && \varphi_{1,2,34}(x_{23}) = x_{23}x_{24}, && \varphi_{1,2,34}(x_{13}) = x_{13}x_{14},
\end{aligned}$$

(A-19)
$$\varphi_{12}(x_{12}) = x_{12}, \quad \varphi_{23}(x_{12}) = x_{23}, \quad \varphi_{12,3}(x_{12}) = x_{13}x_{23}, \quad \varphi_{1,23}(x_{12}) = x_{12}x_{13}.$$

A.5 The profinite completion $\widehat{\mathsf{PaB}}$ of PaB

Let $\mathcal{G}$ be a connected groupoid with finitely many objects and G be the group that represents the isomorphism class of $\mathrm{Aut}(a)$ for some object a of $\mathcal{G}$. We tacitly assume that the group G is residually finite. Following [5], an equivalence relation $\sim$ on $\mathcal{G}$ is called *compatible* if:

(1) $\gamma_1 \sim \gamma_2$ implies the source (resp. the target) of γ_1 coincides with the source (resp. the target) of γ_2.
(2) $\gamma_1 \sim \gamma_2$ implies $\gamma_1 \cdot \gamma \sim \gamma_2 \cdot \gamma$ and $\tau \cdot \gamma_1 \sim \tau \cdot \gamma_2$ (if the compositions are defined).
(3) The set $\mathcal{G}/\sim$ of equivalence classes is finite.

It is clear that, for every compatible equivalence relation $\sim$ on $\mathcal{G}$, the quotient $\mathcal{G}/\sim$ is naturally a finite groupoid (with the same set of objects).

Compatible equivalence relations on $\mathcal{G}$ form a directed poset and the assignment $\sim \mapsto \mathcal{G}/\sim$ gives us a functor from this poset to the category of finite groupoids. In [5], the profinite completion $\widehat{\mathcal{G}}$ of the groupoid $\mathcal{G}$ is defined as the limit of this functor.

In [5], it was also shown that compatible equivalence relations on $\mathcal{G}$ are in bijection with finite-index normal subgroups N of G. This gives us the following "pedestrian" way of thinking about morphisms in $\widehat{\mathcal{G}}(a,b)$: choose[13] $\lambda \in \mathcal{G}(a,b)$, then every morphism in $\gamma \in \widehat{\mathcal{G}}(a,b)$ can be uniquely written as

$$\gamma = \lambda \cdot h,$$

where $h \in \widehat{G}$.

In [5], we also proved that the assignment $\mathcal{G} \mapsto \widehat{\mathcal{G}}$ upgrades to a functor from the category of groupoids to the category of topological groupoids. Moreover, this is a symmetric monoidal functor.

Thus, "putting hats" over $\mathsf{PaB}(n)$ for every $n \geq 0$ gives us an operad $\widehat{\mathsf{PaB}}$ in the category of topological groupoids.

Appendix B Charming GT–shadows in the abelian setting: examples of genuine GT–shadows

Let us prove the following statement:

Proposition B.1 *For* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$, *the following conditions are equivalent*:

(a) *The quotient group* PB_4/N *is abelian.*

(b) *The quotient group* $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ *is abelian.*

(c) *The quotient group* $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ *is abelian.*

Proof Implications (a) $\Longrightarrow$ (b) and (b) $\Longrightarrow$ (c) are straightforward, so we leave them to the reader.

Let us assume that the quotient group $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$ is abelian. Then the images of x_{12} and x_{23} in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ commute. Furthermore, since the image of c in $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ is obviously in the center of $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ and $\mathrm{PB}_3 = \langle x_{12}, x_{23}, c\rangle$, we conclude that the quotient group $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ is also abelian.

To show that the generators $\bar{x}_{ij} := x_{ij}\mathsf{N}$ for $1 \leq i < j \leq 4$ of PB_4/N commute with each other, we consider the group homomorphisms from PB_3 to PB_4 given by formulas (A-18).

Note that, for every homomorphism $\varphi\colon \mathrm{PB}_3 \to \mathrm{PB}_4$ in the set

$$\{\varphi_{234}, \varphi_{12,3,4}, \varphi_{1,23,4}, \varphi_{1,2,34}, \varphi_{234}\}, \tag{B-1}$$

we have $\mathsf{N}_{\mathrm{PB}_3} \leq \varphi^{-1}(\mathsf{N}) \leq \mathrm{PB}_3$. Therefore, since the quotient $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ is abelian, the quotient $\mathrm{PB}_3/\varphi^{-1}(\mathsf{N})$ is also abelian.

[13] $\mathcal{G}(a,b)$ is nonempty because $\mathcal{G}$ is connected.

Applying these observations to every φ in (B-1), we deduce that:

- The elements $\overline{x}_{12}, \overline{x}_{23}, \overline{x}_{13}$ commute with each other.
- The elements $\overline{x}_{23}, \overline{x}_{34}, \overline{x}_{24}$ commute with each other.
- The elements $\overline{x}_{13}\overline{x}_{23}$, $\overline{x}_{34}$ and $\overline{x}_{14}\overline{x}_{24}$ commute with each other.
- The elements $\overline{x}_{12}$, $\overline{x}_{23}\overline{x}_{24}$ and $\overline{x}_{13}\overline{x}_{14}$ commute with each other.
- The elements $\overline{x}_{14}$, $\overline{x}_{12}\overline{x}_{13}$ and $\overline{x}_{24}\overline{x}_{34}$ commute with each other.

Using these observations one can show that $[\overline{x}_{ij}, \overline{x}_{kl}] = 1_{\mathrm{PB}_4/\mathsf{N}}$ for every pair in the set

$$\big\{\{(i,j),(k,l)\} \mid 1 \le i < j \le 4,\ 1 \le k < l \le 4\big\} - \big\{\{(1,2),(3,4)\},\{(1,3),(2,4)\},\{(2,3),(1,4)\}\big\}.$$

Luckily, due to (A-3), we have

$$x_{12}x_{34} = x_{34}x_{12}, \quad x_{23}x_{14} = x_{14}x_{23}, \quad x_{13}^{-1}x_{24}x_{13} = [x_{14}, x_{34}]x_{24}[x_{14}, x_{34}]^{-1}.$$

Thus all generators $\overline{x}_{ij}$ of PB_4/N commute with each other. □

If one of the three equivalent conditions of Proposition B.1 is satisfied then we say that we are in *the abelian setting*.

We can now prove the following analog of the Kronecker–Weber theorem:

Theorem B.2 *Let* $\mathsf{N} \in \mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$. *If the quotient group* PB_4/N *is abelian, then*

$$\mathsf{GT}^{\heartsuit}(\mathsf{N}) = \{(m + N_{\mathrm{ord}}\mathbb{Z}, \overline{1}) \mid 0 \le m \le N_{\mathrm{ord}} - 1,\ \gcd(2m+1, N_{\mathrm{ord}}) = 1\}, \tag{B-2}$$

where $\overline{1}$ *is the identity element of* $\mathsf{F}_2/\mathsf{N}_{\mathsf{F}_2}$. *Furthermore, every* GT*–shadow in* (B-2) *is genuine.*

Proof Since $\overline{1}$ can be represented by the identity element of F_2, every element of the set

$$X_{\mathsf{N}} := \{(m + N_{\mathrm{ord}}\mathbb{Z}, \overline{1}) \mid 0 \le m \le N_{\mathrm{ord}} - 1,\ \gcd(2m+1, N_{\mathrm{ord}}) = 1\} \tag{B-3}$$

satisfies the pentagon relation (2-20).

For every element of X_{N}, the hexagon relations (2-18) and (2-19) boil down to

$$\sigma_1 x_{12}^m \sigma_2 x_{23}^m \,\mathsf{N}_{\mathrm{PB}_3} = \sigma_1\sigma_2(x_{13}x_{23})^m \,\mathsf{N}_{\mathrm{PB}_3}, \tag{B-4}$$

$$\sigma_2 x_{23}^m \,\sigma_1 x_{12}^m \,\mathsf{N}_{\mathrm{PB}_3} = \sigma_2\sigma_1(x_{12}x_{13})^m \,\mathsf{N}_{\mathrm{PB}_3}. \tag{B-5}$$

Equation (B-4) follows easily from the identity

$$\sigma_2^{-1}x_{12}\sigma_2 = x_{23}^{-1}x_{13}x_{23}$$

and the fact that the quotient $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ is abelian.

Similarly, equation (B-5) follows easily from the identity

$$\sigma_1^{-1} x_{23} \sigma_1 = x_{13}$$

and the fact that the quotient $\mathrm{PB}_3/\mathsf{N}_{\mathrm{PB}_3}$ is abelian.

We proved that every element of X_{N} is a GT–pair for N. Moreover, since $2m+1$ represents a unit in the ring $\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$, every GT–pair in X_{N} is friendly, ie the group homomorphism $T^{\mathrm{PB}_2}_{m,1}: \mathrm{PB}_2 \to \mathrm{PB}_2/\mathsf{N}_{\mathrm{PB}_2}$ is onto.

Due to (2-28) and the second identity in (2-29), for every $m \in \mathbb{Z}$ we have

$$T^{\mathrm{PB}_3}_{m,1}(x_{12}) = x_{12}^{2m+1}\mathsf{N}_{\mathrm{PB}_3}, \quad T^{\mathrm{PB}_3}_{m,1}(x_{23}) = x_{23}^{2m+1}\mathsf{N}_{\mathrm{PB}_3}, \quad T^{\mathrm{PB}_3}_{m,1}(c) = c^{2m+1}\mathsf{N}_{\mathrm{PB}_3}.$$

Since the orders of the elements $x_{12}\mathsf{N}_{\mathrm{PB}_3}$, $x_{23}\mathsf{N}_{\mathrm{PB}_3}$ and $c\mathsf{N}_{\mathrm{PB}_3}$ divide N_{ord} and $2m+1$ represents a unit in $\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$, all three cosets $x_{12}\mathsf{N}_{\mathrm{PB}_3}$, $x_{23}\mathsf{N}_{\mathrm{PB}_3}$ and $c\mathsf{N}_{\mathrm{PB}_3}$ belong to the image of $T^{\mathrm{PB}_3}_{m,1}$. Thus, due to Proposition 2.10, every element of X_{N} is a GT–shadow.

Furthermore, every GT–shadow in X_{N} is charming. The first condition of Definition 2.20 is clearly satisfied and the second one follows from the fact that $2m+1$ represents a unit in $\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$ and the orders of the elements $x_{12}\mathsf{N}_{\mathsf{F}_2}$, $x_{23}\mathsf{N}_{\mathsf{F}_2}$ divide N_{ord}.

Since the inclusion $\mathsf{GT}^{\heartsuit}(\mathsf{N}) \subset X_{\mathsf{N}}$ is obvious, the first statement of Theorem B.2 is proved.

Let us now show that every GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is genuine.

By Remark 2.18 and the surjectivity of the cyclotomic character, we know that, for every $\bar{\lambda} \in (\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z})^{\times}$ there should exist at least one genuine GT–shadow $[(m, f)] \in \mathsf{GT}^{\heartsuit}(\mathsf{N})$ such that

$$2\bar{m} + \bar{1} = \bar{\lambda}. \tag{B-6}$$

Let us assume that N_{ord} is odd. In this case $\bar{2} \in (\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z})^{\times}$ and hence, for every fixed $\bar{\lambda} \in (\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z})^{\times}$, equation (B-6) has exactly one solution $\bar{m} \in \mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z}$.

Since, for every $\bar{\lambda} \in (\mathbb{Z}/N_{\mathrm{ord}}\mathbb{Z})^{\times}$, we have exactly one GT–shadow $(\bar{m}, \bar{1})$ in $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ such that $2\bar{m}+1 = \bar{\lambda}$, the surjectivity of the cyclotomic character implies that every GT–shadow in $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is genuine.

The case when $N_{\mathrm{ord}} = 2k$ (for $k \in \mathbb{Z}_{\geq 1}$) requires more work. In this case, equation (B-6) has exactly two solutions for every $\bar{\lambda} \in \left(\mathbb{Z}/2k\mathbb{Z}\right)^{\times}$. More precisely, if $2\bar{m} + \bar{1} = \bar{\lambda}$ then the solution set for (B-6) is $\{\bar{m}, \bar{m} + \bar{k}\}$.

The proof of the desired statement about $\mathsf{GT}^{\heartsuit}(\mathsf{N})$ is based on the fact that the integers $2m+1$ and $2m+2k+1$ represent two distinct units in the ring $\mathbb{Z}/4k\mathbb{Z}$.

Let K be an element of $\mathsf{NFI}_{\mathrm{PB}_4}(\mathrm{B}_4)$ such that

- $\mathsf{K} \leq \mathsf{N}$,
- PB_4/K is abelian, and
- $4k$ divides $K_0 := |\mathrm{PB}_2 : \mathsf{K}_{\mathrm{PB}_2}|$.

One possible way to construct such K is to define a group homomorphism $\psi : \mathrm{PB}_4 \to S_{4k}$ by the formulas

$$\psi(x_{ij}) := (1, 2, \dots, 4k) \quad \text{for all } 1 \le i < j \le 4, \tag{B-7}$$

and set $\mathsf{K} := \mathsf{N} \cap \ker(\psi)$.

Since the natural group homomorphism

$$(\mathbb{Z}/K_0\mathbb{Z})^\times \to (\mathbb{Z}/4k\mathbb{Z})^\times$$

is onto, there exist $\overline{\lambda}_1 \neq \overline{\lambda}_2$ in $(\mathbb{Z}/K_0\mathbb{Z})^\times$ whose images in $(\mathbb{Z}/4k\mathbb{Z})^\times$ are the two distinct units represented by $2m+1$ and $2m+2k+1$, respectively.

Therefore there exist genuine GT–shadows $[(m_1, 1)]$ and $[(m_2, 1)]$ in $\mathsf{GT}^\heartsuit(\mathsf{K})$ such that

$$2m_1 + 1 \equiv \lambda_1 \bmod K_0 \quad \text{and} \quad 2m_2 + 1 \equiv \lambda_2 \bmod K_0.$$

Consequently, m_1 and m_2 satisfy these congruences mod $4k$:

$$2m_1 + 1 \equiv 2m + 1 \bmod 4k \quad \text{and} \quad 2m_2 + 1 \equiv 2m + 2k + 1 \bmod 4k.$$

Thus the images of the genuine GT–shadows $[(m_1, 1)]$ and $[(m_2, 1)]$ in $\mathsf{GT}(\mathsf{N})$ are $[(m, 1)]$ and $[(m+k, 1)]$. □

Remark B.3 In the abelian setting, every charming GT–shadow comes from an element of $G_{\mathbb{Q}}$. The authors *do not know* whether there is a genuine GT–shadow (in the nonabelian setting) that does not come from an element of $G_{\mathbb{Q}}$. Of course, if such a GT–shadow exists then the homomorphism (1-1) is not onto. (Some mathematicians believe that, in modern mathematics, there are no tools for tackling this question.)

References

[1] **D Bar-Natan**, *On associators and the Grothendieck–Teichmüller group, I*, Selecta Math. 4 (1998) 183–212 MR Zbl

[2] **D Bar-Natan**, **Z Dancso**, *Pentagon and hexagon equations following Furusho*, Proc. Amer. Math. Soc. 140 (2012) 1243–1250 MR Zbl

[3] **G V Belyĭ**, *On Galois extensions of a maximal cyclotomic field*, Izv. Akad. Nauk SSSR Ser. Mat. 43 (1979) 267–276 MR Zbl In Russian; translated in Math. USSR-Izv. 14 (1980) 247–256

[4] **V A Dolgushev**, *Documentation for the package GT*, software package documentation (2022) Available at `https://sites.temple.edu/vald/files/2024/05/PackageGT_README.pdf`

[5] **V A Dolgushev**, **A A Lorenz**, *On the profinite completion of a groupoid*, in preparation

[6] **V A Dolgushev**, **C L Rogers**, *Notes on algebraic operads, graph complexes, and Willwacher's construction*, from "Mathematical aspects of quantization" (S Evens, M Gekhtman, B C Hall, X Liu, C Polini, editors), Contemp. Math. 583, Amer. Math. Soc., Providence, RI (2012) 25–145 MR Zbl

[7] **V G Drinfeld**, *On quasitriangular quasi-Hopf algebras and on a group that is closely connected with* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Algebra i Analiz 2 (1990) 149–181 MR Zbl In Russian; translated in Leningrad Math. J. 2 (1991) 829–860

[8] **J S Ellenberg**, *Galois invariants of dessins d'enfants*, from "Arithmetic fundamental groups and noncommutative algebra" (M D Fried, Y Ihara, editors), Proc. Sympos. Pure Math. 70, Amer. Math. Soc., Providence, RI (2002) 27–42 MR Zbl

[9] **B Fresse**, *Homotopy of operads and Grothendieck–Teichmüller groups, I: The algebraic theory and its topological background*, Math. Surv. Monogr. 217, Amer. Math. Soc., Providence, RI (2017) MR Zbl

[10] **H Furusho**, *Pentagon and hexagon equations*, Ann. of Math. 171 (2010) 545–556 MR Zbl

[11] **A Grothendieck**, *Revêtements étales et groupe fondamental*, Lecture Notes in Math. 224, Springer (1971) MR Zbl

[12] **P Guillot**, *The Grothendieck–Teichmüller group of a finite group and G–dessins d'enfants*, from "Symmetries in graphs, maps, and polytopes", Springer Proc. Math. Stat. 159, Springer (2016) 159–191 MR Zbl

[13] **P Guillot**, *The Grothendieck–Teichmüller group of* $\mathrm{PSL}(2,q)$, J. Group Theory 21 (2018) 241–251 MR Zbl

[14] **D Harbater**, **L Schneps**, *Fundamental groups of moduli and the Grothendieck–Teichmüller group*, Trans. Amer. Math. Soc. 352 (2000) 3117–3148 MR Zbl

[15] **Y Ihara**, *Braids, Galois groups, and some arithmetic functions*, from "Proceedings of the International Congress of Mathematicians, I" (I Satake, editor), Math. Soc. Japan, Tokyo (1991) 99–120 MR Zbl

[16] **Y Ihara**, *On the embedding of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *into* $\widehat{\mathrm{GT}}$, from "The Grothendieck theory of dessins d'enfants" (L Schneps, editor), Lond. Math. Soc. Lect. Note Ser. 200, Cambridge Univ. Press (1994) 289–321 MR Zbl

[17] **C Kassel**, **V Turaev**, *Braid groups*, G.raduate Texts in Math. 247, Springer (2008) MR Zbl

[18] **P Lochak**, **L Schneps**, *A cohomological interpretation of the Grothendieck–Teichmüller group*, Invent. Math. 127 (1997) 571–600 MR Zbl

[19] **P Lochak**, **L Schneps**, *Open problems in Grothendieck–Teichmüller theory*, from "Problems on mapping class groups and related topics" (B Farb, editor), Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI (2006) 165–186 MR Zbl

[20] **J-L Loday**, **B Vallette**, *Algebraic operads*, Grundl. Math. Wissen. 346, Springer (2012) MR Zbl

[21] **M Markl**, **S Shnider**, **J Stasheff**, *Operads in algebra, topology and physics*, Math. Surv. Monogr. 96, Amer. Math. Soc., Providence, RI (2002) MR Zbl

[22] **H Nakamura**, **L Schneps**, *On a subgroup of the Grothendieck–Teichmüller group acting on the tower of profinite Teichmüller modular groups*, Invent. Math. 141 (2000) 503–560 MR Zbl

[23] **F Pop**, *Finite tripod variants of I/OM: on Ihara's question/Oda–Matsumoto conjecture*, Invent. Math. 216 (2019) 745–797 MR Zbl

[24] **L Ribes**, **P Zalesskii**, *Profinite groups*, Ergebnisse der Math. 40, Springer (2000) MR Zbl

[25] **L Schneps**, *An introduction to Grothendieck–Teichmüller theory*, slides of lectures (2012) Available at `https://webusers.imj-prg.fr/~leila.schneps/mit.html`

[26] **J Stasheff**, *What is . . . an operad?*, Notices Amer. Math. Soc. 51 (2004) 630–631 MR Zbl

[27] **T Szamuely**, *Galois groups and fundamental groups*, Cambridge Stud. Adv. Math. 117, Cambridge Univ. Press (2009) MR Zbl

[28] **D E Tamarkin**, *Formality of chain operad of little discs*, Lett. Math. Phys. 66 (2003) 65–72 MR Zbl

Department of Mathematics, Temple University
Philadelphia, PA, United States

Department of Mathematics, Rice University
Houston, TX, United States

Department of Mathematics, Vanderbilt University
Nashville, TN, United States

vald@temple.edu, khanh.le@rice.edu, aidan.lorenz@vanderbilt.edu

Received: 25 July 2022 Revised: 8 April 2023

Algebraic & Geometric Topology **24**:5 (2024) 2779–2785
DOI: 10.2140/agt.2024.24.2779
Published: 19 August 2024

A simple proof of the Crowell–Murasugi theorem

THOMAS KINDRED

We give an elementary, self-contained proof of the theorem, proven independently in 1958–1959 by Crowell and Murasugi, that the genus of any oriented nonsplit alternating link equals half the breadth of its Alexander polynomial (with a correction term for the number of link components), and that applying Seifert's algorithm to any oriented connected alternating link diagram gives a surface of minimal genus.

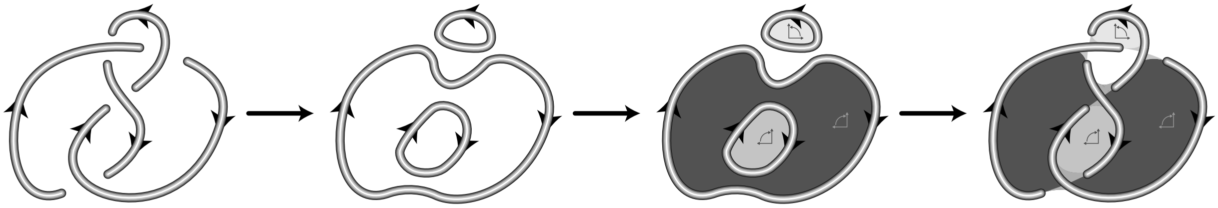
57K10, 57K14

Every oriented link $K \subset S^3$ bounds a connected oriented surface F called a *Seifert surface*. Such F is homeomorphic to an ℓ–punctured surface of some genus $g(F)$, where $\ell = |K|$ (here and throughout, bars count components). The *link genus* $g(K)$ is the minimum genus among all Seifert surfaces for K.

An ordered basis $(a_1, \dots, a_n)$ for $H_1(F)$ determines an $n \times n$ *Seifert matrix* $V = (v_{ij})$, $v_{ij} = \mathrm{lk}(a_i, a_j^+)$, where lk denotes linking number and a_j^+ is the pushoff of (an oriented multicurve representing) a_j in the positive normal direction determined by the orientations on F and S^3.

The polynomial $\det(V - tV^T)$, denoted by $\Delta_K(t)$, is called the *Alexander polynomial* of K. Up to degree shift, it is independent of Seifert surface and basis; see Kauffman [11] and Bar-Natan, Fulman and Kauffman [2]. Writing $\Delta_K(t) = a_r t^r + a_{r+1} t^{r+1} + \cdots + a_{s-1} t^{s-1} + a_s t^s$ with $a_r, a_s \neq 0$, we call $s - r$ the *breadth* of $\Delta_K(t)$ and denote it by $\mathrm{bth}(K)$.

Given any oriented connected diagram $D \subset S^2$ of a link $K \subset S^3$, *Seifert's algorithm* yields a Seifert surface for K as follows. First, "smooth" each crossing of D in the way that respects orientation: , . This gives a disjoint union of oriented circles on S^2 called the *Seifert state* of D; each circle is called a *Seifert circle*. Second, cap all the Seifert circles with disjoint, oriented disks, all on the same side of S^2. Third, attach an oriented half-twisted band at each crossing, so that the resulting surface F is oriented with $\partial F = K$, respecting orientation. Here is an example:



The purpose of this note is to give a short, elementary, self-contained proof of the following theorem, first proven independently in 1958–1959 by Crowell and Murasugi.

Theorem 1 (Crowell [4] and Murasugi [14; 15]) *If F is a surface constructed via Seifert's algorithm from a connected alternating diagram D of an oriented ℓ–component link K, then*

$$g(F) = g(K) = \tfrac{1}{2}(\mathrm{bth}(K) + 1 - \ell).$$

To prove Theorem 1, we will show that a Seifert matrix V for F is invertible. The next two results show that this indeed will suffice.

Proposition 2 *Let F be a Seifert surface for an oriented ℓ–component link K. If $\mathrm{bth}(K) = 2g(F) + \ell - 1$, then $g(K) = g(F) = \frac{1}{2}(\mathrm{bth}(K) + 1 - \ell)$.*

Proof Given an arbitrary Seifert surface F' for K, one may compute $\Delta_K(t)$ from any Seifert matrix for F', so $\mathrm{bth}(K) \le \beta_1(F') = 2g(F') + 1 - \ell$. Hence $g(F) \le g(F')$. □

Proposition 3 (Murasugi [17]) *Let V be a real $n \times n$ matrix, and let $f(t) = \det(V - tV^T)$. If V is invertible, then the breadth of $f(t)$ equals n.*

Proof Denoting the transpose of V^{-1} by V^{-T},

$$f(t) = \det(V^T)\det(VV^{-T} - tI)$$

is a nonzero scalar multiple of the characteristic polynomial of the invertible matrix VV^{-T}, hence has breadth n.[1] □

Next, suppose that $D \subset S^2$ is a connected oriented alternating link diagram such that applying Seifert's algorithm to D yields a *checkerboard*[2] surface F.[3] Then, since D is alternating and connected, all of the crossing bands in F are identical: either they all positive, , or they are all negative, . Let V denote a Seifert matrix for F.

Lemma 4 *With the preceding setup, if the crossing bands in F are positive, then any nonzero $\boldsymbol{x} \in \mathbb{Z}^{\beta_1(F)}$ satisfies $\boldsymbol{x}^T V \boldsymbol{x} > 0$; if the crossing bands in F are negative, then any such $\boldsymbol{x}$ satisfies $\boldsymbol{x}^T V \boldsymbol{x} < 0$. Hence, in either case, V is invertible.*

Here is a self-contained proof. A shorter argument, using Greene [9], follows.

Proof Assume without loss of generality that the crossing bands in F are positive. Among all oriented multicurves in F that represent $\boldsymbol{x}$, choose one, α, that intersects the crossing bands in F in the smallest

[1] The converse is also true. Indeed, if V is singular, then choose an invertible matrix P whose first column is in the nullspace of V. Then $\det(P^TVP - t(P^TVP)^T) = \det^2(P) \cdot f(t)$ has the same breadth as $f(t)$. Further, the first column of P^TVP is $\mathbf{0}$, so only constants appear in the first row of $P^TVP - t(P^TVP)^T$. Hence, the breadth is less than n.

[2] That is, each Seifert circle bounds a disk in S^2 disjoint from the other Seifert circles.

[3] Such a diagram is either *positive* or *negative* and is called *special alternating*.

possible number of components. Then, for each crossing band X in F, $\alpha \cap X$ will consist of a (possibly empty) collection of coherently oriented arcs. Therefore,

$$\boldsymbol{x}^T V \boldsymbol{x} = \operatorname{lk}(\alpha, \alpha^+) = \sum_{\text{crossing bands } X} \frac{|\alpha \cap X|^2}{2} \geq 0. \tag{1}$$

Moreover, the inequality in (1) is strict, or else α would be disjoint from all crossing bands, hence nullhomologous (since D is connected). It follows that V is nonsingular, or else we would have $V\boldsymbol{z} = \boldsymbol{0}$ for some nonzero vector $\boldsymbol{z}$, giving $\boldsymbol{z}^T V \boldsymbol{z} = 0$. □

Alternatively, denote the Gordon–Litherland pairing [8] on F by $\langle \cdot, \cdot \rangle$. Since D is alternating and connected, this pairing is definite; see Greene [9] or Murasugi [16]. Thus,

$$\boldsymbol{x}^T V \boldsymbol{x} = \operatorname{lk}(\alpha, \alpha^+) = \tfrac{1}{2} \operatorname{lk}(\alpha, \alpha_+ \cup \alpha_-) = \langle \boldsymbol{x}, \boldsymbol{x} \rangle \neq 0.$$

To complete the proof of Theorem 1, we need one more definition and lemma. The Murasugi sum, also called generalized plumbing, is a way of gluing together two spanning surfaces along a disk so as to produce another spanning surface. We will prove that if Seifert surfaces F_1 and F_2 have invertible Seifert matrices, then any Murasugi sum of F_1 and F_2 also has invertible Seifert matrix (and conversely).

Definition 5 For $i = 1, 2$, let F_i be a Seifert surface in a 3–sphere S^3_i, and choose a compact 3–ball $B_i \subset S^3_i$ that contains F_i such that

(i) $F_i \cap \partial B_i$ is a disk U_i whose boundary consists alternately of arcs in ∂F_i and arcs in $\operatorname{int}(F_i)$,

(ii) $|\partial U_1 \cap \partial F_1| = |\partial U_2 \cap \partial F_2|$, and

(iii) the positive normal along U_1 (using the orientations on S^3_1 and F_1) points *into* B_1, whereas the positive normal along U_2 points *out of* B_2.

Choose an orientation-reversing homeomorphism $h : \partial B_1 \to \partial B_2$ such that $h(U_1) = U_2$ and $h(\partial U_1 \cap \partial F_1) = \operatorname{cl}(\partial U_2 \cap \operatorname{int}(F_2))$.[4] Then $F = F_1 \cup_h F_2$ is a Seifert surface in the 3–sphere $B_1 \cup_h B_2$. It is a *Murasugi sum* or *generalized plumbing* of F_1 and F_2, denoted by $F = F_1 * F_2$.

Note that there are generally many ways to form a Murasugi sum between two given surfaces. As an aside, we mention that the Murasugi sum construction extends easily to unoriented surfaces, and that both the oriented and unoriented notions of Murasugi sum are natural operations in many respects; see Gabai [5; 6], Ozawa [18], Ozbagci and Popescu-Pampu [19] and Kindred [12]. Here is one such respect:

Lemma 6 *Given a Murasugi sum $F = F_1 * F_2$ of Seifert surfaces with Seifert matrices V, V_1 and V_2, respectively, V is invertible if and only if both V_1 and V_2 are.*

[4] It follows that $h(\operatorname{cl}(\partial U_1 \cap \operatorname{int}(F_1))) = \partial U_2 \cap \partial F_2$.

Proof Write $V=(v_{ij})$. We may assume that V is taken with respect to a basis $(a_1,\dots,a_r,b_1,\dots,b_s)$ for $H_1(F)$, where $(a_1,\dots,a_r)$ is a basis for $H_1(F_1)$ and $(b_1,\dots,b_s)$ is a basis for $H_1(F_2)$. Then V is a block matrix of the form $V=\begin{bmatrix} V_1 & A \\ B & V_2 \end{bmatrix}$. In fact, we claim that $B=0$, ie

$$V=\begin{bmatrix} V_1 & A \\ 0 & V_2 \end{bmatrix}. \tag{2}$$

To see this, let $\alpha_j \subset F_1$ represent a_j and let $\beta_i \subset F_2$ represent b_i for arbitrary $1\le j\le r$ and $1\le i\le s$. Then $v_{ij}=\operatorname{lk}(\beta_i,\alpha_j^+)=0$ because, using the notation and setup from Definition 5, $\alpha_j^+\subset \operatorname{int}(h(B_1))$ and $\beta_i\subset B_2$. From (2), we have $\det(V)=\det(V_1)\det(V_2)$,[5] so the result follows. □

Now we can prove Theorem 1:

Proof of Theorem 1 Let F be a surface constructed via Seifert's algorithm from an alternating diagram D of an oriented link K. Then F is a Murasugi sum of checkerboard Seifert surfaces from connected oriented alternating link diagrams.[6]

Lemma 4 implies that all of these checkerboard surfaces have invertible Seifert matrices, so Lemma 6 implies that F has an invertible Seifert matrix V. Since K has ℓ components, the size of V is $\beta_1(F)=2g(F)+1-\ell$. Thus, by Propositions 2 and 3,

$$g(F)=g(K)=\tfrac{1}{2}(\operatorname{bth}(K)+1-\ell). \qquad \square$$

The preceding proof shows, more generally:

Theorem 7 *Let F be a Seifert surface for an oriented ℓ–component link K. If F is a Murasugi sum of checkerboard surfaces from connected oriented alternating link diagrams, then $g(K)=g(F)=\frac{1}{2}(\operatorname{bth}(K)+1-\ell)$.*

In particular, an oriented connected link diagram is called *homogeneous* if it is a $*$*–product*, ie *diagrammatic Murasugi sum*, of special alternating link diagrams. By definition, Theorem 7 applies to all such diagrams (cf [3] Corollary 4.1):

Corollary 8 *If F is constructed via Seifert's algorithm from a* **homogeneous** *diagram of an ℓ–component oriented link K, then $g(F)=g(K)=\frac{1}{2}(\operatorname{bth}(K)+1-\ell)$.*

We note another consequence of Lemma 6, in combination with:

Theorem 9 (Harer's conjecture [10]; Corollary 3 of [7]) *Any fiber surface in S^3 can be constructed by plumbing and de-plumbing Hopf bands.*

[5] This is due to the formula $\det V=\sum_{\sigma\in S_{r+s}}\operatorname{sign}(\sigma)\prod_{i=1}^{r+s} v_{i\sigma(i)}$ and the pigeonhole principle.

[6] Indeed, D is a $*$*–product* of special alternating diagrams: see [1, Definition 2.37 and Remark 2.38]. For an explicit construction, see [20, page 98].

Corollary 10 *If F is an oriented fiber surface spanning an ℓ–component link $K \subset S^3$ and K has the boundary orientation from F, then*

$$g(F) = g(K) = \tfrac{1}{2}(\mathrm{bth}(K) + 1 - \ell).$$

We close by considering knots K with $g(K) > \frac{1}{2}\,\mathrm{bth}(K)$. The simplest such knots have 11 crossings. There are seven of them [13]: the Conway knot $11n34$ has genus three, as do $11n45$, $11n73$ and $11n152$, while the Kinoshita–Terasaka knot $11n42$ has genus two, as do $11n67$ and $11n97$. Lemma 6 implies that if one takes a minimal genus Seifert surface for any one of these knots and de-plumbs (ie decomposes it as a nontrivial Murasugi sum),[7] then at least one of the resulting factors will have a singular Seifert matrix. Also, by Theorem 1 of [6], all of these surfaces will have minimal genus. This raises the following natural problem.

Problem 11 Characterize or tabulate those Seifert surfaces F which

(i) have minimal genus,

(ii) do not de-plumb,[8] and

(iii) have singular Seifert matrices.

Interestingly, for each of the four aforementioned 11–crossing knots of genus three, de-plumbing a minimal genus Seifert surface gives three Hopf bands and the planar pretzel surface $P_{2,2,-2,-2}$, which has Seifert matrix

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix},$$

and doing this for any of the three aforementioned 11–crossing knots of genus two gives one Hopf band and a surface of genus one that has Seifert matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -2 \\ 0 & -1 & 0 \end{bmatrix}.$$

See Figure 1. Another simple example of the type of surface referenced in Problem 11 is the planar pretzel surface $P_{4,4,-2}$, which has Seifert matrix

$$\begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}.$$

In particular, each of these simplest examples spans a link of multiple components.

Question 12 Does there exist a *knot* K that satisfies $g(K) > \frac{1}{2}\,\mathrm{bth}(K)$ and has a minimal genus Seifert surface F that does not de-plumb?

[7] Beware: surfaces may admit distinct de-plumbings; see Kindred [12]. Still, Lemma 6 implies that this sentence is true for *any* de-plumbing of such a surface.

[8] That is, any decomposition of F as a Murasugi sum $F = F_1 * F_2$ has F_1 or F_2 as a disk.

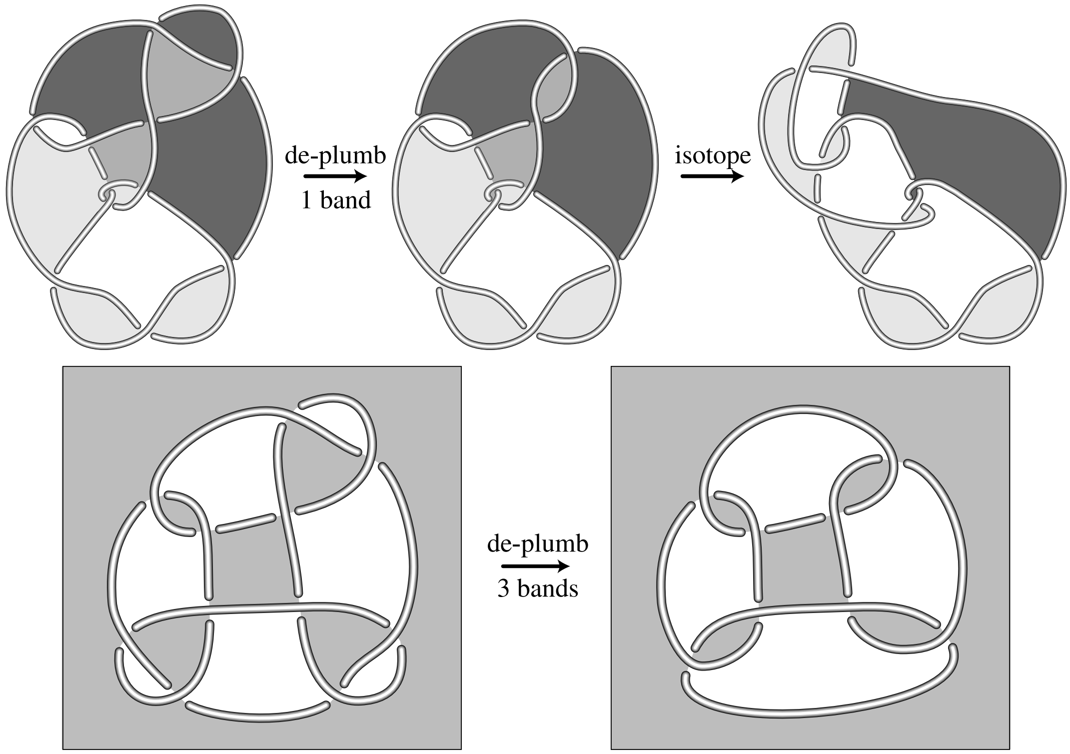


Figure 1: De-plumbing Hopf bands from minimal genus Seifert surfaces for the knots $11n67$ and $11n73$.

Acknowledgements Thank you to the students in topics courses I taught at University of Nebraska–Lincoln and Wake Forest University, especially John Tolbert, for engaging with these arguments and ideas.

References

[1] **J E Banks**, *Homogeneous links, Seifert surfaces, digraphs and the reduced Alexander polynomial*, Geom. Dedicata 166 (2013) 67–98 MR Zbl

[2] **D Bar-Natan**, **J Fulman**, **L H Kauffman**, *An elementary proof that all spanning surfaces of a link are tube-equivalent*, J. Knot Theory Ramifications 7 (1998) 873–879 MR Zbl

[3] **P R Cromwell**, *Homogeneous links*, J. Lond. Math. Soc. 39 (1989) 535–552 MR Zbl

[4] **R Crowell**, *Genus of alternating link types*, Ann. of Math. 69 (1959) 258–275 MR Zbl

[5] **D Gabai**, *The Murasugi sum is a natural geometric operation*, from "Low-dimensional topology" (S J Lomonaco, Jr, editor), Contemp. Math. 20, Amer. Math. Soc., Providence, RI (1983) 131–143 MR Zbl

[6] **D Gabai**, *The Murasugi sum is a natural geometric operation, II*, from "Combinatorial methods in topology and algebraic geometry" (J R Harper, R Mandelbaum, editors), Contemp. Math. 44, Amer. Math. Soc., Providence, RI (1985) 93–100 MR Zbl

[7] **E Giroux**, **N Goodman**, *On the stable equivalence of open books in three-manifolds*, Geom. Topol. 10 (2006) 97–114 MR Zbl

[8] **C M Gordon**, **R A Litherland**, *On the signature of a link*, Invent. Math. 47 (1978) 53–69 MR Zbl

[9] **J E Greene**, *Alternating links and definite surfaces*, Duke Math. J. 166 (2017) 2133–2151 MR Zbl

[10] **J Harer**, *How to construct all fibered knots and links*, Topology 21 (1982) 263–280 MR Zbl

[11] **L H Kauffman**, *The Conway polynomial*, Topology 20 (1981) 101–108 MR Zbl

[12] **T Kindred**, *Checkerboard plumbings*, PhD thesis, University of Iowa (2018) Available at `https://doi.org/10.17077/etd.yv1nlfcz`

[13] **C Livingston**, **A H Moore**, *KnotInfo: table of knot invariants*, online database (2024) Available at `knotinfo.math.indiana.edu`

[14] **K Murasugi**, *On the genus of the alternating knot, I*, J. Math. Soc. Japan 10 (1958) 94–105 MR Zbl

[15] **K Murasugi**, *On the genus of the alternating knot, II*, J. Math. Soc. Japan 10 (1958) 235–248 MR Zbl

[16] **K Murasugi**, *Jones polynomials and classical conjectures in knot theory, II*, Math. Proc. Cambridge Philos. Soc. 102 (1987) 317–318 MR Zbl

[17] **K Murasugi**, *Knot theory and its applications*, Birkhäuser, Boston, MA (1996) MR Zbl

[18] **M Ozawa**, *Essential state surfaces for knots and links*, J. Aust. Math. Soc. 91 (2011) 391–404 MR Zbl

[19] **B Ozbagci**, **P Popescu-Pampu**, *Generalized plumbings and Murasugi sums*, Arnold Math. J. 2 (2016) 69–119 MR Zbl

[20] **C V Quach Hongler**, **C Weber**, *On the topological invariance of Murasugi special components of an alternating link*, Math. Proc. Cambridge Philos. Soc. 137 (2004) 95–108 MR Zbl

Department of Mathematics, Wake Forest University
Winston-Salem, NC, United States

`kindret@wfu.edu`

`www.thomaskindred.com`

Received: 10 October 2022 Revised: 9 February 2023

Algebraic & Geometric Topology **24**:5 (2024) 2787–2805
DOI: 10.2140/agt.2024.24.2787
Published: 19 August 2024

The Burau representation and shapes of polyhedra

ETHAN DLUGIE

We use a geometric approach to show that the reduced Burau representation specialized at roots of unity has another incarnation as the monodromy representation of a moduli space of Euclidean cone metrics on the sphere, as described by Thurston. Using the theory of orbifolds, we leverage this connection to identify the kernels of these specializations in some cases, partially addressing a conjecture of Squier. The 4–strand case is the last case where the faithfulness question for the Burau representation is unknown, a question that is related eg to the question of whether the Jones polynomial detects the unknot. Our results allow us to place the kernel of this representation in the intersection of several topologically natural subgroups of B_4.

20F36, 57K20; 57R18

1 Introduction

We consider two representations of groups arising in low-dimensional topology. First is the (reduced) Burau representation of braid groups

$$\beta_n\colon B_n \to \mathrm{GL}_{n-1}(\mathbb{Z}[t^{\pm}])$$

that has been studied for almost a century [5]. Second is a monodromy representation of punctured sphere mapping class groups coming from a geometric structure on the moduli space of Euclidean cone spheres,

$$\rho_{\vec{k}}\colon \mathrm{Mod}(S_{0,m};\vec{k}) \to \mathrm{PU}(1,m-3),$$

as described by Thurston in [22]. It has been found using algebraic techniques that these seemingly disparate representations are quite closely related in that the latter is, in a sense, a specialization of the former; see McMullen [18] and Venkataramana [24]. Our first theorem is a slight rephrasing of those results, which we will establish via geometric means. See the beginning of Section 2 for an introduction to the terminology of Euclidean cone metrics used in the following statement.

Theorem 1.1 (the Burau representation and polyhedra monodromy) *Fix a choice of curvatures $\vec{k}$, which is to say a tuple of real numbers $\vec{k}=(k_1,\dots,k_m)$ with each $0<k_i<2\pi$ and $\sum_{i=1}^m k_i = 4\pi$. Suppose further that n of these curvatures are equal, say $k_1=\cdots=k_n$ with $n\le m-1$, and write $k_*\in(0,2\pi)$ for this common value. Set $q=\exp(i(\pi-k_*))$. Then the diagram*

$$\begin{array}{ccc} B_n & \xrightarrow{\quad\beta_n\quad} & \beta_n(B_n)\subset \mathrm{GL}_{n-1}(\mathbb{Z}[t^{\pm}]) \\ \downarrow{\iota_*} & & \downarrow{\mathrm{ev}(-q)} \\ \mathrm{Mod}(S_{0,m};\vec{k}) & \xrightarrow{\rho_{\vec{k}}} & \mathrm{PU}(1,m-3)\subset \mathrm{PGL}_{m-2}(\mathbb{C}) \end{array}$$

commutes, where β_n is the n–strand (reduced) Burau representation, $\rho_{\vec{k}}$ is the monodromy representation of the moduli space of cone metrics, ι_* is the map on mapping class groups induced by an inclusion of an n–times marked disk $\iota\colon D_n \hookrightarrow S_{0,m}$, and $\mathrm{ev}(-q)$ is a slight alteration of an evaluation map to be defined in Definition 3.3.

In the case where $m = n+1$ and D_n is included into an $(n+1)$–times punctured sphere, the evaluation map mentioned in this theorem really is just an evaluation. This allows us to realize the "specialized" Burau representation $\beta(-q)$, where the formal variable t is evaluated at a given unit complex number $-q$, as one of these polyhedral monodromy representations.

Corollary 1.2 *Let $\vec{k} = (k_1, \dots, k_{n+1})$ be as in Theorem 1.1 with $k_* = k_1 = \cdots = k_n$. Write $q = \exp(i(\pi - k_*))$ Then the diagram*

$$\begin{array}{ccc} B_n & \xrightarrow{\quad\beta(-q)\quad} & \mathrm{GL}_{n-1}(\mathbb{C}) \\ {\scriptstyle\iota_*}\downarrow & & \downarrow \\ \mathrm{Mod}(S_{0,n+1};\vec{k}) & \xrightarrow{\rho_{\vec{k}}} & \mathrm{PU}(1,n-2) \subset \mathrm{PGL}_{n-1}(\mathbb{C}) \end{array}$$

commutes, where $\mathrm{Mod}(S_{0,n+1};\vec{k})$ is the subgroup of the mapping class group of the $(n+1)$–times punctured sphere that preserves the $(n+1)^{st}$ point and may freely permute the other points.

This yields a containment $\ker(\beta(-q)) \le \mathrm{ncl}_{B_n}(\widetilde{S}) \cdot \langle \tau_n \rangle$, where $\widetilde{S}$ is a lift of a normal generating set for $\ker(\rho_{\vec{k}})$ and $\tau_n \in B_n$ is the full-twist braid on n–strands that generates the center of the braid group.

In the statement, and in the rest of the paper, the notation $\mathrm{ncl}_G(S)$ indicates the normal closure of a set S inside of a group G. We will also write $\tau_p \in B_n$ for a full twist about a curve surrounding p points in the n–punctured disk. Any two such twists are conjugate in the braid group.

In his influential paper [21], Squier briefly considered the specializations of the Burau representation at roots of unity. He made a conjecture about the form that the kernels of such specializations would take. We cannot verify Squier's conjecture in the form that he stated it,[1] but using Corollary 1.2 we are able to identify the kernel of these specializations in several cases.

Theorem 1.3 (Burau at roots of unity) *Suppose q is a primitive d^{th} root of unity and denote by $\beta(-q)\colon B_n \to \mathrm{GL}_{n-1}(\mathbb{C})$ the specialization of the Burau representation at $t = -q$. Then we have*

$$\ker(\beta(-q)) = \mathrm{ncl}_{B_n}(\sigma^d, \tau_{n-1}^j) \cdot \langle \tau_n^\ell \rangle \tag{1}$$

for the following values of n, d, j, and ℓ:

n	4								5				6		7		8	9	10
d	5	6	7	8	9	10	12	18	4	5	6	8	4	5	3	4	3	3	3
j	∞	∞	14	8	6	5	4	3	∞	5	3	2	4	2	∞	2	6	3	2
ℓ	5	3	7	4	9	5	3	9	4	2	3	8	2	5	6	4	3	2	3

[1] See Section 6 for a discussion on this point.

Here $\sigma \in B_n$ denotes one of the half-twist generators of B_n (all of which are conjugate), $\tau_{n-1} \in B_n$ denotes a full twist on a curve surrounding $n-1$ points in the punctured disk (all of which are conjugate), and $\tau_n \in B_n$ denotes the full twist on the boundary of the punctured disk (which generates the center of B_n). In a case with $j = \infty$, we mean that the kernel is $\mathrm{ncl}(\sigma^d) \cdot \langle \tau_n^\ell \rangle$ with no power of τ_{n-1}.

We can also use this method to identify the kernel of $\beta(-q)$ in all cases with $n = 3$ and $d \geq 7$. The result is given in Theorem 5.4 and corrects the statement of Funar and Kohno [13, Theorem 1.2].

Whether or not the Burau representation is faithful is a natural question to ask. At present, the answer is unknown only in the $n = 4$ case, and this question has direct connections to the question of whether the Jones polynomial detects the unknot; see Bigelow [3] and Ito [14]. An element of the kernel of β_4 must also lie in the kernel of every specialization. Thus Theorem 1.3 as a direct corollary restricts the kernel of β_4 to live in the intersection of several topologically natural normal subgroups of the braid group. One should note however that the intersection of these finitely many subgroups is still nontrivial by Long [16, Lemma 2.1], so this alone is not enough to establish faithfulness.

Corollary 1.4 (narrowing $\ker(\beta_4)$) *Let $\beta_4 \colon B_4 \to \mathrm{GL}_3(\mathbb{Z}[t^{\pm}])$ denote the reduced Burau representation of the 4–strand braid group. Then*

$$\ker(\beta_4) \leq \mathrm{ncl}_{B_4}(\sigma^d, \tau_3^\ell) \cdot \langle \tau_4^\ell \rangle$$

for powers d, j, and ℓ as indicated in the table in Theorem 1.3. All eight of these normal subgroups have infinite index in B_4.

In fact all of the normal subgroups of braid groups given by Theorem 1.3 have infinite index in their respective braid groups. We comment on the relationship between this and some remarkable work of Coxeter [9] in Section 6.

Some history and context The question of the faithfulness of the Burau representation has persisted since the representation was first defined nearly a century ago; see Burau [5]. Faithfulness is easily shown for $n = 2$, and 3 (see eg Birman [4, Theorem 3.15]). Faithfulness for other cases remained open for several decades. Squier put forth two conjectures [21, (C1) and (C2)] that, if both true, would yield the faithfulness of the Burau representation. However, Moody found the Burau representation to be nonfaithful for $n \geq 10$ [19], and this result was quickly lowered to $n \geq 6$ by Long and Paton [17]. A few years later, Bigelow found a simpler example of an element in the kernel of β_6 and furthermore found that the Burau representation is not faithful for $n = 5$ [1]. Funar and Kohno proved Squier's conjecture (C2) in [13], so we know, for all $n \geq 5$, that Squier's conjecture (C1) is false for almost all (even) values of d. As of the writing of this article, the faithfulness question is only open in the $n = 4$ case.

Braid groups are already known to be linear by another representation, the Lawrence–Krammer representation. See Krammer [15] for an algebraic treatment of this result and Bigelow [2] for a topological proof. Yet the faithfulness of the Burau representation, especially in the $n = 4$ case, is still of interest due to its connection with the Jones polynomial in knot theory. Nonfaithfulness of β_4 implies that the Jones polynomial fails to detect the unknot; see Bigelow [3] and Ito [14]. There has been work on the $n = 4$ question in the last few decades. For instance, a computer search by Fullarton and Shadrach shows that a nontrivial element in the kernel of β_4 would have to be exceedingly complicated [12], suggesting faithfulness. On the other hand, Cooper and Long found that β_4 is not faithful when taken with coefficients mod 2 and with coefficients mod 3 [7; 8].

Thurston's work in [22] was a geometric reframing of the monodromy of hypergeometric functions considered by Deligne and Mostow in [10]. The algebrogeometric approach to studying these monodromy representations has continued, notably in work of McMullen [18] and Venkataramana [24]. The analysis of Euclidean cone metrics on surfaces was extended by Veech [23] and is still today an active area of research in low-dimensional topology and dynamical systems.

Organization The rest of the paper is organized as follows:

- Section 2 introduces Euclidean cone metrics on the sphere, and we construct explicit complex projective coordinates on the moduli space.

- In Section 3, we prove Theorem 1.1 and Corollary 1.2, which allow us to relate the Burau representation at roots of unity with the monodromy representation of the moduli spaces of Euclidean cone metrics. Our proof uses the complex projective coordinates defined in Section 2.

- In Section 4, we gather several results about the complex hyperbolic geometry of the moduli space and facts about geometric orbifolds.

- In Section 5, we prove Theorem 1.3, identifying the kernel of the Burau representation at some roots of unity. This uses Corollary 1.2 with the results of Section 4. We also present the application of these ideas to the β_3 case in Section 5.2.

- Section 6 contains a discussion of limitations of this work and several possible future directions and connections that I hope can spark further research with these techniques.

Acknowledgements Substantial thanks to my advisor, Ian Agol, who first informed me about the connection between the Burau representation and the polyhedra monodromy of Thurston's work. Special thanks as well to both Nancy Scherich and Sam Freedman for repeated helpful conversations through early drafts about the framing and phrasing of these results. Thanks to many others for taking the time to read and comment on a preprint of this work. Thanks to Louis Funar and Toshitake Kohno for graciously accepting a correction to one of their previous statements. And finally, I thank the referee for a speedy review of and many influential comments on my initial submission. In particular, the referee taught me about the conjectures of Squier, which have come to be a focal point of this work in its current form.

The author's work was supported in part by a grant from the Simons Foundation (Ian Agol, 376200).

2 Euclidean cone metrics on S^2

Here we recall the moduli space of Euclidean cone metrics on the sphere. We describe local coordinates on the moduli space into complex projective space. The construction is used in the proof of Theorem 1.1 in Section 3.

Following [22], we consider Euclidean cone metrics on the sphere. Such a metric is flat everywhere on the sphere away from some number of singular cone points $b_1, \dots, b_m$. Around each cone point b_i one sees some cone angle not equal to the usual 2π that one finds around a smooth point. Define the *curvature* k_i at b_i to be the angular defect of the cone point. The Gauss–Bonnet theorem applies with this notion of curvature to give $\sum_{i=1}^{m} k_i = 4\pi$.

Thurston considers only those cone metrics which are nonnegatively curved,[2] ie all $k_i > 0$. Fixing a tuple of positive real numbers $\vec{k} = (k_1, \dots, k_m)$ with $0 < k_i < 2\pi$ and $\sum_{i=1}^{m} k_i = 4\pi$, Thurston considers the moduli space of Euclidean cone metrics on the sphere with curvatures $\vec{k}$ up to orientation-preserving similarity. We denote this space by $\mathcal{M}(\vec{k})$.

There is a natural map from $\mathcal{M}(\vec{k})$ to (a finite cover of) the usual moduli space of conformal structures on the punctured sphere by simply taking the conformal class of a flat cone metric. There is also an inverse map inspired by the Schwarz–Christoffel mapping of complex analysis. Thurston uses this idea to show that the moduli space $\mathcal{M}(\vec{k})$ is actually orbifold-isomorphic to the moduli space of conformal structures on m–punctured spheres with punctures labeled by the k_i [22, Proposition 8.1]. This more classical moduli space is a complex orbifold of dimension $m-3$. The orbifold fundamental group of the moduli space is $\mathrm{Mod}(S_{0,m}; \vec{k})$, the group of mapping classes of the m–punctured sphere that preserve the labeling by curvatures.

Thurston shows directly that the moduli space of cone metrics has complex dimension $m-3$ by giving local $\mathbb{CP}^{m-3}$ coordinates on $\mathcal{M}(\vec{k})$ in terms of cocycles on the sphere with twisted/local coefficients. Schwartz gave a more geometric interpretation of these coordinates in [20]. In brief, we have the following:

Lemma 2.1 [20] *The moduli space $\mathcal{M}(\vec{k})$ of Euclidean cone metrics on the sphere with m cone points of curvatures $k_1, \dots, k_m$ is a complex projective orbifold of dimension $m-3$.*

Taking a similar approach to Schwartz, we describe more concrete complex projective coordinates. This specific construction is for the sake of our calculation in the proof of Theorem 1.1.

2.1 Local $\mathbb{CP}^{m-3}$ coordinates

A $\mathbb{CP}^{m-3}$ structure on moduli space is encoded by a developing map

$$\mathrm{dev}\colon \mathcal{T}(\vec{k}) \to \mathbb{CP}^{m-3}.$$

[2] A theorem of Alexandrov implies that every such metric arises uniquely as the intrinsic length metric on the boundary of a convex polyhedron in Euclidean space.

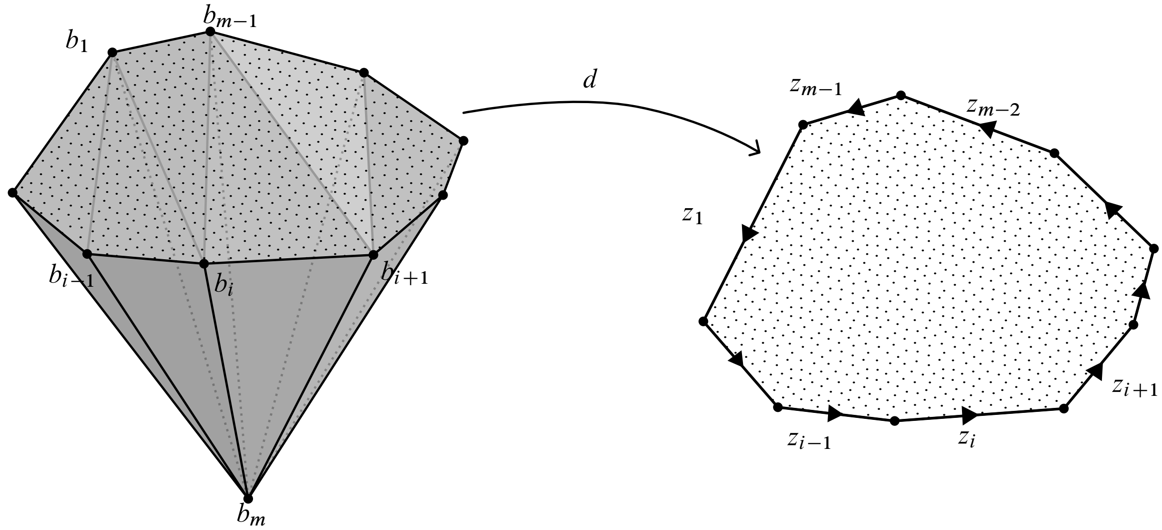


Figure 1: A marked Euclidean cone metric on the sphere and the complex coordinates measured from the boundary of the developing image of a flat disk in the sphere.

Here $\mathcal{T}(\vec{k})$ is the Teichmüller space of Euclidean cone metrics on the sphere up to scaling with curvatures $\vec{k}$, which is the universal cover of the moduli space $\mathcal{M}(\vec{k})$. We describe the map dev in a few steps.

- A *point in Teichmüller space* is a flat cone sphere X and an isotopy class of homeomorphisms $f: S_{0,m} \to X$ from the m–times marked sphere $S_{0,m} = (S^2, \{b_1, \dots, b_m\})$ such that $f(b_i) \in X$ is a cone point of curvature k_i. The isotopy is taken relative to the cone points, and the metric on X is only considered up to scaling.

- A *Euclidean developing* arises as follows. There is a disk $D \subset S_{0,m}$ such that

$$D \cap \{b_1, \dots, b_m\} = \partial D \cap \{b_1, \dots, b_m\} = \{b_1, \dots, b_{m-1}\},$$

and b_1 through b_{m-1} are arranged counterclockwise in order around ∂D. Abusing notation, we also write $b_1, \dots, b_m$ for the images of these points in X. Since $f(D) \subset X$ has no cone points in its interior and is simply connected, it has an isometric immersion (a developing map) to the Euclidean plane $d: f(D) \to \mathbb{E}^2$. This map is only defined up to orientation-preserving planar isometries.

- *Complex coordinates* come from the positions of the cone points in $\mathbb{E}^2 \approx \mathbb{C}$, namely the tuple

$$(d(b_1), \dots, d(b_{m-1})) \in \mathbb{C}^{m-1}.$$

These points are well defined up to isotopy of f, since the isotopy is relative to the cone points. If we instead record the differences $z_i = d(b_{i+1}) - d(b_i)$, $i = 1, \dots, m-2$, then we get a point $(z_1, \dots, z_{m-2}) \in \mathbb{C}^{m-2}$ that is well defined up to a translation of d. The last coordinate $z_{m-1} = d(b_1) - d(b_{m-1})$ can be recovered from the rest; see Figure 1.

- Finally, *projectivization* is required because we can modify d further by a rotation or real scaling (X was only defined up to scaling). This means we have to mod out by the action of $\mathbb{C}^\times$, which is to say we get a well-defined point $[z_1 : \cdots : z_{m-2}] \in \mathbb{CP}^{m-3}$.

All of the choices in the construction are accounted for, and we get a well-defined assignment $(X, [f]) \mapsto [z_1 : \cdots : z_{m-2}]$. This defines $\mathrm{dev}\colon \mathcal{T}(\vec{k}) \to \mathbb{CP}^{m-3}$ and descends to a complex projective structure on moduli space.

3 The Burau representation and polyhedra monodromy

In this section, we show that specializing the Burau representation to roots of unity yields (a portion of) the monodromy representation of Thurston's moduli space of Euclidean cone metrics.

Considering cone metrics with curvatures $\vec{k} = (k_1, \dots, k_m)$, the mapping class group of interest is the subgroup of the m–punctured sphere mapping class group which preserves the labeling of points by their curvatures. Points labeled with the same curvature may be interchanged. We denote this group by $\mathrm{Mod}(S_{0,m}; \vec{k})$.

If $k_1 = \cdots = k_n$ with $n \le m$, then there is an action of the n–strand braid group $B_n \to \mathrm{Mod}(S_{0,m}; \vec{k})$. Theorem 1.1 says that something akin to a specialization of the Burau representation appears in the monodromy representation of $\mathrm{Mod}(S_{0,m}; \vec{k})$ from Thurston's work. To connect the representations in the general case, which involves mapping between matrix groups of different dimensions, we need to define the evaluation map $\mathrm{GL}_{n-1}(\mathbb{Z}[t^\pm]) \to \mathrm{PGL}_{m-2}(\mathbb{C})$ used in the statement of the theorem.

Definition 3.1 Define a map $v\colon \mathrm{GL}_{n-1}(\mathbb{Z}[t^\pm]) \to (\mathbb{Z}(t))^{n-1}$ to the free module over the field of rational functions by sending a matrix A to the column vector

$$v(A) = \frac{1}{1-t^n}(I_{n-1} - A)(1-t, 1-t^2, \dots, 1-t^{n-1})^\top.$$

One can show that this map is a *crossed homomorphism*.[3] We use it to define an *affine extension* $\widetilde{(-)}\colon \mathrm{GL}_{n-1}(\mathbb{Z}[t^\pm]) \to \mathrm{GL}_n(\mathbb{Z}(t))$ by

$$\widetilde{A} = \left(\begin{array}{c|c} A & v(A) \\ \hline 0 & 1 \end{array}\right).$$

One can show that this map is a group homomorphism.

We would like to evaluate this matrix at a complex number, but it might be possible that one of the entries of $v(A)$ is undefined at the desired evaluation. Conveniently, this is never the case for matrices in the image of the Burau representation.

[3]The derivation of the formula for $v(A)$ comes from our understanding of the topological meaning behind the Burau representation. To keep the topic of this work contained, we simply present the formula here and comment on its utility in the course of the proof of Lemma 3.2.

Lemma 3.2 *For any $A \in \beta_n(B_n)$ one has $\widetilde{A} \in \mathrm{GL}_n(\mathbb{Z}[t^{\pm}])$, ie the matrix $\widetilde{A}$ has coefficients in the ring of Laurent polynomials rather than the field of rational functions.*

Proof Recall (see eg [4, Chapter 3]) that under the reduced Burau representation, the half-twist generator σ_i acts as

$$\beta_n(\sigma_i) = I_{i-2} \oplus \begin{pmatrix} 1 & 0 & 0 \\ t & -t & 1 \\ 0 & 0 & 1 \end{pmatrix} \oplus I_{n-i-2}$$

for $1 < i < n-1$, while

$$\beta_n(\sigma_1) = \begin{pmatrix} -t & 1 \\ 0 & 1 \end{pmatrix} \oplus I_{n-3} \quad \text{and} \quad \beta_n(\sigma_{n-1}) = I_{n-3} \oplus \begin{pmatrix} 1 & 0 \\ t & -t \end{pmatrix}.$$

A straightforward computation gives that $v(\beta_n(\sigma_i)) = (0, \dots, 0, 0)$ for $i = 1, \dots, n-2$, and $v(\beta_n(\sigma_{n-1})) = (0, \dots, 0, 1)$. We remark that the effect of the affine extension then is nothing more than to take the matrix $\beta_n(\sigma_i) \in \mathrm{GL}_{n-1}(\mathbb{Z}[t^{\pm}])$ to the matrix $\beta_{n+1}(\sigma_i) \in \mathrm{GL}_n(\mathbb{Z}[t^{\pm}])$.

So the lemma holds for each $\beta_n(\sigma_i)$. Since $\mathrm{GL}_n(\mathbb{Z}[t^{\pm}])$ is a subgroup of $\mathrm{GL}_n(\mathbb{Z}(t))$ and the $\beta_n(\sigma_i)$ generate the image $\beta_n(B_n)$, the lemma follows. □

Since the matrices in the image of the Burau representation have Laurent polynomial coefficients, we can specialize the formal variable t to any nonzero complex number and obtain a matrix with complex coefficients.

Definition 3.3 Let $-q \in \mathbb{C}$ be a nonzero complex number and let $m \geq n+1$. Define the map $\mathrm{ev}(-q)\colon \beta_n(B_n) \to \mathrm{PGL}_{m-2}(\mathbb{C})$ by the composition

$$\beta_n(B_n) \xrightarrow{\widetilde{(-)}} \mathrm{GL}_n(\mathbb{Z}[t^{\pm}]) \xrightarrow{(-)\oplus I_{m-2-n}} \mathrm{GL}_{m-2}(\mathbb{Z}[t^{\pm}]) \xrightarrow{t\mapsto -q} \mathrm{GL}_{m-2}(\mathbb{C}) \twoheadrightarrow \mathrm{PGL}_{m-2}(\mathbb{C}).$$

In the case $m = n+1$, we interpret the second map $(-)\oplus I_{-1}$ as deleting the last row and column of a matrix.

Remark 3.4 In the case $m = n+1$, the second map serves to undo the effect of the first. Thus the map $\mathrm{ev}(-q)$ factors through the evaluation of the Burau representation:

$$\beta_n(B_n) \subset \mathrm{GL}_{n-1}(\mathbb{Z}[t^{\pm}]) \xrightarrow{t\mapsto -q} \mathrm{GL}_{n-1}(\mathbb{C}) \twoheadrightarrow \mathrm{PGL}_{n-1}(\mathbb{C}).$$

This observation is the basis of the proof of Corollary 1.2.

Now we have the algebraic setup required to prove Theorem 1.1.

Proof of Theorem 1.1 We will verify that the diagram

$$\begin{array}{ccc} B_n & \xrightarrow{\quad\beta_n\quad} & \beta_n(B_n) \\ \iota_*\big\downarrow & & \big\downarrow \mathrm{ev}(-q) \\ \mathrm{Mod}(S_{0,m};\vec{k}) & \xrightarrow{\rho_{\vec{k}}} & \mathrm{PU}(1,m-3) \subset \mathrm{PGL}_{m-2}(\mathbb{C}) \end{array}$$

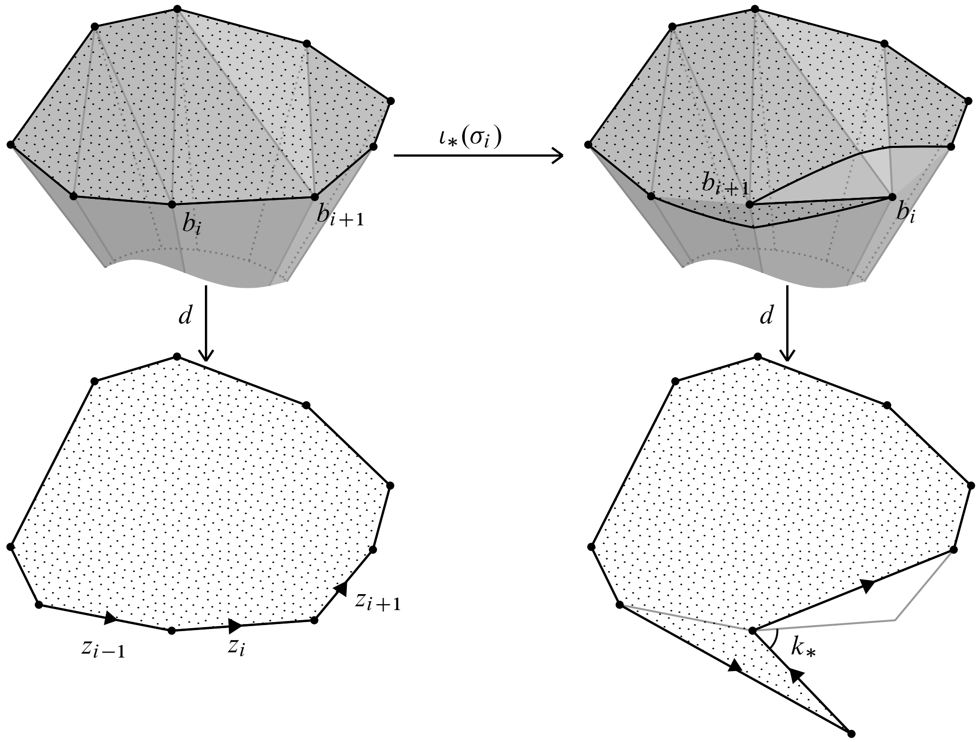


Figure 2: The action of $\iota_*(\sigma_i)$ on the disk in $S_{0,m}$, indicated with polka dots, is depicted on the top. On the bottom is the effect on the developing image, with the outline of the original disk superimposed on the right for comparison.

commutes on a generating set for the braid group. In particular, we will use the $n-1$ Artin generators $\sigma_1, \dots, \sigma_{n-1} \in B_n \approx \mathrm{Mod}(D_n)$ that swap two points in the n–times marked disk with a counterclockwise half-twist.

Label the marked points of $S_{0,m}$ as $b_1, \dots, b_m$ and consider cone metrics which have curvature k_i at b_i for each i. The n points $b_1, \dots, b_n$ can be freely permuted by mapping classes since they have the same curvature $k_* = k_1 = \dots = k_n$, and so there is an inclusion of surfaces $\iota\colon D_n \hookrightarrow S_{0,m}$ sending the n marked points of a marked disk D_n to the points $b_1, \dots, b_n$. We write σ_i for the generator of B_n, the counterclockwise half-twist on D_n that swaps the points $\iota^{-1}(b_i)$ and $\iota^{-1}(b_{i+1})$.

Now we compute the action of $\iota_*(\sigma_i)$ on the complex projective coordinates as constructed in Section 2.1. Recall that our construction was to take a disk in a cone sphere X on whose boundary lay the cone points $b_1, \dots, b_{m-1}$ and to look at the differences $z_1, \dots, z_{m-2}$ between those points under a developing map to $\mathbb{C}$. Applying a half twist $\iota_*(\sigma_i)$ changes that disk as indicated in the top of Figure 2, and we consult the bottom of the figure for the effect on the developing image. The computation is done for one particular choice of homotopy class of disk in the cone sphere, but the induced action on the coordinates is independent of the chosen disk.

The arc between b_i and b_{i+1} is taken to itself with reverse orientation on the cone sphere, but from within the disk it is approached from the opposite side, around the cone point b_i. Thus the developing image of this arc is rotated clockwise by the angle deficit/curvature at b_i, which is k_*, and is then reflected. This means z_i is mapped to $-\exp(-ik_*)z_i = \exp(i(\pi - k_*))z_i$ under the monodromy representation of $\iota_*(\sigma_i)$. Similarly, z_{i-1} maps to $z_{i-1} - \exp(i(\pi - k_*))z_i$, and z_{i+1} maps to $z_i + z_{i+1}$. All other coordinates are fixed.

So writing $q = \exp(i(\pi - k_*))$, we see that the coordinates $z_1, \dots, z_{m-2}$ are transformed linearly via the matrix

$$\rho_{\vec{k}}(\iota_*(\sigma_i)) = I_{i-2} \oplus \begin{pmatrix} 1 & 0 & 0 \\ -q & q & 1 \\ 0 & 0 & 1 \end{pmatrix} \oplus I_{m-i-3}$$

for $1 < i < n-1$, with appropriate modifications for the cases $i = 1$ and $i = n-1$ as necessary. Evidently this is the image of the Burau representation $\beta_n(\sigma_i)$ after our modified evaluation map $\mathrm{ev}(-q)$. Since the diagram commutes when applying the various maps to any $\sigma_i \in B_n$ and the σ_i generate the braid group, we have the desired result. □

In the case $m = n+1$, the evaluation map $\mathrm{ev}(-q)$ is truly an evaluation of the Burau representation (rather than the evaluation of the affine extension of the representation). This is the basis for Corollary 1.2.

Proof of Corollary 1.2 In the case $m = n+1$, Remark 3.4 tells us that the commutative diagram in Theorem 1.1 factors as

$$\begin{array}{ccc} B_n & \xrightarrow{\beta_n} & \beta_n(B_n) \subset \mathrm{GL}_{n-1}(\mathbb{Z}[t^{\pm}]) \\ & \searrow^{\beta(-q)} & \downarrow {\scriptstyle t \mapsto -q} \\ \downarrow {\scriptstyle \iota_*} & & \mathrm{GL}_{n-1}(\mathbb{C}) \\ & & \twoheaddownarrow \\ \mathrm{Mod}(S_{0,n+1}; \vec{k}) & \xrightarrow{\rho_{\vec{k}}} & \mathrm{PU}(1, n-2) \subset \mathrm{PGL}_{n-1}(\mathbb{C}) \end{array}$$

which gives the commutative diagram in the statement of the corollary. Thus an element of $\ker(\beta(-q))$ must lie in the kernel of the composite map

$$B_n \xrightarrow{\iota_*} \mathrm{Mod}(S_{0,n+1}; \vec{k}) \xrightarrow{\rho_{\vec{k}}} \mathrm{PU}(1, n-2).$$

The inclusion of the punctured disk $\iota\colon D_n \hookrightarrow S_{0,n+1}$ induces a surjective map of mapping class groups $\iota_*\colon B_n \twoheadrightarrow \mathrm{Mod}(S_{0,n+1}; \vec{k})$ whose kernel is the central subgroup $\langle \tau_n \rangle$ generated by the full-twist braid (see eg [11, Proposition 3.19]). If $\ker(\rho_{\vec{k}}) = \mathrm{ncl}_{\mathrm{Mod}(S_{0,n+1}; \vec{k})}(S)$, then a product of conjugates of elements of $S \subset \mathrm{Mod}(S_{0,n+1}; \vec{k})$ lifts to a product of conjugates of elements of $\widetilde{S} \subset B_n$. Lifts differ by the kernel of ι_*, so the kernel of the composition $\rho_{\vec{k}} \circ \iota_*$ equals $\mathrm{ncl}_{B_n}(\widetilde{S}, \tau_n) = \mathrm{ncl}_{B_n}(\widetilde{S}) \cdot \langle \tau_n \rangle$, the last equality following because the full twist τ_n is central in the braid group. □

4 The completion of moduli space and orbifolds

In this section we gather several results that we will leverage in combination with Corollary 1.2 to restrict the kernel of the Burau representation. First, we recall Thurston's results about the metric completion of the moduli space of Euclidean cone metrics. Then we gather facts about orbifold fundamental groups and the monodromy representations of geometric orbifolds. Orbifold structures will allow us to exactly identify the kernels of the representations $\rho_{\vec{k}}$.

Lemma 2.1 says that $\mathcal{M}(\vec{k})$ is a complex projective orbifold, yielding a monodromy representation $\rho_{\vec{k}}\colon \mathrm{Mod}(S_{0,m};\vec{k}) \to \mathrm{PGL}_{m-2}(\mathbb{C})$. But in fact there is an extra piece of information we have, a hermitian form on the complex coordinate space whose diagonal part $A\colon \mathbb{C}^{m-2} \to \mathbb{R}$ simply measures the area of a Euclidean cone sphere whose developing map (before scaling) yields the coordinates $z_1, \dots, z_{m-2} \in \mathbb{C}$. Since acting by mapping classes only changes the choice of disk in a Euclidean cone sphere X and not the underlying geometry of X, we see that our monodromy representation lands in the unitary group of the area form, ie $\rho_{\vec{k}}\colon \mathrm{Mod}(S_{0,m};\vec{k}) \to \mathrm{PU}(A)$. Thurston gave a geometric method to find the signature of this form:

Lemma 4.1 [22, Proposition 3.3] *The area form A has signature $(1, m-3)$.*

And so $\mathrm{PU}(A) \approx \mathrm{PU}(1, m-3)$ is the group of holomorphic isometries of complex hyperbolic space $\mathbb{C}\mathbb{H}^{m-3}$. Therefore $\mathcal{M}(\vec{k})$ is in fact a *complex hyperbolic* orbifold. As Thurston explains, the complex hyperbolic metric is not complete. The metric completion $\overline{\mathcal{M}}(\vec{k})$ is obtained by adding several strata: lower-dimensional moduli spaces corresponding to the collision of groups of cone points whose curvatures sum to less than 2π. Around each (real) codimension-2 stratum where two cone points collide there is a cone angle in the complex hyperbolic metric. Thurston examines the geometry of these strata:

Lemma 4.2 [22, Proposition 3.5] *The cone angle around a codimension-2 stratum of $\overline{\mathcal{M}}(\vec{k})$ where cone points of curvature k_1 and k_2 collide is*

(i) $\pi - k_*$ *when* $k_* = k_1 = k_2$, *and*

(ii) $2\pi - (k_1 + k_2)$ *when* $k_1 \neq k_2$.

In general, the completion of the moduli space of Euclidean cone metrics has the structure of a complex hyperbolic *cone manifold*. We won't go into the particulars of cone manifolds, but we note that in certain cases the completed moduli space is actually a complex hyperbolic *orbifold*.

Lemma 4.3 [22, Theorem 4.1] *If the cone angles around all codimension-2 strata of $\overline{\mathcal{M}}(\vec{k})$ are integral submultiples of 2π, then the metric completion $\overline{\mathcal{M}}(\vec{k})$ is a complex hyperbolic orbifold. A codimension-2 stratum with cone angle $2\pi/d$ is an orbifold stratum of order d.*

Corollary 1.2 allows us to understand the kernel of the specialized Burau representation because the monodromy representation $\rho_{\vec{k}}$ has a kernel that is quite explicit thanks to orbifold theory.

First, the proof of [6, Theorem 2.9] and the paragraph following it tell us how to find the orbifold fundamental group of the completed moduli space when this space has an orbifold structure:

Lemma 4.4 [6, Theorem 2.9] *Suppose the completed moduli space $\overline{\mathcal{M}}(\vec{k})$ is a complex hyperbolic orbifold. Then the kernel of the map*

$$\mathrm{Mod}(S_{0,m};\vec{k}) = \pi_1^{\mathrm{orb}}(\mathcal{M}(\vec{k})) \twoheadrightarrow \pi_1^{\mathrm{orb}}(\overline{\mathcal{M}}(\vec{k}))$$

is the normal closure of powers of loops around the added codimension-2 strata of $\overline{\mathcal{M}}(\vec{k})$, the power being the order of the given stratum.

And then, as a special case of [6, Theorem 2.26], we have the following:

Lemma 4.5 [6, Theorem 2.26] *Suppose the completed moduli space $\overline{\mathcal{M}}(\vec{k})$ is a complex hyperbolic orbifold. Then the monodromy representation of the completed moduli space $\pi_1^{\mathrm{orb}}(\overline{\mathcal{M}}(\vec{k})) \to \mathrm{PU}(1, m-3)$ is faithful.*

5 The kernel of Burau at roots of unity

In this section, we explain how Corollary 1.2 and the results gathered in Section 4 can be used to identify the kernel of the Burau representation specialized at certain roots of unity. In particular, we will prove Theorem 1.3 and Corollary 1.4.

Our analysis uses the 94 choices of curvatures $\vec{k}$ enumerated by Thurston [22] for which the completion of the moduli space of Euclidean cone metrics $\overline{\mathcal{M}}(\vec{k})$ is a complex hyperbolic orbifold. In these cases, we know exactly the kernel of the monodromy representation $\rho_{\vec{k}}\colon \mathrm{Mod}(S_{0,m};\vec{k}) \to \mathrm{PU}(1, m-3)$. Searching Thurston's list of orbifolds allows us, via the commutative diagram of Corollary 1.2, to restrict the kernel of the associated specialization of the Burau representation. First we demonstrate how our method applies in the $n = 6$ case and verify it with a certain braid already known to lie in the kernel of β_6.

Example 5.1 (the case of $n = 6$ and $d = 4$) To obtain a containment on the kernel of the Burau representation, we appeal to a particular moduli space of cone structures. Consider the choice of curvatures $\vec{k} = 2\pi(1,1,1,1,1,1,2)\cdot\frac{1}{4}$. The completion of the moduli space $\overline{\mathcal{M}}(\vec{k})$ is formed by adding two codimension-2 strata. One stratum corresponds to the collision of two cone points of curvature $2\pi\cdot\frac{1}{4}$, and by Lemma 4.2 the cone angle around this stratum is $\pi - 2\pi\left(\frac{1}{4}\right) = 2\pi\cdot\frac{1}{4}$. The other stratum corresponds to the collision of a cone point of curvature $2\pi\cdot\frac{1}{4}$ with the one of curvature $2\pi\left(\frac{2}{4}\right)$, and the cone angle here is again $2\pi - 2\pi\left(\frac{1}{4}+\frac{2}{4}\right) = 2\pi\cdot\frac{1}{4}$. Since these angles are submultiples of 2π, Lemma 4.3 says that $\overline{\mathcal{M}}(\vec{k})$ is an orbifold. The added strata are orbifold strata both of order 4.

Now pick two mapping classes: $\sigma \in \mathrm{Mod}(S_{0,7};\vec{k})$ that exchanges two cone points of equal curvature with a half-twist, and $\tau \in \mathrm{Mod}(S_{0,7};\vec{k})$ that performs a full twist of a pair of points of distinct curvatures. Then σ and τ represent the π_1^{orb}–conjugacy classes of loops around the added orbifold strata. Lemmas 4.4

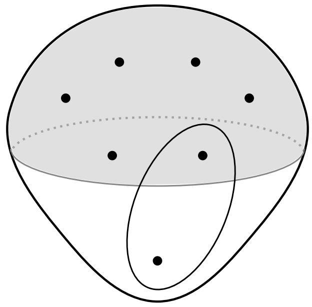


Figure 3: The same curve encircles either one puncture in the disk and the puncture outside of the disk, or $n-1$ punctures in the disk. So the full twist $\tau \in \mathrm{Mod}(S_{0,n+1}; \vec{k})$ about this curve lifts to the twist $\tau_{n-1} \in B_n$ about $n-1$ points.

and 4.5 give that the monodromy representation $\rho_{\vec{k}}\colon \mathrm{Mod}(S_{0,7}; \vec{k}) \to \mathrm{PU}(1,4)$ has kernel exactly the normal subgroup $\mathrm{ncl}_{\mathrm{Mod}(S_{0,7};\vec{k})}(\sigma^4, \tau^4)$. The evident inclusion of a 6–punctured disk $\iota\colon D_6 \hookrightarrow S_{0,7}$ induces a surjective map of mapping class groups $\iota_*\colon B_6 \twoheadrightarrow \mathrm{Mod}(S_{0,7}; \vec{k})$ whose kernel is the central subgroup $\langle \tau_6 \rangle$ (see eg [11, Proposition 3.19]).

Conflating notation, the mapping classes $\sigma, \tau \in \mathrm{Mod}(S_{0,7}; \vec{k})$ lift to a half-twist generator $\sigma \in B_6$ and a full twist on five strands $\tau_5 \in B_6$; see Figure 3. So by Corollary 1.2 we have

$$\ker(\beta(-q)) \le \mathrm{ncl}_{B_6}(\sigma^4, \tau_5^4) \cdot \langle \tau_6 \rangle \tag{2}$$

for $q = \exp\bigl(2\pi i \cdot \frac{1}{4}\bigr)$ a primitive 4^{th} root of unity.

Any braid in the kernel of the Burau representation β_6, before specialization, must also lie in the kernel of any specialization. So (2) also gives a containment on $\ker(\beta_6)$. There is in fact a not-too-complicated, nontrivial braid known to lie in the kernel of β_6, found by Bigelow [1]. The containment (2) implies that this braid is trivial if we allow ourselves the extra relations that $\sigma^4 = \tau_5^4 = \tau_6 = 1$. See Figure 4 for a computation verifying this conclusion. □

5.1 Proof of the main theorem

Following Example 5.1, we have the geometric reasoning we need to prove the theorem under consideration in this section.

Proof of Theorem 1.3 First we note that one of the containments necessary to establish (1) can be established by pure computation, which we now show. Again, we let $\sigma \in B_n$ denote one of the half-twist generators of the braid group, $\tau_{n-1} \in B_n$ a full twist on a curve surrounding $n-1$ points in the punctured disk, and $\tau_n \in B_n$ the full twist on the boundary of the disk. One can show that, up to conjugacy, we have

$$\beta(\sigma^d) = \begin{pmatrix} (-t)^d & 1 - t + t^2 - \cdots + (-t)^{d-1} \\ 0 & 1 \end{pmatrix} \oplus I_{n-3}, \quad \beta(\tau_{n-1}^j) = \widehat{t^{(n-1)j} I_{n-2}}, \quad \beta(\tau_n^\ell) = t^{n\ell} I_{n-1},$$

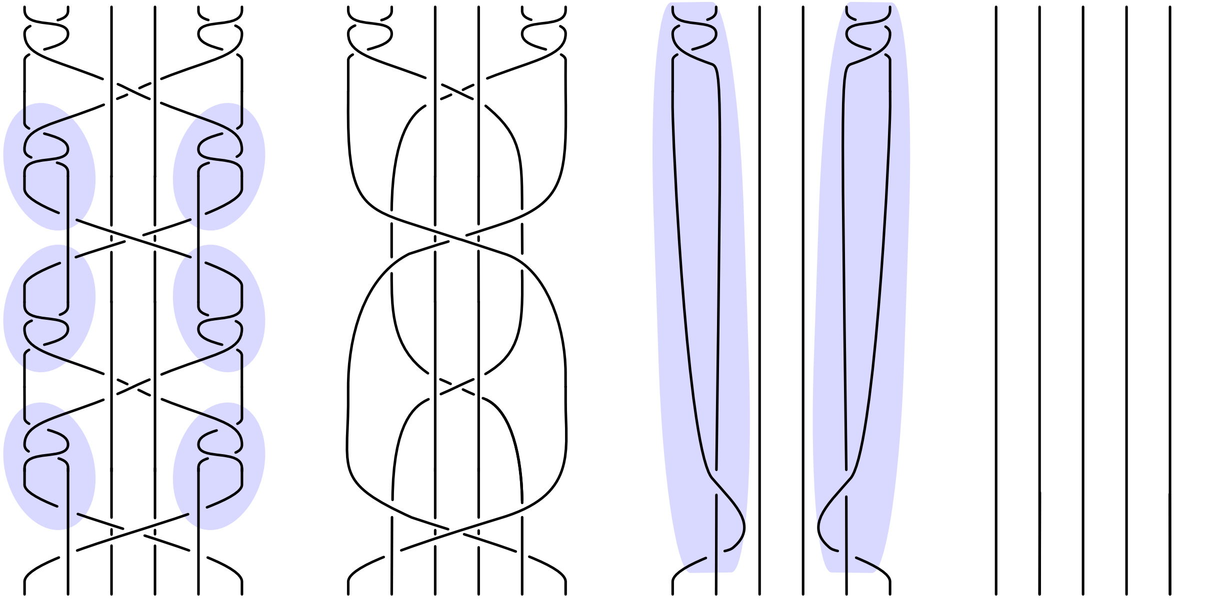


Figure 4: Bigelow's 6–strand braid [1, Figure 4] is nontrivial in B_6, but it is trivial given the extra relation $\sigma^4 = 1$. We use this relation to replace σ^3 with σ^{-1} in the shaded areas of Bigelow's braid (left), isotope (center left), and trivialize instances of σ^4 (center right) to obtain a trivial braid (right).

where $\widetilde{(-)}\colon \mathrm{GL}_{n-2}(\mathbb{Z}[t^{\pm}]) \to \mathrm{GL}_{n-1}(\mathbb{Z}(t))$ is the affine extension from Definition 3.1. Explicitly, we have

$$\beta(\tau_{n-1}^{j}) = \left(\begin{array}{c|c} t^{(n-1)j} & \dfrac{1-t^{(n-1)j}}{1-t^{n-1}} \begin{pmatrix} 1-t \\ \vdots \\ 1-t^{n-2} \end{pmatrix} \\ \hline 0 & 1 \end{array}\right).$$

When t is evaluated at $-q$ for $q \in \mathbb{C}$ a primitive d^{th} root of unity, it is evident that $\beta(\sigma^d)$ evaluates to the identity. For all of the cases n, d, and j indicated in the table of Theorem 1.3, with $j < \infty$, one can verify that $-q$ is an $((n-1)j)^{\text{th}}$ root of unity and not an $(n-1)^{\text{th}}$ root of unity. Thus $\beta(\tau_{n-1}^{j})$ evaluates to the identity in all of these cases.

Finally, the matrix $\beta(\tau_n^{\ell})$ evaluates to $(-q)^{n\ell} I_{n-1}$. The values ℓ indicated in the table are exactly given as $\ell = 2d/\gcd(2d, (d+2)n)$, which is the order of the unit complex number $(-q)^n$ in the multiplicative group of unit complex numbers. So $\beta(\tau_n^{\ell})$ evaluates to the identity for all of the cases given in the table. Thus

$$\mathrm{ncl}_{B_n}(\sigma^d, \tau_{n-1}^{j}) \cdot \langle \tau_n^{\ell} \rangle \leq \ker(\beta(-q)).$$

For the reverse inclusion, we appeal to the geometry of the moduli space of polyhedra and use the same strategy as Example 5.1. The appendix of [22] points us to several choices of curvatures $\vec{k}$ for which the completed moduli space $\overline{\mathcal{M}}(\vec{k})$ is an orbifold. In the cases $\vec{k} = (k_1, \ldots, k_n, k_{n+1})$ where

$k_1 = \cdots = k_n = k_*$, Corollary 1.2 comes into play to give a restriction on $\ker(\beta(-q))$. Here we give our table again, with an extra row giving the number of the relevant entry in Thurston's appendix:

n	4								5				6		7		8	9	10
d	5	6	7	8	9	10	12	18	4	5	6	8	4	5	3	4	3	3	3
j	∞	∞	14	8	6	5	4	3	∞	5	3	2	4	2	∞	2	6	3	2
ℓ	5	3	7	4	9	5	3	9	4	2	3	8	2	5	6	4	3	2	3
#	55	2	77	47	84	9	75	48	5	54	1	44	4	50	16	3	14	12	11

For example, entry number 44 in Thurston's appendix gives $\vec{k} = 2\pi(3,3,3,3,3,1)\cdot\frac{1}{8}$. One added orbifold stratum in the moduli space, corresponding to a half twist σ on two points of the same curvature, has cone angle $\pi - 2\pi\left(\frac{3}{8}\right) = 2\pi\cdot\frac{1}{8}$. The other added orbifold stratum, corresponding to a full twist τ on two points of distinct curvature, has cone angle $2\pi - 2\pi\left(\frac{3}{8}+\frac{1}{8}\right) = 2\pi\cdot\frac{1}{2}$. So by Lemmas 4.4 and 4.5, the kernel of $\rho_{\vec{k}}\colon \mathrm{Mod}(S_{0,6};\vec{k}) \to \mathrm{PU}(1,3)$ equals $\mathrm{ncl}_{\mathrm{Mod}(S_{0,6};\vec{k})}(\sigma^8,\tau^2)$. Since σ lifts to a half-twist generator $\sigma \in B_5$ and τ lifts to a full twist on four strands $\tau_4 \in B_5$, Corollary 1.2 gives

$$\ker(\beta(-q)) \le \mathrm{ncl}_{B_n}(\sigma^8, \tau_4^2)\cdot\langle\tau_5\rangle$$

when $q = \exp\left(i\left(\pi - 2\pi\cdot\frac{3}{8}\right)\right) = \exp\left(2\pi i\cdot\frac{1}{8}\right)$ is a primitive 8^{th} root of unity.

Since conjugates of σ^8 and τ_4^2 are already in the kernel of $\beta(-q)$ by computation and

$$\ell = 2d/\gcd(2d,(d+2)n) = 8$$

is the order of $(-q)^8$, which is the smallest power of τ_5 in the kernel of $\beta(-q)$, we get the slightly stronger restriction

$$\ker(\beta(-q)) \le \mathrm{ncl}_{B_n}(\sigma^8, \tau_4^2)\cdot\langle\tau_5^8\rangle.$$

This establishes the equality in Theorem 1.3. The other cases with $j < \infty$ are similar.

For the four cases in the table with $j = \infty$, one finds that $k_n + k_{n+1} \ge 2\pi$. In these cases, the completion $\overline{\mathcal{M}}(\vec{k})$ has no stratum added corresponding to the mapping class τ, and so $\ker(\rho_{\vec{k}})$ is just given as $\mathrm{ncl}_{\mathrm{Mod}(S_{0,n+1};\vec{k})}(\sigma^d)$. The rest of the proof is the same. □

Remark 5.2 For three of the four cases in the table with $j = \infty$, the image of $\tau_{n-1} \in B_n$ under $\beta(-q)$ in fact has infinite order. In other words, $\tau_{n-1}^j \notin \ker(\beta(-q))$ for any j. However, in the case $n = 4$ and $d = 5$ we see that $\beta(-q)(\tau_3)$ has order 10. Our results thus imply that $\tau_3^{10} \in \mathrm{ncl}_{B_4}(\sigma^5, \tau_4^5)$. This aligns with [9, Section 11], which establishes the order of the central element in the group $B_3/\mathrm{ncl}(\sigma^5)$ (among similar results).

Remark 5.3 For the cases $(n,d) = (5,6)$, $(7,4)$, and $(4,10)$, the relevant choice of curvatures for the above proof has all curvatures equal: $k_1 = \cdots = k_n = k_{n+1}$. In these cases, the relevant moduli space is a *finite cover* of the moduli space considered by Thurston, the cover in which the $(n+1)^{\text{st}}$ point is distinguished. One can check that in these cases the full twist on the n^{th} point and $(n+1)^{\text{st}}$ point still corresponds to an orbifold stratum rather than just a cone stratum.

In the same vein, the observant reader will find that we have neglected to use Thurston's entry number 10, which has $\vec{k} = 2\pi(1, 1, \dots, 1) \cdot \frac{1}{6}$ with 12 cone points of equal curvature. This corresponds to the case $(n, d) = (11, 3)$ in our context. However here, the cover of the moduli space that distinguishes the 12^{th} point is no longer an orbifold. The stratum corresponding to the twist τ has cone angle $2\pi - 2\pi\left(\frac{1}{6} + \frac{1}{6}\right) = 2\pi\left(\frac{2}{3}\right)$. So we cannot say for certain what the kernel of the representation $\rho_{\vec{k}}$ is in this case.

Thinking of the broader context here, recall that the $n = 4$ case is the last remaining case in which we do not know whether the Burau representation is faithful. Theorem 1.3 gives us the restriction on $\ker(\beta_4)$ of Corollary 1.4.

Proof of Corollary 1.4 Any element in the kernel of β_4, before specialization, is an element of the kernel of *every* specialization. So $\ker(\beta_4) \leq \ker(\beta(-q))$ for any q. In particular, the eight entries of Theorem 1.3 with $n = 4$ give restrictions on $\ker(\beta_4)$.

For the statement about infinite index, note that the quotient $B_4 / \operatorname{ncl}(\sigma^d, \tau_3^j, \tau_4)$ is the orbifold fundamental group of some $\overline{\mathcal{M}}(\vec{k})$ in the cases of interest. Part of [22, Theorem 0.2] is that $\overline{\mathcal{M}}(\vec{k})$ is a complex hyperbolic orbifold of *finite volume*. Therefore, the orbifold fundamental group of this space must be infinite and so the map $B_4 \twoheadrightarrow \operatorname{Mod}(S_{0,5}; \vec{k}) \twoheadrightarrow \pi_1^{\text{orb}}(\overline{\mathcal{M}}(\vec{k}))$ has a kernel of infinite index. The normal subgroup $\operatorname{ncl}_{B_4}(\sigma^d, \tau_3^j) \cdot \langle \tau_4^\ell \rangle$, with a higher power of τ_4, is a subgroup of this kernel. □

5.2 Burau$_3$ at roots of unity

Thurston's appendix does not list choices of curvatures $\vec{k}$ with four cone points because there are infinitely many such choices for which the completion $\overline{\mathcal{M}}(\vec{k})$ is a $\mathbb{CH}^1 \approx \mathbb{RH}^2$ orbifold. In fact, there are infinitely many such $\vec{k} = (k_1, k_2, k_3, k_4)$ satisfying $k_1 = k_2 = k_3$, inviting us to include a 3–strand braid group. Using the same method as in the proof of Theorem 1.3, we have the following:

Theorem 5.4 *Let q be a primitive d^{th} root of unity for $d \geq 7$. Then*

$$\ker(\beta_3(-q)) = \operatorname{ncl}_{B_3}(\sigma^d) \cdot \langle \tau_3^\ell \rangle,$$

where $\ell = 2d/\gcd(12, d+6)$ is the order of the unit complex number $(-q)^3$.

Proof Take $k_1 = k_2 = k_3 = \pi - 2\pi/d$ and $k_4 = \pi + 3\pi/d$. With $\vec{k} = (k_1, k_2, k_3, k_4)$, the same argument as in the proof of Theorem 1.3 gives the result. Namely, the completion of the moduli space $\overline{\mathcal{M}}(\vec{k})$ has one orbifold stratum added, corresponding to a half twist σ, of cone angle $\pi - (\pi - 2\pi/d) = 2\pi/d$. Note that in this case one has $k_3 + k_4 > 2\pi$, so there is no added stratum corresponding to the mapping class τ. □

This theorem corrects [13, Theorem 1.2], which is incorrect in the cases when d is a multiple of 3. I would like to note that the correct result was evident in the course of the arguments of that paper. Additionally, when stated in this form, it is clear that Theorem 5.4 combines with their result [13, Theorem 2.1] to give a new proof of the faithfulness of β_3.

6 Future directions

My original motivation for developing these results was to use Thurston's moduli space of polyhedra to place restrictions on the kernel of the Burau representation. Perhaps the restrictions we have found in Corollary 1.4 can be combined with other results about the kernel of β_4 to prove faithfulness of the representation, or to narrow down a search for a nontrivial element of the kernel. The connections with Squier's conjectures yield other questions and future directions.

A mismatch between our results and Squier's conjecture Squier conjectured that the kernel of $\beta(-q)$ would be equal to $\mathrm{ncl}(\sigma^d)$ for any n and for q any d^{th} root of unity [21, (C1)]. From the results of [13] and the typical nonfaithfulness of Burau [1], we know that for any $n \geq 5$ this conjecture is false for almost all even values of d. The form we have acquired in Theorem 1.3 for the finitely many special cases of $\ker(\beta(-q))$ certainly looks more complicated than Squier's very simple normal generating set. It would be interesting to see in which cases the normal subgroup $\mathrm{ncl}(\sigma^d, \tau_{n-1}^j) \cdot \langle \tau_n^\ell \rangle$ is or is not equal to $\mathrm{ncl}(\sigma^d)$, ie to see when Squier's conjecture is correct as originally stated. Part of investigating this question could be a calculation along the lines of Figure 4.

The case with more than $n+1$ cone points One restriction in this work is that we have only considered choices of curvatures $\vec{k} = (k_1, \dots, k_m)$ with $m = n+1$. There are two reasons for this. First, taking $m = n+1$ allows us to identify the kernel of the composite map

$$B_n \xrightarrow{\iota_*} \mathrm{Mod}(S_{0,n+1}; \vec{k}) \xrightarrow{\rho_{\vec{k}}} \mathrm{PU}(1, n-2)$$

in terms of $\ker(\rho_{\vec{k}})$ because the map of mapping class groups ι_* is surjective. When $m > n+1$, the induced map on mapping class groups is injective (and not surjective) [11, Theorem 3.18],

$$B_n \xhookrightarrow{\iota_*} \mathrm{Mod}(S_{0,m}; \vec{k}) \xrightarrow{\rho_{\vec{k}}} \mathrm{PU}(1, m-3),$$

but I was unable to rigorously identify the preimage of $\ker(\rho_{\vec{k}})$ in B_n in this case. I would like to see how to identify the kernel in this case and whether it could give any more information about the Burau representation.

Second, the condition $m = n+1$ tells us that the "evaluation map" in Theorem 1.1 is actually an evaluation, and so we can place the evaluation of the Burau representation in the commutative diagram as in Corollary 1.2. When $m > n+1$, this is not the case. It is not hard to find braids $b \in B_n$ and roots of unity q for which $\beta(b)|_{t=-q}$ is trivial while $\mathrm{ev}(-q)(\beta(b))$ is not. For instance, powers of the full-twist braid exhibit this behavior. So it is not immediately clear to me how the moduli space of polyhedra could be used to place restrictions on $\ker(\beta(-q))$, aside from the cases explored in this paper.

Similar results for the Gassner representation To get further mileage out of Thurston's list, one might consider the Gassner representation of the *pure* braid group

$$\mathcal{G}_n \colon \mathrm{PB}_n \to \mathrm{GL}_{n-1}(\mathbb{Z}[t_1^{\pm}, \dots, t_n^{\pm}]),$$

which specializes to the Burau representation by sending t_i to t for all i. See [4, Chapter 3] for a construction. Since the Gassner representation is defined on a subgroup of the braid group and takes values in a larger matrix group, one might expect it to be "more faithful" than the Burau representation. Yet it is not known whether or not $\mathcal{G}_n$ is faithful for any value of $n \geq 4$. A version of Theorem 1.1 for the Gassner representation exists, for instance in [24] via an algebraic approach. A geometric or topological construction would be desirable along the lines of our Theorem 1.1, though our method of proof would be much more cumbersome in this context due to the more complicated generating set of the pure braid group. Yet armed with a version of Theorem 1.1 and Corollary 1.2 for the Gassner representation, more of Thurston's 94 moduli spaces could help to identify the kernels of some specializations of various $\mathcal{G}_n$. Though finitely many restrictions on $\ker \mathcal{G}_n$ could never establish faithfulness alone (see [16, Lemma 2.1]), perhaps this could shed some light on the faithfulness question for these representations.

Relationship with some remarkable work of Coxeter In [9], Coxeter investigated the quotients of braid groups defined by $B_n(d) = B_n / \operatorname{ncl}(\sigma^d)$. For $n = 2$, the quotient is a finite cyclic group. For $d = 2$, the quotient is isomorphic to the symmetric group on n letters. For all but five other choices of (n, d), the quotient $B_n(d)$ is an infinite group. Coxeter established infiniteness in these cases using hyperbolic geometry, and I think this is along the lines of the argument for infinite index in the proof of Corollary 1.4. The five sporadic cases of (n, d) for which $B_n(d)$ is finite correspond to the Platonic solids, and Coxeter gave a remarkable formula for the order of $B_n(d)$ in terms of the combinatorics of the associated Platonic solid. Coxeter proved his formula by individually computing the orders of these five quotient groups and checking that the formula works in each case. I hope that the geometric perspective of the moduli spaces considered here might give more insight into — and perhaps a more revealing proof of — Coxeter's result.

References

[1] **S Bigelow**, *The Burau representation is not faithful for $n = 5$*, Geom. Topol. 3 (1999) 397–404 MR Zbl

[2] **S J Bigelow**, *Braid groups are linear*, J. Amer. Math. Soc. 14 (2001) 471–486 MR Zbl

[3] **S Bigelow**, *Does the Jones polynomial detect the unknot?*, J. Knot Theory Ramifications 11 (2002) 493–505 MR Zbl

[4] **J S Birman**, *Braids, links, and mapping class groups*, Ann. of Math. Stud. 82, Princeton Univ. Press (1974) MR Zbl

[5] **W Burau**, *Über Zopfgruppen und gleichsinnig verdrillte Verkettungen*, Abh. Math. Sem. Univ. Hamburg 11 (1935) 179–186 MR Zbl

[6] **D Cooper**, **C D Hodgson**, **S P Kerckhoff**, *Three-dimensional orbifolds and cone-manifolds*, MSJ Memoirs 5, Math. Soc. Japan, Tokyo (2000) MR Zbl

[7] **D Cooper**, **D D Long**, *A presentation for the image of* $\operatorname{Burau}(4) \otimes Z_2$, Invent. Math. 127 (1997) 535–570 MR Zbl

[8] **D Cooper**, **D D Long**, *On the Burau representation modulo a small prime*, from "The Epstein birthday schrift" (I Rivin, C Rourke, C Series, editors), Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry (1998) 127–138 MR Zbl

[9] **H S M Coxeter**, *Factor groups of the braid group*, from "Proceedings of the fourth Canadian Mathematical Congress", Univ. Toronto Press (1959) 95–122 MR Zbl

[10] **P Deligne**, **G D Mostow**, *Monodromy of hypergeometric functions and nonlattice integral monodromy*, Inst. Hautes Études Sci. Publ. Math. 63 (1986) 5–89 MR Zbl

[11] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Ser. 49, Princeton Univ. Press (2012) MR Zbl

[12] **N J Fullarton**, **R Shadrach**, *Observed periodicity related to the four-strand Burau representation*, Exp. Math. 28 (2019) 57–66 MR Zbl

[13] **L Funar**, **T Kohno**, *On Burau's representations at roots of unity*, Geom. Dedicata 169 (2014) 145–163 MR Zbl

[14] **T Ito**, *A kernel of a braid group representation yields a knot with trivial knot polynomials*, Math. Z. 280 (2015) 347–353 MR Zbl

[15] **D Krammer**, *Braid groups are linear*, Ann. of Math. 155 (2002) 131–156 MR Zbl

[16] **D D Long**, *A note on the normal subgroups of mapping class groups*, Math. Proc. Cambridge Philos. Soc. 99 (1986) 79–87 MR Zbl

[17] **D D Long**, **M Paton**, *The Burau representation is not faithful for $n \geq 6$*, Topology 32 (1993) 439–447 MR Zbl

[18] **C T McMullen**, *Braid groups and Hodge theory*, Math. Ann. 355 (2013) 893–946 MR Zbl

[19] **J Moody**, *The faithfulness question for the Burau representation*, Proc. Amer. Math. Soc. 119 (1993) 671–679 MR Zbl

[20] **R E Schwartz**, *Notes on shapes of polyhedra*, preprint (2015) arXiv 1506.07252

[21] **C C Squier**, *The Burau representation is unitary*, Proc. Amer. Math. Soc. 90 (1984) 199–202 MR Zbl

[22] **W P Thurston**, *Shapes of polyhedra and triangulations of the sphere*, from "The Epstein birthday schrift" (I Rivin, C Rourke, C Series, editors), Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry (1998) 511–549 MR Zbl

[23] **W A Veech**, *Flat surfaces*, Amer. J. Math. 115 (1993) 589–689 MR Zbl

[24] **T N Venkataramana**, *Monodromy of cyclic coverings of the projective line*, Invent. Math. 197 (2014) 1–45 MR Zbl

Department of Mathematics, University of California, Berkeley
Berkeley, CA, United States

dlugie.e@math.berkeley.edu

Received: 12 October 2022 Revised: 31 May 2023

Algebraic & Geometric Topology **24**:5 (2024) 2807–2849
DOI: 10.2140/agt.2024.24.2807
Published: 19 August 2024

Turning vector bundles

DIARMUID CROWLEY
CSABA NAGY
BLAKE SIMS
HUIJUN YANG

We define a turning of a rank-$2k$ vector bundle $E \to B$ to be a homotopy of bundle automorphisms ψ_t from $\mathbb{1}_E$, the identity of E, to $-\mathbb{1}_E$, minus the identity, and call a pair (E, ψ_t) a turned bundle. We investigate when vector bundles admit turnings and develop the theory of turnings and their obstructions. In particular, we determine which rank-$2k$ bundles over the $2k$–sphere are turnable.

If a bundle is turnable, then it is orientable. In the other direction, complex bundles are turned bundles and for bundles over finite CW–complexes with rank in the stable range, Bott's proof of his periodicity theorem shows that a turning of E defines a homotopy class of complex structure on E. On the other hand, we give examples of rank-$2k$ bundles over $2k$–dimensional spaces, including the tangent bundles of some $2k$–manifolds, which are turnable but do not admit a complex structure. Hence turned bundles can be viewed as generalisations of complex bundles.

We also generalise the definition of turning to other settings, including other paths of automorphisms, and we relate the generalised turnability of vector bundles to the topology of their gauge groups and the computation of certain Samelson products.

57R22; 55R15, 55R25

1 Introduction

Let $\pi : E \to B$ be a real Euclidean vector bundle over a base space B, which for simplicity we assume is connected. The bundle E has two canonical automorphisms: $\mathbb{1}_E$, the identity, and $-\mathbb{1}_E$, the automorphism which takes a vector to its negative. A *turning* of E is a continuous path ψ_t of bundle automorphisms from $\mathbb{1}_E$ to $-\mathbb{1}_E$: if a turning of E exists, we call E *turnable* and the pair (E, ψ_t) a *turned* vector bundle. The *turning problem* for E is to determine whether E is turnable.

While the turning problem is a natural topological problem amenable to classical methods in algebraic topology, to the best of our knowledge it has not been explicitly discussed in the literature. Our primary interest in turnings stems from the fact that they generalise complex structures. As we explain in Section 5, Bott's original proof of his periodicity theorem shows for bundles over finite CW–complexes that stable turnings are equivalent to stable complex structures. On the other hand, we discovered that there are unstable bundles which are turnable but do not admit a complex structure; eg see the discussion following Theorem 1.3. Hence there is a sequence of strict inclusions

$$\{\text{complex bundles}\} \subsetneq \{\text{turned bundles}\} \subsetneq \{\text{stably complex bundles}\}.$$

The turning problem and its generalisations also arise naturally in the study of gauge groups. Turnings are closely related to the group of components and the fundamental group of the gauge group associated to E, and by studying turnings we can gain information about the low-dimensional topology of gauge groups; see eg Theorem 1.9. The generalised turning problem for loops in the structure group is also related to certain Samelson products and our results on turnings lead to new computations of Samelson products, which have implications for the high-dimensional topology of certain gauge groups; see Proposition 3.31.

1.1 Turnability of vector bundles

We begin with some elementary remarks on the turning problem. Given $b \in B$, let $E_b := \pi^{-1}(b)$ be the fibre over b, which is a Euclidean vector space. Since a turning of E restricts to a path from $\mathbb{1}_{E_b}$ to $-\mathbb{1}_{E_b}$, if E is turnable, then the rank of E must be even. Moreover, a turning of E can be used to continuously orient each fibre E_b. Hence if E is turnable, then E is orientable; see the discussion prior to Lemma 2.4. On the other hand, if E admits a complex structure then the map $t \mapsto e^{i\pi t}\mathbb{1}_E$ defines a turning of E and so complex bundles are turned bundles. Since oriented rank-2 bundles are equivalent to complex line bundles, they are turnable and so we assume $k > 1$, unless stated otherwise.

We next discuss the turning problem stably. Suppose that the base space B is (homotopy equivalent to) a finite CW–complex, let $\underline{\mathbb{R}}^j$ denote the trivial rank-j bundle over B and let $E \oplus \underline{\mathbb{R}}^j$ denote the Whitney sum of E and $\underline{\mathbb{R}}^j$. We say that E is *stably turnable* if $E \oplus \underline{\mathbb{R}}^j$ is turnable for some $j \geq 0$ and similarly we say that E admits a stable complex structure if $E \oplus \underline{\mathbb{R}}^j$ admits a complex structure for some $j \geq 0$. Then we have the following result; see Theorem 5.10 for a more refined statement.

Theorem 1.1 *Let $E \to B$ be a vector bundle over a finite-dimensional CW–complex. Then E is stably turnable if and only if E admits a stable complex structure.*

The question of whether E admits a stable complex structure, while in general difficult, can be characterised entirely using the ring structure in real K–theory; see Proposition 5.11. We therefore turn our attention to the turning problem for bundles outside the stable range and in this paper we pay close attention to rank-$2k$ bundles over $2k$–dimensional CW–complexes. Such bundles are just outside the stable range and provide a large class of interesting examples, including the tangent bundles TM of orientable smooth $2k$–manifolds M and rank-$2k$ bundles over S^{2k}. Starting with TS^{2k}, we recall that Kirchoff [12] proved that if TS^{2k} admits a complex structure then TS^{2k+1} is trivial. Shortly afterwards Borel and Serre [3] showed that TS^{2k} admits a complex structure if and only if $k = 0, 1$ or 3 and a little later Bott and Milnor [6] showed that TS^{2k+1} is trivial if and only if $k = 0, 1$ or 3. An important first result on the turning problem is the following strengthening of Kirchoff's theorem; see Theorem 4.3.

Theorem 1.2 *If TS^{2k} is turnable then TS^{2k+1} is trivial.*

Combined with the results of Borel and Serre and Bott and Milnor, Theorem 1.2 shows that TS^{2k} is turnable if and only if it admits a complex structure. However, such a statement does not hold generally, even for rank-$2k$ bundles E over S^{2k}, as Theorem 1.3 below shows.

We next consider the turnability of all oriented rank-$2k$ bundles over S^{2k}. Stable vector bundles over S^{2k} are classified by the real K–theory groups $\widetilde{\mathrm{KO}}(S^{2k})$ which are, respectively, isomorphic to $\mathbb{Z}$, $\mathbb{Z}/2$ and 0 for k respectively even, congruent to 1 mod 4 or congruent to 3 mod 4. Given an oriented rank-$2k$ vector bundle $E \to S^{2k}$, we let $\xi_E \in \widetilde{\mathrm{KO}}(S^{2k})$ denote the reduced real K–theory class defined by E. When $k = 2$, oriented vector bundles over S^4 admit unique homotopy classes of spin structures. By Kervaire [10], the spin characteristic class $p = p_1/2$ defines an isomorphism $p\colon \widetilde{\mathrm{KO}}(S^4) \to H^4(S^4;\mathbb{Z})$ and by Wall [17], $\rho_2(p(\xi_E)) = \rho_2(\mathrm{e}(E))$ for all oriented rank-4 bundles $E \to S^4$, where ρ_d denotes reduction mod d and for any base space B, $\mathrm{e}(E) \in H^{2k}(B;\mathbb{Z})$ denotes the Euler class of an oriented rank-$2k$ bundle $E \to B$. As a corollary of Theorem 4.1 we obtain:

Theorem 1.3 *For $k > 1$, let $E \to S^{2k}$ be an oriented rank-$2k$ bundle. Then E is turnable if and only if one of the following holds*:

(a) $k = 2$ *and* $\rho_4(\mathrm{e}(E) + p(\xi_E)) = 0$ *or* $\rho_4(\mathrm{e}(E) - p(\xi_E)) = 0$.

(b) $k = 3$.

(c) $k > 2$ *is even*, $\rho_4(\mathrm{e}(E)) = 0$ *and* $\rho_2(\xi_E) = 0$.

(d) $k > 3$ *is odd and* $\rho_4(\mathrm{e}(E)) = 0$.

Theorem 1.3 gives many examples of bundles which are turnable but do not admit a complex structure. For example, for $m \geq 2$ let $\tau_m \in \pi_{m-1}(\mathrm{SO}_m)$ denote the homotopy class of the clutching function of TS^m and for $n \in \mathbb{Z}$ let nTS^m denote the bundle corresponding to $n\tau_m \in \pi_{m-1}(\mathrm{SO}_m)$. Then for $m = 4j$, nTS^{4j} is turnable if and only if n is even, whereas by a theorem of Thomas [16, Theorem 1.7], nTS^{4j} admits a complex structure if and only if $n = 0$.

Theorem 1.3 leads to a general result on the turnability of rank-$2k$ bundles over general finite $2k$–dimensional CW–complexes. If J is a complex structure on $E \oplus \underline{\mathbb{R}}^{2j}$ for some $j \geq 0$, denote by $c_k(J) \in H^{2k}(B;\mathbb{Z})$ the k^{th} Chern class of J. We define the subgroup $I^{2k}(B) \subseteq H^{2k}(B;\mathbb{Z}/4)$ by

$$I^{2k}(B) := \begin{cases} ((\times 2) \circ \mathrm{Sq}^2 \circ \rho_2)(H^{2k-2}(B;\mathbb{Z})) & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even,} \end{cases}$$

where Sq^2 is the second Steenrod square and $\times 2$ is the natural map induced by the inclusion of coefficients $\times 2\colon \mathbb{Z}/2 \to \mathbb{Z}/4$.

The following result is a simple consequence of Theorems 6.1 and 6.2.

Theorem 1.4 *Let $E \to B$ be an oriented rank-$2k$ vector bundle over a finite CW–complex of dimension at most $2k$, and if k is even, assume that $H^{2k}(B;\mathbb{Z})$ contains no 2–torsion. Then E is turnable if and only if there is a $j \geq 0$ and a complex structure J on $E \oplus \underline{\mathbb{R}}^{2j}$ such that*

$$[\rho_4(c_k(J))] = \pm[\rho_4(\mathrm{e}(E))] \in H^{2k}(B;\mathbb{Z}/4)/I^{2k}(B).$$

Remark 1.5 When k is even and $H^{2k}(B;\mathbb{Z})$ contains 2–torsion, the condition in Theorem 1.4 remains necessary but is no longer sufficient; see Theorem 6.1 and Example 6.3.

Theorem 1.4 shows that there are many examples of manifolds whose tangent bundles are turnable but do not admit a complex structure. For example, if $M_l = \sharp_l(S^4 \times S^4)$ is the l–fold connected sum of $S^4 \times S^4$ with itself, then for any $j > 0$, the bundle $TM_l \oplus \underline{\mathbb{R}}^{2j}$ admits two homotopy classes of complex structures J, each with $c_4(J) = 0$. But $\mathrm{e}(TM_l) = \pm 2(l+1)$ by the Poincaré–Hopf theorem; see [8, page 113]. It follows that TM_l does not admit a complex structure and that TM_l is turnable if and only if l is odd; eg $T(S^4 \times S^4)$ is turnable but does not admit a complex structure. For a more general statement about when TM is turnable but does not admit a complex structure, see Corollary 6.4.

1.2 The turning obstruction for bundles over suspensions

In order to study the turning problem and obtain most of our results above, we define a complete obstruction to the existence of turnings for bundles over suspensions. For this we need to refine our definition of turning by specifying the homotopy class of the turning in a fibre. If ψ_t is a turning of an oriented rank-$2k$ bundle $E \to B$, then by restricting ψ_t to a fibre E_b we obtain a path of isometries from $\mathbb{1}_{E_b}$ to $-\mathbb{1}_{E_b}$. When E_b is identified with $\mathbb{R}^{2k}$ via an orientation-preserving isomorphism, this path is identified with a path in SO_{2k} from $\mathbb{1}$ to $-\mathbb{1}$, which is well defined up to path homotopy. Given a path γ in SO_{2k} from $\mathbb{1}$ to $-\mathbb{1}$, we shall call a turning a *γ–turning* if its restriction to each fibre is path homotopic to γ. When $k > 1$, as we generally assume, $\pi_1(\mathrm{SO}_{2k}) \cong \mathbb{Z}/2$, so there are precisely two path homotopy classes of possible paths.

Suppose that the base space $B = SX$ is a suspension, ie it is the union of two copies of the cone on X. The restriction of E to each cone admits a γ–turning, which is unique up to homotopy and E admits a γ–turning if and only if the γ–turnings over the cones agree over X up to homotopy. We can then define the *γ–turning obstruction of E* by measuring the difference of the γ–turnings over the cones and there are several equivalent ways to do this, which we present in Section 3.1. Here we discuss what we later call the *adjointed γ–turning obstruction*. Recall that the set of isomorphism classes of oriented rank-$2k$ bundles over SX form a group, which is naturally isomorphic to $[X, \mathrm{SO}_{2k}]$ via the map which sends the isomorphism class of E to the homotopy class of its clutching function $g\colon X \to \mathrm{SO}_{2k}$. We define the adjointed γ–turning obstruction

$$\text{(1-1)} \qquad \mathrm{TO}_\gamma\colon [X, \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}], \quad [g] \mapsto \big[[x,t] \mapsto g(x)\gamma(t)g(x)^{-1}\big],$$

where $[x,t] \in SX$ is the point defined by $(x,t) \in X \times I$. The following result, which follows from Proposition 3.2 and Lemma 3.6, justifies calling TO_γ the γ–turning obstruction.

Proposition 1.6 *Let $E \to SX$ be an oriented rank-$2k$ vector bundle with clutching function $g: X \to \mathrm{SO}_{2k}$. Then E is γ–turnable if and only if $\mathrm{TO}_\gamma([g]) = 0$. Moreover, if X is itself a suspension, then TO_γ is a homomorphism of abelian groups.*

Proposition 1.6 states that the γ–turning obstruction is additive for bundles over double suspensions. This is an essential input to the Theorem 4.1, which largely computes TO_γ for rank-$2k$ bundles over the $2k$–sphere and both homotopy classes of paths γ. Theorem 1.3 above is an immediate corollary of Theorem 4.1.

The final element in the proof of Theorem 4.1 involves generalising the turning problem. The definition of the γ–turning obstruction naturally leads us to consider the turning obstruction for an essential loop $\eta: I \to \mathrm{SO}_{2k}$ with $\eta(0) = \eta(1) = \mathbb{1}$. If we replace $\gamma(t)$ by $\eta(t)$ in (1-1) above, we obtain the function

$$\mathrm{TO}_\eta: [X, \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}], \quad [g] \mapsto \big[[x,t] \mapsto g(x)\eta(t)g(x)^{-1}\big].$$

If $E \to SX$ is a bundle with clutching function $g: X \to \mathrm{SO}_{2k}$, then $\mathrm{TO}_\eta([g])$ is a complete obstruction to finding a loop ψ_t of bundle automorphisms of E based at the identity such that the restriction of ψ_t to a fibre E_b is an essential loop of isometries of E_b. Moreover, if γ is a path in SO_{2k} from $\mathbb{1}$ to $-\mathbb{1}$, then the concatenation of paths $\eta * \gamma$ represents the other path homotopy class of such paths. Hence a bundle $E \to SX$ with clutching function g is turnable if and only if one of $\mathrm{TO}_\gamma([g])$ or $\mathrm{TO}_{\eta*\gamma}([g])$ vanishes. The following result relates TO_η and TO_γ and states that TO_γ is in general 4–torsion; see also Theorem 3.23.

Theorem 1.7 *Let E be an oriented rank-$2k$ vector bundle with clutching function $g: X \to \mathrm{SO}_{2k}$. Then:*

(a) $2\,\mathrm{TO}_\eta([g]) = 0$.

(b) $\mathrm{TO}_{\eta*\gamma}([g]) = \mathrm{TO}_\eta([g]) + \mathrm{TO}_\gamma([g])$.

(c) *If k is even, then $2\,\mathrm{TO}_\gamma([g]) = 0$.*

(d) *If k is odd, then $2\,\mathrm{TO}_\gamma([g]) = \mathrm{TO}_\eta([g])$ and $4\,\mathrm{TO}_\gamma([g]) = 0$.*

Remark 1.8 Notwithstanding Theorem 1.7(d), we know of no example of a bundle $E \to SX$ with clutching function g, where $2\,\mathrm{TO}_\gamma([g]) \neq 0$. In particular, by Theorem 4.1, $2\,\mathrm{TO}_\gamma(\tau_{4k+2}) = \mathrm{TO}_\eta([\tau_{4k+2}]) = 0$ for all $k \geq 1$. The proof we give of this result is computational and somewhat surprising to us. It would be interesting to know if there is a space X and a clutching function $g: X \to \mathrm{SO}_{2k}$ with $2\,\mathrm{TO}_\gamma([g]) \neq 0$.

1.3 General turnings, the topology of gauge groups and Samelson products

Let $\mathrm{Fr}(E)$ denote the frame bundle of an oriented vector bundle $E \to B$, which is a principal SO_{2k}–bundle over B. The group of automorphisms of E is canonically homeomorphic to the gauge group of $\mathrm{Fr}(E)$, and so the turning problem can be viewed as a problem in the topology of gauge groups: we are asking when a topological feature of the structure group extends to the whole gauge group.

To describe general turning problems, we let G be a path-connected topological group, for example a connected Lie group, and $P \to B$ be a principal G–bundle with gauge group $\mathcal{G}_P$: if $G = \mathrm{SO}_{2k}$ and $P = \mathrm{Fr}(E)$, then we shall write $\mathcal{G}_E$ in place of $\mathcal{G}_{\mathrm{Fr}(E)}$. If $Z(G)$ denotes the centre of G, then multiplication by $z \in Z(G)$ defines an element $z_P \in \mathcal{G}_P$. Given a path $\gamma\colon I \to G$ between elements of $Z(G)$, the *γ–turning problem* for P is to determine whether there is a path ψ_t in $\mathcal{G}_P$ with $\psi(0) = \gamma(0)_P$ and $\psi(1) = \gamma(1)_P$ and whose restriction to a fibre is path-homotopic to γ.

When $B = SX$ is a suspension, then principal G–bundles $P \to SX$ are determined up to isomorphism by their clutching functions $g\colon X \to G$ and the definition and properties of the γ–turning obstruction for vector bundles generalise in the obvious way. The (adjointed) γ–turning obstruction is the map

$$\mathrm{TO}_\gamma\colon [X, G] \to [SX, G], \quad [g] \mapsto \big[[x,t] \mapsto g(x)\gamma(t)g(x)^{-1}\big],$$

and P is γ–turnable if and only if $\mathrm{TO}_\gamma([g]) = 0$; see Remark 3.22. If we allow γ to vary among all paths between central elements of G, the path homotopy classes of the possible paths γ form a groupoid, which is a full subcategory of the fundamental groupoid of G. We call this groupoid the *central groupoid of G* and denote it by $\pi^Z(G)$. If we fix $[g] \in [X, G]$, then we can regard $\mathrm{TO}_\gamma([g])$ as a function of γ. The resulting map

$$\pi^Z(G) \to [SX, G]$$

is a morphism of groupoids (where the group $[SX, G]$ is regarded as a groupoid on one object) and this general point of view allows us to prove Theorem 1.7.

Returning to vector bundles E and the topology of their gauge groups $\mathcal{G}_E$, the η–turning problem has the most direct implications (see Theorem 2.26):

Theorem 1.9 *If $B = SX$ is a suspension and $n_{SX} := |[SX, \mathrm{SO}_{2k}]|$ is finite, then for any rank-$2k$ vector bundle $E \to SX$,*

$$|\pi_0(\mathcal{G}_E)| = \begin{cases} n_{SX} & \text{if } E \text{ is } \eta\text{–turnable},\\ \tfrac{1}{2} n_{SX} & \text{if } E \text{ is not } \eta\text{–turnable}.\end{cases}$$

Theorem 1.9 shows that when $[SX, \mathrm{SO}_{2k}]$ is finite, for example when $SX = S^{2k}$, then the η–turnability of a vector bundle E is a homotopy invariant of its gauge group $\mathcal{G}_E$. While the turnability of E is not a priori a homotopy invariant of $\mathcal{G}_E$, recent work of Kishimoto, Membrillo-Solis and Theriault [13] on the homotopy classification of the gauge groups of rank-4 bundles $E \to S^4$, when combined with our results in Theorem 4.1, does show that the turnability of these bundles is a homotopy invariant of their gauge groups; see Proposition 4.8 for a more detailed statement.

We compute TO_η for all rank-$2k$ bundles over S^{2k} in Theorem 4.1. In fact in this case $\mathrm{TO}_\eta([g]) = \langle [g], \eta\rangle$ is the Samelson product of $[g] \in \pi_{2k-1}(\mathrm{SO}_{2k})$ and $\eta \in \pi_1(\mathrm{SO}_{2k})$; see Lemma 3.24. On the other hand, Samelson products are in general delicate to calculate and so the computations of $\mathrm{TO}_\eta([g])$ in Theorem 4.1, which are carried out using the point of view of the turning obstruction, may be of independent interest. For example (see Proposition 3.29), for $\eta_{4j-1}\colon S^{4j} \to S^{4j-1}$ an essential map, we have:

Corollary 1.10 *The Samelson product* $\langle \tau_{2k}, \eta \rangle$ *is given by* $\langle \tau_{4j+2}, \eta \rangle = 0$ *and* $\langle \tau_{4j}, \eta \rangle = \tau_{4j}\eta_{4j-1} \neq 0$.

Corollary 1.10 also has implications for the high-dimensional homotopy groups of certain gauge groups; see Proposition 3.31 in Section 3.3.

Organisation The rest of this paper is organised as follows. In Section 2 we set up the necessary preliminaries to discuss the turning problem. We define turnings and γ–turnings, universal bundles which classify turnings and relate the turning problem to the topology of the gauge group. In Section 3 we define the γ–turning obstruction for bundles over suspensions and develop the theory of the γ–turning obstruction, regarded as a map from the central groupoid of a path-connected topological group G. We also show that the η–turning obstruction is given by certain Samelson products. In Section 4, we consider rank-$2k$ vector bundles over the $2k$–sphere and compute their turning obstructions in detail, proving Theorem 1.3. In Section 5, we consider the turning problem for bundles in the stable range and prove Theorem 1.1. Finally, in Section 6 we combine the results of Sections 4 and 5 on rank-$2k$ vector bundles over the $2k$–sphere and stable vector bundles to prove Theorem 1.4 on rank-$2k$ vector bundles over $2k$–dimensional CW–complexes.

Background information This paper is based on the MSc thesis of Sims, also entitled "*Turning vector bundles*", which was submitted to the University of Melbourne in May 2020 under the supervision of Crowley and Nagy.

Acknowledgements Crowley and Yang acknowledge the sponsorship of the China Scholarship Council for the 2018 visit of Yang to the University of Melbourne. Yang was sponsored by the Natural Science Foundation of Henan (#242300421380). Nagy was supported by EPSRC New Investigator grant EP/T028335/1.

2 Turnings and gauge groups

In this section we set up the necessary definitions and notation for the turning problem and establish some basic results. In Section 2.1 we define turnings and γ–turnings and introduce the terminology to describe the relationship between turnings and orientations of a vector bundle. A more general notion of turning, for principal G–bundles, is defined in Section 2.4. In Section 2.2 we define the associated turning bundle of a vector bundle and construct a universal turned bundle. We also establish some equivalent conditions to turnability in terms of the associated turning bundle and the universal turned bundle. In Sections 2.3 and 2.5 we study the connection between the turnability of a vector bundle and the low-dimensional homotopy groups of its gauge group.

2.1 Turnings

All vector spaces V in this paper are real and Euclidean. The connected component of the group of isometries of V is denoted by $\mathrm{SO}(V)$, $\mathbb{1} \in \mathrm{SO}(V)$ is the identity and $-\mathbb{1}\colon V \to V$ is defined by $-\mathbb{1}(v) = -v$ for all $v \in V$. We use $\mathbb{R}^j$ to denote j–dimensional Euclidean space with its standard metric and as usual we set $\mathrm{SO}_j := \mathrm{SO}(\mathbb{R}^j)$.

All vector bundles $\pi\colon E \to B$ are real and Euclidean and for simplicity we assume that the base space B is connected. We denote the trivial bundle $\mathbb{R}^j \times B \to B$ by $\underline{\mathbb{R}}^j$; the base space will either be specified or clear from the context. We shall use the symbol E to ambiguously denote both the total space of the bundle and the bundle itself. For $b \in B$, $E_b := \pi^{-1}(b)$ is the fibre of E over b, which is a vector space. Let $I := [0,1]$ be the unit interval.

Definition 2.1 Let V be an even-dimensional real vector space, so that $-\mathbb{1} \in \mathrm{SO}(V)$. A *turning* of V is a path $\gamma\colon I \to \mathrm{SO}(V)$ from $\mathbb{1}$ to $-\mathbb{1}$.

In particular, a turning of $\mathbb{R}^{2k}$ is a path in SO_{2k} from $\mathbb{1}$ to $-\mathbb{1}$ and we write $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ for the mapping space $\mathrm{Map}\big((I,(\{0\},\{1\})),(\mathrm{SO}_{2k},(\{\mathbb{1}\},\{-\mathbb{1}\}))\big)$ consisting of all turnings of $\mathbb{R}^{2k}$, with the compact–open topology. Note that $\Omega\mathrm{SO}_{2k}$, the space of loops based at $\mathbb{1}$, acts freely and transitively on $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ by pointwise multiplication. Hence choosing $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ defines a homeomorphism from $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ to $\Omega\mathrm{SO}_{2k}$ and we will use this homeomorphism to compute the homotopy groups of $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$.

Definition 2.2 (standard turning of $\mathbb{R}^{2k}$) Let $\mathbb{R}^{2k} = \mathbb{C}^k$ define the standard complex structure on $\mathbb{R}^{2k}$ and let $U_k \subseteq \mathrm{SO}_{2k}$ be the unitary subgroup. The *standard turning* of $\mathbb{R}^{2k}$ is the path

$$\beta\colon I \to \mathrm{SO}_{2k}, \quad t \mapsto e^{i\pi t}\mathbb{1} \in U_k \subseteq \mathrm{SO}_{2k}.$$

If $2k > 2$, then $\pi_0(\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}) \cong \pi_1(\mathrm{SO}_{2k}) \cong \mathbb{Z}/2$, so there are two turnings of $\mathbb{R}^{2k}$ up to homotopy. Indeed, if $\bar\beta$ is a representative of the other homotopy class and

$$\eta\colon (I,\{0,1\}) \to (\mathrm{SO}_{2k},\mathbb{1})$$

is a loop representing the generator of $\pi_1(\mathrm{SO}_{2k})$, then $[\bar\beta] = [\eta * \beta]$, where $*$ denotes concatenation of paths and $[\gamma]$ denotes the path homotopy class of a path γ. If we pointwise conjugate β with a fixed element of $O_{2k} \setminus \mathrm{SO}_{2k}$, then we obtain a path in $[\bar\beta]$; equivalently, an orientation-reversing isomorphism $\mathbb{R}^{2k} \to \mathbb{R}^{2k}$ pulls back β to a turning that is path homotopic to $\bar\beta$. Note that the turning defined by the formula $t \mapsto e^{-i\pi t}\mathbb{1}$ is path homotopic to β if k is even and to $\bar\beta$ if k is odd.

Let V be a vector space of dimension $2k$ equipped with a turning and an orientation. If $2k > 2$, then we say that the turning and the orientation are compatible if the turning is homotopic to β under an orientation-preserving identification $V \cong \mathbb{R}^{2k}$. If $2k = 2$, then they are compatible if the turning is homotopic to the path $t \mapsto e^{ri\pi t}\mathbb{1}$ for some positive (odd) r under an orientation-preserving identification $V \cong \mathbb{R}^2$. In both cases there is a unique orientation of V which is compatible with a given turning, hence we obtain a well-defined map from the homotopy classes of turnings of V to its orientations. This map is a bijection if $2k > 2$ and surjective if $2k = 2$.

Definition 2.3 (turning, turnable and turned) Let $\pi\colon E \to B$ be a rank-$2k$ vector bundle. A *turning* of E is a path ψ_t in the space of automorphisms of E from $\mathbb{1}_E$ to $-\mathbb{1}_E$. If a turning exists, we say that E is *turnable*, and a *turned* vector bundle is a pair (E,ψ_t), where ψ_t is a turning of E.

Clearly any trivial bundle is turnable and since bundle automorphisms can be pulled back along continuous maps, the pullback of a turnable bundle is turnable. Furthermore, every complex bundle E is turnable via the path $t \mapsto e^{i\pi t}\mathbb{1}_E$.

A turning of a bundle restricts to a turning of each fibre, so by our earlier observations it determines an orientation on each fibre. Therefore we have:

Lemma 2.4 *Every turnable bundle is orientable.* □

If a rank-2 bundle is orientable, then it admits a complex structure, so we have:

Proposition 2.5 *A rank-2 bundle is turnable if and only if it is orientable.* □

From now on we will focus on oriented bundles $\pi : E \to B$ of rank $2k$ and we assume that $2k > 2$ unless otherwise stated. Since we are assuming that B is connected, it follows that an orientable bundle has precisely two possible orientations and we let $\bar{E}$ denote the same bundle with opposite orientation.

Definition 2.6 For a path $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$, let $\Omega_\gamma\mathrm{SO}_{2k} \subset \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ denote the connected component of γ.

Definition 2.7 (γ–turnable, positive/negative turnable) Let E be a rank-$2k$ bundle and $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$. We say that E is *γ–turnable* if it has a turning whose restriction to each fibre E_b lies in $\Omega_\gamma\mathrm{SO}_{2k}$ under an orientation-preserving identification $E_b \cong \mathbb{R}^{2k}$.

We also say that E is *positive/negative turnable* if it has a turning which determines the positive/negative orientation on E.

Obviously, if γ is homotopic to γ', then γ–turnability is equivalent to γ'–turnability. By definition positive turnability is equivalent to β–turnability, and negative turnability is equivalent to $\bar{\beta}$–turnability. A bundle is turnable if and only if it is positive turnable or negative turnable. Finally, a bundle E is positive turnable if and only if $\bar{E}$ is negative turnable. For the next definition, recall that E is called *chiral* if E is not isomorphic to $\bar{E}$.

Definition 2.8 (bi-turnable, strongly chiral) If E is both positive and negative turnable, we call it *bi-turnable*. If E is turnable but not bi-turnable, we say that E is *strongly chiral*.

If $E \cong \bar{E}$, then E cannot be strongly chiral. This shows that strong chirality implies chirality.

Definition 2.9 (turning type) The *turning type* of an orientable rank-$2k$ bundle is the property of being either bi-turnable, strongly chiral or not turnable.

2.2 The associated turning bundle

We can also think of a turning of a bundle as a continuous choice of turning in each fibre. To make this precise we define the associated turning bundle below, in analogy with the associated automorphism bundle.

Every oriented rank-$2k$ vector bundle $\pi\colon E \to B$ has an associated principal SO_{2k}–bundle, namely the frame bundle $\mathrm{Fr}(E)$, whose fibres consist of oriented, orthonormal frames of the fibres of E. We will view such frames as linear isomorphisms $\phi_b\colon \mathbb{R}^{2k} \to E_b$. Then SO_{2k} acts on the right on the total space of $\mathrm{Fr}(E)$ via precomposition.

Definition 2.10 (the associated automorphism bundle and the associated turning bundle) For an oriented, rank-$2k$ vector bundle $E \to B$ we define, via the Borel construction,

(a) the associated *automorphism bundle*

$$\mathrm{Aut}(E) := \mathrm{Fr}(E) \times_{SO_{2k}} SO_{2k} \to B,$$

where SO_{2k} acts on itself by conjugation, and

(b) the associated *turning bundle*

$$\mathrm{Turn}(E) := \mathrm{Fr}(E) \times_{SO_{2k}} \Omega_{\pm\mathbb{1}} SO_{2k} \to B,$$

where SO_{2k} acts on $\Omega_{\pm\mathbb{1}} SO_{2k}$ by pointwise conjugation.

Remark 2.11 The fibre of $\mathrm{Aut}(E)$ over $b \in B$ can be identified with $SO(E_b)$, with the equivalence class $[\phi_b, A] \in \mathrm{Fr}(E) \times_{SO_{2k}} SO_{2k}$ corresponding to $\phi_b \circ A \circ \phi_b^{-1}\colon E_b \to E_b$. Similarly, the fibre of $\mathrm{Turn}(E)$ over b consists of the turnings of E_b, with $[\phi_b, \gamma]$ corresponding to the path $t \mapsto \phi_b \circ \gamma(t) \circ \phi_b^{-1}$.

A turning of a bundle E restricts to a turning of each fibre and so determines a section of $\mathrm{Turn}(E)$. In this way we obtain a homeomorphism between the space of turnings of E and the space of sections of $\mathrm{Turn}(E)$. In particular, we have:

Lemma 2.12 *A vector bundle E is turnable if and only if $\mathrm{Turn}(E) \to B$ has a section.* □

Definition 2.13 Fix a model BSO_{2k} for the classifying space of oriented rank-$2k$ vector bundles and let $VSO_{2k} \to BSO_{2k}$ denote the universal rank-$2k$ bundle. We define $BT_{2k} := \mathrm{Turn}(VSO_{2k})$ to be the total space of the associated turning bundle, and let $\pi_{2k}\colon BT_{2k} \to BSO_{2k}$ be its projection.

Remark 2.14 The symbol BT_{2k} should be read as a single unit. Defining a topological monoid T_{2k} whose classifying space is the associated turning bundle of the universal bundle $VSO_{2k} \to BSO_{2k}$ is an interesting question, but we will not address it in this paper.

Below we explain how BT_{2k} acts as a classifying space for turned vector bundles.

Proposition 2.15 *A rank-$2k$ bundle E over a CW–complex X is turnable if and only if its classifying map $f\colon X \to BSO_{2k}$ can be lifted over $\pi_{2k}\colon BT_{2k} \to BSO_{2k}$.*

Proof Since $E \cong f^*(V\mathrm{SO}_{2k})$, we have $\mathrm{Turn}(E) \cong f^*(\mathrm{Turn}(V\mathrm{SO}_{2k}))$; ie there is a pullback diagram

$$\begin{array}{ccc} \mathrm{Turn}(E) & \longrightarrow & BT_{2k} \\ \big\downarrow & & \big\downarrow \pi_{2k} \\ X & \xrightarrow{f} & B\mathrm{SO}_{2k} \end{array}$$

It follows from the universal property of pullbacks that f can be lifted to BT_{2k} if and only if $\mathrm{Turn}(E)$ has a section, which is equivalent to the turnability of E. □

We now show how BT_{2k} classifies rank-$2k$ turned vector bundles.

Definition 2.16 Let $VT_{2k} := \pi_{2k}^*(V\mathrm{SO}_{2k}) \to BT_{2k}$.

Note that VT_{2k} has a canonical turning, which we denote by ψ_t^{c}. Since VT_{2k} is defined as a pullback of $V\mathrm{SO}_{2k}$, its fibre over $x \in BT_{2k}$ can be identified with $(V\mathrm{SO}_{2k})_y$, the fibre of $V\mathrm{SO}_{2k}$ over $y = \pi_{2k}(x)$. By Remark 2.11, x itself can be regarded as a turning of $(V\mathrm{SO}_{2k})_y$ and hence of $(VT_{2k})_x$. That is, each fibre $(VT_{2k})_x$ of VT_{2k} comes equipped with a turning (which varies continuously with x), showing that $\mathrm{Turn}(VT_{2k})$ has a canonical section.

The turned bundle $(VT_{2k} \to BT_{2k}, \psi_t^{\mathrm{c}})$ is universal in the two senses explained in Theorem 2.18 below.

Definition 2.17 For a space X, let $\mathrm{TB}_{2k}(X)$ be the set of isomorphism classes of rank-$2k$ turned bundles over X: it consists of equivalence classes of turned bundles over X, where two turned bundles are equivalent if there is an isomorphism between them under which their turnings are homotopic.

Theorem 2.18 ($(VT_{2k}, \psi_t^{\mathrm{c}})$ is a universal rank-$2k$ turned bundle)

(a) *If a rank-$2k$ bundle E over a CW–complex X is equipped with a turning ψ_t, then there is a homotopically unique map $X \to BT_{2k}$ which induces (E, ψ_t) from $(VT_{2k}, \psi_t^{\mathrm{c}})$.*

(b) *For every CW–complex X there is a bijection* $\mathrm{TB}_{2k}(X) \cong [X, BT_{2k}]$.

Proof (a) A stronger statement holds: the space of pairs $(g, \overline{g})$, where $g\colon X \to BT_{2k}$ is a continuous map and $\overline{g}\colon E \to g^*(VT_{2k})$ is an isomorphism respecting the given turnings, is contractible. To such a pair $(g, \overline{g})$ we assign a pair $(f, \overline{f})$, where $f\colon X \to B\mathrm{SO}_{2k}$ is a continuous map and $\overline{f}\colon E \to f^*(V\mathrm{SO}_{2k})$ is an isomorphism, by letting $f = \pi_{2k} \circ g$ and $\overline{f} = \overline{g}$ (using that $f^*(V\mathrm{SO}_{2k}) = g^*(\pi_{2k}^*(V\mathrm{SO}_{2k})) = g^*(VT_{2k})$). Each pair $(f, \overline{f})$ determines a pullback diagram as in the proof of Proposition 2.15. It follows from the pullback property that f has a unique lift $g\colon X \to BT_{2k}$ corresponding to the given turning of E (section of $\mathrm{Turn}(E)$) and if we regard $\overline{f}$ as an isomorphism $\overline{g}\colon E \to g^*(VT_{2k})$, then this $\overline{g}$ respects the turnings. This shows that the assignment $(g, \overline{g}) \mapsto (f, \overline{f})$ is a homeomorphism. And since $V\mathrm{SO}_{2k}$ is a universal bundle, the space of pairs $(f, \overline{f})$ is contractible.

(b) To a map $g\colon X \to BT_{2k}$ we assign $g^*(VT_{2k})$ with its induced turning. This way we obtain a well-defined map $[X, BT_{2k}] \to \mathrm{TB}_{2k}(X)$, because a homotopy of g induces a bundle over $X \times I$ with a turning and after identifying this bundle with $g^*(VT_{2k}) \times I$ its turning determines a homotopy between the turnings over $X \times \{0\}$ and $X \times \{1\}$. It follows from (a) that this map is surjective.

Suppose that two maps $g_1, g_2\colon X \to BT_{2k}$ determine the same element in $\mathrm{TB}_{2k}(X)$. This means that, after identifying $g_1^*(VT_{2k})$ with $E := g_2^*(VT_{2k})$ via some isomorphism, the induced turnings on E are homotopic, ie there is a homotopy between the corresponding sections of $\mathrm{Turn}(E)$. This homotopy then determines a homotopy (via lifts of $\pi_{2k} \circ g_1\colon X \to B\mathrm{SO}_{2k}$) between g_1 and another lift $g_1'\colon X \to BT_{2k}$ such that under the isomorphism $g_2^*(VT_{2k}) = E \cong g_1^*(VT_{2k}) = (\pi_{2k} \circ g_1)^*(V\mathrm{SO}_{2k}) = (\pi_{2k} \circ g_1')^*(V\mathrm{SO}_{2k}) = (g_1')^*(VT_{2k})$ the same turning is induced on $g_2^*(VT_{2k})$ and $(g_1')^*(VT_{2k})$. By (a), this implies that g_1' is homotopic to g_2. Therefore the map $[X, BT_{2k}] \to \mathrm{TB}_{2k}(X)$ is also injective, hence it is a bijection. □

Remark 2.19 In the constructions of this section, instead of $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ we could use one of its connected components, $\Omega_\gamma\mathrm{SO}_{2k}$ for a $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$. Then the turning bundle $\mathrm{Turn}(E)$ would be replaced with its subbundle $\mathrm{Turn}^\gamma(E)$ and BT_{2k} with its connected component $BT_{2k}^\gamma = \mathrm{Turn}^\gamma(V\mathrm{SO}_{2k})$. A bundle E over a CW–complex X is γ–turnable if and only if $\mathrm{Turn}^\gamma(E)$ has a section and if and only if its classifying map $f\colon X \to B\mathrm{SO}_{2k}$ can be lifted to BT_{2k}^γ. Moreover, $VT_{2k}^\gamma := VT_{2k}|_{BT_{2k}^\gamma}$ is universal among bundles equipped with a γ–turning.

2.3 The gauge group

For an oriented rank-$2k$ vector bundle $\pi\colon E \to B$, recall that $\mathrm{Fr}(E)$ denotes the frame bundle of E, which is a principal SO_{2k}–bundle over B. As an elementary exercise in linear algebra shows, the space of automorphisms of a vector bundle $E \to B$ is canonically homeomorphic to the gauge group of $\mathrm{Fr}(E)$, as defined in [9, Chapter 7] and we will use these topological groups interchangeably, denoting them by $\mathcal{G}_E$. In this section we relate the existence of turnings on E to the topology of $\mathcal{G}_E$.

The automorphisms $\mathbb{1}_E$ and $-\mathbb{1}_E$ define elements of $\mathcal{G}_E$. Specifically, since $-\mathbb{1}$ lies in $Z(\mathrm{SO}_{2k})$, the centre of SO_{2k}, we obtain the global map $-\mathbb{1}_E \in \mathcal{G}_E$ given by $p \mapsto p(-\mathbb{1})$. Considering $[\mathbb{1}_E], [-\mathbb{1}_E] \in \pi_0(\mathcal{G}_E)$, we see from Definition 2.3 that E is turnable if and only if $[-\mathbb{1}_E] = [\mathbb{1}_E] \in \pi_0(\mathcal{G}_E)$. Indeed, somewhat more is true, as we now explain.

Fixing a frame $p \in \mathrm{Fr}(E)$ over $b = \pi(p)$ and restricting to the fibre of $\mathrm{Fr}(E) \to B$ over b, we obtain a continuous homomorphism of topological groups $r_p\colon \mathcal{G}_E \to \mathrm{SO}_{2k}$. Replacing SO_{2k} by the mapping cylinder of r_p, we regard r_p as an inclusion and consider the pair $(\mathrm{SO}_{2k}, \mathcal{G}_E)$. A path $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ defines an element $[\gamma]_{\mathcal{G}} \in \pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E)$, by identifying $\mathbb{1}_E, -\mathbb{1}_E \in \mathcal{G}_E$ with $r_p(\mathbb{1}_E) = \mathbb{1}, r_p(-\mathbb{1}_E) = -\mathbb{1} \in \mathrm{SO}_{2k}$ respectively, and viewing γ as a path in SO_{2k} connecting $\mathbb{1}_E$ and $-\mathbb{1}_E$. Since $r_p\colon \mathcal{G}_E \to \mathrm{SO}_{2k}$ is a homomorphism of topological groups, $\pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E)$ inherits a group structure from the group structures on $\mathcal{G}_E$ and SO_{2k} and we denote the unit by e. Then we have:

Lemma 2.20 *A bundle E is γ–turnable if and only if $[\gamma]_{\mathcal{G}} = e \in \pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E)$.* □

Given the above, it is natural to consider the final segment of the homotopy long exact sequence of the pair $(\mathrm{SO}_{2k}, \mathcal{G}_E)$, which runs as follows:

$$\cdots \to \pi_1(\mathcal{G}_E) \xrightarrow{(r_p)_*} \pi_1(\mathrm{SO}_{2k}) \to \pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E) \to \pi_0(\mathcal{G}_E) \to 0. \tag{2-1}$$

Now $[\overline{\beta}]_{\mathcal{G}} = [\beta]_{\mathcal{G}} + [\eta]$, where $+$ denotes the natural action of $\pi_1(\mathrm{SO}_{2k})$ on $\pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E)$ and $[\eta]$ in $\pi_1(\mathrm{SO}_{2k})$ is the generator. We see that $[\beta]_{\mathcal{G}} = [\overline{\beta}]_{\mathcal{G}}$ if and only if $(r_p)_* \colon \pi_1(\mathcal{G}_E) \to \pi_1(\mathrm{SO}_{2k})$ is onto. For example, in Section 4.2 we shall see that there are rank-4 bundles $E \to S^4$ which are β–turnable but not $\overline{\beta}$–turnable. Applying Lemma 2.20 we see that for these bundles $[\beta]_{\mathcal{G}} \neq [\overline{\beta}]_{\mathcal{G}} \in \pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E)$ and hence the map $(r_p)_* \colon \pi_1(\mathcal{G}_E) \to \pi_1(\mathrm{SO}_{2k})$ is zero. In fact, the homomorphism $\pi_1(\mathcal{G}_E) \to \pi_1(\mathrm{SO}_{2k})$ is closely related to the "turning obstruction for the essential loop in SO_{2k}", and we next discuss turnings in a more general setting.

2.4 The central groupoid and general turnings

In this subsection we generalise the definition of a turning. Let G be a path-connected topological group with centre $Z(G)$. For the computations in this paper the groups SO_{2k}, their double covers Spin_{2k} and their quotients $\mathrm{PSO}_{2k} := \mathrm{SO}_{2k}/\{\pm\mathbb{1}\}$ will be relevant, and we will consider these groups in more detail at the end of this subsection.

Let $\pi \colon P \to B$ be a principal G–bundle over a path-connected space B. The gauge group of P, denoted by $\mathcal{G}_P$, is the group of G–equivariant fibrewise automorphisms of P. Given $z \in Z(G)$, fibrewise multiplication by z defines a central element $z_P \in Z(\mathcal{G}_P)$, where for all $p \in P$,

$$z_P(p) := p \cdot z.$$

We note that if $Z(G)$ is discrete, then the map $Z(G) \to Z(\mathcal{G}_P)$, $z \mapsto z_P$, is an isomorphism. We shall be interested in paths $\gamma \colon I \to G$ which start and end at elements of the centre $Z(G)$, and whether they can be lifted to paths in $\mathcal{G}_P$ which start and end at $\gamma(0)_P$ and $\gamma(1)_P$. Hence we make the following definition:

Definition 2.21 (central groupoid) The *central groupoid* of G is the restriction of the fundamental groupoid of G to paths which start and end in the centre of G. We will use $\pi^Z(G)$ to ambiguously denote the central groupoid of G or the set of its morphisms.

Remark 2.22 Pointwise multiplication gives $\pi^Z(G)$, the set of morphisms of the central groupoid, a group structure, and there is short exact sequence

$$1 \to \pi_1(G, e) \to \pi^Z(G) \to Z(G) \times Z(G) \to 1,$$

where $\pi^Z(G) \to Z(G) \times Z(G)$ is defined by $[\gamma] \mapsto (\gamma(0), \gamma(1))$ and $e \in G$ is the identity. While we do not use this group structure in what follows, it may be helpful for understanding $\pi^Z(G)$; eg it shows that $\pi^Z(\mathrm{SO}_{2k})$ has eight morphisms.

For a point $p \in P$, let $b = \pi(p)$ and $P_b := p \cdot G$ be the fibre of $P \to B$ over b. Define the *restriction map*

$$r_p \colon \mathcal{G}_P \to G$$

by restricting elements of the gauge group to the fibre over b and using the equation

$$\phi(p) = p \cdot r_p(\phi)$$

for all $\phi \in \mathcal{G}_P$. If we vary $p \in P_b$, then $r_{p \cdot g}(\phi) = g^{-1} r_p(\phi) g$ for all $\phi \in \mathcal{G}_P$ and $g \in G$. Recalling that G and B are path-connected, we see for any path $\phi_t \colon I \to \mathcal{G}_P$ with $\phi_0 = z_P$ and $\phi_1 = z'_P$, that $[r_p(\phi_t)] \in \pi^Z(G)$ is independent of the choice of p.

Definition 2.23 Let $\phi_t \colon I \to \mathcal{G}_P$ be a path such that $\phi_0 = z_P$ and $\phi_1 = z'_P$ for some $z, z' \in Z(G)$. We define $r(\phi_t) \in \pi^Z(G)$ to be $[r_p(\phi_t)]$ for any $p \in P$.

Definition 2.24 (γ–turning and γ–turnable) Let $[\gamma] \in \pi^Z(G)$ be represented by a path $\gamma \colon I \to G$. A *γ–turning* of a principal G–bundle P is a path $\phi_t \colon I \to \mathcal{G}_P$ with $\phi_0 = \gamma(0)_P, \phi(1) = \gamma(1)_P$ and $r(\phi_t) = [\gamma] \in \pi^Z(G)$. If P admits a γ–turning then P is called *γ–turnable*.

We end this subsection by considering the groups SO_{2k}, Spin_{2k} and PSO_{2k}. When $2k > 2$, we have $\mathrm{PSO}_{2k} = \mathrm{SO}_{2k}/Z(\mathrm{SO}_{2k}) \cong \mathrm{Spin}_{2k}/Z(\mathrm{Spin}_{2k})$ and we list the centres and fundamental groups of these groups in the following tables (where $j \geq 1$), which follow from Lemma 2.25 below:

G	$Z(G)$	$\pi_1(G)$
Spin_{4j}	$\mathbb{Z}/2 \oplus \mathbb{Z}/2$	$\{e\}$
SO_{4j}	$\mathbb{Z}/2$	$\mathbb{Z}/2$
PSO_{4j}	$\{e\}$	$\mathbb{Z}/2 \oplus \mathbb{Z}/2$

G	$Z(G)$	$\pi_1(G)$
Spin_{4j+2}	$\mathbb{Z}/4$	$\{e\}$
SO_{4j+2}	$\mathbb{Z}/2$	$\mathbb{Z}/2$
PSO_{4j+2}	$\{e\}$	$\mathbb{Z}/4$

The next lemma is well known but we include its proof to further illustrate the structure of the central groupoid of SO_{2k}.

Lemma 2.25 *If $k \geq 2$, then $Z(\mathrm{PSO}_{2k}) = \{e\}$, $Z(\mathrm{Spin}_{2k}) \cong \pi_1(\mathrm{PSO}_{2k})$ and*

$$\pi_1(\mathrm{PSO}_{2k}) \cong \begin{cases} \mathbb{Z}/2 \oplus \mathbb{Z}/2 & \text{if } k \text{ is even,} \\ \mathbb{Z}/4 & \text{if } k \text{ is odd.} \end{cases}$$

Proof To see that the centre of PSO_{2k} is trivial, let $x \in \mathrm{SO}_{2k}$ lie in the preimage of $Z(\mathrm{PSO}_{2k})$. Then the commutator $[x, \cdot\,]$ defines a map $\mathrm{SO}_{2k} \to Z(\mathrm{SO}_{2k})$. Since $Z(\mathrm{SO}_{2k})$ is discrete and $[x, \mathbb{1}] = \mathbb{1}$, this is the constant $\mathbb{1}$ map. Hence $x \in Z(\mathrm{SO}_{2k})$ and thus $Z(\mathrm{PSO}_{2k}) = \{e\}$.

If $q \colon \mathrm{Spin}_{2k} \to \mathrm{SO}_{2k}$ denotes the nontrivial double covering, we see that $Z(\mathrm{Spin}_{2k}) = q^{-1}(Z(\mathrm{SO}_{2k}))$. Now $Z(\mathrm{SO}_{2k}) = \{\pm\mathbb{1}\}$, the covering $q' \colon \mathrm{Spin}_{2k} \to \mathrm{PSO}_{2k}$ is the universal covering of PSO_{2k} and it is the composition of q and $\mathrm{SO}_{2k} \to \mathrm{PSO}_{2k}$. It follows that $Z(\mathrm{Spin}_{2k}) = (q')^{-1}([\mathbb{1}]) \cong \pi_1(\mathrm{PSO}_{2k})$.

To compute $\pi_1(\mathrm{PSO}_{2k})$ we first consider the central groupoid $\pi^Z(\mathrm{SO}_{2k})$. It is generated by the morphisms $[\beta], [\bar{\beta}]\colon \mathbb{1} \to -\mathbb{1}$, subject to the relation $([\bar{\beta}]^{-1} \circ [\beta])^2 = \mathrm{Id}_{\mathbb{1}}$; ie $[\bar{\beta}]^{-1} \circ [\beta] = [\eta]$ is the generator of $\pi_1(\mathrm{SO}_{2k}) = \mathbb{Z}/2$. The following diagram shows the named morphisms in $\pi^Z(\mathrm{SO}_{2k})$:

$$-\mathbb{1} \underset{[\bar{\beta}]}{\overset{[\beta]}{\leftleftarrows}} \mathbb{1} \circlearrowleft [\eta]$$

The projection $\mathrm{SO}_{2k} \to \mathrm{PSO}_{2k}$ induces a surjective map of groupoids $\pi^Z(\mathrm{SO}_{2k}) \to \pi_1(\mathrm{PSO}_{2k})$, which sends two morphisms $[\gamma], [\gamma'] \in \pi^Z(\mathrm{SO}_{2k})$ into the same element of $\pi_1(\mathrm{PSO}_{2k})$ if and only if $[\gamma'] = [-\gamma]$, where $-\gamma$ denotes the path γ multiplied pointwise by $-\mathbb{1}$. Since

$$[-\beta] = \begin{cases} [\beta]^{-1} & \text{if } k \text{ is even,} \\ [\bar{\beta}]^{-1} & \text{if } k \text{ is odd,} \end{cases}$$

the computation of $\pi_1(\mathrm{PSO}_{2k})$ follows. □

2.5 η–turnings and the path components of $\mathcal{G}_E$

Recall that η denotes the generator of $\pi_1(\mathrm{SO}_{2k})$ and that by definition a rank-$2k$ vector bundle $E \to B$ is η–turnable if and only if the restriction induces a surjective map $\pi_1(\mathcal{G}_E) \to \pi_1(\mathrm{SO}_{2k})$. We return to our discussion of the exact sequence (2-1) from Section 2.3 and first identify it with an isomorphic exact sequence. Assuming that $b \in B$ is nondegenerate, the homomorphism $r_p\colon \mathcal{G}_E \to \mathrm{SO}_{2k}$ is onto and there is a short exact sequence of topological groups

$$\mathcal{G}_{E,0} \to \mathcal{G}_E \xrightarrow{r_p} \mathrm{SO}_{2k}, \tag{2-2}$$

where, by definition, $\mathcal{G}_{E,0} := \mathrm{Ker}(r_p) \subset \mathcal{G}_E$. Regarding (2-2) as a principal $\mathcal{G}_{E,0}$–bundle, it is classified by a map $\mathrm{SO}_{2k} \to B\mathcal{G}_{E,0}$ such that

$$\mathcal{G}_E \xrightarrow{r_p} \mathrm{SO}_{2k} \to B\mathcal{G}_{E,0} \tag{2-3}$$

is a fibration sequence. The homotopy long exact sequence of (2-1) maps isomorphically to the homotopy long exact sequences of (2-3) and (2-2) as follows:

$$\begin{array}{ccccccccc}
\pi_1(\mathcal{G}_E) & \xrightarrow{(r_p)_*} & \pi_1(\mathrm{SO}_{2k}) & \longrightarrow & \pi_1(\mathrm{SO}_{2k}, \mathcal{G}_E) & \longrightarrow & \pi_0(\mathcal{G}_E) & \longrightarrow & 0 \\
\| & & \| & & \downarrow \cong & & \| & & \\
\pi_1(\mathcal{G}_E) & \xrightarrow{(r_p)_*} & \pi_1(\mathrm{SO}_{2k}) & \longrightarrow & \pi_1(B\mathcal{G}_{E,0}) & \longrightarrow & \pi_0(\mathcal{G}_E) & \longrightarrow & 0 \\
\| & & \| & & \downarrow \cong & & \| & & \\
\pi_1(\mathcal{G}_E) & \xrightarrow{(r_p)_*} & \pi_1(\mathrm{SO}_{2k}) & \longrightarrow & \pi_0(\mathcal{G}_{E,0}) & \longrightarrow & \pi_0(\mathcal{G}_E) & \longrightarrow & 0
\end{array} \tag{2-4}$$

Now we fix the base space B and suppose that $B = SX$ is a suspension. For any rank-$2k$ vector bundle $E \to SX$, there is a homotopy equivalence $\mathcal{G}_{E,0} \simeq \mathrm{Map}((SX, *), (\mathrm{SO}_{2k}, \mathbb{1}))$ and in particular the homotopy type of $\mathcal{G}_{E,0}$ does not depend on the vector bundle E. If $|\pi_0(\mathcal{G}_{E,0})| = |[SX, \mathrm{SO}_{2k}]|$ is finite then the exact sequences above in (2-4) show that $|\pi_0(\mathcal{G}_E)|$ depends on the η–turnability of E.

Specifically, we have the following:

Theorem 2.26 *If $B = SX$ is a suspension and $n_{SX} := |[SX, \mathrm{SO}_{2k}]|$ is finite, then for any rank-$2k$ vector bundle $E \to SX$,*

$$|\pi_0(\mathcal{G}_E)| = \begin{cases} n_{SX} & \text{if } E \text{ is } \eta\text{–turnable}, \\ \frac{1}{2} n_{SX} & \text{if } E \text{ is not } \eta\text{–turnable.} \end{cases}$$

In particular, the η–turnability of rank-$2k$ vector bundles over SX with n_{SX} finite is a homotopy invariant of the gauge groups of these bundles. □

Remark 2.27 As one might expect, if a rank-$2k$ vector bundle E admits a spin structure, then the η–turnability of E is equivalent to the γ_{z_0}–turnability of the associated principal Spin_{2k}–bundle for a certain path γ_{z_0} in Spin_{2k}. Let $e \in \mathrm{Spin}_{2k}$ be the identity and define $z_0 \in Z(\mathrm{Spin}_{2k}) \setminus \{e\}$ to be the unique element mapping to $\mathbb{1} \in \mathrm{SO}_{2k}$. Since Spin_{2k} is simply connected, there is a unique path homotopy class of paths from e to z_0, and we denote this path by γ_{z_0}. Given a principal Spin_{2k}–bundle $P \to B$, we can consider the γ_{z_0}–turning problem for P. As Spin_{2k} acts on $\mathbb{R}^{2k}$ via the double covering $\mathrm{Spin}_{2k} \to \mathrm{SO}_{2k}$ and the standard action of SO_{2k}, there is a rank-$2k$ vector bundle $E_P := P \times_{\mathrm{Spin}_{2k}} \mathbb{R}^{2k}$ associated to P and it is not hard to see that the following statements are equivalent:

(1) The principal Spin_{2k}–bundle $P \to B$ is γ_{z_0}–turnable.

(2) The vector bundle E_P associated to P is η–turnable.

(3) The map $(r_p)_* \colon \pi_1(\mathcal{G}_{E_P}) \to \pi_1(\mathrm{SO}_{2k})$ is onto.

3 The turning obstruction

In this section we define the turning obstruction for bundles over suspensions.

First we consider rank-$2k$ oriented vector bundles. Over a suspension SX such a bundle corresponds to an element of $[X, \mathrm{SO}_{2k}]$. Given a path $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ we first define a map

$$\mathrm{to}_\gamma \colon [X, \mathrm{SO}_{2k}] \to [X, \Omega_\gamma \mathrm{SO}_{2k}]$$

(where $\Omega_\gamma \mathrm{SO}_{2k} \subset \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ denotes the connected component of γ, see Definition 2.6) and prove that it is a complete obstruction to the γ–turnability of a bundle. We also introduce variants

$$\overline{\mathrm{to}}_\gamma \colon [X, \mathrm{SO}_{2k}] \to [X, \Omega_0 \mathrm{SO}_{2k}] \quad \text{and} \quad \mathrm{TO}_\gamma, \overline{\mathrm{TO}}_\gamma \colon [X, \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}]$$

and show that they are equivalent to to_γ. Finally we prove that these maps are homomorphisms if X is a suspension.

In Section 3.2 we consider bundles with a path-connected structure group G. Given a path γ between elements of the centre $Z(G)$, we define a generalised turning obstruction map $\mathrm{to}_\gamma \colon [X, G] \to [X, \Omega_\gamma G]$. If we fix an element $[g] \in [X, G]$, then we can regard $\mathrm{to}_\gamma([g])$ as a function of γ and we show that it is compatible with concatenation of paths. We also consider a normalised version of the turning obstruction, $\overline{\mathrm{to}}_\gamma \colon [X, G] \to [X, \Omega_0 G]$. This allows us to compare turning obstructions for different paths and we find that $\overline{\mathrm{to}}_\gamma = \overline{\mathrm{to}}_{a\gamma}$ for any $a \in Z(G)$ (where $(a\gamma)(t) = a\gamma(t)$). When $Z(G)$ is discrete, we introduce the quotient $PG = G/Z(G)$ and use these observations to show that the turning obstructions are determined by a map $\widehat{\mathrm{to}}_\cdot(\cdot)\colon \pi_1(PG) \times [X, G] \to [X, \Omega_0 G]$, which is a homomorphism in the first variable (and also in the second one, if X is a suspension). As an application we prove Theorem 1.7.

3.1 The turning obstruction for vector bundles

Let X be a CW–complex and C_0X and C_1X two copies of the cone on X, so that $SX = C_0X \cup_X C_1X$. By [9, Chapter 8, Theorem 8.2], the set of isomorphism classes of oriented rank-$2k$ bundles over SX is in bijection with $[X, \mathrm{SO}_{2k}]$, the set of homotopy classes of maps from X to SO_{2k}. A bundle E corresponds to its *clutching function* $g\colon X \to \mathrm{SO}_{2k}$ between two local trivialisations $\varphi_i\colon C_iX \times \mathbb{R}^{2k} \to E|_{C_iX}$, defined by $\varphi_0^{-1}|_X \circ \varphi_1|_X(x, v) = (x, g(x)v)$.

Definition 3.1 (the γ–turning obstruction) Let $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$.

(a) We define the map $\rho_\gamma\colon \mathrm{SO}_{2k} \to \Omega_\gamma \mathrm{SO}_{2k}$ by $\rho_\gamma(A) = (t \mapsto A\gamma(t)A^{-1})$.

(b) For any CW–complex X the *γ–turning obstruction map* is $\mathrm{to}_\gamma := (\rho_\gamma)_*\colon [X, \mathrm{SO}_{2k}] \to [X, \Omega_\gamma \mathrm{SO}_{2k}]$.

Let $0 \in [X, \Omega_\gamma \mathrm{SO}_{2k}]$ denote the homotopy class of the constant map. Definition 3.1 is justified by the following:

Proposition 3.2 *Let E be an oriented rank-$2k$ bundle over SX with clutching function $g\colon X \to \mathrm{SO}_{2k}$. Then E is γ–turnable if and only if* $\mathrm{to}_\gamma([g]) = 0 \in [X, \Omega_\gamma \mathrm{SO}_{2k}]$.

Proof Recall that γ–turnings of E can be identified with sections of the associated γ–turning bundle $\mathrm{Turn}^\gamma(E) = \mathrm{Fr}(E) \times_{\mathrm{SO}_{2k}} \Omega_\gamma \mathrm{SO}_{2k}$; see Lemma 2.12 and Remark 2.19. The local trivialisations φ_i of E induce local trivialisations $\overline{\varphi}_i\colon C_iX \times \Omega_\gamma \mathrm{SO}_{2k} \to \mathrm{Turn}^\gamma(E)|_{C_iX}$ of $\mathrm{Turn}^\gamma(E)$. By construction, the clutching function $g\colon X \to \mathrm{SO}_{2k}$ is also the clutching function of $\mathrm{Turn}^\gamma(E)$ — recall that SO_{2k} acts on $\Omega_\gamma \mathrm{SO}_{2k}$ by pointwise conjugation.

Since C_iX is contractible and $\Omega_\gamma \mathrm{SO}_{2k}$ is connected, each restriction $\mathrm{Turn}^\gamma(E)|_{C_iX}$ has a unique section $s_i\colon C_iX \to \mathrm{Turn}^\gamma(E)|_{C_iX}$ up to homotopy, given by $s_i(y) = \overline{\varphi}_i(y, \gamma)$. Hence a global section of $\mathrm{Turn}^\gamma(E)$ exists if and only if $s_0|_X$ and $s_1|_X$ are homotopic sections. For $x \in X$ we have

$$s_0(x) = \overline{\varphi}_0(x, \gamma) \quad \text{and} \quad s_1(x) = \overline{\varphi}_1(x, \gamma) = \overline{\varphi}_0 \circ \overline{\varphi}_0^{-1} \circ \overline{\varphi}_1(x, \gamma) = \overline{\varphi}_0(x, \gamma^{g(x)}) = \overline{\varphi}_0(x, \rho_\gamma(g(x))),$$

where $\gamma^{g(x)}$ denotes the action of $g(x) \in \mathrm{SO}_{2k}$ on $\gamma \in \Omega_\gamma \mathrm{SO}_{2k}$. These sections are homotopic if and only if $\rho_\gamma \circ g\colon X \to \Omega_\gamma \mathrm{SO}_{2k}$ is homotopic to the constant map with value γ, ie if and only if $\mathrm{to}_\gamma([g]) = 0$. □

The turning obstruction is a map of pointed sets, but $[X, \mathrm{SO}_{2k}]$ is a group and we will show that $[X, \Omega_\gamma \mathrm{SO}_{2k}]$ can also be equipped with a group structure and that to_γ and related maps are often group homomorphisms; see Lemma 3.6.

Let $\Omega_0 \mathrm{SO}_{2k} \subset \Omega \mathrm{SO}_{2k}$ denote the component of contractible loops and define the homeomorphism

$$p_\gamma \colon \Omega_\gamma \mathrm{SO}_{2k} \to \Omega_0 \mathrm{SO}_{2k}, \quad \delta \mapsto (t \mapsto \delta(t)\gamma(t)^{-1}),$$

as well as the commutator map $\overline{\rho}_\gamma := p_\gamma \circ \rho_\gamma$,

$$\overline{\rho}_\gamma \colon \mathrm{SO}_{2k} \to \Omega_0 \mathrm{SO}_{2k}, \quad A \mapsto (t \mapsto A\gamma(t)A^{-1}\gamma(t)^{-1}).$$

Definition 3.3 (normalised γ–turning obstruction) Let $\gamma \in \Omega_{\pm\mathbb{1}} \mathrm{SO}_{2k}$. For any CW–complex X, the *normalised γ–turning obstruction map* is $\overline{\mathrm{to}}_\gamma := (\overline{\rho}_\gamma)_* \colon [X, \mathrm{SO}_{2k}] \to [X, \Omega_0 \mathrm{SO}_{2k}]$.

Since p_γ is a homeomorphism, the induced map $(p_\gamma)_* \colon [X, \Omega_\gamma \mathrm{SO}_{2k}] \to [X, \Omega_0 \mathrm{SO}_{2k}]$, $[h] \mapsto [p_\gamma \circ h]$, is a bijection which preserves 0, and hence an oriented rank-$2k$ bundle $E \to SX$ with clutching function $g \colon X \to \mathrm{SO}_{2k}$ is γ–turnable if and only if $\overline{\mathrm{to}}_\gamma([g]) = 0$. For computing to_γ and $\overline{\mathrm{to}}_\gamma$ it is useful to consider their adjointed versions, which we will define below.

Definition 3.4 (forgetful adjoints) Let $h \colon X \to \Omega_\gamma \mathrm{SO}_{2k}$ and $h' \colon X \to \Omega_0 \mathrm{SO}_{2k}$ be continuous maps. By taking their adjoints

$$\mathrm{ad}(h) \colon SX \to \mathrm{SO}_{2k}, \quad [x,t] \mapsto h(x)(t), \qquad \text{and} \qquad \mathrm{ad}(h') \colon SX \to \mathrm{SO}_{2k}, \quad [x,t] \mapsto h'(x)(t),$$

we define the *forgetful adjoint maps* $\mathrm{ad} \colon [X, \Omega_\gamma \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}]$ and $\mathrm{ad} \colon [X, \Omega_0 \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}]$. (We call these maps "forgetful" because $[\mathrm{ad}(h)]$ and $[\mathrm{ad}(h')]$ are regarded as elements of $[SX, \mathrm{SO}_{2k}]$ rather than of the more restricted sets of homotopy classes on which the inverse adjoint maps $[\mathrm{ad}(h)] \mapsto [h]$ and $[\mathrm{ad}(h')] \mapsto [h']$ are naturally defined.)

If X is connected, then the forgetful adjoint maps are bijections and they preserve 0, the homotopy class of the constant map.

Definition 3.5 (adjointed γ–turning obstructions) Define $\mathrm{TO}_\gamma := \mathrm{ad} \circ \mathrm{to}_\gamma \colon [X, \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}]$ and $\overline{\mathrm{TO}}_\gamma := \mathrm{ad} \circ \overline{\mathrm{to}}_\gamma \colon [X, \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}]$.

Lemma 3.6 (a) *Let X be a CW–complex. Then $\mathrm{TO}_\gamma = \overline{\mathrm{TO}}_\gamma \colon [X, \mathrm{SO}_{2k}] \to [SX, \mathrm{SO}_{2k}]$.*

(b) *If X is a suspension, then $\mathrm{to}_\gamma, \overline{\mathrm{to}}_\gamma, \mathrm{TO}_\gamma$ and $\overline{\mathrm{TO}}_\gamma$ are each homomorphisms of abelian groups.*

Proof (a) We show that the diagram

$$\begin{array}{ccccc} & & [X, \Omega_\gamma \mathrm{SO}_{2k}] & \xrightarrow{\mathrm{ad}} & [SX, \mathrm{SO}_{2k}] \\ & \overset{\mathrm{to}_\gamma}{\nearrow} & \Big\downarrow (p_\gamma)_* & & \Big\| \\ [X, \mathrm{SO}_{2k}] & & & & \\ & \underset{\overline{\mathrm{to}}_\gamma}{\searrow} & & & \\ & & [X, \Omega_0 \mathrm{SO}_{2k}] & \xrightarrow{\mathrm{ad}} & [SX, \mathrm{SO}_{2k}] \end{array}$$

commutes. The left-hand triangle commutes be definition. For the right-hand square, consider the path of paths $s \mapsto \gamma_s$, where $\gamma_s \colon I \to \mathrm{SO}_{2k}$ is defined by $\gamma_s(t) = \gamma(st)$ for $s, t \in I$. Then the map

$$H \colon SX \times I \to \mathrm{SO}_{2k}, \quad ([x,t], s) \mapsto g(x)\gamma(t)g(x)^{-1}\gamma_s(t)^{-1},$$

is a homotopy from $\mathrm{ad}(\rho_\gamma \circ g)$ to $\mathrm{ad}(\overline{\rho}_\gamma \circ g)$, which proves that the square commutes.

(b) First note that $\Omega_0\mathrm{SO}_{2k}$ is a topological group (via pointwise multiplication of loops), so we can use the homeomorphism p_γ to get a topological group structure on $\Omega_\gamma\mathrm{SO}_{2k}$. Hence for $H = \mathrm{SO}_{2k}$, $\Omega_0\mathrm{SO}_{2k}$ or $\Omega_\gamma\mathrm{SO}_{2k}$ and any space Y the set $[Y, H]$ inherits a group structure from H (and with these group structures $(p_\gamma)_*$ is automatically an isomorphism).

If Y is pointed, then the set $[Y, H]_*$ of homotopy classes of basepoint-preserving maps is also a group and the forgetful map $[Y, H]_* \to [Y, H]$ is an isomorphism. If Y is a suspension, then for any space Z the set $[Y, Z]_*$ has a group structure coming from the co-H-space structure on Y and on the sets $[Y, H]_*$ the two group structures coincide and they are abelian. Since the group structure on $[Y, Z]_*$ is natural in Z, it follows that if X is a suspension, then the maps to_γ and $\overline{\mathrm{to}}_\gamma$ (which are induced by maps of spaces) are homomorphisms (and all groups involved are abelian).

Finally, loop concatenation gives $\Omega_0\mathrm{SO}_{2k}$ an H-space structure and the induced group structure on $[X, \Omega_0\mathrm{SO}_{2k}]$ coincides with the previously defined one. By comparing this with the group structure on $[SX, \mathrm{SO}_{2k}]$ coming from the suspension SX, we obtain that the forgetful adjoint maps and hence TO_γ and $\overline{\mathrm{TO}}_\gamma$, are also homomorphisms. □

Remark 3.7 By Lemma 3.6(b), if $X = SY$ is a suspension, then the set of isomorphism classes of γ–turnable bundles over $SX = S^2Y$ can be identified with a subgroup of $[X, \mathrm{SO}_{2k}]$.

Question 3.8 (a) The sets $[X, \mathrm{SO}_{2k}]$ and $[SX, \mathrm{SO}_{2k}]$ have natural group structures even when X is not a suspension (or co-H-space). Is TO_γ a group homomorphism for an arbitrary X?

(b) Isomorphism classes of rank-$2k$ oriented bundles over a space B are in bijection with $[B, B\mathrm{SO}_{2k}]$, so when B is a suspension, TO_γ can be regarded as a map $[B, B\mathrm{SO}_{2k}] \to [B, \mathrm{SO}_{2k}]$. Is there a similar γ–turning obstruction map for bundles over an arbitrary space B?

We next briefly discuss the behaviour of the turning obstruction under stabilisation: we will return to this topic in greater detail in Section 5. Let $i \colon \mathrm{SO}_{2k} \to \mathrm{SO}_{2k+2}$ denote the standard inclusion and let $S = i_* \colon [X, \mathrm{SO}_{2k}] \to [X, \mathrm{SO}_{2k+2}]$ denote the stabilisation map induced by i. Given a path $\gamma_0 \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_2$, taking the orthogonal sum with γ_0 defines a map $i_{\gamma_0} \colon \Omega_\gamma\mathrm{SO}_{2k} \to \Omega_{\gamma\oplus\gamma_0}\mathrm{SO}_{2k+2}$. It is clear from the definitions that the turning obstructions satisfy $\mathrm{to}_{\gamma\oplus\gamma_0}([i \circ g]) = i_{\gamma_0 *}(\mathrm{to}_\gamma([g]))$ and indeed we have:

Lemma 3.9 *Let X be a CW–complex and $g \colon X \to \mathrm{SO}_{2k}$ a map. Then the adjointed γ–turning obstruction satisfies* $\mathrm{TO}_{\gamma\oplus\gamma_0}([i \circ g]) = S(\mathrm{TO}_\gamma([g]))$. *In particular,* $\mathrm{TO}_\beta([i \circ g]) = \mathrm{TO}_{\overline{\beta}}([i \circ g])$.

Proof Write $A \oplus B \in \mathrm{SO}_{2k+2}$ for the block sum of matrices $A \in \mathrm{SO}_{2k}$ and $B \in \mathrm{SO}_2$ and consider the path of paths $s \mapsto (\gamma_0)_s$, where $(\gamma_0)_s\colon I \to \mathrm{SO}_2$ is defined by $(\gamma_0)_s(t) = \gamma_0(st)$ for $s, t \in I$. Then

$$H\colon SX \times I \to \mathrm{SO}_{2k+2}, \quad ([x,t], s) \mapsto g(x)\gamma(t)g(x)^{-1} \oplus (\gamma_0)_{(1-s)}(t),$$

is a homotopy from $\mathrm{ad}(\rho_{\gamma\oplus\gamma_0} \circ (i \circ g))$ to $i \circ \mathrm{ad}(\rho_\gamma(g))$, which proves the first statement of the lemma. In particular, $\mathrm{TO}_{\gamma\oplus\gamma_0}([i \circ g])$ is independent of the choice of $\gamma_0 \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_2$. Since for any $\gamma \in \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ the map $\pi_0(\Omega_{\pm\mathbb{1}}\mathrm{SO}_2) \to \pi_0(\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k+2})$ given by $[\gamma_0] \mapsto [\gamma \oplus \gamma_0]$ is surjective, the second statement follows. □

We conclude this subsection with a remark on a related point of view on the turning obstruction.

Remark 3.10 Consider the associated γ–turning bundle of the universal bundle $V\mathrm{SO}_{2k}$ (see Definition 2.13 and Remark 2.19) and the Puppe sequence

$$\cdots \to \Omega BT^\gamma_{2k} \to \Omega B\mathrm{SO}_{2k} \to \Omega_\gamma \mathrm{SO}_{2k} \to BT^\gamma_{2k} \to B\mathrm{SO}_{2k},$$

where Ω denotes the based loops functor. After applying the functor $[X, -]_*$ to this sequence and the adjunction $[X, \Omega Y]_* \cong [\Sigma X, Y]_*$, where ΣX denotes the reduced suspension, we obtain an exact sequence

$$\cdots \to [\Sigma X, BT^\gamma_{2k}]_* \to [\Sigma X, B\mathrm{SO}_{2k}]_* \xrightarrow{\partial} [X, \Omega_\gamma \mathrm{SO}_{2k}]_* \to [X, BT^\gamma_{2k}]_* \to [X, B\mathrm{SO}_{2k}]_*.$$

The arguments in the proof of Proposition 3.2 also show that the γ–turning obstruction can be identified with the boundary map ∂ (note that $[\Sigma X, B\mathrm{SO}_{2k}]_* \cong [X, \mathrm{SO}_{2k}]_* \cong [X, \mathrm{SO}_{2k}]$ and $[X, \Omega_\gamma \mathrm{SO}_{2k}]_* \cong [X, \Omega_\gamma \mathrm{SO}_{2k}]$). So by using the exactness of the sequence we obtain an alternative proof of Proposition 2.15 for bundles over suspensions.

3.2 General turning obstructions

In this subsection we define the turning obstruction for the general turnings of principal bundles, as in Section 2.4. This will help us establish some basic properties of the turning obstruction for vector bundles.

Definition 3.11 Let G be a path-connected topological group, $a, b \in G$ arbitrary elements and $\gamma\colon I \to G$ a path in G. We introduce the following notation:

- ΩG denotes the space of loops in G based at the identity element.
- $\Omega_0 G \subseteq \Omega G$ is the space of nullhomotopic loops, ie the connected component of the constant loop.
- $\Omega_{a,b} G$ is the space of paths in G from a to b.
- $\Omega_\gamma G$ is the space of paths homotopic (rel ∂I) to γ, ie the connected component of γ in $\Omega_{\gamma(0),\gamma(1)} G$.
- $\Omega_Z G = \bigcup_{a,b \in Z(G)} \Omega_{a,b} G$, where $Z(G)$ denotes the centre of G.

Definition 3.12 (parametrised central groupoid of G) Let G be a path-connected topological group. The *parametrised central groupoid of* G has objects the elements of $Z(G)$. The set of all morphisms is $[G, \Omega_Z G]$. The set of morphisms from a to b is $[G, \Omega_{a,b} G]$. Given objects $a, b, c \in Z(G)$ and maps $f_1 \colon G \to \Omega_{a,b} G$ and $f_2 \colon G \to \Omega_{b,c} G$, the composition $[f_2] \circ [f_1]$ is represented by $f_1 * f_2 \colon G \to \Omega_{a,c} G$, defined by $(f_1 * f_2)(x) = f_1(x) * f_2(x)$, where $*$ denotes concatenation of paths.

Let G be a path-connected topological group and recall that for a CW–complex X, isomorphism classes of principal G–bundles over SX with structure group G are in bijection with $[X, G]$.

Definition 3.13 (general γ–turning obstruction) Let $\gamma \in \Omega_Z G$.

(a) We define the map $\rho_\gamma \colon G \to \Omega_\gamma G$ by $\rho_\gamma(x) = (t \mapsto x\gamma(t)x^{-1})$.

(b) For any CW–complex X the *γ–turning obstruction map* is $\mathrm{to}_\gamma := (\rho_\gamma)_* \colon [X, G] \to [X, \Omega_\gamma G]$.

Note that the image of ρ_γ is contained in $\Omega_{\gamma(0),\gamma(1)} G$, because $\gamma(0), \gamma(1) \in Z(G)$, and in particular in the component $\Omega_\gamma G$, because G is path-connected and ρ_γ sends the identity element to γ. We let $0 \in [X, \Omega_\gamma G]$ denote the homotopy class of the constant map. The proof of the following proposition is entirely analogous to the proof of Proposition 3.2.

Proposition 3.14 *Suppose that P is a G–bundle over SX with clutching function $g \colon X \to G$. Then P is γ–turnable if and only if* $\mathrm{to}_\gamma([g]) = 0 \in [X, \Omega_\gamma G]$. □

Definition 3.15 Let $\mathrm{to}_G \colon \pi^Z(G) \to [G, \Omega_Z G]$ be defined by $\mathrm{to}_G([\gamma]) = [\rho_\gamma]$.

This map is well defined, because a path homotopy between γ and γ' determines a homotopy between the maps ρ_γ and $\rho_{\gamma'}$. Since composition is defined in terms of concatenation both in $\pi^Z(G)$ and $[G, \Omega_Z G]$, it is a map of groupoids (with the identity map on the objects). With this notation,

$$\mathrm{to}_\gamma = \mathrm{to}_G([\gamma])_* \colon [X, G] \to [X, \Omega_\gamma G] \subseteq [X, \Omega_Z G].$$

Definition 3.16 Let $PG = G/Z(G)$. Let $p_G \colon \pi^Z(G) \to \pi_1(PG)$ denote the map of groupoids induced by the projection $G \to PG$ (where $\pi_1(PG)$ is regarded as a groupoid on one object).

Every $\gamma \in \Omega_Z G$ determines a homeomorphism $p'_\gamma \colon \Omega_\gamma G \to \Omega_0 G$ which sends a path δ to the loop $t \mapsto \delta(t)\gamma(t)^{-1}$. If $[\gamma] = [\gamma']$, so that $\Omega_\gamma G = \Omega_{\gamma'} G$, then these homeomorphisms are homotopic; hence they induce a well-defined map $p'_{[\gamma]} \colon [G, \Omega_\gamma G] \to [G, \Omega_0 G]$. On the other hand, if $[\gamma] \neq [\gamma']$, then $\Omega_\gamma G$ and $\Omega_{\gamma'} G$ (and hence $[G, \Omega_\gamma G]$ and $[G, \Omega_{\gamma'} G]$) are disjoint, so p'_G below is well defined:

Definition 3.17 Let $p'_G \colon [G, \Omega_Z G] \to [G, \Omega_0 G]$ be the union of the maps $p'_{[\gamma]} \colon [G, \Omega_\gamma G] \to [G, \Omega_0 G]$.

Since $\Omega_0 G$ is an H-space, $[G, \Omega_0 G]$ is a group, ie a groupoid on one object. For every pair γ, γ' of composable paths in $\Omega_Z G$, the diagram

$$\begin{array}{ccc} \Omega_\gamma G \times \Omega_{\gamma'} G & \xrightarrow{\ *\ } & \Omega_{\gamma * \gamma'} G \\ {\scriptstyle p'_\gamma \times p'_{\gamma'}} \downarrow & & \downarrow {\scriptstyle p'_{\gamma*\gamma'}} \\ \Omega_0 G \times \Omega_0 G & \xrightarrow{\ *\ } & \Omega_0 G \end{array}$$

commutes. Hence p'_G is a map of groupoids.

Proposition 3.18 *Suppose that G is a path-connected topological group and $Z(G)$ is discrete. Then to_G descends to a homomorphism $\widehat{\mathrm{to}}_G\colon \pi_1(PG) \to [G, \Omega_0 G]$ of abelian groups, ie there is a commutative diagram of groupoids:*

$$\begin{array}{ccc} \pi^Z(G) & \xrightarrow{\mathrm{to}_G} & [G, \Omega_Z G] \\ {\scriptstyle p_G} \downarrow & & \downarrow {\scriptstyle p'_G} \\ \pi_1(PG) & \xrightarrow{\widehat{\mathrm{to}}_G} & [G, \Omega_0 G] \end{array}$$

Proof First we define $\widehat{\mathrm{to}}_G$. Since $Z(G)$ is discrete, the projection $G \to PG$ is a covering, so a loop $\gamma \in \Omega PG$ can be lifted to a path $\widetilde{\gamma} \in \Omega_Z G$ and we define $\widehat{\mathrm{to}}_G([\gamma]) = p'_G \circ \mathrm{to}_G([\widetilde{\gamma}])$. If $\widetilde{\gamma}'$ is another lift of γ, then $\widetilde{\gamma}' = a\widetilde{\gamma}$ for $a = \widetilde{\gamma}'(0)\widetilde{\gamma}(0)^{-1} \in Z(G)$. Then

$$(p'_{\widetilde{\gamma}'} \circ \rho_{\widetilde{\gamma}'}(x))(t) = x\widetilde{\gamma}'(t)x^{-1}\widetilde{\gamma}'(t)^{-1} = xa\widetilde{\gamma}(t)x^{-1}\widetilde{\gamma}(t)^{-1}a^{-1} = x\widetilde{\gamma}(t)x^{-1}\widetilde{\gamma}(t)^{-1} = (p'_{\widetilde{\gamma}} \circ \rho_{\widetilde{\gamma}}(x))(t),$$

and so $p'_G \circ \mathrm{to}_G([\widetilde{\gamma}']) = p'_G \circ \mathrm{to}_G([\widetilde{\gamma}])$. Therefore $\widehat{\mathrm{to}}_G([\gamma])$ does not depend on the choice of the lift $\widetilde{\gamma}$. A homotopy of γ can be lifted to a homotopy of $\widetilde{\gamma}$, so $\widehat{\mathrm{to}}_G([\gamma])$ is also independent of the choice of the representative γ of the homotopy class $[\gamma]$. Therefore $\widehat{\mathrm{to}}_G$ is well defined. By its construction, the diagram commutes. Finally, a lift of the concatenation of two loops in ΩPG is the concatenation of lifts of the loops, so $\widehat{\mathrm{to}}_G$ is a map of groupoids, ie a group homomorphism. □

In light of the above, it is useful define the normalised turning obstruction map in the general setting:

Definition 3.19 Suppose that G is a path-connected topological group and $\gamma \in \Omega_Z G$.

(a) Let $\bar{\rho}_\gamma\colon G \to \Omega_0 G$ be defined by $\bar{\rho}_\gamma(x) = (t \mapsto x\gamma(t)x^{-1}\gamma(t)^{-1})$.

(b) For any CW–complex X, let $\overline{\mathrm{to}}_\gamma = (\bar{\rho}_\gamma)_*\colon [X, G] \to [X, \Omega_0 G]$.

That is, $\bar{\rho}_\gamma = p'_\gamma \circ \rho_\gamma$ and hence $\overline{\mathrm{to}}_\gamma = (p'_\gamma)_* \circ \mathrm{to}_\gamma\colon [X, G] \to [X, \Omega_0 G]$. Since p'_γ is a homeomorphism, $(p'_\gamma)_*\colon [X, \Omega_\gamma G] \to [X, \Omega_0 G]$ is a bijection, so computing $\overline{\mathrm{to}}_\gamma$ is equivalent to computing to_γ. In particular, $\overline{\mathrm{to}}_\gamma$ and to_γ vanish for the same bundles. From our earlier arguments we have:

Proposition 3.20 *Suppose that G is a path-connected topological group and $g\colon X \to G$ is a map.*

(a) *If $\gamma, \gamma' \in \Omega_Z G$ are composable paths, then $\overline{\mathrm{to}}_{\gamma*\gamma'}([g]) = \overline{\mathrm{to}}_\gamma([g]) + \overline{\mathrm{to}}_{\gamma'}([g])$.*

(b) *For any $\gamma \in \Omega_Z G$ and $a \in Z(G)$, we have $\overline{\mathrm{to}}_{a\gamma}([g]) = \overline{\mathrm{to}}_\gamma([g])$.* □

If $Z(G)$ is discrete, then we can use Proposition 3.18 to describe $\overline{\mathrm{to}}_\gamma([g])$, regarded as a two-variable function in γ and $[g]$, in terms of a simpler function. Hence we define

$$\widehat{\mathrm{to}}_{\bullet}(\cdot)\colon \pi_1(PG)\times[X,G]\to[X,\Omega_0 G],\quad \widehat{\mathrm{to}}_{[\gamma]}([g])=\widehat{\mathrm{to}}_G([\gamma])\circ[g].$$

Equivalently, $\widehat{\mathrm{to}}_{[\gamma]}([g])=g^*(\widehat{\mathrm{to}}_G([\gamma]))=\widehat{\mathrm{to}}_G([\gamma])_*([g])$.

Proposition 3.21 *Suppose that G is a path-connected topological group and $Z(G)$ is discrete. For every CW–complex X, the maps $\overline{\mathrm{to}}_{\bullet}(\cdot)$ and $\widehat{\mathrm{to}}_{\bullet}(\cdot)$ satisfy:*

(a) $\overline{\mathrm{to}}_\gamma([g])=\widehat{\mathrm{to}}_{p_G([\gamma])}([g])$ *for every* $\gamma\in\Omega_Z G$ *and* $[g]\in[X,G]$.

(b) $\widehat{\mathrm{to}}_{\bullet}([g])\colon\pi_1(PG)\to[X,\Omega_0 G]$ *is a homomorphism of abelian groups for every* $[g]\in[X,G]$.

(c) *If X is a suspension, then the map* $\widehat{\mathrm{to}}_{[\gamma]}\colon[X,G]\to[X,\Omega_0 G]$ *is a homomorphism of abelian groups for every* $[\gamma]\in\pi_1(PG)$.

Proof Part (a) follows from the commutativity of the diagram in Proposition 3.18.

Part (b) holds, because $\widehat{\mathrm{to}}_G$ is a homomorphism and the induced map $g^*\colon[G,\Omega_0 G]\to[X,\Omega_0 G]$ is a homomorphism for every $[g]\in[X,G]$, because $\Omega_0 G$ is an H-space.

Part (c) holds, because $\widehat{\mathrm{to}}_{[\gamma]}$ is induced by a map of spaces. □

Remark 3.22 Just as with the turning obstruction for vector bundles, we can take the forgetful adjoints of to_γ and $\overline{\mathrm{to}}_\gamma$ and define

$$\mathrm{TO}_\gamma:=\mathrm{ad}\circ\mathrm{to}_\gamma\colon[X,G]\to[SX,G]\quad\text{and}\quad\overline{\mathrm{TO}}_\gamma:=\mathrm{ad}\circ\overline{\mathrm{to}}_\gamma\colon[X,G]\to[SX,G].$$

We leave the reader to formulate and verify the obvious generalisation of Lemma 3.6. Then by Proposition 3.14, a principal G–bundle $P\to SX$ with clutching function $g\colon X\to G$ is γ–turnable if and only if $\mathrm{TO}_\gamma([g])=0$, equivalently if $\overline{\mathrm{TO}}_\gamma([g])=0$.

We now deduce some consequences of Proposition 3.21. Recall that $[\eta]\in\pi_1(\mathrm{SO}_{2k})$ is the generator.

Theorem 3.23 *Let $\gamma\in\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$, let X be a CW–complex and $E\to SX$ a rank-$2k$ vector bundle with clutching function $g\colon X\to\mathrm{SO}_{2k}$. Then:*

(a) $2\,\overline{\mathrm{to}}_\eta([g])=0$.

(b) $\overline{\mathrm{to}}_{\eta*\gamma}([g])=\overline{\mathrm{to}}_\eta([g])+\overline{\mathrm{to}}_\gamma([g])$.

(c) *If k is even, then* $2\,\overline{\mathrm{to}}_\gamma([g])=0$.

(d) *If k is odd, then* $2\,\overline{\mathrm{to}}_\gamma([g])=\overline{\mathrm{to}}_\eta([g])$ *and hence* $4\,\overline{\mathrm{to}}_\gamma([g])=0$.

Proof Parts (a) and (b) are direct applications of Proposition 3.20 a).

Parts (c) and (d) follow from Proposition 3.21(a), Proposition 3.21(b) and the fact that $\pi_1(\mathrm{PSO}_{2k})\cong(\mathbb{Z}/2)^2$ when k is even and $\pi_1(\mathrm{PSO}_{2k})\cong\mathbb{Z}/4$ when $k>1$ is odd; see Lemma 2.25. □

3.3 Samelson products and turning obstructions

In this subsection we relate turning obstructions to Samelson products. The Samelson product is a classical operation in algebraic topology [15], and Samelson products can be delicate to compute. First, we show that taking the Samelson product with some loop $[\gamma] \in \pi_1(G)$ coincides with the normalised turning obstruction map $\overline{\mathrm{TO}}_\gamma$ (after suitable identifications); see Lemma 3.24. Second, we show that turning obstructions in G are determined by Samelson products in PG; see Corollary 3.27. As an application we determine some Samelson products based on our calculations of turning obstructions in Section 4.3; see Proposition 3.29. Finally, we show that our results on the η–turning obstruction have consequences for the high-dimensional topology of related gauge groups.

We start by recalling the definition of the Samelson product, in the special case when one of the operands is a loop. Assume that X is connected with $x_0 \in X$ a basepoint and let $g\colon (X, x_0) \to (G, e)$ be a based map. Let $\gamma\colon (S^1, 1) \to (G, e)$ be a map representing $[\gamma] \in \pi_1(G, e)$. Then there is a well-defined map

$$\mathrm{comm}_{g,\gamma}\colon X \wedge S^1 = \Sigma X \to G, \quad [x, t] \mapsto g(x)\gamma(t)g(x)^{-1}\gamma(t)^{-1}.$$

This construction gives rise to the *Samelson product*

$$[X, G]_* \times \pi_1(G, e) \to [\Sigma X, G]_*, \quad ([g], [\gamma]) \mapsto \langle [g], [\gamma] \rangle := [\mathrm{comm}_{g,\gamma}].$$

We can identify the set $[\Sigma X, G]_*$ with $[SX, G]$ via the forgetful map $[\Sigma X, G]_* \to [\Sigma X, G]$ and the map $[\Sigma X, G] \to [SX, G]$ induced by the collapse map $SX \to \Sigma X$ (which are both isomorphisms). The following lemma is a direct consequence of the definitions, where $\overline{\mathrm{TO}}_\gamma$ is defined in Remark 3.22.

Lemma 3.24 *For any $[\gamma] \in \pi_1(G, e)$ and $g\colon (X, x_0) \to (G, e)$, we have*

$$\overline{\mathrm{TO}}_\gamma([g]) = \langle [g], [\gamma] \rangle \in [SX, G]. \qquad \square$$

Lemma 3.24 implies that certain Samelson products can be computed as a special case of turning obstructions. On the other hand, we next show that turning obstructions can be computed from certain Samelson products.

Definition 3.25 Let G be a path-connected topological group. Define the map $\mathrm{sp}_G\colon \pi_1(G) \to [G, \Omega_0 G]$ by $\mathrm{sp}_G([\gamma]) = [x \mapsto (t \mapsto x\gamma(t)x^{-1}\gamma(t)^{-1})]$.

The map sp_G is a homomorphism, because the group structure of $[G, \Omega_0 G]$ can be defined via concatenation in $\Omega_0 G$. This homomorphism encodes the Samelson product (similarly to how to_G encodes the turning obstruction): if $[\gamma] \in \pi_1(G)$ and $g\colon X \to G$, then $\langle [g], [\gamma] \rangle \in [\Sigma X, G]_*$ is the adjoint of $\mathrm{sp}_G([\gamma]) \circ [g] \in [X, \Omega_0 G]_*$.

Let $\pi\colon G \to PG$ denote the projection.

Proposition 3.26 *Suppose that G is a path-connected topological group and $Z(G)$ is discrete. Then there is a commutative diagram of groups*

$$\begin{array}{ccc} \pi_1(PG) & \xrightarrow{\widehat{\mathrm{to}}_G} & [G, \Omega_0 G] \\ \Big\downarrow{\scriptstyle \mathrm{sp}_{PG}} & & \Big\downarrow{\scriptstyle (\Omega_0\pi)_*} \\ [PG, \Omega_0 PG] & \xrightarrow{\pi^*} & [G, \Omega_0 PG] \end{array}$$

where $(\Omega_0\pi)_*$ is an isomorphism.

Proof Let $\gamma \in \Omega PG$. Since $Z(G)$ is discrete, the projection $\pi\colon G \to PG$ is a covering, so γ can be lifted to a path $\tilde{\gamma} \in \Omega_Z G$. By definition we have $\widehat{\mathrm{to}}_G([\gamma]) = [x \mapsto (t \mapsto x\tilde{\gamma}(t)x^{-1}\tilde{\gamma}(t)^{-1})]$. Its image in $[G, \Omega_0 PG]$ is

$$[x \mapsto (t \mapsto \pi(x\tilde{\gamma}(t)x^{-1}\tilde{\gamma}(t)^{-1}))] = [x \mapsto (t \mapsto \pi(x)\gamma(t)\pi(x)^{-1}\gamma(t)^{-1})],$$

using that $\pi \circ \tilde{\gamma} = \gamma$. The image of $\mathrm{sp}_G([\gamma]) = [y \mapsto (t \mapsto y\gamma(t)y^{-1}\gamma(t)^{-1})]$ in $[G, \Omega_0 PG]$ is also $[x \mapsto (t \mapsto \pi(x)\gamma(t)\pi(x)^{-1}\gamma(t)^{-1})]$, therefore the diagram commutes.

Since $\pi\colon G \to PG$ is a covering, every nullhomotopic loop in PG can be lifted to a nullhomotopic loop in G, hence $\Omega_0\pi\colon \Omega_0 G \to \Omega_0 PG$ is a homeomorphism. Moreover, this homeomorphism respects the H-space structures, and therefore $(\Omega_0\pi)_*\colon [G, \Omega_0 G] \to [G, \Omega_0 PG]$ is an isomorphism. □

Recall that by Proposition 3.21, the turning obstruction map $\overline{\mathrm{to}}_\gamma$ can be computed from $\widehat{\mathrm{to}}_G$; namely $\overline{\mathrm{to}}_\gamma([g]) = \widehat{\mathrm{to}}_G(p_G([\gamma])) \circ [g] \in [X, \Omega_0 G]$ for every $\gamma \in \Omega_Z G$ and $g\colon X \to G$. By Proposition 3.26, we have $\widehat{\mathrm{to}}_G(p_G([\gamma])) = [(\Omega_0\pi)^{-1}] \circ \mathrm{sp}_{PG}(p_G([\gamma])) \circ [\pi]$, hence $\overline{\mathrm{to}}_\gamma([g]) = [(\Omega_0\pi)^{-1}] \circ \mathrm{sp}_{PG}(p_G([\gamma])) \circ [\pi \circ g]$. This shows that $\overline{\mathrm{to}}_\gamma$ is determined by $\mathrm{sp}_{PG}(p_G([\gamma]))$; ie turning obstructions in G are determined by Samelson products in PG. We can also express this in terms of the adjointed versions:

Corollary 3.27 *Suppose that G is a path-connected topological group and that $Z(G)$ is discrete. Let $\gamma \in \Omega_Z G$ and $g\colon X \to G$. Then*

$$\overline{\mathrm{TO}}_\gamma([g]) = (\pi_*)^{-1}(\langle [\pi \circ g], p_G([\gamma]) \rangle) \in [SX, G],$$

where $(\pi_*)^{-1}$ is the inverse of the isomorphism $\pi_*\colon [SX, G] \to [SX, PG]$. □

Remark 3.28 If X is simply connected, then $\pi_*\colon [X, G] \to [X, PG]$ is also an isomorphism, which allows us to take the reverse point of view and compute Samelson products in PG from turning obstructions in G: suppose that $\gamma \in \pi_1(PG)$ and $g\colon X \to PG$, then $\langle [g], [\gamma] \rangle = \pi_*(\overline{\mathrm{TO}}_{\tilde{\gamma}}((\pi_*)^{-1}([g]))) \in [SX, PG]$, where $\tilde{\gamma} \in \Omega_Z G$ is a lift of γ.

In the next section we will compute various turning obstructions; see Theorem 4.1. By Lemma 3.24 those results give a variety of information about Samelson products $\langle [g], \eta \rangle$ for $\eta \in \pi_1(\mathrm{SO}_{2k})$ the generator, for example we get the following proposition. Recall that $\tau_{2k} \in \pi_{2k-1}(\mathrm{SO}_{2k})$ is the homotopy class of a clutching function of the tangent bundle of S^{2k}, and for $m > 2$ let $\eta_m\colon S^{m+1} \to S^m$ be essential.

Proposition 3.29 *The Samelson product $\langle \tau_{2k}, \eta \rangle \in \pi_{2k}(\mathrm{SO}_{2k})$ is given as follows:*

(a) *If $k = 2j+1$ is odd, $\langle \tau_{4j+2}, \eta \rangle = 0$.*

(b) *If $k = 2j$ is even, $\langle \tau_{4j}, \eta \rangle = \tau_{4j}\eta_{4j-1} \neq 0$.* □

Remark 3.30 Proposition 3.29 can be viewed as an extension of an odd-primary theorem of Hamanaka and Kono [7, Theorem A] to the prime 2.

As another application of Lemma 3.24, we consider the situation where η is not the turning datum in a turning problem, but instead the clutching function of a bundle. Let $E^{2k}_\eta \to S^2$ be a fixed nontrivial oriented rank-$2k$ bundle over S^2. Then $\mathrm{Fr}(E)$ is a nontrivial principal SO_{2k}–bundle over S^2, we write $\mathcal{G}^{2k}_\eta$ for the gauge group of $\mathrm{Fr}(E^{2k}_\eta)$ and consider the fibration sequence (2-2) for $\mathcal{G}^{2k}_\eta$, which we write as $\mathcal{G}^{2k}_{\eta,0} \to \mathcal{G}^{2k}_\eta \to \mathrm{SO}_{2k}$. As discussed in Section 2.5, $\mathcal{G}^{2k}_{\eta,0} \cong \mathrm{Map}((S^2,*),(\mathrm{SO}_{2k},\mathrm{Id}))$. Hence there is a natural isomorphism $\pi_i(\mathcal{G}^{2k}_{\eta,0}) \cong \pi_{i+2}(\mathrm{SO}_{2k})$ and by a theorem of Wockel [19, Theorem 2.3] the boundary map

$$\partial^{2k}_\eta \colon \pi_i(\mathrm{SO}_{2k}) \to \pi_{i-1}(\mathcal{G}^{2k}_{\eta,0}) = \pi_{i+1}(\mathrm{SO}_{2k})$$

in the associated long exact sequence is given by

$$\partial^{2k}_\eta([g]) = -\langle [g], \eta \rangle$$

for all $[g] \in \pi_i(\mathrm{SO}_{2k})$. Combining Lemma 3.24 and Theorem 4.1 therefore gives information about the map ∂^{2k}_η. In particular, for τ_{2k} and η_{4j-1} as in Proposition 3.29 we have:

Proposition 3.31 *The boundary map $\partial^{2k}_\eta \colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO}_{2k})$ satisfies*

$$\partial^{4j}_\eta(\tau_{4j}) = \tau_{4j}\eta_{4j-1} \neq 0 \quad \textit{and} \quad \partial^{4j+2}_\eta(\tau_{4j+2}) = 0. \qquad \square$$

Remark 3.32 If we let E^∞_η denote the stabilisation of the E^{2k}_η, then its frame bundle $\mathrm{Fr}(E^\infty_\eta)$ is a nontrivial principal SO–bundle over S^2 and we let $\mathcal{G}^\infty_\eta$ denote the gauge group of E^∞_η. Since the stable group SO is a homotopy abelian H–space, it follows that there is a weak homotopy equivalence

$$\mathcal{G}^\infty_\eta \simeq \mathrm{Map}(S^2,\mathrm{SO}) \cong \mathrm{Map}_*(S^2,\mathrm{SO}) \times \mathrm{SO}.$$

By comparing the homotopy long exact sequences of the fibrations

$$\mathcal{G}^{4j}_{\eta,0} \to \mathcal{G}^{4j}_\eta \to \mathrm{SO}_{4j} \quad \text{and} \quad \mathcal{G}^\infty_{\eta,0} \to \mathcal{G}^\infty_\eta \to \mathrm{SO},$$

where $\mathcal{G}^\infty_{\eta,0} \subset \mathcal{G}^\infty_\eta$ is the group of gauge transformations which are the identity in the fibre over the basepoint, we see that $\partial^{4j}_\eta \colon \pi_i(\mathrm{SO}_{4j}) \to \pi_{i+1}(\mathrm{SO}_{4j})$ is zero for $i < 4j-2$. When $i = 4j-2$, the domain of ∂^{4j}_η is $\pi_{4j-2}(\mathrm{SO}_{4j}) \cong \pi_{4j-2}(\mathrm{SO}) = 0$, so ∂^{4j}_η vanishes for $i \leq 4j-2$. Hence Proposition 3.31 shows that the first possibly nonzero boundary map in the homotopy long exact sequence of $\mathcal{G}^{4j}_{\eta,0} \to \mathcal{G}^{4j}_\eta \to \mathrm{SO}_{4j}$ is in fact nonzero.

4 Turning rank-$2k$ bundles over the $2k$–sphere

In this section we determine the turning obstructions for oriented rank-$2k$ bundles over the $2k$–sphere for all $k \geq 2$. To state our results, it will be convenient to use the notation $\mathrm{TO}_+ := \mathrm{TO}_\beta$ and $\mathrm{TO}_- := \mathrm{TO}_{\overline{\beta}}$ and when we wish to discuss these obstructions together, we will write $\mathrm{TO}_\pm$. We also define the adjointed η–turning obstruction $\mathrm{TO}_\eta := \mathrm{ad} \circ \mathrm{to}_\eta \colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO}_{2k})$. With this notation, the goal of this section is to compute the homomorphisms

$$\mathrm{TO}_\pm \colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO}_{2k}) \quad \text{and} \quad \mathrm{TO}_\eta \colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO}_{2k}).$$

Thus, if $E \to S^{2k}$ is a rank-$2k$ vector bundle with clutching function $g \colon S^{2k-1} \to \mathrm{SO}_{2k}$, then E is positive turnable if and only if $\mathrm{TO}_+([g]) = 0$, E is negative turnable if and only if $\mathrm{TO}_-([g]) = 0$ and E is η–turnable if and only if $\mathrm{TO}_\eta([g]) = 0$.

In order to state the computations of $\mathrm{TO}_\pm$ and TO_η we record some facts we need about the source and target groups of these homomorphisms, which can be found in [11]. We also introduce notation for generators of these groups. Recall that $\mathrm{e}(\xi) = \mathrm{e}(E_\xi) \in \mathbb{Z}$ is the Euler class and that $\mathrm{e}(\xi)$ is even unless $k = 2, 4$; see [6, Theorem, page 87]. Let $\tau_{2k} \in \pi_{2k-1}(\mathrm{SO}_{2k})$ denote the homotopy class of the clutching function of TS^{2k}. There is an isomorphism

$$\pi_{2k-1}(\mathrm{SO}_{2k}) \cong \mathbb{Z}(\tau_{2k}) \oplus C(\sigma_{2k}),$$

where $C(\sigma_{2k})$ is a cyclic group isomorphic to $\pi_{2k-1}(\mathrm{SO})$ and $S(\sigma_{2k}) \in \pi_{2k-1}(\mathrm{SO})$ is a generator. When $k = 2, 4$, we take $\mathrm{e}(\sigma_{2k}) = 1$, and when $k = 2$, we assume that σ_4 admits a complex structure; see Definition 4.6 and Theorem 4.7. When $k \neq 2, 4$, by Lemma 4.13 below, we assume that $C(\sigma_{2k}) = S(\pi_{2k-1}(\mathrm{SO}_{2k-2}))$; in particular, $\mathrm{e}(\sigma_{2k}) = 0$.

There are isomorphisms

$$\pi_{2k}(\mathrm{SO}_{2k}) \cong \begin{cases} 0 & \text{if } k = 3, \\ \mathbb{Z}/4 & \text{if } k \geq 5 \text{ is odd}, \\ (\mathbb{Z}/2)^2 & \text{if } k \equiv 2 \bmod 4, \\ (\mathbb{Z}/2)^3 & \text{if } k \equiv 0 \bmod 4. \end{cases}$$

When k is odd, we let $\zeta \in \pi_{2k}(\mathrm{SO}_{2k})$ be a generator and note that $\pi_{2k}(\mathrm{SO}) = 0$. When k is even, the stabilisation homomorphism $S \colon \pi_{2k}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO})$ is split onto, where $\pi_{2k}(\mathrm{SO}) = 0$ if $k \equiv 2 \bmod 4$ and $\mathbb{Z}/2$ if $k \equiv 0 \bmod 4$. Moreover, for all $k \neq 3$ there is a short exact sequence

$$0 \to \mathbb{Z}/2(\tau_{2k}\eta_{2k-1}) \to \pi_{2k}(\mathrm{SO}_{2k}) \xrightarrow{\mathrm{ev}_* \oplus S} \pi_{2k}(S^{2k-1}) \oplus \pi_{2k}(\mathrm{SO}) \to 0, \tag{4-1}$$

with $\mathrm{ev} \colon \mathrm{SO}_{2k} \to S^{2k-1}$ given by evaluation at a point in S^{2k-1} and $\eta_{2k-1} \colon S^{2k} \to S^{2k-1}$ essential. The sequence (4-1) is nonsplit when $k \geq 5$ is odd and splits when k is even.

Theorem 4.1 *The turning obstructions* $\mathrm{TO}_\pm \colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO}_{2k})$ *satisfy the following*:

(a) *If* k *is odd*, *then* $\mathrm{TO}_\pm(\xi) = \mathrm{e}(\xi)\zeta$.

(b) *If* $k \equiv 2 \bmod 4$, *then* $\mathrm{ev}_*(\mathrm{TO}_\pm(\tau_{2k})) = 1$, $\mathrm{TO}_+(\sigma_{2k}) = 0$ *and* $\mathrm{TO}_-(\sigma_{2k}) = \mathrm{e}(\sigma_{2k})\,\mathrm{TO}_-(\tau_{2k})$.

(c) *If* $k \equiv 0 \bmod 4$, *then* $\mathrm{ev}_*(\mathrm{TO}_\pm(\tau_{2k})) = 1$, $S(\mathrm{TO}_\pm(\tau_{2k})) = 0$ *and* $S(\mathrm{TO}_\pm(\sigma_{2k})) = 1$; *in particular*, $\mathrm{TO}_\pm \otimes \mathrm{Id}_{\mathbb{Z}/2}$ *is injective.*

In particular, if k *is odd, then* $\mathrm{TO}_\eta = \mathrm{TO}_+ - \mathrm{TO}_- = 0$. *If* $k = 2j$ *is even, then* TO_η *satisfies the following*:

(d) $\mathrm{TO}_\eta(\tau_{4j}) = \tau_{4j}\eta_{4j-1} \neq 0$.

(e) *If* $j = 1$ *or* 2, *then* $\mathrm{ev}_*(\mathrm{TO}_\eta(\sigma_{4j})) = 1$ *and* $\mathrm{TO}_\eta \otimes \mathrm{Id}_{\mathbb{Z}/2}$ *is injective.*

(f) *If* $j \geq 3$, *then* $\mathrm{TO}_\eta(\sigma_{4j}) = 0$.

Remark 4.2 Theorem 4.1 shows that unless $k = 2$, for all $[g] \in \pi_{2k-1}(\mathrm{SO}_{2k})$ we have $\mathrm{TO}_+([g]) = 0$ if and only if $\mathrm{TO}_-([g]) = 0$. Hence for $k \neq 2$ rank-$2k$ bundles $E \to S^{2k}$ are either bi-turnable or not turnable and so these bundles are not strongly chiral. On the other hand, when $k = 2$, a bundle $E \to S^4$ is strongly chiral if and only if $\mathrm{e}(E)$ is odd.

The remainder of this section is devoted to the proof of Theorem 4.1. In Section 4.1 we consider the turnability of the tangent bundle of the $2k$–sphere, which is an essential input to the proof. In Section 4.2 we consider the exceptional case of the 4–sphere. In Section 4.3 we consider $\mathrm{TO}_\eta = \mathrm{TO}_+ - \mathrm{TO}_-$. In Section 4.4 we assemble the previous work to prove Theorem 4.1.

4.1 Turning the tangent bundle of the $2k$–sphere

Let TS^n denote the tangent bundle of the n–sphere. We fix the standard orientation on the n–sphere and this orients TS^n. In [12], Kirchoff proved that if TS^{2k} admits a complex structure then TS^{2k+1} is trivial. Later, it was proven in [6] that TS^{2k+1} is trivial if and only if $2k+1 = 1, 3$ or 7. Since elementary calculations show that TS^2 and TS^6 admit complex structures, Kirchoff's theorem implies that TS^{2k} admits a complex structure if and only if TS^{2k+1} is trivial. Here we prove a strengthening of Kirchoff's theorem, which only assumes that TS^{2k} is turnable.

Theorem 4.3 (Kirchoff's theorem for turnings) *If* TS^{2k} *is turnable, then* TS^{2k+1} *is trivial.*

Corollary 4.4 TS^{2k} *is turnable if and only if it admits a complex structure, which is the case if and only if* $2k = 2$ *or* 6. □

Proof of Theorem 4.3 We first recall the following well-known definition of a clutching function c_m for TS^m; see [9, Chapter 8, Corollary 9.9]. Given $x \in S^{m-1}$, write $\mathbb{R}^m = \langle x \rangle \oplus \langle x \rangle^\perp$ as the sum of the line spanned by x and its orthogonal complement and write $v \in \mathbb{R}^m$ as $v = (w, y)$, where $w \in \langle x \rangle$ and $y \in \langle x \rangle^\perp$. Let $c_m \colon S^{m-1} \to O_m$ be the function which assigns to $x \in S^{m-1}$ the reflection to the hyperplane orthogonal to x:

$$c_m(x) \colon \mathbb{R}^m \to \mathbb{R}^m, \quad (w, y) \mapsto (-w, y).$$

Suppose that TS^{2k} is turnable. We will show that the clutching function $c_{2k+1} : S^{2k} \to O_{2k+1}$ is nullhomotopic, proving that TS^{2k+1} is trivial. Using the notation above, we see that

$$TS^{2k} = \{((0, y), x) \mid y \in \langle x \rangle^\perp\} \subset \mathbb{R}^{2k+1} \times S^{2k}.$$

Since TS^{2k} is turnable, there exists a turning α_t on TS^{2k} with $\alpha_0 = \mathbb{1}$ and $\alpha_1 = -\mathbb{1}$. We use α_t to define the homotopy of automorphisms of the trivial bundle given by

$$H\colon (\mathbb{R}^{2k+1} \times S^{2k}) \times I \to \mathbb{R}^{2k+1} \times S^{2k}, \quad (((w, y), x), t) \mapsto ((-w, \alpha_t(y)), x).$$

We see that $H_0 = c_{2k+1}$ and $H_1 = -\mathbb{1}$. Hence H is the required homotopy of clutching functions from c_{2k+1} to a constant map. □

4.2 Rank-4 bundles over the 4–sphere

The set of isomorphism classes of rank-4 bundles over S^4 is in bijection with $\pi_3(\mathrm{SO}_4) \cong \mathbb{Z} \oplus \mathbb{Z}$. We recall the canonical double covering

$$q\colon S^3 \times S^3 \to \mathrm{SO}_4, \quad (x, y) \mapsto (v \mapsto x \cdot v \cdot y),$$

where we regard $x, y \in S^3$ as unit quaternions, $v \in \mathbb{H}$ and $\cdot$ denotes quaternionic multiplication. If we define $g_{(k_1,k_2)}\colon S^3 \to \mathrm{SO}_4$ by $x \mapsto q(x^{k_1}, x^{k_2})$, then the map

$$\mathbb{Z} \oplus \mathbb{Z} \to \pi_3(\mathrm{SO}_4), \quad (k_1, k_2) \mapsto [g_{(k_1,k_2)}],$$

is an isomorphism, which we use as coordinates for $\pi_3(\mathrm{SO}_4)$. By [9, Chapter 8, Proposition 12.10], for example, the map $g_{(1,1)}$ is a clutching function for TS^4 and so $\tau_4 = [g_{(1,1)}]$.

Definition 4.5 For $(k_1, k_2) \in \mathbb{Z}^2$, let $E_{k_1,k_2} \to S^4$ be the oriented rank-4 vector bundle with clutching function $g_{(k_1,k_2)}$; eg $TS^4 \cong E_{1,1}$.

Definition 4.6 We define $\sigma_4 := [g_{(0,1)}]$.

The turning obstructions TO_+, TO_- and TO_η take values in the group $\pi_4(\mathrm{SO}_4)$, and we use the isomorphism $q_*\colon \pi_4(S^3 \times S^3) \to \pi_4(\mathrm{SO}_4)$ to identify $\pi_4(\mathrm{SO}_4) \cong \pi_4(S^3) \oplus \pi_4(S^3) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$.

Theorem 4.7 *The turning obstructions for rank-4 bundles over S^4 are given follows*:

(a) $\mathrm{TO}_+(a, b) = (\rho_2(a), 0)$.
(b) $\mathrm{TO}_-(a, b) = (0, \rho_2(b))$.
(c) $\mathrm{TO}_\eta(a, b) = (\rho_2(a), \rho_2(b))$.

Proof (a) Let $i, j, k \in \mathbb{H}$ be the standard purely imaginary unit quaternions. If we take the standard complex structure on $\mathbb{H} = \mathbb{C} \oplus \mathbb{C} j$ to be given by left multiplication by $i \in \mathbb{H}$, then $g_{(0,1)}(x)$ commutes with i for every $x \in S^3$, and so $g_{(0,1)}(x) \in U_2 \subset \mathrm{SO}_4$ for all $x \in S^3$. Thus $E_{0,1}$ admits a complex structure and so $\mathrm{TO}_+(0, 1) = 0$. By Corollary 4.4, $\mathrm{TO}_\pm(1, 1) \neq 0$, hence $\mathrm{TO}_+(1, 0) \neq 0$.

Consider the map $g_{(1,0)}\colon S^3 \to \mathrm{SO}_4$. In S^3 there is a unique homotopy class $[\gamma]$ of paths from 1 to -1, and we have $(g_{(1,0)})_*([\gamma]) = [\beta]$. Obviously, $(g_{(1,0)})_*([\mathrm{Id}_{S^3}]) = [g_{(1,0)}] = (1, 0) \in \pi_3(\mathrm{SO}_4)$ and $(g_{(1,0)})_*(\pi_4(S^3)) = \mathbb{Z}/2 \oplus 0 \le \pi_4(\mathrm{SO}_4)$. Since $g_{(1,0)}$ is a continuous group homomorphism, we have $(g_{(1,0)})_*(\mathrm{TO}_\gamma([\mathrm{Id}_{S^3}])) = \mathrm{TO}_\beta([g_{(1,0)}]) = \mathrm{TO}_+(1, 0)$. This implies that $\mathrm{TO}_+(1, 0)$ is a nontrivial element of $\mathbb{Z}/2 \oplus 0$, ie $\mathrm{TO}_+(1, 0) = (1, 0)$. Therefore $\mathrm{TO}_+(a, b) = a\,\mathrm{TO}_+(1, 0) + b\,\mathrm{TO}_+(0, 1) = (\rho_2(a), 0)$.

(b) Since $ji = -k$, right multiplication by i defines a complex structure on $\mathbb{H}$ whose induced orientation is opposite to the standard orientation. Since $g_{(1,0)}(x)$ commutes with right multiplication by i for all $x \in S^3$, we see that $\bar{E}_{1,0}$ admits a complex structure. Hence $\mathrm{TO}_-(1,0) = 0$, and therefore $\mathrm{TO}_-(0,1) \neq 0$. Let $\bar{S}^3$ denote S^3 with the opposite group structure, then $g_{(0,1)}\colon \bar{S}^3 \to \mathrm{SO}_4$ is a continuous group homomorphism. The unique homotopy class of paths in $\bar{S}^3$ from 1 to -1 maps to $[\bar{\beta}]$ under $g_{(0,1)}$. Therefore, similarly to part (a), we get that $\mathrm{TO}_-(0,1) = (0,1)$.

(c) By Theorem 3.23(b), $\mathrm{TO}_\eta(a,b) = \mathrm{TO}_+(a,b) - \mathrm{TO}_-(a,b)$. □

We now discuss the relationship between the turning type of rank-4 bundles and the homotopy classification of their gauge groups, due to Kishimoto, Membrillo-Solis and Theriault [13]. Following the notation of [13], let $\mathcal{G}_{k_1,k_2}$ denote the gauge group of E_{k_1,k_2}. Let $\{\{a,b\}\}$ denote the multiset consisting of the elements a, b and for integers a and b write (a,b) for their greatest common divisor. Then, for integers r, a and b, write $M^r(a,b)$ for the multiset $\{\{(a,r),(b,r)\}\}$. By [13, Theorem 1.1(b)] if $\mathcal{G}_{k_1,k_2} \simeq \mathcal{G}_{l_1,l_2}$ then $M^4(k_1,k_2) = M^4(l_1,l_2)$.

Recall that the turning type of an orientable bundle is characterised by whether it is either bi-turnable, strongly chiral or not turnable. By Theorem 4.7, E_{k_1,k_1} is bi-turnable if $M^2(k_1,k_2) = \{\{2,2\}\}$, strongly chiral if $M^2(k_1,k_2) = \{\{1,2\}\}$ and not turnable if $M^2(k_1,k_2) = \{\{1,1\}\}$. Hence, combining [13, Theorem 1.1(b)] and Theorem 4.7, we have:

Proposition 4.8 *The turning type of E_{k_1,k_2} is a homotopy invariant of $\mathcal{G}_{k_1,k_2}$.* □

4.3 The turning obstruction TO_η

In this subsection we cover some preliminaries for the computation of $\mathrm{TO}_\eta\colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k}(\mathrm{SO}_{2k})$.

For any $m \geq 2$, let $V_{m,2}$ be the Stiefel manifold of ordered pairs of orthonormal vectors in $\mathbb{R}^m$. Given $\underline{v} = (v_1, v_2) \in V_{m,2}$ we define $V = \langle v_1, v_2 \rangle$ and write $x \in \mathbb{R}^m$ as $x = (v, w)$ where $v \in V$ and $w \in V^\perp$. The isomorphism $\mathbb{C} \to V$ defined by $1 \mapsto v_1$ and $i \mapsto v_2$ defines a complex structure on V. We define $\gamma_{\underline{v}}$ in $\Omega\mathrm{SO}_m$ by

$$\gamma_{\underline{v}}(t)(v,w) = (e^{2\pi i t} v, w),$$

and we define the map

$$L = L_m\colon V_{m,2} \to \Omega\mathrm{SO}_m, \quad \underline{v} \mapsto \gamma_{\underline{v}}.$$

Next we consider the canonical projection $p\colon \mathrm{SO}_m \to V_{m,2}$ and the composition

$$L \circ p\colon \mathrm{SO}_m \to \Omega\mathrm{SO}_m.$$

It is clear from the definitions that $L \circ p$ is the map ρ_η of Definition 3.13(a), so after the identification $\pi_{m-1}(\Omega\mathrm{SO}_m) = \pi_m(\mathrm{SO}_m)$ we obtain the following:

Lemma 4.9 *For all $[g] \in \pi_{m-1}(\mathrm{SO}_m)$, $(L \circ p)_*([g]) = \mathrm{TO}_\eta([g])$.* □

Remark 4.10 Combining Lemmas 4.9, 3.6 and 3.24, we get

$$(L \circ p)_*([g]) = \mathrm{TO}_\eta([g]) = \overline{\mathrm{TO}}_\eta([g]) = \langle [g], \eta \rangle,$$

and this equation can be generalised to give a method for computing similar Samelson products as follows.

For $2 \leq i \leq m$, let $V_{m,i}$ denote the Stiefel manifold of mutually orthonormal ordered i–tuples $\underline{v} = (v_1, \dots, v_i)$ of vectors in $\mathbb{R}^m$, set $V = \langle v_1, \dots, v_i \rangle = \mathbb{R}^i$ and write $x \in \mathbb{R}^m$ as $x = (v, w)$, where $v \in V$ and $w \in V^\perp$. Then given any map $\alpha\colon S^{i-1} \to \mathrm{SO}_i$ we define $\alpha_{\underline{v}} \in \Omega^{i-1}\mathrm{SO}_m$, the $(i-1)$–fold based loop space of SO_m, by

$$\alpha_{\underline{v}}(s)(v, w) = (\alpha(s)v, w)$$

for all $s \in S^{i-1}$ and $(v, w) \in \mathbb{R}^m$. Allowing $\underline{v}$ to vary, we obtain the map

$$L(\alpha)\colon V_{m,i} \to \Omega^{i-1}\mathrm{SO}_m, \quad \underline{v} \mapsto \alpha_{\underline{v}},$$

and note that $L\colon V_{m,2} \to \Omega\mathrm{SO}_m$ above is $L(\alpha)$ for the special case of $\alpha\colon S^1 \to \mathrm{SO}_2 = U(1), t \mapsto e^{2\pi i t}$. If $\iota\colon \mathrm{SO}_i \to \mathrm{SO}_m$ denotes the standard inclusion, and $p\colon \mathrm{SO}_m \to V_{m,i}$ the standard projection, then after the identification $\pi_j(\Omega^{i-1}\mathrm{SO}_m) = \pi_{i+j-1}(\mathrm{SO}_m)$, a higher-dimensional version of Lemma 3.6 leads to the equation

$$(L(\alpha) \circ p)_*([g]) = \langle [g], \iota_*([\alpha]) \rangle$$

for all $[g] \in \pi_j(\mathrm{SO}_m)$.

Now we consider the case $m = 2k$ and the induced homomorphisms $p_*\colon \pi_{2k-1}(\mathrm{SO}_{2k}) \to \pi_{2k-1}(V_{2k,2})$ and $L_*\colon \pi_{2k-1}(V_{2k,2}) \to \pi_{2k}(\mathrm{SO}_{2k})$. Let $\mathrm{ev}\colon \mathrm{SO}_{2k} \to S^{2k-1}$ be the map defined by evaluation at a point in S^{2k-1}.

Definition 4.11 Define $a_k \in \mathbb{Z}$ by $a_k := 1$ if k is even and $a_k := 2$ if k is odd.

Lemma 4.12 *Let $k \geq 3$. For any isomorphism $\pi_{2k-1}(V_{2k,2}) \to \mathbb{Z}/2 \oplus \mathbb{Z}$ we have:*

(a) $p_*(\tau_{2k}) = (\rho_2(a_k), 2)$.

(b) $L_*(1, 0) = \tau_{2k}\eta_{2k-1}$.

(c) $\mathrm{ev}_*(L_*(0, 1)) = \rho_2(a_k)$.

Proof (a) The sequence $\pi_{2k-1}(S^{2k-2}) \to \pi_{2k-1}(V_{2k,2}) \to \pi_{2k-1}(S^{2k-1})$ is split short exact, with $p_*(\tau_{2k})$ mapping to $2 \in \pi_{2k-1}(S^{2k-1}) = \mathbb{Z}$. It follows that the map $f = p \circ \tau_{2k}\colon S^{2k-1} \to V_{2k,2}$ vanishes on mod 2 cohomology. Also for $x \in H^{2k-2}(V_{2k,2}; \mathbb{Z}/2)$ the generator, $\mathrm{Sq}^2(x) \in H^{2k}(V_{2k,2}; \mathbb{Z}/2) = 0$ and so the functional Steenrod square $\mathrm{Sq}^2_x(g)$ is defined for all maps $g\colon S^{2k-1} \to V_{2k,2}$ which vanish on mod 2 cohomology. Moreover, $g = 2g'$ for some map $g'\colon S^{2k-1} \to V_{2k,2}$ if and only if $\mathrm{Sq}^2_x(g) = 0$.

Now the map $\tau_{2k}\colon S^{2k-1} \to \mathrm{SO}_{2k}$ factors over the double covering $q\colon S^{2k-1} \to \mathbb{R}P^{2k-1}$ and a map $\tau'_{2k}\colon \mathbb{R}P^{2k-1} \to \mathrm{SO}_{2k-1}$. Since q vanishes on mod 2 cohomology and $\mathrm{Sq}^2(t^{2k-2}) = 0$, for $t \in H^1(\mathbb{R}P^{2k-1}; \mathbb{Z}/2)$ a generator, it follows that the functional Steenrod square $\mathrm{Sq}^2_{t^{2k-2}}$ is defined on q.

We consider the composition

$$S^{2k-1} \xrightarrow{q} \mathbb{R}P^{2k-1} \xrightarrow{\tau'_{2k}} \mathrm{SO}_{2k} \xrightarrow{p} V_{2k,2}.$$

A degree argument shows that the map $p \circ \tau'_{2k} \colon \mathbb{R}P^{2k-1} \to V_{2k,2}$ satisfies $(p \circ \tau'_{2k})^*(x) = t^{2k-2}$ and naturality of functional Steenrod squares gives that

$$\mathrm{Sq}^2_x(p \circ \tau_{2k}) = \mathrm{Sq}^2_{t^{2k-2}}(q).$$

But q is the attaching map of the top cell of $\mathbb{R}P^{2k}$ and so $\mathrm{Sq}^2_{t^{2k-2}}(q) = \mathrm{Sq}^2(t^{2k-2}) = \rho_2(a_k)$. Hence we have $\mathrm{Sq}^2_x(p \circ \tau_{2k}) = \rho_2(a_k)$ and so $p_*(\tau_{2k}) = (\rho_2(a_k), 2) \in \pi_{2k-1}(V_{2k,2})$.

(b) Let $\iota_{2k-2} \colon S^{2k-2} \to V_{2k,2}$ be the inclusion of a fibre of the projection $V_{2k,2} \to S^{2k-1}$. Then we have $(1,0) = [\iota_{2k-2} \circ \eta_{2k-2}] \in \pi_{2k-1}(V_{2k,2})$. Hence it suffices to prove that $L_*([\iota_{2k-2}]) = \tau_{2k} \in \pi_{2k-1}(\mathrm{SO}_{2k})$. Now a degree argument shows that $\mathrm{ev}_*(L_*([\iota_{2k-2}])) = 1 + (-1)^{2k-2} = 2 \in \pi_{2k-1}(S^{2k-1}) = \mathbb{Z}$ and we consider the commutative diagram

$$\begin{array}{ccc} V_{2k,2} & \xrightarrow{L_{2k}} & \Omega\mathrm{SO}_{2k} \\ \downarrow & & \downarrow S \\ V_{2k+2,2} & \xrightarrow{L_{2k+2}} & \Omega\mathrm{SO}_{2k+2} \end{array}$$

Since $\pi_{2k-2}(V_{2k+2,2}) = 0$, we see that $S(L_*([\iota_{2k-2}])) = 0$. Hence $L_*([\iota_{2k-2}])$ is the clutching function of a stably trivial bundle with Euler class 2, so $L_*([\iota_{2k-2}]) = \tau_{2k}$, as required.

(c) The standard complex structure on $\mathbb{R}^{2k} = \mathbb{C}^k$ defines a section $s \colon S^{2k-1} \to V_{2k,2}$ of the projection $V_{2k,2} \to S^{2k-1}$ by $s(v) = (v, iv)$. Taking induced maps on π_{2k-1} gives a splitting $\pi_{2k-1}(V_{2k,2}) \cong \mathbb{Z}/2 \oplus \mathbb{Z}$, where $[s] = (0,1)$ and $[\iota_{2k-2} \circ \eta_{2k-2}] = (1,0)$. Since $\mathrm{ev}_*(L_*(1,0)) = 0$, it suffices to prove $\mathrm{ev}_*(L_*([s])) = \rho_2(a_k)$.

It is clear that $L \circ s$ factors as the composition of the Hopf map $H \colon S^{2k-1} \to \mathbb{C}P^{k-1}$ and a map $L' \colon \mathbb{C}P^{k-1} \to \Omega\mathrm{SO}_{2k}$. Another degree argument shows that the adjoint of $\Omega(\mathrm{ev}) \circ L' \colon \mathbb{C}P^{k-1} \to \Omega S^{2k-1}$ has degree one and so the homotopy class of $S^{2k-1} \to \Omega S^{2k-1}$ is determined by the functional Steenrod square $\mathrm{Sq}^2_{z^{k-1}}(H)$, where $z \in H^2(\mathbb{C}P^{k-1}; \mathbb{Z}/2)$ is the generator. Since H is the attaching map of the top cell of $\mathbb{C}P^k$, we have

$$\mathrm{Sq}^2_{z^{k-1}}(H) = \mathrm{Sq}^2(z^{k-1}) = \rho_2(a_k) \in \mathbb{Z}/2 \cong H^{2k}(\mathbb{C}P^k; \mathbb{Z}/2),$$

which completes the proof of part (c). □

Lemma 4.13 *If $k \neq 2, 4$, we may choose $\sigma_{2k} \in S(\pi_{2k-1}(\mathrm{SO}_{2k-2})) \subset \pi_{2k-1}(\mathrm{SO}_{2k})$.*

Proof If $k \neq 2, 4$, then $\sigma_{2k} \in \pi_{2k-1}(\mathrm{SO}_{2k})$ is such that $S(\sigma_{2k})$ generates $\pi_{2k-1}(\mathrm{SO})$ and $\mathrm{e}(\sigma_{2k}) = 0$. The map $\pi_{2k-1}(\mathrm{SO}_{2k-2}) \to \pi_{2k-1}(\mathrm{SO})$ is onto for $k \geq 7$ by [2], which proves the lemma when $k \geq 7$.

For the remaining cases, $k \in \{3,5,6\}$. If $k=3$, then $\pi_{2k}(\mathrm{SO})=0$, and so $\sigma_6=0$. When $k=5,6$ we consider part of the homotopy long exact sequence of the fibration $\mathrm{SO}_{2k-2} \to \mathrm{SO}_{2k-1} \to S^{2k-2}$:

$$\pi_{2k-1}(\mathrm{SO}_{2k-2}) \to \pi_{2k-1}(\mathrm{SO}_{2k-1}) \to \pi_{2k-1}(S^{2k-2}) \to \pi_{2k-2}(\mathrm{SO}_{2k-2}) \to \pi_{2k-2}(\mathrm{SO}_{2k-1}).$$

Applying results of Kervaire [11], we deduce that the boundary map $\pi_{2k-1}(S^{2k-2}) \to \pi_{2k-2}(\mathrm{SO}_{2k-2})$ is injective and so $\pi_{2k-1}(\mathrm{SO}_{2k-2}) \to \pi_{2k-1}(\mathrm{SO}_{2k-1})$ is onto. Since $\mathrm{e}(\sigma_{2k})=0$, $\sigma_{2k} \in S(\pi_{2k-1}(\mathrm{SO}_{2k-1}))$ and so $\sigma_{2k} \in S(\pi_{2k-1}(\mathrm{SO}_{2k-2}))$. □

Lemma 4.14 *If k is odd, then $\mathrm{TO}_\eta([g])=0$ for all $[g] \in \pi_{2k-1}(\mathrm{SO}_{2k})$.*

Proof By Lemma 4.12(a), $p_*(\tau_{2k})=(0,2)$ and by Lemma 4.12(c), $L_*(0,2)=0$. Hence by Lemma 4.9, $\mathrm{TO}_\eta(\tau_{2k})=(L\circ p)_*(\tau_{2k})=0$. By Lemma 4.13, $\sigma_{2k} \in S(\pi_{2k-1}(\mathrm{SO}_{2k-2}))$, ie $\sigma_{2k}=i\circ\sigma'$ for some $\sigma' \in \pi_{2k-1}(\mathrm{SO}_{2k-2})$, with $i\colon \mathrm{SO}_{2k-2} \to \mathrm{SO}_{2k}$ the inclusion. Using the analogue of Lemma 3.9 for closed loops, $\mathrm{TO}_\eta(\sigma_{2k})=\mathrm{TO}_{\mathbb{1}\oplus\eta}(i\circ\sigma')=S(\mathrm{TO}_{\mathbb{1}}(\sigma'))=0$, where $\mathbb{1}$ denotes the constant loop at $\mathbb{1} \in \mathrm{SO}_{2k-2}$. □

4.4 The proof of Theorem 4.1

In this subsection we complete the proof of Theorem 4.1.

Proof of Theorem 4.1 (d) The $j=1$ case follows from Theorem 4.7. If $j \geq 2$, then $p_*(\tau_{4j})=(1,2)$ by Lemma 4.12(a) and since $\pi_{4j}(\mathrm{SO}_{4j})$ is a 2–torsion group, $L_*(1,2)=L_*(1,0)$. By Lemma 4.12(b), $L_*(1,0)=\tau_{4j}\eta_{4j-1}$ and so $\mathrm{TO}_\eta(\tau_{4j})=(L\circ p)_*(\tau_{4j})=\tau_{4j}\eta_{4j-1}$.

(e) If $j=1$, then $\mathrm{TO}_\eta(\sigma_{4j})=(0,1)$ by Theorem 4.7, and $\mathrm{ev}_*(a,b)=a+b$. If $j=2$, then $\mathrm{e}(\sigma_{4j})=1$ and so $p_*(\sigma_{4j})=(\epsilon,1)$ for some $\epsilon \in \mathbb{Z}/2$. Since $\mathrm{ev}_*(L_*(1,0))=0$, Lemma 4.12(c) ensures that $\mathrm{ev}_*(\mathrm{TO}_\eta(\sigma_{4j}))=\mathrm{ev}_*((\epsilon,1))=1$.

(f) If $j \geq 3$, then by Lemma 4.13, $\sigma_{4j} \in S(\pi_{4j-1}(\mathrm{SO}_{4j-2}))$ and so $\mathrm{TO}_\eta(\sigma_{4j})=0$ (as in Lemma 4.14).

(a) If k is odd, the fact that $\mathrm{TO}_+(\xi)=\mathrm{TO}_-(\xi)$ follows from Theorem 3.23(b) and Lemma 4.14. If $k=1,3$, then $\pi_{2k}(\mathrm{SO}_{2k})=0$ and the statement holds trivially. If $k \geq 5$ is odd, then by Corollary 4.4, $\mathrm{TO}_+(\tau_{2k}) \neq 0$. Since k is odd, $2\,\mathrm{TO}_+(\tau_{2k})=\mathrm{TO}_\eta(\tau_{2k})=0$ by Theorem 3.23(d) and Lemma 4.14. Since $\pi_{2k}(\mathrm{SO}_{2k}) \cong \mathbb{Z}/4$ and $\mathrm{e}(\tau_{2k})=2$, the result holds for $\mathbb{Z}(\tau_{2k}) \subseteq \pi_{2k-1}(\mathrm{SO}_{2k})$. If $k \equiv 3 \bmod 4$ then $\pi_{2k-1}(\mathrm{SO}_{2k})=\mathbb{Z}(\tau_{2k})$. If $k \equiv 1 \bmod 4$, then $\pi_{2k-1}(\mathrm{SO}_{2k})=\mathbb{Z}(\tau_{2k})\oplus\mathbb{Z}/2(\sigma_{2k})$, provided $k \geq 5$ as we are assuming. Hence it suffices to show that $\mathrm{TO}_+(\sigma_{2k})=0$. Now $\pi_{2k-1}(U) \to \pi_{2k-1}(\mathrm{SO})$ is onto and so $S(\sigma_{2k})$ is stably complex. By [9, Chapter 20, Corollary 9.8] $S(\sigma_{2k})$ admits a complex structure with $c_k(\sigma_{2k})=(k-1)!$ and since $k \geq 5$, $(k-1)!$ is divisible by 4. Hence $\sigma_{2k}-\frac{1}{2}(k-1)!\,\tau_{2k}$ admits a complex structure; see Theorem 6.6. Then

$$0=\mathrm{TO}_+\big(\sigma_{2k}-\tfrac{1}{2}(k-1)!\,\tau_{2k}\big)=\mathrm{TO}_+(\sigma_{2k})-\tfrac{1}{4}(k-1)!(2\,\mathrm{TO}_+(\tau_{2k}))=\mathrm{TO}_+(\sigma_{2k}).$$

(b) The special case $k=2$ is proven in Theorem 4.7. For $k\geq 6$, we first prove that $\mathrm{ev}_*(\mathrm{TO}_\pm(\tau_{2k}))=1$. By Corollary 4.4, $\mathrm{TO}_\pm(\tau_{2k})\neq 0$. By Theorem 3.23(b) and Theorem 4.1(d), we have $\mathrm{TO}_-(\tau_{2k})=\mathrm{TO}_+(\tau_{2k})+\tau_{2k}\eta_{2k-1}$. Since $\mathrm{Ker}(\mathrm{ev}_*)$ is generated by $\tau_{2k}\eta_{2k-1}$, it follows that $\mathrm{ev}_*(\mathrm{TO}_\pm(\tau_{2k}))=1$. As $k\geq 6$, we have $\mathrm{e}(\sigma_{2k})=0$ and we must show that $\mathrm{TO}_\pm(\sigma_{2k})=0$. By Lemma 4.13, $\sigma_{2k}\in S(\pi_{2k-1}(\mathrm{SO}_{2k-2}))$ and so $\mathrm{TO}_\eta(\sigma_{2k})=0$, so it suffices to show that $\mathrm{TO}_+(\sigma_{2k})=0$. The argument is analogous to the case $k\equiv 1 \bmod 4$.

(c) We first prove that $\mathrm{ev}_*(\mathrm{TO}_\pm(\tau_{2k}))=1$. Since τ_{2k} is stably trivial, so is $\mathrm{TO}_\pm(\tau_{2k})$. The proof is now the same as the proof when $k\equiv 2 \bmod 4$. To see that $S(\mathrm{TO}_\pm(\sigma_{2k}))=1$, we note that $S(\sigma_{2k})$ generates $\pi_{2k-1}(\mathrm{SO})\cong\mathbb{Z}$ and that the natural map $\pi_{2k-1}(U)\to\pi_{2k-1}(\mathrm{SO})$ has image the subgroup of index 2 by [4]. Hence σ_{2k} does not admit a stable complex structure and so σ_{2k} is not stably turnable by Theorem 5.10. Now by Lemma 3.9, $S(\mathrm{TO}_\pm(\sigma_{2k}))=\mathrm{TO}_\pm(S(\sigma_{2k}))=1$. □

5 Stable turnings and stable complex structures

In this section we define stable turnings of vector bundles $E\to B$. When B has the homotopy type of a finite CW–complex, we will see that Bott's proof of Bott periodicity shows that the space of stable complex structures on E is weakly homotopy equivalent to the space of stable turnings of E. In particular, in this case E is stably turnable if and only if E admits a stable complex structure.

Recall that $\underline{\mathbb{R}}^j$ denotes the trivial vector bundle over B of rank j.

Definition 5.1 (stably turnable) A rank-n vector bundle $E\to B$ is *stably turnable* if $E\oplus\underline{\mathbb{R}}^j$ is turnable for some $j\geq 0$.

Of course, if E is turnable, then E is stably turnable.

Remark 5.2 In the definition of stably turnable, $n+j$ must be even but n need not be even.

Remark 5.3 It is clear from the definition of turnings that if $E\oplus\underline{\mathbb{R}}^j$ is turnable, then $E\oplus\underline{\mathbb{R}}^{j+2l}$ is turnable for any nonnegative integer l.

For any rank-n vector bundle $E\to B$, recall that $\mathrm{Fr}(E)$, the frame bundle of E, is the principal SO_n–bundle associated to E and for any nonnegative integer j with $n+j$ even,

$$\mathrm{Turn}(E\oplus\underline{\mathbb{R}}^j)=\mathrm{Fr}(E\oplus\underline{\mathbb{R}}^j)\times_{\mathrm{SO}_{n+j}}\Omega_{\pm\mathbb{1}}\mathrm{SO}_{n+j}$$

is the associated turning bundle of $E\oplus\underline{\mathbb{R}}^j$. Now orthogonal sum with the path $\beta\in\Omega_{\pm\mathbb{1}}\mathrm{SO}_2$ defines the injective map $i_\beta\colon\Omega_{\pm\mathbb{1}}\mathrm{SO}_{n+j}\to\Omega_{\pm\mathbb{1}}\mathrm{SO}_{n+j+2}$, which we regard as an inclusion. Thus we regard $\mathrm{Turn}(E\oplus\underline{\mathbb{R}}^j)$ as a subbundle of $\mathrm{Turn}(E\oplus\underline{\mathbb{R}}^{j+2})$ and we set

$$\mathrm{Turn}(E^\infty):=\bigcup_{n+j\ \text{even}}\mathrm{Turn}(E\oplus\underline{\mathbb{R}}^j),$$

which is a fibre bundle over B with fibre $\Omega_{\pm\mathbb{1}}\mathrm{SO}:=\bigcup_{j=1}^\infty\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2j}$.

Lemma 5.4 *Let $E \to B$ be a vector bundle over a space homotopy equivalent to a finite CW–complex. Then E is stably turnable if and only if $\mathrm{Turn}(E^\infty) \to B$ admits a section.*

Proof Remark 5.3 and Lemma 2.12 tell us that a vector bundle $E \to B$ being stably turnable implies that $\mathrm{Turn}(E^\infty) \to B$ admits a section. Conversely, noting that the inclusion map $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2j} \to \Omega_{\pm\mathbb{1}}\mathrm{SO}$ is $(2j-2)$–connected and B is homotopy equivalent to a finite CW–complex, it follows from the obstruction theory (cf [18, Chapter VI, Section 5]) that if $\mathrm{Turn}(E^\infty) \to B$ admits a section, then there must exist a nonnegative integer j such that $\mathrm{Turn}(E \oplus \underline{\mathbb{R}}^j)$ admits a section, which means that $E \to B$ is stably turnable and the proof is complete. □

Given Lemma 5.4, an efficient way to define the notion of a stable turning is via a section of $\mathrm{Turn}(E^\infty)$.

Definition 5.5 (stable turning and the space of stable turnings) A *stable turning* of a vector bundle $E \to B$ is a section of the fibre bundle $\mathrm{Turn}(E^\infty) \to B$. The *space of stable turnings* of E, denoted by $\Gamma(\mathrm{Turn}(E^\infty))$, is the space of sections of $\mathrm{Turn}(E^\infty) \to B$, equipped with the restriction of the compact–open topology.

We next consider *minimal turnings*, which are turnings that restrict to minimal geodesics in each fibre. The manifold SO_{2k} has a canonical Lie invariant metric, which allows us to consider geodesics in SO_{2k}. We write $\Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}_{2k} \subset \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ for the subspace of paths which are minimal geodesics in SO_{2k} from $\mathbb{1}$ to $-\mathbb{1}$. Since the conjugation action of SO_{2k} on itself is by isometries, it preserves geodesics and so $\Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}_{2k}$ is an SO_{2k}–subspace of $\Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k}$. For any nonnegative integer j with $n+j$ even, denote by $\mathrm{Turn}^{\min}(E \oplus \underline{\mathbb{R}}^j) \subset \mathrm{Turn}(E \oplus \underline{\mathbb{R}}^j)$ the subbundle of minimal turnings and set

$$\mathrm{Turn}^{\min}(E^\infty) := \bigcup_{n+j\ \text{even}} \mathrm{Turn}^{\min}(E \oplus \underline{\mathbb{R}}^j),$$

which is a fibre bundle over B with fibre $\Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO} = \bigcup_{j=1}^{\infty} \Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}_{2j}$.

Now $\Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}$ is an SO–subset of $\Omega_{\pm\mathbb{1}}\mathrm{SO}$ and hence $\mathrm{Turn}^{\min}(E^\infty)$ is a subbundle of $\mathrm{Turn}(E^\infty)$. Since the inclusion $\Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO} \to \Omega_{\pm\mathbb{1}}\mathrm{SO}$ is a weak homotopy equivalence [14, Theorem 24.5], part (a) of the next lemma follows immediately. Part (b) follows from Lemma 5.4.

Lemma 5.6 *Let $E \to B$ be a vector bundle over a space homotopy equivalent to a finite CW–complex. Then the following hold*:

(a) *The fibrewise inclusion* $\mathrm{Turn}^{\min}(E^\infty) \to \mathrm{Turn}(E^\infty)$ *is a weak homotopy equivalence.*

(b) *E is stably turnable if and only if* $\mathrm{Turn}^{\min}(E^\infty) \to B$ *admits a section.* □

Now we recall the relationship of complex structures on a bundle and minimal turnings. A complex structure on a rank-$2k$ vector bundle E is an element $J \in \mathcal{G}_E$ such that $J^2 = -\mathbb{1}$. In particular, a complex structure on E endows each fibre of E with the structure of a complex vector space. If E has rank n, then a stable complex structure on E is a complex structure on $E \oplus \underline{\mathbb{R}}^j$ for some $j \geq 0$ with $n+j$ even.

Now let

$$\mathcal{J}_{2k} := \{J \in \mathrm{SO}_{2k} \mid J^2 = -\mathbb{1}\}$$

be the space of (special orthogonal) complex structures on $\mathbb{R}^{2k}$. The space $\mathcal{J}_{2k}$ is an SO_{2k}–space, where SO_{2k} acts on $\mathcal{J}_{2k}$ by conjugation. For any nonnegative integer j with $n+j$ even, define

$$\mathcal{J}(E \oplus \underline{\mathbb{R}}^j) := \mathrm{Fr}(E \oplus \underline{\mathbb{R}}^j) \times_{\mathrm{SO}_{n+j}} \mathcal{J}_{n+j} \subset \mathrm{Aut}(E \oplus \underline{\mathbb{R}}^j)$$

to be the bundle of fibrewise complex structures on $E \oplus \underline{\mathbb{R}}^j$. Regarding $\mathcal{G}_E = \Gamma(\mathrm{Aut}(E))$ we see that a complex structure on E is equivalent to a section of $\mathcal{J}(E) \to B$ and that a stable complex structure on E is equivalent to a section of $\mathcal{J}(E \oplus \underline{\mathbb{R}}^j) \to B$. Letting j tend to infinity, we define

$$\mathcal{J}(E^\infty) := \bigcup_{n+j \text{ even}} \mathcal{J}(E \oplus \underline{\mathbb{R}}^j)$$

to be the bundle of fibrewise stable complex structures on E. The space $\mathcal{J}(E^\infty)$ is the total space of a bundle over B with fibre $\mathcal{J}_\infty := \bigcup_{j=1}^\infty \mathcal{J}_{2j}$, and we have:

Lemma 5.7 *A vector bundle $E \to B$ over a space homotopy equivalent to a finite CW–complex admits a stable complex structure if and only if $\mathcal{J}(E^\infty) \to B$ admits a section.* □

In light of Lemma 5.7, an efficient way to define the notion of a stable complex structure is via a section of $\mathcal{J}(E^\infty)$.

Definition 5.8 (stable complex structure and the space of stable complex structures) A *stable complex structure* on an oriented vector bundle $E \to B$ is a section of the fibre bundle $\mathcal{J}(E^\infty) \to B$. The *space of stable complex structures* of E, denoted by $\Gamma(\mathcal{J}(E^\infty))$, is the space of sections of $\mathcal{J}(E^\infty) \to B$, equipped with the restriction of the compact–open topology.

Now a complex structure on $\mathbb{R}^{2k}$ defines a minimal geodesic in SO_{2k} via complex multiplication with unit complex numbers in the upper half-plane. Explicitly, we define the map

$$\varphi_{2k} \colon \mathcal{J}_{2k} \to \Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}_{2k}, \quad J \mapsto (t \mapsto \exp(\pi t J)),$$

where $\exp\colon T_{\mathbb{1}}\mathrm{SO}_{2k} \to \mathrm{SO}_{2k}$ is the exponential map from the tangent space over the identity. Then φ_{2k} is an SO_{2k}–equivariant homeomorphism; see eg Milnor [14, Lemma 24.1]. It follows for any rank-$2k$ bundle $E \to B$ that φ_{2k} induces a fibrewise homeomorphism $\varphi_{2k} \colon \mathcal{J}(E) \to \mathrm{Turn}^{\min}(E)$. Stably, we define the map $\varphi_\infty := \lim_{j\to\infty} \varphi_{2j} \colon \mathcal{J}_\infty \to \Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}$ and we have:

Lemma 5.9 *Let $E \to B$ be a vector bundle. The map $\varphi_\infty\colon \mathcal{J}_\infty \to \Omega^{\min}_{\pm\mathbb{1}}\mathrm{SO}$ induces a fibrewise homeomorphism $\varphi_\infty \colon \mathcal{J}(E^\infty) \to \mathrm{Turn}^{\min}(E^\infty)$.* □

Now the fibre bundle map

$$\mathcal{J}(E^\infty) \to \mathrm{Turn}^{\min}(E^\infty) \to \mathrm{Turn}(E^\infty)$$

induces a map $\Gamma(\mathcal{J}(E^\infty)) \to \Gamma(\mathrm{Turn}(E^\infty))$. By combining Lemmas 5.9, 5.6(a), 5.4 and 5.7 we obtain:

Theorem 5.10 *Let $E \to B$ be a vector bundle over a space homotopy equivalent to a finite CW–complex. The induced map $\Gamma(\mathscr{J}(E^\infty)) \to \Gamma(\mathrm{Turn}(E^\infty))$ from the space of stable complex structures on E to the space of stable turnings on E is a weak homotopy equivalence. Hence E is stably turnable if and only if E admits a stable complex structure.* □

In the remainder of this section we present an alternative proof of the final sentence of Theorem 5.10 using K–theory. Let BU (resp. BO) be the classifying space of the stable unitary group U (resp. stable orthogonal group O). Since O/U is homotopy equivalent to $\Omega^2 BO$ (cf [4]), the canonical fibration

$$O/U \hookrightarrow BU \to BO$$

gives rise to the Bott exact sequence (cf Bott [5, (12.2)] or Atiyah [1, (3.4)])

$$\cdots \to \mathrm{KO}^{-2}(B) \to \widetilde{\mathrm{KU}}(B) \xrightarrow{r} \widetilde{\mathrm{KO}}(B) \xrightarrow{\partial} \widetilde{\mathrm{KO}}^{-1}(B) \to \cdots . \tag{5-1}$$

Here r is the real reduction homomorphism and ∂ is the homomorphism given by $\partial(\xi) = \eta \cdot \xi$, where η is the generator of $\mathrm{KO}^{-1}(\mathrm{pt}) = \mathbb{Z}/2$ and $\cdot$ denotes the product in real K–theory.

Now for a rank-n vector bundle $E \to B$, let $\xi_E \in \widetilde{\mathrm{KO}}(B)$ be the real K–theory class $\xi_E := E \ominus \underline{\mathbb{R}}^n$, which is represented by the virtual bundle obtained as the formal difference of E and the trivial rank-n bundle over B. When B is homotopy equivalent to a finite CW–complex, the bundle E admits a stable complex structure if and only if the real K–theory class ξ_E lies in the image of the real reduction homomorphism r. Hence the next proposition follows from the Bott exact sequence above.

Proposition 5.11 *Let $E \to B$ be a vector bundle over a space homotopy equivalent to a finite CW–complex. Then E admits a stable complex structure if and only if $\eta \cdot \xi_E = 0$.* □

We now relate the boundary map ∂ in (5-1) to turnings of vector bundles. Let $\psi\colon E \to E$ be an automorphism of a vector bundle $\pi\colon E \to B$. The *mapping torus* of ψ is the vector bundle $T(\psi) \to B \times S^1$, where

$$T(\psi) := (E \times I)/\simeq$$

with $(v, 0) \simeq (\psi(v), 1)$, and the bundle map is given by $[(v,t)] \mapsto (\pi(v), [t])$. We note that $T(\psi)$ is orientable if and only if ψ is orientation-preserving, in which case $T(\psi)$ inherits an orientation from E. Let $\widehat{\otimes}$ denote the exterior tensor product of vector bundles.

Lemma 5.12 *Let $\underline{\mathbb{R}}$ denote the trivial line bundle over S^1. The bundle $T(\psi)$ is isomorphic to $E \mathbin{\widehat{\otimes}} \underline{\mathbb{R}}$ if and only if ψ is homotopic to the identity.*

Proof The classification of vector bundles [9, Chapter 3, Section 4] shows that a vector bundle isomorphism $E \to E'$ is equivalent to a vector bundle $F \to B \times I$ such that $F|_{B\times\{0\}} = E$ and $F|_{B\times\{1\}} = E'$. Similarly, a vector bundle automorphism $E \to E$ is equivalent to a vector bundle $F \to B \times S^1$ such that $F|_{B\times\{1\}} = E$. In particular, the bundle $E \mathbin{\widehat{\otimes}} \underline{\mathbb{R}}$ corresponds to the identity automorphism $\mathbb{1}_E\colon E \to E$. It follows that a vector bundle isomorphism $T(\psi) \to E \mathbin{\widehat{\otimes}} \underline{\mathbb{R}}$ is equivalent to a bundle over $B \times S^1 \times I$ and so is equivalent to a path of bundle automorphisms from ψ to $\mathbb{1}_E$. □

An alternative proof of the final sentence of Theorem 5.10 Let $L \to S^1$ be the Möbius bundle, ie the nontrivial rank-1 bundle. Then $\eta \in \mathrm{KO}^{-1}(\mathrm{pt}) = \widetilde{\mathrm{KO}}(S^1)$ is represented by the virtual bundle $L \ominus \underline{\mathbb{R}}$. Let ξ_E be represented by the virtual bundle $E' \ominus \underline{\mathbb{R}}^{2k}$, where $E' \to B$ is a vector bundle of rank $2k$, which is larger than the formal dimension of B. Then $\eta \cdot \xi_E \in \mathrm{KO}^{-1}(B) = \mathrm{KO}(B \wedge S^1)$ is represented by the virtual bundle

$$(E' \mathbin{\widehat{\otimes}} L) \oplus (\underline{\mathbb{R}}^{2k} \mathbin{\widehat{\otimes}} \underline{\mathbb{R}}) \ominus (\underline{\mathbb{R}}^{2k} \mathbin{\widehat{\otimes}} L) \ominus (E' \mathbin{\widehat{\otimes}} \underline{\mathbb{R}})$$

over $B \times S^1$, which is canonically zero over $B \vee S^1$. Since L is the mapping torus of $-\mathbb{1}_{\mathbb{R}} \colon \mathbb{R} \to \mathbb{R}$, it follows that $\underline{\mathbb{R}}^{2k} \mathbin{\widehat{\otimes}} L$ is the mapping torus of $-\mathbb{1}_{\mathbb{R}^{2k}} \colon \mathbb{R}^{2k} \to \mathbb{R}^{2k}$ and so is trivial by Lemma 5.12. Hence $\eta \cdot \xi_E = 0$ if and only if $E' \mathbin{\widehat{\otimes}} L$ is stably isomorphic to $E' \mathbin{\widehat{\otimes}} \underline{\mathbb{R}}$.

Since L is the mapping torus of $-\mathbb{1}_{\mathbb{R}} \colon \mathbb{R} \to \mathbb{R}$, the bundle $E' \mathbin{\widehat{\otimes}} L$ is the mapping tours of $-\mathbb{1}_{E'} \colon E' \to E'$. Now by Lemma 5.12, $E' \mathbin{\widehat{\otimes}} L$ is isomorphic to $E' \mathbin{\widehat{\otimes}} \underline{\mathbb{R}}$ if and only if E' is turnable. Hence $\eta_E \cdot \xi = 0$ if and only if E' is stably turnable and so the final sentence of Theorem 5.10 follows directly from Proposition 5.11. □

6 Turning rank-$2k$ bundles over $2k$–complexes

In Section 5 we saw that a turning of bundle E induces a stable complex structure on E and in Section 4 we computed the turning obstruction for rank-$2k$ bundles over the $2k$–sphere. In this section we combine these results to gain useful information about the turning obstruction for rank-$2k$ bundles over $2k$–dimensional complexes. Throughout this section, B will be a space that is homotopy equivalent to a connected finite CW–complex of dimension $2k$ or less.

Theorem 6.1 below gives a necessary condition for an oriented rank-$2k$ bundle E over B to be positive-turnable. Its statement requires some preliminary definitions. We will say that a complex structure J on $E \oplus \underline{\mathbb{R}}^{2j}$ for some $j \geq 0$ is *compatible with* E if J induces the same orientation on $E \oplus \underline{\mathbb{R}}^{2j}$ as E does. Recall that $\times 2 \colon H^*(B; \mathbb{Z}/2) \to H^*(B; \mathbb{Z}/4)$ is the map induced by the inclusion of coefficients $\times 2 \colon \mathbb{Z}/2 \to \mathbb{Z}/4$ and the subgroup $I^{2k}(B) \subseteq H^{2k}(B; \mathbb{Z}/4)$, which is defined by

$$I^{2k}(B) = \begin{cases} ((\times 2) \circ \mathrm{Sq}^2 \circ \rho_2)(H^{2k-2}(B; \mathbb{Z})) & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even.} \end{cases}$$

Recall also that ρ_4 denotes reduction mod 4.

Theorem 6.1 *Let $E \to B$ be an oriented rank-$2k$ vector bundle. If E is positive-turnable, then for some $j \geq 0$, $E \oplus \underline{\mathbb{R}}^{2j}$ admits a complex structure J such that J is compatible with E and $c_k(J)$ satisfies*

$$[\rho_4(c_k(J))] = [\rho_4(\mathrm{e}(E))] \in H^{2k}(B; \mathbb{Z}/4)/I^{2k}(B).$$

Example 6.3 below shows that there are turnable bundles which do not satisfy the condition of Theorem 6.1. However, this condition is sufficient for the bundle to be turnable if k is odd, and in many cases if k is even.

Theorem 6.2 *Let E be an oriented rank-$2k$ vector bundle over B with either k odd, or k even and $H^{2k}(B;\mathbb{Z})$ 2–torsion free. If E admits a stable complex structure J such that*

$$[\rho_4(c_k(J))] = [\rho_4(\mathrm{e}(E))] \in H^{2k}(B;\mathbb{Z}/4)/I^{2k}(B),$$

then E is positive-turnable.

Example 6.3 Let $M(\mathbb{Z}/2, 4k-1) := S^{4k-1} \cup_2 D^{4k}$ be the mod 2 Moore space with

$$H^{4k}(M(\mathbb{Z}/2, 4k-1); \mathbb{Z}) = \mathbb{Z}/2$$

and let $c\colon M(\mathbb{Z}/2, 4k-1) \to S^{4k}$ be the map collapsing the $(4k-1)$–cell. Since c is the suspension of the map $c'\colon M(\mathbb{Z}/2, 4k-2) \to S^{4k-1}$, which collapses the $(4k-2)$–cell of $M(\mathbb{Z}/2, 4k-2)$, we see that the γ–turning obstruction of $E := c^*(TS^{4k})$ is the pullback $c'^*(\mathrm{TO}_\gamma(\tau_{4k})) \in [M(\mathbb{Z}/2, 4k-2), \mathrm{SO}_{2k}]$. Since $\tau_{4k} \notin 2\pi_{4k-1}(\mathrm{SO}_{4k})$, it follows that $c'^*(\mathrm{TO}_\gamma(\tau_{4k})) \neq 0$. By Proposition 3.2, E is not γ–turnable for any path γ and so E is not turnable. However, E is stably parallelisable and so admits a stable complex structure J with $c_{2k}(J) = 0$. Moreover $\mathrm{e}(E) = 0$, since $\mathrm{e}(TS^{4k}) \in 2H^{4k}(S^{4k};\mathbb{Z})$. Hence E satisfies the condition of Theorem 6.1.

Before proving Theorems 6.1 and 6.2, we give an application of Theorem 6.2 which shows, in particular, that for all $l \geq 1$ there are $8l$–manifolds M whose tangent bundles are turnable but not complex, eg $M = S^4 \times S^4$.

Corollary 6.4 *For $i > 0$, let M be an orientable $4i$–manifold such that the following hold:*

(a) *M is stably parallelisable.*

(b) $\chi(M) \neq 0$.

(c) $\mathrm{KU}(M) \to \mathrm{KO}(M)$ *is injective.*

Then TM does not admit a complex structure but TM is turnable if and only if $\chi(M) \equiv 0 \bmod 4$. In particular, for all $m \geq 1$ and $l \geq 0$, the manifolds $M_l := \sharp_l(S^{4m} \times S^{4m})$ are such that TM_l does not admit a complex structure but TM_l is turnable if and only if l is odd.

Proof TM is stably trivial and $\mathrm{KU}(M) \to \mathrm{KO}(M)$ is injective, so E_J, the complex bundle underlying J, is trivial, so $c_{4j}(J) = 0$. On the other hand, $\mathrm{e}(TM) = \chi(M) = 2 + 2l$ by the Poincaré–Hopf theorem [8, page 113]. Hence by Theorem 6.6, TM does not admit a complex structure. However, by Theorems 6.1 and 6.2, TM is turnable if and only if l is odd. □

We now turn to the proofs of Theorems 6.1 and 6.2. Without loss of generality, we may assume that B is a finite CW–complex of dimension at most $2k$. It will be useful to modify a rank-$2k$ bundle $E \to B$ over the $2k$–cells of B and we first describe this process. Let $F \to S^{2k}$ be an oriented rank-$2k$ bundle with clutching function g. Given a $2k$–cell $e^{2k}_\alpha \subset B$ and a $2k$–disc D^{2k}_α embedded in the interior of e^{2k}_α, we define the bundle $E \sharp_\alpha F \to B$ as follows: Write $B = B^\circ \cup_{S^{2k-1}_\alpha} D^{2k}_\alpha$, where $B^\circ := B \setminus \mathrm{Int}(D^{2k}_\alpha)$ and

$$E = E^\circ \cup_\phi (\mathbb{R}^{2k} \times D^{2k}_\alpha),$$

where E° is the restriction of E to B° and $\phi\colon E|_{S^{2k-1}_\alpha} \to \mathbb{R}^{2k} \times S^{2k-1}_\alpha$ is a bundle isomorphism. Define

$$E \sharp_\alpha F := E^\circ \cup_{g\circ\phi} (\mathbb{R}^{2k} \times D^{2k}_\alpha).$$

We let $e^{2k*}_\alpha \in C^{2k}(B)$ be the $2k$–dimensional cellular cochain of B which evaluates to 1 on the $2k$–cell e^{2k}_α, and 0 on every other $2k$–cell. The next result follows from an elementary application of obstruction theory.

Lemma 6.5 *Let E_1 and E_2 be rank-$2k$ oriented bundles over B. Then E_1 is stably equivalent to E_2 if and only if there are $2k$–cells $\alpha_1, \dots, \alpha_n$ of B and integers $j_1, \dots, j_n$ such that*

$$E_2 \cong E_1 \#_{\alpha_1} j_1 TS^{2k} \#_{\alpha_2} \cdots \#_{\alpha_n} j_n TS^{2k}.$$

Moreover, in this case the Euler classes of E_1 and E_2 are represented by cocycles $\mathrm{cc}(\mathrm{e}(E_1))$ and $\mathrm{cc}(\mathrm{e}(E_2))$ such that $\mathrm{cc}(\mathrm{e}(E_2)) = \mathrm{cc}(\mathrm{e}(E_1)) + \sum_{i=1}^n 2 j_i e^{2k*}_{\alpha_i}$. □

Lemma 6.5 can be used to prove the following theorem of Thomas.

Theorem 6.6 [16, Theorem 1.7] *Let E be a rank-$2k$ vector bundle over B, where $H^{2k}(B;\mathbb{Z})$ is 2–torsion free. Then E admits a complex structure if and only if E admits a stable complex structure J such that $c_k(J) = \mathrm{e}(E)$.* □

Proof of Theorem 6.1 Recall the universal rank-$2k$ turning bundle

$$BT_{2k} = E\mathrm{SO}_{2k} \times_{\mathrm{SO}_{2k}} \Omega_{\pm\mathbb{1}}\mathrm{SO}_{2k} \to B\mathrm{SO}_{2k}$$

from Definition 2.13 and define $BT := E\mathrm{SO} \times_{\mathrm{SO}} \Omega_{\pm\mathbb{1}}\mathrm{SO} \to B\mathrm{SO}$ to be the universal stable turning bundle. There is a natural map $BT_{2k} \to BT$ and by the results of Section 5, $BT \simeq BU$. Hence we consider the commutative diagram

$$\begin{array}{ccccc} & & BU_k & \longrightarrow & BU \\ & & \downarrow & & \downarrow \simeq \\ & \overset{\bar f}{\nearrow} & BT_{2k} & \xrightarrow{S} & BT \\ & & \downarrow & & \downarrow \\ B & \xrightarrow{f} & B\mathrm{SO}_{2k} & \longrightarrow & B\mathrm{SO} \end{array}$$

where f classifies E and $\bar f$ classifies a positive-turning on E. Since the natural map $BU \to BT$ is a fibre homotopy equivalence over $B\mathrm{SO}$, the turning on E induces a stable complex structure J on E. Consider the lifting problem

$$\begin{array}{ccc} & & BU_k \\ & \nearrow & \downarrow \\ B & \xrightarrow{S\circ\bar f} & BT \end{array}$$

Since the homotopy fibre of $BU_k \to BT$ is $2k$–connected, the map $J := S \circ \bar{f}\colon B \to BT$ has a lift to $J'\colon B \to BU_k$, which is unique up to homotopy. The map J' defines a complex rank-k bundle (E', J') over B, where E' is stably equivalent to E.

By Lemma 6.5, there is a bundle isomorphism

$$\alpha\colon E \to E' \#_{\alpha_1} j_1 TS^{2k} \# \cdots \#_{\alpha_n} j_n TS^{2k}. \tag{6-1}$$

We set $B^\circ := B \setminus \bigl(\bigcup_{i=1}^n \operatorname{Int}(D^{2k}_{\alpha_i})\bigr)$, $E^\circ := E|_{B^\circ}$, $E'^\circ := E'|_{B^\circ}$ and $\alpha^\circ := \alpha|_{E^\circ}$. The lift $\bar{f}$ defines a turning ψ_t on E, which restricts to a turning ψ_t° on E°, and the complex structure J' defines a turning ψ'_t on E', which restricts to a turning $\psi_t'^\circ$ on E'° that pulls back along α° to a turning $(\alpha^\circ)^*(\psi_t'^\circ)$ on E°.

If ψ_t° and $(\alpha^\circ)^*(\psi_t'^\circ)$ are equivalent turnings on E°, then the obstruction to extending $(\alpha^\circ)^*(\psi_t'^\circ)$ to E vanishes. On the other hand, the obstruction to extending $\psi_t'^\circ$ to E' vanishes. It follows that the cocycle $\sum_{i=1}^n j_i\, \mathrm{TO}_+(\tau_{2k}) e^{2k*}_{\alpha_i}$ represents 0 in $H^{2k}(B; \pi_{2k}(\mathrm{SO}_{2k}))$, so the cocycle $\sum_{i=1}^n j_i 2 e^{2k*}_{\alpha_i}$ represents 0 in $H^{2k}(B; \mathbb{Z}/4)$ and hence $\rho_4(c_k(J)) = \rho_4(\mathrm{e}(E))$.

If ψ_t° and $(\alpha^\circ)^*(\psi_t'^\circ)$ are not equivalent turnings on E°, then they differ over the $(2k-2)$–cells and $(2k-1)$–cells of B°. If k is even, this variation does not effect the cohomology class represented by $\sum_{i=1}^n j_i\, \mathrm{ev}_*(\mathrm{TO}_+(\tau_{2k}) e^{2k*}_{\alpha_i})$. If k is odd, then changing the turning can alter the values of the turning obstruction over the cells $e^{2k}_{\alpha_i}$ by any element of $((\times 2) \circ \mathrm{Sq}^2 \circ \rho_2)(H^{2k-2}(B; \mathbb{Z}))$. □

Proof of Theorem 6.2 Suppose that E admits a stable complex structure J. Then, as in the proof of Theorem 6.1, J descends to a complex rank-k bundle (E', J') on B such that E and E' are stably isomorphic. Since (E', J') is a complex bundle, E' is positive-turnable. Consider the bundle isomorphism α from (6-1). Assume first k is even and that $H^{2k}(B; \mathbb{Z})$ contains no 2–torsion. Since J' stabilises to J, $c_k(J') = c_k(J)$ and so $\rho_4(c_k(J')) = \rho_4(c_k(J)) = \rho_4(\mathrm{e}(E))$. It follows that each j_i in (6-1) is even. Now $2\,TS^{2k}$ is positive-turnable by Theorem 4.1. Since E' is positive-turnable, it follows that E is positive-turnable.

When k is odd the argument is similar to the case where k is even, but requires adjustments. We first note that $\rho_2(c_k(J')) = w_{2k}(E) = \rho_2(\mathrm{e}(E))$. Moreover, for k odd, $\mathrm{TO}_+(\tau_{2k}) = \tau_{2k}\eta_{2k-1} \in 2\pi_{2k}(\mathrm{SO}_{2k})$. It follows that if some j_i in (6-1) is odd, then we can modify the turning of E' to ensure that the pullback of its restriction to E'° extends to all of E. □

Proof of Theorem 1.4 and Remark 1.5 These statements follow from Theorems 6.1 and 6.2 and the fact that a turnable bundle over a connected space is either positively turnable or negatively turnable. □

Question 6.7 From the proof of Theorem 6.1 we see that there is a map $BT_{2k} \to BU$, obtained by composing the stabilisation map $BT_{2k} \to BT$ with a homotopy inverse of the natural map $BU \to BT$. This map $BT_{2k} \to BU$ encodes the fact that a turned bundle (E, ψ_t) over a finite CW–complex defines a stable turning of E and a stable complex structure J on $E \oplus \underline{\mathbb{R}}^j$ for some j. It follows that the universal bundle $VT_{2k} \to BT_{2k}$ has a well-defined homotopy class of stable complex structures and hence there are well-defined Chern classes $c_i(VT_{2k}) \in H^{2i}(BT_{2k};\mathbb{Z})$.

We pose the following question:

> What special properties, if any, do the Chern classes of VT_{2k} possess?

For example, one would expect that $c_i(VT_{2k}) = 0$ for all $i > k$, and presumably also that there is an equality $[\rho_4(c_k(VT_{2k}))] = [\rho_4(\mathrm{e}(VT_{2k}))] \in H^{2k}(BT_{2k};\mathbb{Z}/4)/I^{2k}(BT_{2k})$.

References

[1] **M F Atiyah**, *K–theory and reality*, Q. J. Math. 17 (1966) 367–386 MR Zbl

[2] **M G Barratt**, **M E Mahowald**, *The metastable homotopy of* O(n), Bull. Amer. Math. Soc. 70 (1964) 758–760 MR Zbl

[3] **A Borel**, **J-P Serre**, *Détermination des p–puissances réduites de Steenrod dans la cohomologie des groupes classiques*: *applications*, C. R. Acad. Sci. Paris 233 (1951) 680–682 MR Zbl

[4] **R Bott**, *The stable homotopy of the classical groups*, Ann. of Math. 70 (1959) 313–337 MR Zbl

[5] **R Bott**, *Lectures on* $K(X)$, Benjamin, New York (1969) MR Zbl

[6] **R Bott**, **J Milnor**, *On the parallelizability of the spheres*, Bull. Amer. Math. Soc. 64 (1958) 87–89 MR Zbl

[7] **H Hamanaka**, **A Kono**, *A note on the Samelson products in* $\pi_*(\mathrm{SO}(2n))$ *and the group* $[\mathrm{SO}(2n), \mathrm{SO}(2n)]$, Topology Appl. 154 (2007) 567–572 MR Zbl

[8] **H Hopf**, *Differential geometry in the large*, 2nd edition, Lecture Notes in Math. 1000, Springer (1989) MR Zbl

[9] **D Husemoller**, *Fibre bundles*, 3rd edition, Graduate Texts in Math. 20, Springer (1994) MR Zbl

[10] **M A Kervaire**, *A note on obstructions and characteristic classes*, Amer. J. Math. 81 (1959) 773–784 MR

[11] **M A Kervaire**, *Some nonstable homotopy groups of Lie groups*, Illinois J. Math. 4 (1960) 161–169 MR Zbl

[12] **A Kirchoff**, *Sur l'existence des certains champs tensoriels sur les sphères à n dimensions*, C. R. Acad. Sci. Paris 225 (1947) 1258–1260 MR Zbl

[13] **D Kishimoto**, **I Membrillo-Solis**, **S Theriault**, *The homotopy types of* SO(4)*–gauge groups*, Eur. J. Math. 7 (2021) 1245–1252 MR Zbl

[14] **J Milnor**, *Morse theory*, Ann. of Math. Stud. 51, Princeton Univ. Press (1963) MR Zbl

[15] **H Samelson**, *A connection between the Whitehead and the Pontryagin product*, Amer. J. Math. 75 (1953) 744–752 MR Zbl

[16] **E Thomas**, *Complex structures on real vector bundles*, Amer. J. Math. 89 (1967) 887–908 MR Zbl

[17] **C T C Wall**, *Classification of* $(n-1)$*–connected* $2n$*–manifolds*, Ann. of Math. 75 (1962) 163–189 MR Zbl

[18] **G W Whitehead**, *Elements of homotopy theory*, Graduate Texts in Math. 61, Springer (1978) MR Zbl

[19] **C Wockel**, *The Samelson product and rational homotopy for gauge groups*, Abh. Math. Sem. Univ. Hamburg 77 (2007) 219–228 MR Zbl

School of Mathematics & Statistics, University of Melbourne
Parkville, VIC, Australia

School of Mathematics and Statistics, University of Glasgow
Glasgow, United Kingdom

School of Mathematics and Statistics, University of Melbourne
Parkville, VIC, Australia

School of Mathematics and Statistics, Henan University
Kaifeng, Henan, China

dcrowley@unimelb.edu.au, csaba.nagy@glasgow.ac.uk, blake.sims@unimelb.edu.au, yhj@amss.ac.cn

Received: 21 November 2022

Algebraic & Geometric Topology **24**:5 (2024) 2851–2888
DOI: 10.2140/agt.2024.24.2851
Published: 19 August 2024

Rigidification of cubical quasicategories

PIERRE-LOUIS CURIEN
MURIEL LIVERNET
GABRIEL SAADIA

We construct a cubical analogue of the rigidification functor from quasicategories to simplicial categories present in the work of Joyal and Lurie. We define a functor $\mathfrak{C}^{\square}$ from the category $c\mathcal{S}et$ of cubical sets of Doherty, Kapulkin, Lindsey, and Sattler to the category $s\mathcal{C}at$ of (small) simplicial categories. We show that this rigidification functor establishes a Quillen equivalence between the Joyal model structure on $c\mathcal{S}et$ (as it is called by the four authors) and Bergner's model structure on $s\mathcal{C}at$. We follow the approach to rigidification of Dugger and Spivak, adapting their framework of necklaces to the cubical setting.

18N60; 06A07, 20B30

Introduction

The last decades have seen an explosion of the use of ∞–categories in various fields such as algebraic topology, algebraic geometry and homotopy type theory. In the early 2000s, various definitions of ∞–categories have emerged, starting from the notion of quasicategories developed by Joyal [2008] and Lurie [2009] based on the definition of Boardman and Vogt [1973]. Other definitions have been explored such as enriched categories in spaces (or Kan complexes), or complete Segal spaces to name a few. By model of ∞–categories we mean a category with a Quillen model category structure whose fibrant-cofibrant objects are the ∞–categories in consideration and whose notion of weak equivalence corresponds to a good notion of equivalence of ∞–categories. Bergner's book [2018] clearly explains these different models and the Quillen equivalences relating them.

For instance, the model for quasicategories is the Joyal model category structure on the category $s\mathcal{S}et$ of simplicial sets, while the model for categories enriched in Kan complexes is the Bergner model structure on the category $s\mathcal{C}at$ of (small) simplicial categories. There exists a so-called rigidification functor $\mathfrak{C}^\Delta$ from $s\mathcal{S}et$ to the category $s\mathcal{C}at$ which is a Quillen equivalence between these two models. This functor is called rigidification because simplicial categories have a strict composition of 1–morphisms, as opposed to quasicategories where only weak compositions exist. The construction of the rigidification as well as the proof that it yields a Quillen equivalence have been achieved first in an unpublished manuscript of Joyal [2007], then by Lurie [2009], and then by Dugger and Spivak [2011a]. Dugger and Spivak build their rigidification functor using a technical tool, necklaces, and prove that there is a zigzag of weak equivalences of simplicial categories between their construction and Lurie's one. The key idea of this construction is the following: given an ordered simplicial set X, the simplicial set $\mathfrak{C}^\Delta(X)(a,b)$ is the nerve of a poset whose objects are directed paths and relations are generated by 2–simplices in X. If $X = \Delta^n$, one obtains the subset lattice of an ordered set.

Cubical sets have been often considered as an alternative to simplicial sets in combinatorial topology, including in the early work of Kan and Serre (see eg [Serre 1951]). It has been also developed in computer science, in particular in concurrency theory (see eg [Fajstrup et al. 2016; Gaucher 2008; Pratt 1991]) and in homotopy type theory (see eg [Cohen et al. 2018]). Based on the work of Kapulkin, Lindsey and Wong [Kapulkin et al. 2019], Doherty, Kapulkin, Lindsey, and Sattler [Doherty et al. 2024] defined a notion of cubical quasicategory, and have constructed a model category structure on cubical sets, analogous to the Joyal structure on simplicial sets, whose fibrant-cofibrant objects are cubical quasicategories. They also show that the categories $c\mathcal{S}et$ of cubical sets and $s\mathcal{S}et$ are related by two adjunctions. The first one is $T \dashv U$, where $T\colon c\mathcal{S}et \to s\mathcal{S}et$ is a triangulation functor, and the second one is $Q \dashv \int$, where $Q\colon s\mathcal{S}et \to c\mathcal{S}et$ is a "cubification" functor implementing simplices as cubes with some degenerate faces. Both give rise to Quillen equivalences between these model category structures, so cubical quasicategories provide another definition for the notion of ∞–category. By plugging together the Quillen equivalences $T\colon c\mathcal{S}et \to s\mathcal{S}et$ and $\mathfrak{C}^\Delta\colon s\mathcal{S}et \to s\mathcal{C}at$ of *triangulation* and rigidification, we get a Quillen equivalence between Joyal model structure of $c\mathcal{S}et$ and Bergner's model structure on $s\mathcal{C}at$.

The goal of this paper is to build a different, direct Quillen equivalence $\mathfrak{C}^\square\colon c\mathcal{S}et \to s\mathcal{C}at$ using directed paths in the spirit of Dugger and Spivak. Note that the same notion of directed path is used in directed homotopy theory with applications to computer science. We refer the interested reader to the papers by Ziemiański [2017; 2020]. In particular, for a representable cubical set $\square^n$ and for two vertices a and b, the simplicial set $\mathfrak{C}^\square(\square^n)(a,b)$ is the nerve of a poset whose objects are directed paths from a to b in the n–cube, and relations are generated by 2–cubes in $\square^n$. We prove a result of independent interest, namely that this poset is isomorphic to the weak Bruhat order on a symmetric group. Following closely the techniques developed by Dugger and Spivak, we prove that the functor $\mathfrak{C}^\square$ is the left adjoint of a Quillen equivalence between two models of ∞–categories, making use this time of the *cubification* equivalence of [Doherty et al. 2024], by showing that $\mathfrak{C}^\Delta$ factorises through cubification via our rigidification, up to natural homotopy, ie $\mathfrak{C}^{\square} \circ Q \overset{\sim}{\Longrightarrow} \mathfrak{C}^{\Delta}$, and then concluding by the two-out-of-three property.

Plan of the paper

In Section 1, we recall Bergner's model structure and the material from [Doherty et al. 2024] needed for our purposes. Section 2 is devoted to the study of paths and necklaces (adapted from [Dugger and Spivak 2011a; 2011b]). We define our rigidification functor and study its properties in Section 3. The Quillen equivalence is established in Section 4. Appendices A and B deal with relevant categorical and combinatorial matters, respectively.

1 Recollection and notation

1.1 Simplicial rigidification

We recall the *Bergner model structure* of $s\mathcal{C}at$, where $s\mathcal{C}at$ denotes the category of simplicial categories, that is, categories enriched in simplicial sets. We recall that given a simplicial category $\mathcal{C}$, the category $\pi_0(\mathcal{C})$ has as objects those of $\mathcal{C}$ and that (for all a, b) $\pi_0(\mathcal{C})(a, b)$ is the set of connected components of the simplicial set $\mathcal{C}(a, b)$.

The Bergner model structure of $s\mathcal{C}at$ is the enriched model structure coming from the usual Kan–Quillen model structure on $s\mathcal{S}et$.

- Weak equivalences are *Dwyer–Kan equivalences*, that is, functors $F\colon \mathcal{C} \to \mathcal{D}$ such that
  - $\pi_0(F)\colon \pi_0(\mathcal{C}) \to \pi_0(\mathcal{D})$ is essentially surjective, and
  - for all $x, y \in \mathrm{Ob}(\mathcal{C})$, the map $F_{x,y}\colon \mathcal{C}(x, y) \to \mathcal{D}(Fx, Fy)$ is a Kan–Quillen equivalence.
- Fibrations are *Dwyer–Kan fibrations*, that is, functors $F\colon \mathcal{C} \to \mathcal{D}$ such that
  - $\pi_0(F)\colon \pi_0(\mathcal{C}) \to \pi_0(\mathcal{D})$ is an isofibration between categories, and
  - for all $x, y \in \mathrm{Ob}(\mathcal{C})$, the map $F_{x,y}\colon \mathcal{C}(x, y) \to \mathcal{D}(Fx, Fy)$ is a Kan fibration.

See [Bergner 2018] for more details.

Note that if F happens to be a bijection on objects, then $\pi_0(F)$ is a fortiori essentially surjective. This will be the case in our main result, so we will have to focus only on the second condition for DK-equivalences.

For the next proposition we use notation of Section A.1.

Proposition 1.1.1 *The following two functors form a Quillen adjunction*

$$s\mathcal{S}et \underset{\mathrm{Hom}}{\overset{\Sigma}{\underset{\longleftarrow}{\overset{\longrightarrow}{\bot}}}} s\mathcal{C}at_{*,*}$$

where

- $s\mathcal{C}at_{*,*}$ *stands for the category of bipointed (small) simplicial categories, with the model structure induced by the Bergner structure on* $s\mathcal{C}at$,

- *the model structure on* $s\mathcal{S}et$ *is the Joyal structure*,
- $\Sigma(S)$ *is the simplicial category with two objects*, α *and* ω, *and with only one nontrivial mapping space* $\mathrm{Hom}(\alpha,\omega) = S$, *and*
- $\mathrm{Hom}(\mathcal{C}_{x,y}) = \mathcal{C}(x,y)$.

Proof The functor Hom is a right adjoint and sends fibrations and acyclic fibrations to fibrations and acyclic fibrations for the Kan–Quillen model category structure on $s\mathcal{S}et$. Since the Kan–Quillen model category structure is a left Bousfield localisation of the Joyal model category structure, we get the result. □

The simplicial rigidification functor $\mathfrak{C}^\Delta : s\mathcal{S}et \to s\mathcal{C}at$ is obtained as a left Kan extension along the Yoneda functor. On the representables, $\mathfrak{C}^\Delta$ is defined as follows:

- $\mathrm{Ob}(\mathfrak{C}^\Delta(\Delta^n)) = \{0,\dots,n\}$.
- For $i \le j$, $\mathfrak{C}^\Delta(\Delta^n)(i,j)$ is the nerve of the poset $\mathcal{P}(]i,j[)$, where $]i,j[$ is the set
$$\{i+1, i+2, \dots, j-1\}.$$
The poset structure is given by subset inclusion. Note that this is the one-point simplicial set if $j = i$ or $j = i+1$. For $i > j$, $\mathfrak{C}^\Delta(\Delta^n)(i,j) = \varnothing$.
- Composition $N(\mathcal{P}(]j,k[)) \times N(\mathcal{P}(]i,j[)) \to N(\mathcal{P}(]i,k[))$ is induced by the function mapping Y, X to $(X \cup \{j\} \cup Y) \setminus \{i,k\}$.

The nerve functor $N : \mathcal{C}at \to s\mathcal{S}et$ being monoidal, it induces a functor from categories enriched in categories to categories enriched in simplicial sets (see [Riehl 2014, Chapter 3] for basics on enriched category theory). We also call this functor the nerve functor and denote it by N. In particular, the simplicial category $\mathfrak{C}^\Delta(\Delta^n)$ is obtained as the nerve of a poset-enriched category.

Remark 1.1.2 The simplicial rigidification functor is built by left Kan extension and so is cocontinuous, which implies in particular that the set of objects of $\mathfrak{C}^\Delta(X)$ is in bijection with X_0. This is a general fact. Indeed, for a functor $F : I \to s\mathcal{C}at$, the set of objects of the simplicial category $\operatorname{colim} F$ is in bijection with the colimit (in $\mathcal{S}et$) of the object functor $\mathrm{Ob} \circ F : I \to \mathcal{S}et$, since the object functor is cocontinuous: it is left adjoint to the codiscrete functor $\mathcal{S}et \xrightarrow{\mathrm{coDisc}} s\mathcal{C}at$ sending a set X to the simplicial category whose set of objects is X and whose simplicial set of morphisms between any two objects is Δ^0.

Theorem 1.1.3 [Bergner 2018, Corollary 7.8.17] *The functor* $\mathfrak{C}^\Delta : s\mathcal{S}et \to s\mathcal{C}at$ *is the left adjoint of a Quillen equivalence between the Joyal model structure on* $s\mathcal{S}et$ *and the Bergner model structure on* $s\mathcal{C}at$.

1.2 Cubical quasicategories

We next present the material of [Doherty et al. 2024] needed for our purposes. There are different notions of cubical sets depending on whether one considers all or part of the negative and positive connections, the diagonals, and the symmetries. In this paper we consider the category of cubical sets with negative connections only, and shall denote it simply □. Note that as in [Doherty et al. 2024], our results certainly hold if we consider the category of cubical sets with positive connections instead, or both connections. We refer the reader interested by the reason why at least one connection is needed for the constructions of the Quillen functors to the introductions of [Doherty et al. 2024; Maltsiniotis 2009], as well as Liang Ze Wong's slides [2020].

The category $\square$ is the subcategory of the category of posets whose objects are $[1]^n, n \geq 0$, and whose morphisms are generated by

- the faces $\partial^n_{i,\epsilon}\colon [1]^{n-1} \to [1]^n$ ($1 \leq i \leq n$ and $\epsilon \in \{0,1\}$), consisting in inserting ϵ at the i–coordinate,
- the degeneracies $\sigma^n_i\colon [1]^n \to [1]^{n-1}$ ($1 \leq i \leq n$), consisting in forgetting the i–coordinate, and
- the negative connections $\gamma^n_{i,0}\colon [1]^n \to [1]^{n-1}$ ($1 \leq i \leq n-1$), mapping $(x_1, \ldots, x_n)$ to

$$(x_1, \ldots, x_{i-1}, \max(x_i, x_{i+1}), x_{i+2}, \ldots, x_n).$$

Adapting [Grandis and Mauri 2003, Theorem 5.1] (see also [Maltsiniotis 2009]) to our case, we have that every map in the category $\square$ can be factored uniquely as a composite

$$(\partial_{c_1,\epsilon_1} \cdots \partial_{c_r,\epsilon_r})(\gamma_{b_1,0} \cdots \gamma_{b_q,0})(\sigma_{a_1} \cdots \sigma_{a_p})$$

with $1 \leq a_1 < \cdots < a_p$, $1 \leq b_1 < \cdots < b_q$ and $c_1 > \cdots > c_r \geq 1$.

In particular, it factors uniquely as an epimorphism followed by a monomorphism. Relying on this factorisation, one can give an alternative presentation of $\square$ by generators (as above) and relations given by cubical identities, as listed in [Doherty et al. 2024] just before Proposition 1.16.

The category of presheaves on $\square$ is called the category of cubical sets and denoted by $c\mathcal{S}et$. The representable presheaves are denoted by $\square^n$, and are called the *n–cubes*.

In addition, the factorisation of Grandis and Mauri in $\square$ induces the existence of the standard form of an n–cube x in a cubical set S. We recollect here [Doherty et al. 2024, Proposition 1.18 and Corollaries 1.19 and 1.20], where, as usual, "nondegenerate" stands for "not in the image of a degeneracy or a connection".

Proposition 1.2.1 *Let S and T be two cubical sets.*

1. *For any n–cube $x\colon \square^n \to S$, there exists a unique decomposition $x = y \circ \varphi$, where $\varphi\colon \square^n \to \square^m$ is an epimorphism and $y\colon \square^m \to S$ is a nondegenerate m–cube.*
2. *Any map $\varphi\colon S \to T$ in $c\mathcal{S}et$ is determined by its action on nondegenerate cubes.*
3. *A map $\varphi\colon S \to T$ is a monomorphism if and only if it maps nondegenerate cubes of S to nondegenerate cubes of T and does so injectively.*

A *vertex* of a cubical set S is an element of S_0 (where $S_0 = S([1]^0)$). The vertices of $\square^n$ are in one-to-one correspondence with the n–tuples $(a_1, \ldots, a_n)$ of $[1]^n$, or equivalently with the subsets of $\{1, \ldots, n\}$. We will use either point of view, depending on the context.

Notation 1.2.2 The order of $[1]^n$ induces an order $\preccurlyeq$ on the vertices of $\square^n$:

$$(a_1,\dots,a_n)\preccurlyeq(b_1,\dots,b_n)\quad\Longleftrightarrow\quad a_i\le b_i\quad\text{for all }1\le i\le n.$$

It is isomorphic to the subset lattice of $\{1,\dots,n\}$ via $(a_1,\dots,a_n)\mapsto\{i\mid a_i=1\}$. Hence it has a least element $\alpha=\varnothing$ and a greatest element $\omega=\{1,\dots,n\}$ (or $\alpha=(0,\dots,0)$ and $\omega=(1,\dots,1)$).

For $a\preccurlyeq b$, let $d(a,b)$ be the cardinality of $b\setminus a$ and let $\iota^n_{a,b}$ be the face map $\square^{d(a,b)}\hookrightarrow\square^n$ satisfying $\iota^n_{a,b}(\alpha)=a$ and $\iota^n_{a,b}(\omega)=b$ (see Lemma 1.2.4).

Example 1.2.3 If $n=5$, $a=(1,0,0,0,0)$ and $b=(1,0,1,0,1)$, then $b\setminus a=\{3,5\}$, and $\iota^5_{a,b}\colon\square^2\to\square^5$ is given by $\iota^5_{a,b}(x,y)=(1,0,x,0,y)$.

Lemma 1.2.4 *A map $\varphi\colon\square^n\to\square^m$ satisfies $d(\varphi(\alpha),\varphi(\omega))\le n$. The map φ is a monomorphism if and only if $d(\varphi(\alpha),\varphi(\omega))=n$, and in this case φ is determined by $\varphi(\alpha)$ and $\varphi(\omega)$. In particular, if $n=1$, then $\varphi(\alpha)=\varphi(\omega)$ or $\varphi(\omega)\setminus\varphi(\alpha)=\{i\}$ for some i.*

Proof We decompose $\varphi=u\circ v$ with $v\colon\square^n\to\square^p$ a composition of degeneracies and connections and $u\colon\square^p\to\square^m$ a composition of faces. We have $p\le n$, and $v(\alpha)=\alpha$ and $v(\omega)=\omega$. A composition of faces inserts some 0 and 1 at some places and thus leaves the distance between two vertices invariant. In particular $d(u(\alpha),u(\omega))=p\le n$. In addition, since degeneracies and connections always decrease $d(\alpha,\omega)$ strictly, we get that φ is a monomorphism if and only if φ is a composition of faces, if and only if $d(\varphi(\alpha),\varphi(\omega))=n$. A face is uniquely determined by its value on α and ω, so is a composition of faces. The second part of the statement is immediate. $\square$

We next recall two model category structures on cubical sets. The first one, the Grothendieck model structure, models homotopy types and is described by Cisinski [2014], and the second one models $(\infty,1)$–categories and is described in [Doherty et al. 2024].

Definition 1.2.5 We recall here some useful definitions of [Doherty et al. 2024, Section 4].

- The boundary of $\square^n$, that is, the union of all the faces of $\square^n$, is denoted by $\partial\square^n$ and the canonical inclusion by $\partial^n\colon\partial\square^n\to\square^n$.
- The union of all the faces except $\partial_{i,\epsilon}$ is denoted by $\sqcap^n_{i,\epsilon}$, and the inclusion $\sqcap^n_{i,\epsilon}\to\square^n$ is called an *open box inclusion*.
- Given a face $\partial_{i,\epsilon}$ of $\square^n$, its *critical edge* $e_{i,\epsilon}$ is the unique edge of $\square^n$ that is adjacent to $\partial_{i,\epsilon}$ and contains the vertex α or ω which is not in $\partial_{i,\epsilon}$. Namely, this is the edge between the vertices $(1-\epsilon,\dots,1-\epsilon)$ and $(1-\epsilon,\dots,1-\epsilon,\epsilon,1-\epsilon,\dots,1-\epsilon)$, where ϵ is placed at the i–coordinate. Equivalently, for $\epsilon=1$, this is the edge from α to $\{i\}$ and if $\epsilon=0$ this is the edge from $\{1,\dots,n\}\setminus\{i\}$ to ω.
- For $n\ge 2$, quotienting by the critical edge results in the (i,ϵ)–*inner cube* $\widehat{\square}^n_{i,\epsilon}$, the (i,ϵ)–*inner open box* $\widehat{\sqcap}^n_{i,\epsilon}$, and the (i,ϵ)–*inner open box inclusion* $h^n_{i,\epsilon}\colon\widehat{\sqcap}^n_{i,\epsilon}\to\widehat{\square}^n_{i,\epsilon}$.

- A (cubical) *Kan fibration* is a map having the right lifting property with respect to all open box inclusions.
- A (cubical) *inner fibration* is a map having the right lifting property with respect to all inner open box inclusions.
- A *cubical quasicategory* is a cubical set X such that $X \to *$ is an inner fibration.

Theorem 1.2.6 (Cisinski [Doherty et al. 2024, Theorem 1.34]) *The category $c\mathcal{S}et$ carries a cofibrantly generated model structure, referred to as the Grothendieck model structure, in which*

- *cofibrations are the monomorphisms, and*
- *fibrations are Kan fibrations.*

We next sum up [Doherty et al. 2024, Theorems 4.2 and 4.16] for the Joyal model structure.

Theorem 1.2.7 *The category $c\mathcal{S}et$ carries a cofibrantly generated model structure, referred to as the Joyal model structure, in which*

- *cofibrations are the monomorphisms, and*
- *fibrant objects are cubical quasicategories.*

Moreover, fibrations between fibrant objects are inner fibrations having the right lifting property with respect to the two endpoint inclusions $j_0 \colon \{0\} \to K$ and $j_1 \colon \{1\} \to K$, where K is the cubical set

$$\begin{array}{ccccc} 1 & \longrightarrow & 0 & = & 0 \\ \| & & \downarrow & & \| \\ 1 & = & 1 & \longrightarrow & 0 \end{array}$$

We next recall the notion of equivalence and of special open box; see [Doherty et al. 2024, Section 4].

Definition 1.2.8 Let X be a cubical set.

- An edge $f \colon \square^1 \to X$ is an *equivalence* if it factors through the inclusion of the middle edge $\square^1 \to K$.
- For $n \geq 2$, $1 \leq i \leq n$ and $\epsilon \in \{0, 1\}$, a *special open box* in X is a map $\sqcap^n_{i,\epsilon} \to X$ which sends the critical edge $e_{i,\epsilon}$ to an equivalence.

Intuitively, in reference to the above drawing of K, the definition of equivalence says that f has a left and a right inverse (the images of the nondegenerate horizontal edges), witnessed as such by the images of the two 2–cubes.

Finally, we collect results of [Doherty et al. 2024, Sections 5 and 6] on the comparison between cubical and simplicial models.

Definition 1.2.9 We can construct a monoidal product $\otimes\colon c\mathcal{S}et \times c\mathcal{S}et \to c\mathcal{S}et$ by taking the left Kan extension of the monoidal product on $\square$ given by $[1]^n \times [1]^m \mapsto [1]^{n+m}$ postcomposed with the Yoneda morphism. In particular $\square^n \otimes \square^m \cong \square^{n+m}$.

Note that this monoidal product is not symmetric.

In [Doherty et al. 2024] the authors provide four different but analogous functors $s\mathcal{S}et \to c\mathcal{S}et$, each of them labelled by a face of the 2–cube. We choose the one labelled by the face $\partial_{2,1}$ (corresponding to $Q_{R,1}$ in [Doherty et al. 2024]) and denote it by Q throughout the paper. Note that one of these constructions appeared first in [Kapulkin et al. 2019]. It is obtained as a left Kan extension of a functor $\Delta \to c\mathcal{S}et$, which we describe in the following definition.

Definition 1.2.10 Let $n \geq 0$. The cubical set Q^n is the quotient of the n–cube $\square^n$ obtained as the pushout

$$\begin{array}{ccc}
(\square^0 \otimes \square^{n-1}) \sqcup (\square^1 \otimes \square^{n-2}) \sqcup \cdots \sqcup (\square^{n-1} \otimes \square^0) & \xrightarrow{(\partial_{1,1},\partial_{2,1},\ldots,\partial_{n,1})} & \square^n \\
\downarrow & & \downarrow \pi_n \\
\square^{n-1} \sqcup \square^{n-2} \sqcup \cdots \sqcup \square^0 & \xrightarrow{\qquad\qquad} & Q^n
\end{array}$$

The family $(Q^n)_{n\geq 0}$ assembles to a functor $Q\colon \Delta \to c\mathcal{S}et$, where faces and degeneracies are induced by the generating maps of $\square$ as follows:

- the i^{th} face $Q(d_i)\colon Q^{n-1} \to Q^n$ is the map induced by $\square^{n-1} \xrightarrow{\partial_{i,0}} \square^n$ if $i > 0$ and by $\partial_{1,1}$ if $i = 0$, and
- the i^{th} degeneracy $Q(s_i)\colon Q^{n+1} \to Q^n$ is the map induced by $\square^{n+1} \xrightarrow{\gamma_{i,0}} \square^n$ if $i > 0$ and by σ_1 if $i = 0$.

Lemma 1.2.11 *The set of vertices of Q^n is in bijection with the set $\{0,\ldots,n\}$ and the map π_n sends $a \subseteq \{1,\ldots,n\}$ to $\sup a$, setting $\sup\varnothing = 0$. Furthermore, the action of the faces and degeneracies on the vertices of Q^n coincides with the action on the vertices of the simplicial set Δ^n.*

Proof Since colimits in $c\mathcal{S}et$ are computed dimensionwise, the set $(Q^n)_0$ of vertices is obtained as the pushout of the diagram above evaluated at $[1]^0$. We claim that the set $\{0,\ldots,n\}$, together with the map

$$\pi_n\colon \mathcal{P}(\{1,\ldots,n\}) \to \{0,\ldots,n\}, \quad a \mapsto \sup a,$$

satisfies the universal property of the pushout. Consider $I_1 \subseteq \{1,\ldots,i-1\}$ and $I_2 \subseteq \{i+1,\ldots,n\}$. Then (I_1, I_2) is mapped horizontally to $I_1 \cup \{i\} \cup I_2$ and vertically to I_2; hence $I_1 \cup \{i\} \cup I_2$ is identified to $I_1' \cup \{i\} \cup I_2$ for any other $I_1' \subseteq \{1,\ldots,i-1\}$. The claim follows easily from this observation. The rest of the statement is also checked easily. ◻

The left Kan extension of $Q\colon \Delta \to c\mathcal{S}et$ along the Yoneda morphism is also denoted by $Q\colon s\mathcal{S}et \to c\mathcal{S}et$ and admits a right adjoint $\int$ defined as $(\int S)_n := \mathrm{Hom}_{c\mathcal{S}et}(Q^n, S)$. We have the following Quillen equivalences [Doherty et al. 2024, Corollary 6.24 and Proposition 6.25].

Theorem 1.2.12 *The adjunction $Q : s\mathcal{S}et \rightleftarrows c\mathcal{S}et : \int$ is both a Quillen equivalence*

- *between the Joyal model structure on $s\mathcal{S}et$ and the Joyal model structure on $c\mathcal{S}et$, and*
- *between the Kan–Quillen model structure on $s\mathcal{S}et$ and the Grothendieck model structure on $c\mathcal{S}et$.*

2 Necklaces and paths

In this section and the following one, we follow closely the steps taken by Dugger and Spivak [2011a] in order to understand more concretely the simplicial rigidification functor. We adapt their approach to define the simplicial rigidification of cubical sets.

2.1 Necklaces

Let $c\mathcal{S}et_{*,*} = \partial\square^1 \downarrow c\mathcal{S}et$ be the category of double pointed cubical sets. Given a cubical set S and two vertices $a, b \in S_0$, the notation $S_{a,b}$ stands for the double pointed cubical set corresponding to the morphism $(\partial\square^1 \to S) \in c\mathcal{S}et_{*,*}$ mapping 0 to a and 1 to b. We refer to Section A.1 for general constructions. When there is no ambiguity on the double pointing, we omit the indices and write $S \in c\mathcal{S}et_{*,*}$. For example, the cube $\square^n$ is naturally double pointed by α and ω (see Notation 1.2.2), and if not specified otherwise we will consider this double pointing.

Definition 2.1.1

- A (*cubical*) *necklace* is an object T of $c\mathcal{S}et_{*,*}$ of the form $\square^{n_1} \vee \cdots \vee \square^{n_k}$, for some sequence $(n_1, \dots, n_k)$ of positive integers. The double pointing is induced by $\alpha \in \square^{n_1}$ and $\omega \in \square^{n_k}$. The empty sequence corresponds to the necklace $T = \square^0$ and it is the unique one satisfying $\alpha = \omega$.
- For $k \geq 1$, the canonical morphism $B_i : \square^{n_i} \to T$ in $c\mathcal{S}et$ is called the i^{th} *bead* of T, so that $\mathrm{id}_T = B_1 * \cdots * B_k$ (see Definition A.1.2 for the notation).
- We denote by $\mathcal{N}ec$ the full subcategory of $c\mathcal{S}et_{*,*}$, whose objects are cubical necklaces. Objects will be identified with sequences $(n_1, \dots, n_k)$ of positive integers. Note that if S is a necklace and $(a, b) \neq (\alpha, \omega)$, then an object of the slice category $\mathcal{N}ec \downarrow S_{a,b}$, ie a morphism $T \to S_{a,b}$ with T a necklace, is not a morphism in $\mathcal{N}ec$ since the double pointing in $\mathcal{N}ec$ is given by (α, ω).
- Given two sequences $(n_1, \dots, n_k)$ and $(m_1, \dots, m_l)$, their *concatenation* is the sequence
$$(n_1, \dots, n_k, m_1, \dots, m_l).$$
- A *decomposition* of a nonempty sequence $(n_1, \dots, n_k)$ in l blocks is a collection $(A_1, \dots, A_l)$ of nonempty sequences such that their concatenation is $(n_1, \dots, n_k)$.

The following proposition describes the morphisms in $\mathcal{N}ec$.

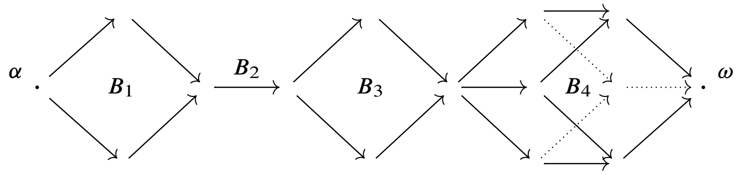


Figure 1: The necklace associated to the sequence (2, 1, 2, 3).

Proposition 2.1.2 (1) *In the category $\mathcal{N}ec$, a morphism φ from $(n_1, \dots, n_k)$ to (m) decomposes uniquely as $\varphi = \varphi_1 * \cdots * \varphi_k$, where all $\varphi_i : \square^{n_i} \to \square^m$ are morphisms in $\square$, and satisfy*

$$\varphi_1(\alpha) = \alpha,$$
$$\varphi_i(\omega) = \varphi_{i+1}(\alpha) \quad \text{for all } 1 \le i \le k-1,$$
$$\varphi_k(\omega) = \omega.$$

In particular, $m \le n_1 + \cdots + n_k$.

(2) *Given a morphism $f : (n_1, \dots, n_k) \to (m_1, \dots, m_l)$ in $\mathcal{N}ec$, there is a decomposition $(A_1, \dots, A_l)$ of the sequence $(n_1, \dots, n_k)$ into l parts and morphisms $f_j : A_j \to (m_j)$ in $\mathcal{N}ec$ such that*

$$f = f_1 \vee \cdots \vee f_l.$$

This decomposition is unique if, for any $1 \le i \le k$, the restriction of f to the bead (n_i) is not constant.

Proof The first part of the proposition is a direct consequence of the definition of the concatenation. Note that a map from $(n_1, \dots, n_k)$ to (m) yields a chain $\alpha = a_0 \preccurlyeq a_1 \preccurlyeq \cdots \preccurlyeq a_{k-1} \preccurlyeq a_k = \omega$ in $(\square^m)_0$ with $d(a_{i-1}, a_i) \le n_i$, by Lemma 1.2.4. Hence $m = d(\alpha, \omega) \le n_1 + \cdots + n_k$.

Let us prove the second part. For the sake of clarity, we denote by (α_i, ω_i) the initial and terminal vertices of $\square^{m_i}$. For any cubical set S and any $n \ge 0$, we denote by $\sigma : S_0 \to S_n$ the map induced by the unique map $\square^n \to \square^0$ in $\square$. Let $T = \square^{m_1} \vee \cdots \vee \square^{m_l}$. By definition of concatenation, the set of n–cubes in the cubical set T is the quotient of the disjoint union of the n–cubes of $\square^{m_i}$ by the relation $\sigma(\omega_i) = \sigma(\alpha_{i+1})$ for $1 \le i \le l-1$. Let $\varphi : (n_1, \dots, n_k) \to T$ be a morphism in $\mathcal{N}ec$ and let $\varphi_i : (n_i) \to (m_1, \dots, m_l)$ be its components, that is, φ_i is an n_i–cube of T, and $\varphi = \varphi_1 * \cdots * \varphi_k$. Since $\varphi_1(\alpha) = \alpha_1$, necessarily φ_1 is an n_1–cube of $\square^{m_1}$. Since $\varphi_1(\omega) = \varphi_2(\alpha)$, there are two possibilities: either $\varphi_1(\omega) \ne \omega_1$ and then φ_2 is an n_2–cube of $\square^{m_1}$, or $\varphi_1(\omega) = \omega_1$ and φ_2 is an n_2–cube of $\square^{m_2}$. Inductively, we get a decomposition $(A_1, \dots, A_l)$, where A_i is a sequence of consecutive n_j such that φ_j is an n_j–cube of $\square^{m_i}$. The decomposition is not unique in general. For example, if above we had $\varphi_2 = \sigma(\omega_1)$, then we could have chosen to keep n_2 in A_1. But under the assumption that each φ_i is not constant, we do have uniqueness, since we can identify unambiguously in which component of T it lies. □

Example 2.1.3 (i) There is no morphism from (2, 1, 3) to (3, 2, 1): there is a unique decomposition of (2, 1, 3) into three parts and, by Lemma 1.2.4, there is no morphism in $\mathcal{N}ec$ from $\square^2$ to $\square^3$.

(ii) A morphism $f: (2,1,3) \to (2,1)$ in $\mathcal{N}ec$ is either of the form $g \vee h$ with $g: (2,1) \to (2)$ and $h: (3) \to (1)$ in $\mathcal{N}ec$ (first type), or is $\mathrm{id}_{(2)} \vee (\mathrm{id}_{(1)} * c_\omega)$, where $c_\omega: (3) \to (1)$ is the constant map with value ω. Indeed, there are two decompositions of $(2,1,3)$ into two blocks. The decomposition $((2,1),(3))$ gives the first decomposition. The decomposition $((2),(1,3))$ gives $f = k \vee (l_1 * l_2)$, where $k: (2) \to (2) \in \mathcal{N}ec$ is necessarily the identity and $l_1 * l_2: (1,3) \to (1)$. The only maps $l_1: (1) \to (1)$ such that $l_1(\alpha) = \alpha$ are c_α and the identity. If $l_1 = c_\alpha$, then $(k * c_\omega) \vee l_2: (2,1) \vee (3) \to (2) \vee (1)$ is also a decomposition of f of the first type. If $l_1 = \mathrm{id}_{(1)}$, then $l_2 = c_\omega$.

Notation 2.1.4 If $T = (n_1, \dots, n_k)$ is a cubical necklace, then $T_0 = (\square^{n_1})_0 \vee \cdots \vee (\square^{n_k})_0$ is a bounded poset (see Definition A.2.1). We will also denote the order in T_0 by $\preccurlyeq$. Moreover, setting $n = n_1 + \cdots + n_k$, any monomorphism in $\mathcal{N}ec$ from T to $\square^n$ is a morphism of posets on vertices, which justifies the notation $\preccurlyeq$.

The following lemma is an easy consequence of Lemma 1.2.4.

Lemma 2.1.5 *Let $T = (n_1, \dots, n_k)$ be a cubical necklace.*

- *Any monomorphism $\varphi: T \hookrightarrow \square^n$ in $\mathcal{N}ec$ is uniquely determined by a sequence $a_0 \prec a_1 \prec \cdots \prec a_k$ of vertices in $\square^n$ satisfying $a_0 = \alpha, a_k = \omega$ and $d(a_{i-1}, a_i) = n_i$. The sequence*
$$\varnothing \prec \{1, \dots, n_1\} \prec \cdots \prec \{1, \dots, n_1 + \cdots + n_{k-1}\} \prec \{1, \dots, n\}$$
corresponds to an embedding $T \hookrightarrow \square^n$ that we will call the ***standard embedding****.*
- *If $n_1 = \cdots = n_k = 1$, then any morphism $\varphi: T \to \square^n$ in $\mathcal{N}ec$ is uniquely determined by a sequence $a_0 \preccurlyeq a_1 \preccurlyeq \cdots \preccurlyeq a_k$ of vertices in $\square^n$ satisfying $a_0 = \alpha$, $a_k = \omega$ and $d(a_{i-1}, a_i) \le 1$.*
- *If $n_1 = \cdots = n_k = 1$, then any monomorphism $\varphi: T \hookrightarrow \square^n$ in $\mathcal{N}ec$ is uniquely determined by a sequence $a_0 \prec a_1 \prec \cdots \prec a_k$ of vertices in $\square^n$ satisfying $a_0 = \alpha$, $a_k = \omega$ and $d(a_{i-1}, a_i) = 1$.*

Definition 2.1.6 Let $T = (n_1, \dots, n_k)$ be a necklace and $a \preccurlyeq b$ be vertices of T. We define the cubical set $T_{[a,b]}$ and the morphism $\iota^T_{a,b}: T_{[a,b]} \hookrightarrow T$ in $c\mathcal{S}et$ as follows:

- If $a, b \in \square^{n_i}$ then $T_{[a,b]} := \square^{d(a,b)}$ and $\iota^T_{a,b} := B_i \circ \iota^{n_i}_{a,b}$ (cf Notation 1.2.2).
- If $a \in \square^{n_i}$ and $b \in \square^{n_j}$ with $i < j$, then $T_{[a,b]} := \square^{d(a,\omega)} \vee \square^{n_{i+1}} \vee \cdots \vee \square^{n_{j-1}} \vee \square^{d(\alpha,b)}$ and $\iota^T_{a,b} := (B_i \circ \iota^{n_i}_{a,\omega}) * B_{i+1} * \cdots * B_{j-1} * (B_j \circ \iota^{n_j}_{\alpha,b})$.

Hence $T_{[a,b]}$ is a necklace and $\iota^T_{a,b}: T_{[a,b]} \hookrightarrow T$ is a monomorphism in $c\mathcal{S}et$. We call $T_{[a,b]}$ the *subnecklace* of T between a and b.

Remark 2.1.7 This is well defined as $\square^0$ is the unit of the monoidal product $\vee$, so the construction of $T_{[a,b]}$ does not depend on the bead chosen for containing a or b. Recall that $T_{a,b}$ denotes a double pointed version of T while $T_{[a,b]}$ is a necklace whose underlying cubical set is different from T. The next proposition makes the link between these two cubical sets.

Proposition 2.1.8 *Let T be a necklace and $a \preccurlyeq b \in T_0$. The object $\iota^T_{a,b}\colon T_{[a,b]} \hookrightarrow T$ is terminal in $\mathcal{N}ec \downarrow T_{a,b}$.*

Proof Let $f\colon X_{\alpha,\omega} \to T_{a,b}$ be a map in $c\mathcal{S}et_{*,*}$ with X a necklace. Proceeding like in the proof of Proposition 2.1.2, we get that f factors uniquely through $T_{[a,b]}$ as $f = \iota^T_{a,b} \circ \hat{f}$ with $\hat{f}\colon X_{\alpha,\omega} \to T_{[a,b]}$ a morphism in $\mathcal{N}ec$. $\square$

The next lemma states properties that will be needed in the proof of Proposition 3.3.2.

Lemma 2.1.9 *Let S be a cubical subset of $\square^n$. Let $a \preccurlyeq b$ be two vertices of S.*

1. *An m–cube $x\colon \square^m \to S$ is nondegenerate if and only if x is a monomorphism.*
2. *A map $f\colon T \to S$ with T a necklace is a monomorphism if and only if it is a monomorphism on every bead.*
3. *Any object $f\colon T \to S_{a,b}$ of $\mathcal{N}ec \downarrow S_{a,b}$ factors uniquely as $f = \iota(f)\pi(f)$, where $\pi(f)\colon T \to T^f$ is an epimorphism in $\mathcal{N}ec$ and $\iota(f)\colon T^f \to S_{a,b}$ is a monomorphism in $c\mathcal{S}et$.*

Proof We make use of Proposition 1.2.1.

(1) If x is nondegenerate in S, then x is nondegenerate in $\square^n$, hence a monomorphism, and so is x. Conversely, if x is a monomorphism, we can factor uniquely $x = ip$ with p an epimorphism and i a nondegenerate map. In particular p is a monomorphism, thus an isomorphism of cubes, that is, it is the identity.

(2) Let $B_i\colon \square^{n_i} \to T = (n_1, \dots, n_k)$ be the inclusion of the i^{th} bead of T. If f is a monomorphism, then fB_i is. For the converse, we use Proposition 1.2.1(3). Let $x\colon \square^m \to T$ be a nondegenerate m–cube of T. There exists a bead B_i such that x factors through it; thus $f(x) = (fB_i)(x)$ is nondegenerate. Let x and y be two nondegenerate cubes in T such that $f(x) = f(y)$. Assume x factors through B_i, and y through B_j, with $i < j$. Then $fB_i(\alpha_i) \prec fB_i(\omega_i) \preccurlyeq fB_j(\alpha_j)$. The inequalities hold because S is a cubical subset of $\square^n$ and the left one is strict because fB_i is a monomorphism. Hence $f(x) = f(y)$ is not possible. Therefore, x and y factor through the same bead, and hence $x = y$ by our assumption.

(3) For every bead B_i of T, fB_i factors uniquely as $\iota_i(f) \circ \pi_i(f)$ with $\pi_i(f)$ an epimorphism and $\iota_i(f)$ a monomorphism (by (1)). Setting $\pi(f) = \pi_1(f) \vee \dots \vee \pi_k(f)$ and $\iota(f) = \iota_1(f) * \dots * \iota_k(f)$, we get the desired factorisation (by (2)). It is unique since f writes uniquely as $f_1 * \dots * f_k$ and each f_i factors uniquely. $\square$

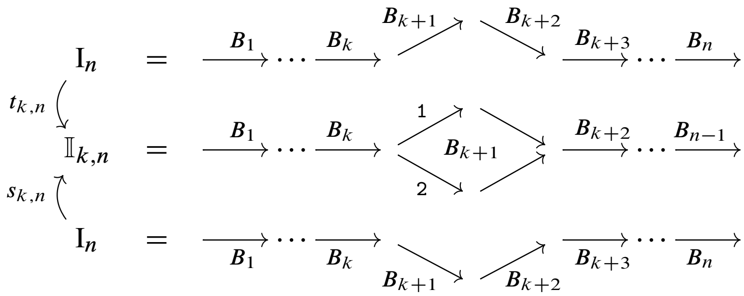
2.2 The path category of a cubical set

In this section we associate to a cubical set S a category enriched in prosets (ie preordered sets) $\mathfrak{C}^{\square}_{\text{path}}(S)$ in Proset-$\mathcal{C}at$. The idea is that $\mathfrak{C}^{\square}_{\text{path}}(S)(a,b)$ has for objects concatenations of nondegenerate 1–cubes joining a to b and that the preorder is induced by the 2–cubes of S.

Notation 2.2.1 For $n \geq 0$, let I_n be the necklace $(\square^1)^{\vee n}$. For $n \geq 2$ and $0 \leq k \leq n-2$, let $\mathbb{I}_{k,n}$ be the concatenation $\mathrm{I}_k \vee \square^2 \vee \mathrm{I}_{n-2-k}$. The source and target maps $s_{k,n}, t_{k,n} \colon \mathrm{I}_n \to \mathbb{I}_{k,n}$ are the morphisms in $\mathcal{N}ec$ defined by

$$s_{k,n} = \mathrm{id}^{\vee k} \vee (\partial_{1,0} * \partial_{2,1}) \vee \mathrm{id}^{\vee n-2-k} \quad \text{and} \quad t_{k,n} = \mathrm{id}^{\vee k} \vee (\partial_{2,0} * \partial_{1,1}) \vee \mathrm{id}^{\vee n-2-k}$$

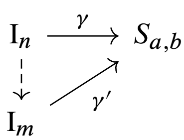
as presented in the following diagram:



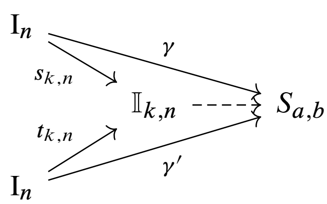
Definition 2.2.2 Let $S_{a,b} \in c\mathcal{S}et_{*,*}$. The set $\mathfrak{C}^{\square}_{\mathrm{path}}(S)(a,b)$ of *paths joining a to b* is defined as

$$\mathfrak{C}^{\square}_{\mathrm{path}}(S)(a,b) = \bigcup_n \mathrm{Hom}(\mathrm{I}_n, S_{a,b})/\sim,$$

where $\sim$ is the equivalence relation generated by $\gamma \sim \gamma'$ if there exists a factorisation in $c\mathcal{S}et_{*,*}$



For $\gamma \colon \mathrm{I}_n \to S_{a,b}$ and $\gamma' \colon \mathrm{I}_n \to S_{a,b}$, we write $[\gamma] \rightsquigarrow [\gamma']$ if there exists $0 \leq k \leq n-2$ and a factorisation in $c\mathcal{S}et_{*,*}$



We then define the preorder structure $\mathfrak{C}^{\square}_{\mathrm{path}}(S)(a,b)$ as the reflexive transitive closure of $\rightsquigarrow$, which we also denote by $\rightsquigarrow$.

The next proposition lifts the definition at the level of categories enriched in preordered sets, that is, Proset-categories.

Proposition 2.2.3 *Any cubical set S gives rise to a Proset-category $\mathfrak{C}^{\square}_{\mathrm{path}}(S)$ whose objects are the vertices of S and whose homsets are given by Definition 2.2.2, with composition given by concatenation of paths. In addition, the assignment $S \mapsto \mathfrak{C}^{\square}_{\mathrm{path}}(S)$ upgrades to a functor $\mathfrak{C}^{\square}_{\mathrm{path}} \colon c\mathcal{S}et \to$ Proset-$\mathcal{C}at$.*

Proof For $\gamma\colon \mathrm{I}_n \to S_{a,b}$ and $\beta\colon \mathrm{I}_m \to S_{b,c}$, the class of the concatenation $\gamma * \beta\colon \mathrm{I}_{n+m} = \mathrm{I}_n \vee \mathrm{I}_m \to S_{a,c}$ does not depend on the choice of the representatives γ and β. This defines a composition

$$\mathfrak{C}^{\square}_{\mathrm{path}}(S)(b,c) \times \mathfrak{C}^{\square}_{\mathrm{path}}(S)(a,b) \to \mathfrak{C}^{\square}_{\mathrm{path}}(S)(a,c)$$

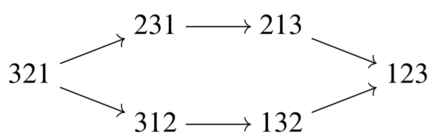
by $([\beta],[\gamma]) \mapsto [\gamma * \beta]$. Similarly if $[\gamma] \rightsquigarrow [\gamma']$ and $[\beta] \rightsquigarrow [\beta']$ then $[\gamma * \beta] \rightsquigarrow [\gamma' * \beta']$, and everything is functorial in S. □

2.3 The path category of a necklace

Let A be a totally ordered set with k elements. An element in the set of bijections Σ_A of A is represented by a sequence $(a_1,\dots,a_k)$ such that $\{a_1,\dots,a_k\} = A$. We consider the (reverse right) weak Bruhat order on Σ_A, that is, the order generated by

$$(a_1,\dots,a_i,a_{i+1},\dots,a_k) \rightsquigarrow_B (a_1,\dots,a_{i+1},a_i,\dots,a_k) \quad \text{if } 1 \le i < k \text{ and } a_i > a_{i+1}.$$

For example, for $A = \{1,2,3\}$ the Hasse diagram of $\rightsquigarrow_B$ is given by



Note that, given two disjoint subsets A and B of $\{1,\dots,n\}$, the concatenation $*\colon \Sigma_A \times \Sigma_B \to \Sigma_{A \sqcup B}$ of sequences is a map of posets (for the order $\rightsquigarrow_B$), where the total orders on A, B and $A \sqcup B$ are induced by that of $\{1,\dots,n\}$. We refer to the book by Björner [1984] for more on orders on Coxeter groups.

Lemma 2.3.1 *For all n, a and b, each element in $\mathfrak{C}^{\square}_{\mathrm{path}}(\square^n)(a,b)$ has a unique representative γ, which is a monomorphism, corresponding to a sequence $a_0 = a \prec a_1 \prec \cdots \prec a_l = b$, with $d(a_i,a_{i+1}) = 1$. The same holds replacing $\square^n$ with any cubical subset S of $\square^n$.*

Proof Let $[\gamma']$ be an element of $\mathfrak{C}^{\square}_{\mathrm{path}}(\square^n)(a,b)$. By Lemma 2.1.5, γ' corresponds to some sequence $s' = (a'_0 = a \preccurlyeq x_1 \preccurlyeq \cdots \preccurlyeq a'_{k'} = b)$ such that $d(a'_i,a'_{i+1}) \le 1$. We claim that the desired γ is the monomorphism corresponding to the sequence $\kappa(s') = (a_0 = a \prec a_1 \prec \cdots \prec a_l = b)$ obtained by eliminating the repetitions in the sequence s'. This is a consequence of the following two easy facts:

(i) for γ as just defined, $[\gamma] \sim [\gamma']$, and

(ii) if $[\gamma'_1] \sim [\gamma'_2]$, with corresponding sequences s'_1 and s'_2, then $\kappa(s'_1) = \kappa(s'_2)$.

The last part of the statement follows from the observation that if s' above lies in S, then so does $\kappa(s')$. □

Proposition 2.3.2 *For every pair of vertices $a \preccurlyeq b$ in $\square^n$, there is an isomorphism of preordered sets*

$$\mathfrak{C}^{\square}_{\mathrm{path}}(\square^n)(a,b) \to \Sigma_{b \setminus a},$$

compatible with concatenation. As a consequence, the preorder $\rightsquigarrow$ on paths of a cube is a partial order, isomorphic to the weak Bruhat order $\rightsquigarrow_B$ on the symmetric group.

Proof Assume $a \preccurlyeq b$. With the notation of Lemma 2.3.1, we can associate with each $[\gamma] \in \mathfrak{C}^{\square}_{\text{path}}(\square^n)(a,b)$ a sequence $a_0 = a \prec a_1 \prec \cdots \prec a_l = b$, with $d(a_i, a_{i+1}) = 1$. We denote by x_i the unique element in $a_{i+1} \setminus a_i = \{x_i\}$, so that $\{x_1, \ldots, x_l\} = b \setminus a$. Then the map $\Psi \colon \mathfrak{C}^{\square}_{\text{path}}(\square^n)(a,b) \to \Sigma_{b\setminus a}$ sending $[\gamma]$ to the sequence $(x_1, \ldots, x_l)$ in $\Sigma_{b\setminus a}$ is well defined and bijective.

Let $f \colon \mathbb{I}_{k,m} \to \square^n_{a,b}$ in $c\mathcal{S}et_{*,*}$, witnessing $[f \circ s_{k,m}] \rightsquigarrow [f \circ t_{k,m}]$. Let

$$a_0 \preccurlyeq \cdots \preccurlyeq a_k \preccurlyeq a_{k+1} \preccurlyeq a_{k+2} \preccurlyeq \cdots \preccurlyeq a_m,$$
$$a_0 \preccurlyeq \cdots \preccurlyeq a_k \preccurlyeq a'_{k+1} \preccurlyeq a_{k+2} \preccurlyeq \cdots \preccurlyeq a_m$$

be the sequences corresponding to $f \circ s_{k,m}$ and $f \circ t_{k,m}$, respectively. If $d(a_k, a_{k+2}) = 2$, then there exists $u < v$ such that $a_{k+2} \setminus a_k = \{u, v\}$, $a_{k+1} \setminus a_k = \{v\}$ and $a'_{k+1} \setminus a_k = \{u\}$, hence $\Psi(f \circ s_{k,m}) \rightsquigarrow_B \Psi(f \circ t_{k,m})$. If $d(a_k, a_{k+2}) < 2$, then

$$[f \circ s_{k,m}] = [\gamma] = [f \circ t_{k,m}],$$

where γ corresponds to $a_0 \preccurlyeq \cdots \preccurlyeq a_k \prec a_{k+2} \preccurlyeq \cdots \preccurlyeq a_m$. Hence Ψ is a morphism of preordered sets. Similarly, and even more straightforwardly, we see that Ψ^{-1} is also a morphism of prosets, which in particular implies that $\mathfrak{C}^{\square}_{\text{path}}(\square^n)(a,b)$ is a poset. $\square$

We also observe that if $a \not\preccurlyeq b$, there is no morphism $\mathrm{I}_m \to (\square^n)_{a,b}$, and that the constant path is the unique path in $\mathfrak{C}^{\square}_{\text{path}}(\square^n)(a,a)$. Hence $\mathfrak{C}^{\square}_{\text{path}}(\square^n)$ is a P–shaped poset-category, with P the subset lattice of $\{1, \ldots, n\}$. We refer to Section A.2 for this notion, and for the description of the concatenation product $\vee$ on such categories used in the following proposition.

Corollary 2.3.3 *Let $T = \square^{n_1} \vee \cdots \vee \square^{n_k}$ be a necklace. Then*

(1) *for $a \preccurlyeq b \in T_0$, the inclusion $T_{[a,b]} \subseteq T_{a,b}$ induces an isomorphism of posets*

$$\mathfrak{C}^{\square}_{\text{path}}(T)(a,b) \cong \mathfrak{C}^{\square}_{\text{path}}(T_{[a,b]})(\alpha, \omega),$$

(2) *if $T = U \vee V$, the composition in the poset category $\mathfrak{C}^{\square}_{\text{path}}(T)$ provides a morphism*

$$\mathfrak{C}^{\square}_{\text{path}}(V)(\alpha_V, \omega_V) \times \mathfrak{C}^{\square}_{\text{path}}(U)(\alpha_U, \omega_U) \to \mathfrak{C}^{\square}_{\text{path}}(T)(\alpha_T, \omega_T),$$

which is an isomorphism of posets, and

(3) *$\mathfrak{C}^{\square}_{\text{path}}(T)$ is a poset-category and we have an isomorphism of poset-categories*

$$\mathfrak{C}^{\square}_{\text{path}}(T) \cong \mathfrak{C}^{\square}_{\text{path}}(\square^{n_1}) \vee \cdots \vee \mathfrak{C}^{\square}_{\text{path}}(\square^{n_k}).$$

Proof (1) By Definition 2.1.6 and Proposition 2.1.8, a path γ joining a to b in T is equivalent to a morphism $\mathrm{I}_n \to T_{[a,b]}$, where $T_{[a,b]}$ is a necklace. By Proposition 2.1.8, any map $\mathbb{I}_{k,n} \to T_{a,b}$ factorises through $T_{[a,b]}$, hence the result.

(2) Let us prove that the morphism of prosets induced by composition/concatenation

$$\mathfrak{C}^{\square}_{\text{path}}(\square^{n_2})(\alpha_2, \omega_2) \times \mathfrak{C}^{\square}_{\text{path}}(\square^{n_1})(\alpha_1, \omega_1) \to \mathfrak{C}^{\square}_{\text{path}}(\square^{n_1} \vee \square^{n_2})(\alpha_1, \omega_2)$$

is an isomorphism of prosets. By Lemma 2.3.1, and viewing $\square^{n_1} \vee \square^{n_2}$ as a cubical subset of $\square^{n_1+n_2}$ via the standard embedding, any element in the right-hand side admits a unique representative

$$\gamma\colon \mathrm{I}_l \to \square^{n_1} \vee \square^{n_2},$$

which is a monomorphism.

Since γ preserves α and ω, we have $l = n_1 + n_2$ and thus $\gamma = \gamma_1 \vee \gamma_2$ is the unique decomposition provided by Proposition 2.1.2. Hence the morphism is a bijection. We have to prove that $[\gamma] \rightsquigarrow [\gamma']$ implies $[\gamma_1] \rightsquigarrow [\gamma_1']$ and $[\gamma_2] \rightsquigarrow [\gamma_2']$. It is enough to prove it for the "elementary moves" that generate the relation $\rightsquigarrow$ by reflexive and transitive closure. Any $f\colon \mathbb{I}_{k,m} \to \square^{n_1} \vee \square^{m_1}$ factors as $f = f_1 \vee f_2$, where either f_1 or f_2 is a path. It implies that if $[\gamma] \rightsquigarrow [\gamma']$ then either $[\gamma_1] \rightsquigarrow [\gamma_1']$ and $[\gamma_2] = [\gamma_2']$, or $[\gamma_2] \rightsquigarrow [\gamma_2']$ and $[\gamma_1] = [\gamma_1']$. In conclusion, (2) holds, since the left-hand side of the morphism is a poset.

(3) It is clear that this generalises to any finite wedge product of cubes. In particular, if

$$T_{[a,b]} := \square^{d(a,\omega_i)} \vee \square^{n_{i+1}} \vee \cdots \vee \square^{n_{j-1}} \vee \square^{d(\alpha_j,b)},$$

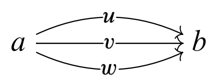
then

$$\begin{aligned} \mathfrak{C}^{\square}_{\mathrm{path}}(T)(a,b) &= \mathfrak{C}^{\square}_{\mathrm{path}}(T_{[a,b]})(\alpha,\omega) && \text{(by (1))}\\ &\cong \mathfrak{C}^{\square}_{\mathrm{path}}(\square^{d(a,\omega_i)})(\alpha_i,\omega_i) \times \cdots \times \mathfrak{C}^{\square}_{\mathrm{path}}(\square^{d(\alpha_j,b)})(\alpha_j,\omega_j) && \text{(by (2))}\\ &= \mathfrak{C}^{\square}_{\mathrm{path}}(\square^{n_i})(a,\omega_i) \times \cdots \times \mathfrak{C}^{\square}_{\mathrm{path}}(\square^{n_j})(\alpha_j,b) && \text{(by (1))}\\ &\cong (\mathfrak{C}^{\square}_{\mathrm{path}}(\square^{n_i}) \vee \cdots \vee \mathfrak{C}^{\square}_{\mathrm{path}}(\square^{n_j}))(a,b) && \text{(by Proposition A.2.5).} \end{aligned}$$

Finally, we note that having established this isomorphism a fortiori implies that $\mathfrak{C}^{\square}_{\mathrm{path}}(T)$ is poset-enriched, since all $\mathfrak{C}^{\square}_{\mathrm{path}}(\square^{n_i})$ are. □

Applying the nerve functor from Proset-categories to simplicial categories, we get the functor $N \circ \mathfrak{C}^{\square}_{\mathrm{path}}\colon c\mathcal{S}et \to s\mathcal{C}at$. Unfortunately it is not cocontinuous, as we show in Example 2.3.4, so that it cannot serve as a left functor in a Quillen equivalence. The next section is devoted to build such a functor and to study its properties.

Example 2.3.4 In this example, all simplicial categories involved have only one (possibly) nontrivial mapping space, and hence reduce to simplicial sets. Consider the quotient $\tilde{\square}^2$ of $\square^2$ obtained by collapsing the edges between $(0,0)$ and $(0,1)$, and between $(1,0)$ and $(1,1)$, and consider the pushout X of the two horizontal inclusions $\partial_{2,0}, \partial_{2,1}\colon \square^1 \to \tilde{\square}^2$. The cubical set X can be represented as



with two nondegenerate 2–cubes inducing $u \rightsquigarrow v \rightsquigarrow w$ in $\mathfrak{C}^{\square}_{\mathrm{path}}(X)$. It follows that $N(\mathfrak{C}^{\square}_{\mathrm{path}}(X))$ is not 1–skeletal. However $N(\mathfrak{C}^{\square}_{\mathrm{path}}(\square^1))$ and $N(\mathfrak{C}^{\square}_{\mathrm{path}}(\tilde{\square}^2))$ are 1–skeletal. But a pushout of 1–skeletal simplicial sets is 1–skeletal, so the pushout of $N(\mathfrak{C}^{\square}_{\mathrm{path}}(\partial_{2,0}))$ and $N(\mathfrak{C}^{\square}_{\mathrm{path}}(\partial_{2,1}))$ cannot be $N(\mathfrak{C}^{\square}_{\mathrm{path}}(X))$. The point of this counterexample is that taking a colimit in $c\mathcal{S}et$ may result in merging of orders: X sees $u \rightsquigarrow v \rightsquigarrow w$, while each copy of $\tilde{\square}^2$ sees only $u \rightsquigarrow v$ or $v \rightsquigarrow w$. By applying the functor $N \circ \mathfrak{C}^{\square}_{\mathrm{path}}$ before the colimit functor, we lose this piece of magic!

3 The rigidification functor $\mathfrak{C}^{\square}$

In this section, we define rigidification as a left Kan extension of the restriction of $N \circ \mathfrak{C}^{\square}_{\text{path}}$ to the cubes $\square^n$, and provide concrete descriptions of its simplicial homsets, making an essential use of necklaces (see Remark 3.2.5).

3.1 Definition of the rigidification

The rigidification functor is defined as the left Kan extension along the Yoneda functor $Y : \square \to c\mathcal{S}et$ of the composition

$$\square \xrightarrow{Y} c\mathcal{S}et \xrightarrow{\mathfrak{C}^{\square}_{\text{path}}} \text{Proset-}\mathcal{C}at \xrightarrow{N} s\mathcal{C}at.$$

By usual means, we obtain an adjunction $\mathfrak{C}^{\square} : c\mathcal{S}et \rightleftarrows s\mathcal{C}at : N^{\square}$. The simplicial category $\mathfrak{C}^{\square}(S)$ is obtained as the colimit over the category of elements of S of some $\mathfrak{C}^{\square}(\square^n)$.

Lemma 3.1.1 *For every cubical set S, the set of objects of the simplicial category $\mathfrak{C}^{\square}(S)$ is in bijection with S_0, and thus will be identified with it.*

Proof By Remark 1.1.2, the functor $s\mathcal{C}at \xrightarrow{\text{Ob}} \mathcal{S}et$ is cocontinuous. We conclude since the statement holds on cubes by definition (cf Proposition 2.2.3). □

Notation 3.1.2 The previous lemma implies that the rigidification functor lifts to a functor

$$\mathfrak{C}^{\square} : c\mathcal{S}et_{*,*} \to s\mathcal{C}at_{*,*}.$$

We denote by $\mathfrak{C}^{\square}_t$ the functor from $c\mathcal{S}et_{*,*}$ to $s\mathcal{S}et$ defined on objects by

$$\mathfrak{C}^{\square}_t(S_{a,b}) = \mathfrak{C}^{\square}(S)(a,b) \quad \text{for all } a, b \in S_0.$$

Lemma 3.1.3 *The space $\mathfrak{C}^{\square}_t(\square^n)$ is contractible.*

Proof We have $\mathfrak{C}^{\square}_t(\square^n) = \mathfrak{C}^{\square}(\square^n)(\alpha, \omega) \cong N(\Sigma_{\{1,\dots,n\}})$ by Proposition 2.3.2 with the weak Bruhat order on $\Sigma_{\{1,\dots,n\}}$. The latter is a bounded poset (see Definition A.2.1); hence its nerve is contractible. □

A direct application of Corollaries 2.3.3 and A.2.6 is the following theorem.

Theorem 3.1.4 *For all necklaces T, there is an isomorphism of simplicial categories*

$$\mathfrak{C}^{\square}(T) \cong N(\mathfrak{C}^{\square}_{\text{path}}(T)).$$

In addition, if $T = U \vee V$, the composition in the simplicial category $\mathfrak{C}^{\square}(T)$ provides a morphism $\mathfrak{C}^{\square}_t(V) \times \mathfrak{C}^{\square}_t(U) \to \mathfrak{C}^{\square}_t(T)$, which is an isomorphism of simplicial sets.

Corollary 3.1.5 *Let T be a necklace. For every $a \preccurlyeq b \in T_0$, the simplicial set $\mathfrak{C}^{\square}(T)(a,b)$ is contractible. In particular, the simplicial set $\mathfrak{C}^{\square}_t(T)$ is contractible.*

Proof By Corollary 2.3.3, $\mathfrak{C}^{\square}_{\text{path}}(T)(a,b)$ is a product of bounded posets; hence its nerve is contractible. We conclude using Theorem 3.1.4. □

3.2 Computing the rigidification functor

The construction by left Kan extension gives us a way to express $\mathfrak{C}^{\square}(S)$ as a colimit in $s\mathcal{C}at$, which is difficult to compute. In this section, we use necklaces as "paths of higher dimension" to obtain a handy way to compute $\mathfrak{C}^{\square}(S)$. Indeed, we follow step by step the techniques developed by Dugger and Spivak [2011a, Proposition 4.3].

Each object $f\colon T \to S_{a,b}$ in $\mathcal{N}ec \downarrow S_{a,b}$ induces a morphism $\mathfrak{C}_t^{\square}(T) \to \mathfrak{C}_t^{\square}(S_{a,b})$ in $s\mathcal{S}et$ and this construction gives a morphism $\operatorname{colim}_{\mathcal{N}ec\downarrow S_{a,b}} \mathfrak{C}_t^{\square}(T) \to \mathfrak{C}_t^{\square}(S_{a,b})$ in $s\mathcal{S}et$. We prove that it is an isomorphism, so that $\mathfrak{C}^{\square}(S)(a,b) = \mathfrak{C}_t(S_{a,b})$ is computed as a colimit in $s\mathcal{S}et$.

Notation 3.2.1 Let $S_{a,b} \in c\mathcal{S}et_{*,*}$.

- We set $E_t(S_{a,b}) := \operatorname{colim}_{\mathcal{N}ec\downarrow S_{a,b}} \mathfrak{C}_t^{\square}(T)$.
- Let $(\alpha_t)_{S_{a,b}}\colon E_t(S_{a,b}) \to \mathfrak{C}_t^{\square}(S_{a,b})$ be the structure map from the colimit to the cocone $\mathfrak{C}_t^{\square}(S_{a,b})$.

The following facts are left to the reader.

- The definition of E_t is functorial, hence defines a functor $E_t\colon c\mathcal{S}et_{*,*} \to s\mathcal{S}et$.
- The morphisms $(\alpha_t)_{S_{a,b}}$ in $s\mathcal{S}et$ form a natural transformation $\alpha_t\colon E_t \Rightarrow \mathfrak{C}_t^{\square}$.

Our goal is to prove that α_t is a natural isomorphism. In fact, we will prove that E_t can be upgraded to a functor $E\colon c\mathcal{S}et \to s\mathcal{C}at$ that is naturally isomorphic to $\mathfrak{C}^{\square}$.

Proposition 3.2.2 *There exists a functor $E\colon c\mathcal{S}et \to s\mathcal{C}at$ and a natural transformation $\alpha\colon E \Rightarrow \mathfrak{C}^{\square}$ such that*

- $\operatorname{Ob}(E(S)) = S_0$,
- $E(S)(a,b) = E_t(S_{a,b})$, *and*
- α_S *is the identity on objects and* $\alpha_S(a,b) = (\alpha_t)_{S_{a,b}}\colon E(S)(a,b) \to \mathfrak{C}^{\square}(S)(a,b)$.

Proof We take $* \to S_{a,a} \in \mathcal{N}ec \downarrow S_{a,a}$ as identity morphism $\mathrm{id}_a \in E(S)(a,a)$. The composition in the simplicial category $E(S)$ is defined to be the composite featured as the left arrow in the diagram

$$\begin{array}{ccc}
E_t(S_{b,c}) \times E_t(S_{a,b}) = \operatorname*{colim}\limits_{V\to S_{b,c}} \mathfrak{C}_t^{\square}(V) \times \operatorname*{colim}\limits_{U\to S_{a,b}} \mathfrak{C}_t^{\square}(U) & \xleftarrow{\ \cong\ } & \operatorname*{colim}\limits_{\substack{V\to S_{b,c}\\ U\to S_{a,b}}} (\mathfrak{C}_t^{\square}(V) \times \mathfrak{C}_t^{\square}(U)) \\
\Big\downarrow {\scriptstyle \circ_E} & & \Big\downarrow {\scriptstyle \cong\ \text{Theorem 3.1.4}} \\
E_t(S_{a,c}) = \operatorname*{colim}\limits_{T\to S_{a,c}} \mathfrak{C}_t^{\square}(T) & \longleftarrow & \operatorname*{colim}\limits_{\substack{V\to S_{b,c}\\ U\to S_{a,b}}} \mathfrak{C}_t^{\square}(U \vee V)
\end{array}$$

where the top arrow is invertible, as $\times$ is cocontinuous in $s\mathcal{S}et$. Then the monoidal structure of $(c\mathcal{S}et_{*,*}, \vee, *)$ ensures that $E(S)$ with identities and composition as above is a simplicial category. The functoriality of E comes from that of E_t.

Let us check that the $(\alpha_t)_{S_{a,b}}$ induce an enriched functor $\alpha_S\colon E(S) \to \mathfrak{C}^{\square}(S)$, which is the identity on objects. We have to prove that the following diagram commutes:

$$\begin{array}{ccccccc}
\operatorname*{colim}_{V\to S_{b,c}} \mathfrak{C}^{\square}_t(V) \times \operatorname*{colim}_{U\to S_{a,b}} \mathfrak{C}^{\square}_t(U) & \overset{\cong}{\Leftarrow} & \operatorname*{colim}_{\substack{V\to S_{b,c}\\ U\to S_{a,b}}} (\mathfrak{C}^{\square}_t(V)\times\mathfrak{C}^{\square}_t(U)) & \overset{\cong}{\Rightarrow} & \operatorname*{colim}_{\substack{V\to S_{b,c}\\ U\to S_{a,b}}} \mathfrak{C}^{\square}_t(U\vee V) & \to & \operatorname*{colim}_{T\to S_{a,c}} \mathfrak{C}^{\square}_t(T) \\
\downarrow{\scriptstyle (\alpha_t)_{S_{b,c}}\times(\alpha_t)_{S_{a,b}}} & \swarrow & & & & \searrow & \downarrow{\scriptstyle (\alpha_t)_{S_{a,c}}} \\
\mathfrak{C}^{\square}_t(S_{b,c})\times\mathfrak{C}^{\square}_t(S_{a,b}) & & \longrightarrow & & & & \mathfrak{C}^{\square}_t(S_{a,c})
\end{array}$$

It suffices to notice that for all $V \to S_{b,c}$ and $U \to S_{a,b}$, $\mathfrak{C}^{\square}(U \vee V) \to \mathfrak{C}^{\square}(S)$ is a simplicially enriched functor and so the following square commutes:

$$\begin{array}{ccc}
\mathfrak{C}^{\square}(U\vee V)(\alpha_V,\omega_V)\times\mathfrak{C}^{\square}(U\vee V)(\alpha_U,\omega_U) & \longrightarrow & \mathfrak{C}^{\square}(U\vee V)(\alpha_U,\omega_V) \\
\downarrow & & \downarrow \\
\mathfrak{C}^{\square}(S)(b,c)\times\mathfrak{C}^{\square}(S)(a,b) & \longrightarrow & \mathfrak{C}^{\square}(S)(a,c)
\end{array}$$

Theorem 3.1.4 applied to our case gives

$$\mathfrak{C}^{\square}(U\vee V)(\alpha_V,\omega_V) = \mathfrak{C}^{\square}(V)(\alpha_V,\omega_V),$$
$$\mathfrak{C}^{\square}(U\vee V)(\alpha_U,\omega_U) = \mathfrak{C}^{\square}(U)(\alpha_U,\omega_U),$$

so the diagram becomes

$$\begin{array}{ccc}
\mathfrak{C}^{\square}_t(V)\times\mathfrak{C}^{\square}_t(U) & \longrightarrow & \mathfrak{C}^{\square}_t(U\vee V) \\
\downarrow & & \downarrow \\
\mathfrak{C}^{\square}_t(S_{b,c})\times\mathfrak{C}^{\square}_t(S_{a,b}) & \longrightarrow & \mathfrak{C}^{\square}_t(S_{a,c})
\end{array}$$

and we conclude by universality of colimits. The naturality of $\alpha_S\colon E(S) \to \mathfrak{C}^{\square}(S)$ comes from the naturality of α_t. □

Proposition 3.2.3 *The natural transformation $\alpha\colon E \Rightarrow \mathfrak{C}^{\square}$ is a natural isomorphism.*

Proof Let $S \in c\mathcal{S}et$. We know that α_S is the identity on objects. Assume first that $S = T$ is a necklace. Let $a, b \in T_0$. If $a \preccurlyeq b$, by Proposition 2.1.8, $\iota^T_{a,b}\colon T_{[a,b]} \hookrightarrow T$ is terminal in $\mathcal{N}ec \downarrow T_{a,b}$; hence $E(T)(a,b) \cong \mathfrak{C}^{\square}_t(T_{[a,b]}) = \mathfrak{C}^{\square}(T)(a,b)$, and the isomorphism is precisely induced by α_T. If $a \not\preccurlyeq b$, then both categories are empty; hence the result holds for necklaces.

We prove that for all vertices a and b of S, the morphism $\alpha_S(a,b)\colon E(S)(a,b) \to \mathfrak{C}^{\square}(S)(a,b)$ is an isomorphism of simplicial sets, by providing an inverse $\beta_S(a,b)$. Recall that

$$\alpha_S(a,b) = \alpha_t(S_{a,b})\colon E_t(S_{a,b}) \to \mathfrak{C}^{\square}_t(S_{a,b})$$

is the (unique) map from the colimit to the cocone $\mathfrak{C}_t^{\square}(S_{a,b})$. Define $(\beta_t)_{S_{a,b}}\colon \mathfrak{C}_t^{\square}(S_{a,b}) \to E_t(S_{a,b})$ as the composite

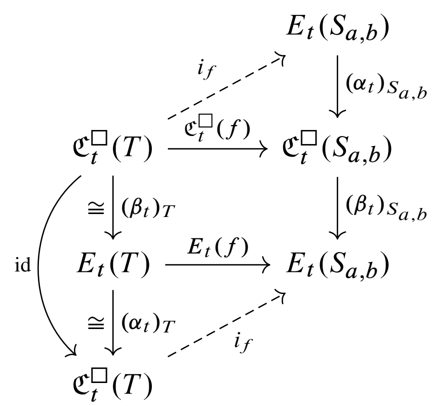
$$\begin{array}{ccc} E_t(S_{a,b}) & \longleftarrow & \big(\operatorname*{colim}_{\square^k\to S} E(\square^k)\big)(a,b) \\ {\scriptstyle (\beta_t)_{S_{a,b}}} \uparrow & & \downarrow \cong \ {\scriptstyle \square^k \text{ is a necklace}} \\ \mathfrak{C}_t^{\square}(S_{a,b}) & \xleftarrow{\ \cong\ } & \big(\operatorname*{colim}_{\square^k\to S} \mathfrak{C}^{\square}(\square^k)\big)(a,b) \end{array}$$

By naturality of all the maps involved in the diagram, the family $(\beta_t)_{S_{a,b}}$ assembles to a natural transformation $\beta_t\colon \mathfrak{C}_t^{\square} \Rightarrow E_t$, giving rise to $\beta\colon \mathfrak{C}^{\square} \Rightarrow E$.

We show, in this order, that β is a right inverse, and a left inverse of α. To show the former, it is enough to show $\alpha_S \circ \beta_S \circ j_f = j_f$, for all $f\colon \square^k \to S_{a,b}$, where j_f is the characteristic map $E(\square^k) \to \operatorname{colim}_{\square^k\to S} E(\square^k)$, and where we identify $\mathfrak{C}^{\square}(S)$ with $\operatorname{colim}_{\square^k\to S} E(\square^k)$. Indeed, we have

$$\begin{aligned} \alpha_S \circ \beta_S \circ j_f &= \alpha_S \circ E(f) && \text{(by definition of } \beta) \\ &= \mathfrak{C}^{\square}(f) \circ \alpha_{\square^k} && \text{(by naturality of } \alpha) \\ &= j_f && \text{(by the identification above).} \end{aligned}$$

We prove now that $(\beta_t)_{S_{a,b}} \circ (\alpha_t)_{S_{a,b}}$ is the identity in a similar way. By the universal property of the colimit, it is enough to prove $(\beta_t)_{S_{a,b}} \circ (\alpha_t)_{S_{a,b}} \circ i_f = i_f$ for all $T \xrightarrow{f} S_{a,b}$, where $i_f\colon \mathfrak{C}_t^{\square}(T) \to E_t(S_{a,b})$ is the characteristic morphism. This comes from the commutative diagram



(the above triangle commutes by definition of α_t, the middle square commutes by naturality of β_t, and the bottom triangle commutes by definition of $E_t(f)$). $\square$

As a direct corollary, we get the main theorem of the section.

Theorem 3.2.4 *Let S be a cubical set and $a, b \in S_0$. We have the following isomorphism of simplicial sets:*

$$\mathfrak{C}_t^{\square}(S_{a,b}) = \mathfrak{C}^{\square}(S)(a,b) \cong \operatorname*{colim}_{(T\to S_{a,b})\in \mathcal{N}ec\downarrow S_{a,b}} \mathfrak{C}_t^{\square}(T).$$

In the following remark, we point out that necklaces were instrumental in getting the characterisation given in the previous theorem.

Remark 3.2.5 We *do not* have an isomorphism between $\mathfrak{C}^{\square}(S)(a,b)$ and the colimit above restricted to cubes. This already fails when S is a necklace. Consider $S = (1,1)$ and $(a,b) = (\alpha,\omega)$. There is no morphism from a cube to $S_{\alpha,\omega}$ in $c\mathcal{S}et_{*,*}$, whereas there is one from the necklace S to itself (the identity morphism). Hence the restricted colimit is empty, while $\mathfrak{C}^{\square}(S)(a,b)$ is not.

Remark 3.2.6 It can be shown, using the result above, that $\pi_0(\mathfrak{C}^{\square}(S)) \cong \pi_0(\mathfrak{C}^{\square}_{\text{path}}(S))$ for any cubical set S.

3.3 Case of cubical subsets of a cube

Definition 3.3.1 Let $S_{a,b}$ in $c\mathcal{S}et_{*,*}$, with S a cubical subset of an n–cube. The subcategory of $\mathcal{N}ec \downarrow S_{a,b}$ whose objects are monomorphisms $T \to S_{a,b}$ and arrows are monomorphisms between necklaces is denoted $\mathrm{SubNeck}(S_{a,b})$. The category $\mathrm{SubNeck}(S_{a,b})$ is actually a poset, as shown in Proposition B.1.1.

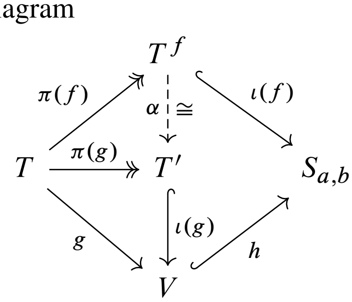
Proposition 3.3.2 *Let $S_{a,b} \in c\mathcal{S}et_{*,*}$, with S a cubical subset of an n–cube. The rigidification functor has the expression*

$$\mathfrak{C}^{\square}(S)(a,b) = \mathfrak{C}^{\square}_t(S_{a,b}) \cong \operatorname*{colim}_{\mathrm{SubNeck}(S_{a,b})} \mathfrak{C}^{\square}_t(T).$$

Proof We use Lemma 2.1.9 and its notation. Recall from Theorem 3.2.4 that

$$\mathfrak{C}^{\square}_t(S_{a,b}) \cong \operatorname*{colim}_{\mathcal{N}ec\downarrow S_{a,b}} \mathfrak{C}^{\square}_t(T).$$

Consider the inclusion functor $U\colon \mathrm{SubNeck}(S_{a,b}) \hookrightarrow \mathcal{N}ec \downarrow S_{a,b}$ and fix $f \in \mathcal{N}ec \downarrow S_{a,b}$. The category $f \downarrow U$ has $\pi(f)\colon f \to \iota(f)$ as object, and hence is not empty. Let $g\colon f \to h$ be an object in $f \downarrow U$, so that $hg = f = \iota(f)\pi(f)$. The morphism g in $\mathcal{N}ec$ admits the factorisation $g = \iota(g)\pi(g)$. By the unique decomposition of f as an epimorphism followed by a monomorphism, there exists an isomorphism $\alpha\colon T^f \to T'$, as illustrated by the diagram



In conclusion, the morphism $\iota(g)\alpha$ is a monomorphism, hence a morphism in $f \downarrow U$ from $\pi(f)$ to g. Thus the category $f \downarrow U$ is connected. We have proved that U is a final functor, from which the statement follows (cf [Mac Lane 1998, Section IX.3]). □

4 Quillen equivalence

In this section, we prove that the adjunction $\mathfrak{C}^{\square} \dashv N^{\square}$ is a Quillen equivalence between the Joyal model structure on $c\mathcal{S}et$ and the Bergner model structure on $s\mathcal{C}at$.

4.1 Properties of the functor $\mathfrak{C}^{\square}$

Proposition 4.1.1 *The functor $\mathfrak{C}^{\square}$ preserves cofibrations.*

Proof Since cofibrations of $c\mathcal{S}et$ are generated by $\partial^n\colon \partial\square^n \to \square^n$, it is enough to prove that

$$\mathfrak{C}^{\square}(\partial^n)\colon \mathfrak{C}^{\square}(\partial\square^n) \to \mathfrak{C}^{\square}(\square^n)$$

is a cofibration. We claim that the diagram

$$\begin{array}{ccc} \Sigma(\mathfrak{C}^{\square}_t(\partial\square^n)) & \longrightarrow & \mathfrak{C}^{\square}(\partial\square^n) \\ {\scriptstyle \Sigma(\mathfrak{C}^{\square}_t(\partial^n))}\downarrow & & \downarrow{\scriptstyle \mathfrak{C}^{\square}(\partial^n)} \\ \Sigma(\mathfrak{C}^{\square}_t(\square^n)) & \longrightarrow & \mathfrak{C}^{\square}(\square^n) \end{array}$$

where the horizontal morphisms are given by the counit of the adjunction $\Sigma \dashv \mathrm{Hom}$ of Proposition 1.1.1, is a pushout diagram. The set of objects of the simplicial categories on the right-hand side of the diagram is in bijection with $(\square^n)_0 = (\partial\square^n)_0$, which is a bounded poset. Let a and b be two such objects. If $(a,b) \neq (\alpha,\omega)$, then $\square^{d(a,b)} \subseteq \partial\square^n$, so that the functor $\mathcal{N}ec \downarrow (\partial\square^n)_{a,b} \to \mathcal{N}ec \downarrow (\square^n)_{a,b}$ is an isomorphism of categories. Theorem 3.2.4 implies then that the map $\mathfrak{C}^{\square}(\partial^n)(a,b)\colon \mathfrak{C}^{\square}(\partial\square^n)(a,b) \to \mathfrak{C}^{\square}(\square^n)(a,b)$ is an isomorphism. We conclude by Proposition A.2.7. We show next that $\mathfrak{C}^{\square}_t(\partial\square^n) \to \mathfrak{C}^{\square}_t(\square^n)$ is a cofibration. Indeed $\partial\square^n$ is a cubical subset of $\square^n$; hence $\mathfrak{C}^{\square}_t(\partial\square^n) \cong \mathrm{colim}_{T \in \mathrm{SubNeck}(\partial\square^n)} \mathfrak{C}^{\square}_t(T)$ by Proposition 3.3.2, so $\mathfrak{C}^{\square}_t(\partial\square^n) \to \mathfrak{C}^{\square}_t(\square^n)$ is a cofibration by Lemma B.1.5. The functor Σ preserves cofibrations by Proposition 1.1.1, and so do pushout diagrams. $\square$

We refer to Definition 1.2.5 for the notation in the next lemma.

Lemma 4.1.2 *In the diagram*

$$\begin{array}{ccc} \mathfrak{C}^{\square}(\sqcap^n_{i,\epsilon})(a,b) & \xrightarrow[\cong]{\mathfrak{C}^{\square}(p)} & \mathfrak{C}^{\square}(\hat{\sqcap}^n_{i,\epsilon})(pa,pb) \\ \downarrow & & \downarrow \\ \mathfrak{C}^{\square}(\square^n)(a,b) & \xrightarrow[\cong]{\mathfrak{C}^{\square}(p)} & \mathfrak{C}^{\square}(\hat{\square}^n_{i,\epsilon})(pa,pb) \end{array}$$

the horizontal arrows (where p is the quotient map) are isomorphisms of simplicial sets when $a \neq \{i\}$ if $\epsilon = 1$, and $b \neq \{1,\dots,n\} \setminus \{i\}$ if $\epsilon = 0$.

Proof We prove the isomorphism $\mathfrak{C}^{\square}(p)\colon \mathfrak{C}^{\square}(\square^n) \to \mathfrak{C}^{\square}(\widehat{\square}^n_{i,\epsilon})$ with $\epsilon = 1$ and $i = 1$, so that the critical edge $e_{1,1}\colon \square^1 \to \square^n$ corresponds to the edge from $\alpha = \varnothing$ to $\{1\}$. The other cases are similar. Since $\mathfrak{C}^{\square}$ preserves colimits, we have the pushout diagram

$$\begin{array}{ccc} \mathfrak{C}^{\square}(\square^1) & \xrightarrow{\mathfrak{C}^{\square}(e_{1,1})} & \mathfrak{C}^{\square}(\square^n) \\ \downarrow & & \downarrow \\ \mathfrak{C}^{\square}(\square^0) & \longrightarrow & \mathfrak{C}^{\square}(\widehat{\square}^n_{1,1}) \end{array}$$

Let us define the following simplicial category S. The set of objects of S is identified with that of $\widehat{\square}^n_{1,1}$. We denote by $\bar{\alpha} = p(\alpha) = p(\{1\})$. Define $S(\bar{\alpha},\bar{\alpha}) = *$, $S(\bar{\alpha}, b) = \mathfrak{C}^{\square}(\square^n)(\alpha, b)$ for $b \neq \bar{\alpha}$ and $S(a,b) = \mathfrak{C}^{\square}(\square^n)(a,b)$ for $a \neq \bar{\alpha}$. The composition is induced by that of $\mathfrak{C}^{\square}(\square^n)$. Let $\pi\colon \mathfrak{C}^{\square}(\square^n) \to S$ be the map which coincides with p on objects, and is the identity on morphisms except for the case $\pi\colon \mathfrak{C}^{\square}(\square^n)(\{1\}, b) \to S(\bar{\alpha}, b)$, for which we use the composite

$$\mathfrak{C}^{\square}(\square^n)(\{1\}, b) \to \mathfrak{C}^{\square}(\square^n)(\{1\}, b) \times \mathfrak{C}^{\square}(\square^n)(\alpha, \{1\}) \xrightarrow{\circ} \mathfrak{C}^{\square}(\square^n)(\alpha, b) \cong S(\bar{\alpha}, b),$$

which is well defined since $\mathfrak{C}^{\square}(\square^n)(\alpha, \{1\}) = *$. One checks easily that S satisfies the universal pushout property; hence $\mathfrak{C}^{\square}(\widehat{\square}^n_{1,1}) \cong S$. We conclude, since by definition of S,

$$\mathfrak{C}^{\square}(\widehat{\square}^n_{1,1})(pa, pb) \cong S(pa, pb) = \mathfrak{C}^{\square}(\square^n)(a,b) \quad \text{if } a \neq \{1\}.$$

The side condition is needed since, for $a = \{1\}$ and $b \succ 1$,

$$S(p(\{1\}), b) = S(\bar{\alpha}, b) = \mathfrak{C}^{\square}(\square^n)(\alpha, b) \not\cong \mathfrak{C}^{\square}(\square^n)(\{1\}, b).$$

The proof for the inner open box is exactly the same since $\mathfrak{C}^{\square}(\sqcap^n_{1,1})(\alpha, \{1\}) = *$. □

Proposition 4.1.3 *$\mathfrak{C}^{\square}(h^n_{i,\epsilon})\colon \mathfrak{C}^{\square}(\widehat{\sqcap}^n_{i,\epsilon}) \to \mathfrak{C}^{\square}(\widehat{\square}^n_{i,\epsilon})$ is an acyclic cofibration.*

Proof The proof is analogous to the proof of Proposition 4.1.1. Note that the set of vertices of $\widehat{\sqcap}^n_{i,\epsilon}$ coincides with that of $\widehat{\square}^n_{i,\epsilon}$ and is a bounded poset. By Proposition A.2.7, in order to establish that the diagram

$$\begin{array}{ccc} \Sigma(\mathfrak{C}^{\square}_t(\widehat{\sqcap}^n_{i,\epsilon})) & \longrightarrow & \mathfrak{C}^{\square}(\widehat{\sqcap}^n_{i,\epsilon}) \\ {\scriptstyle \Sigma(\mathfrak{C}^{\square}_t(h^n_{i,\epsilon}))}\downarrow & & \downarrow{\scriptstyle \mathfrak{C}^{\square}(h^n_{i,\epsilon})} \\ \Sigma(\mathfrak{C}^{\square}_t(\widehat{\square}^n_{i,\epsilon})) & \longrightarrow & \mathfrak{C}^{\square}(\widehat{\square}^n_{i,\epsilon}) \end{array}$$

is a pushout diagram, we have to show that the maps

$$\mathfrak{C}^{\square}(h^n_{i,\epsilon})(pa, pb)\colon \mathfrak{C}^{\square}(\widehat{\sqcap}^n_{i,\epsilon})(pa, pb) \to \mathfrak{C}^{\square}(\widehat{\square}^n_{i,\epsilon})(pa, pb)$$

are isomorphisms, for every (a,b) such that $(pa, pb) \neq (p\alpha, p\omega)$, or equivalently using Lemma 4.1.2, that $\mathfrak{C}^{\square}(h^n_{i,\epsilon})(a,b)\colon \mathfrak{C}^{\square}(\sqcap^n_{i,\epsilon})(a,b) \to \mathfrak{C}^{\square}(\square^n_{i,\epsilon})(a,b)$ is an isomorphism. The latter is established exactly as in the proof of Proposition 4.1.1, noticing that $\square^{d(a,b)} \subseteq \sqcap^n_{i,\epsilon}$ for such (a,b). What remains to prove is that $\mathfrak{C}^{\square}_t(\widehat{\sqcap}^n_{i,\epsilon}) \to \mathfrak{C}^{\square}_t(\widehat{\square}^n_{i,\epsilon})$ is an acyclic cofibration, or equivalently, by Lemma 4.1.2 again, that $\mathfrak{C}^{\square}_t(\sqcap^n_{i,\epsilon}) \to \mathfrak{C}^{\square}_t(\square^n_{i,\epsilon})$ is an acyclic cofibration for the Kan–Quillen structure on simplicial sets. We already know that it is a cofibration, since $\mathfrak{C}^{\square}$ preserves cofibrations by Proposition 4.1.1 and so does $\mathfrak{C}^{\square}_t$. Since $\mathfrak{C}^{\square}_t(\square^n)$ is contractible (Lemma 3.1.3), showing that $\mathfrak{C}^{\square}_t(\sqcap^n_{i,\epsilon}) \to \mathfrak{C}^{\square}_t(\square^n)$ is acyclic amounts to proving that $\mathfrak{C}^{\square}_t(\sqcap^n_{i,\epsilon})$ is contractible. From Proposition 3.3.2, we have

$$\mathfrak{C}^{\square}_t(\sqcap^n_{i,\epsilon}) = \operatorname*{colim}_{T \in \mathrm{SubNeck}(\sqcap^n_{i,\epsilon})} \mathfrak{C}^{\square}_t(T).$$

Furthermore, we show in Section B.1 that the category $\mathrm{SubNeck}(\sqcap^n_{i,\epsilon})$ is direct (and hence is Reedy). The diagram $(T \to \sqcap^n_{i,\epsilon}) \mapsto \mathfrak{C}^{\square}_t(T)$ is Reedy cofibrant: for every $T \in \mathrm{SubNeck}(\sqcap^n_{i,\epsilon})$, the latching morphism

$$\operatorname*{colim}_{\substack{U \in \mathrm{SubNeck}(T) \\ U \neq T}} \mathfrak{C}^{\square}_t(U) \to \mathfrak{C}^{\square}_t(T)$$

is a monomorphism by Lemma B.1.5. It follows from [Hirschhorn 2003, Theorem 19.9.1], and from the fact that any direct category has fibrant constants, that the natural map

$$\operatorname*{hocolim}_{T \in \mathrm{SubNeck}(\sqcap^n_{i,\epsilon})} \mathfrak{C}^{\square}_t(T) \to \operatorname*{colim}_{T \in \mathrm{SubNeck}(\sqcap^n_{i,\epsilon})} \mathfrak{C}^{\square}_t(T)$$

is a weak equivalence of simplicial sets. In conclusion,

$$\mathfrak{C}^{\square}_t(\sqcap^n_{i,\epsilon}) \sim \operatorname*{hocolim}_{T \in \mathrm{SubNeck}(\sqcap^n_{i,\epsilon})} \mathfrak{C}^{\square}_t(T) \sim \operatorname*{hocolim}_{T \in \mathrm{SubNeck}(\sqcap^n_{i,\epsilon})} * \sim N(\mathrm{SubNeck}(\sqcap^n_{i,\epsilon})) \sim *$$

since $\mathfrak{C}^{\square}_t(T)$ and $N(\mathrm{SubNeck}(\sqcap^n_{i,\epsilon}))$ are contractible, by Corollary 3.1.5 and Proposition B.2.1. □

4.2 Quillen adjunction

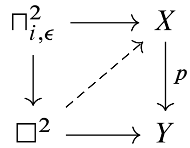
We shall use the following result of Joyal [2008, E.2.14].

Proposition 4.2.1 *An adjunction $L \dashv R$ between two model categories is Quillen if and only if L preserves cofibrations and R preserves fibrations between fibrant objects.*

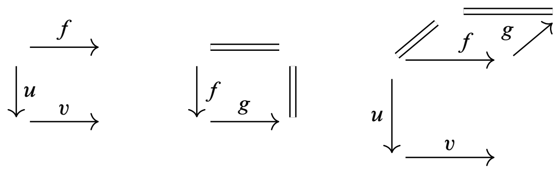
In view of this proposition, what remains to prove is that $N^{\square}$ sends fibrations between fibrant categories to fibrations between cubical quasicategories. We shall use two lemmas, which we now present. We refer to Definition 1.2.8 for the definition of equivalence and that of special open box. The first lemma is tautological.

Lemma 4.2.2 *Let $\mathcal{A}$ be a fibrant simplicial category, $v : \mathfrak{C}^{\square}(\square^1) \to \mathcal{A}$ and $\tilde{v} : \square^1 \to N^{\square}(\mathcal{A})$ its transpose. Since $\mathfrak{C}^{\square}(\square^1)$ is the simplicial category with only one nontrivial arrow, we see v as an arrow in $\mathcal{A}$. Then $\tilde{v}$ is an equivalence in the cubical quasicategory $N^{\square}(\mathcal{A})$ if and only if $\pi_0(v)$ is an isomorphism in $\pi_0(\mathcal{A})$.*

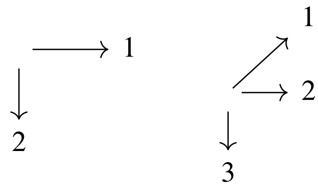
Lemma 4.2.3 *Let $p\colon X \to Y$ be an inner fibration between cubical quasicategories and $\sqcap^2_{i,\epsilon} \to X$ a special open box. Any commutative square of the following form has a lift*:



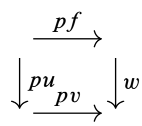
Proof It is a special case of [Doherty et al. 2024, Lemma 4.14]. We give a proof for the case $(i,\epsilon) = (1,1)$, the other cases being similar. We represent the map $\sqcap^2_{1,1} \to X$ on the left below, with f an equivalence. The 1–cube f being an equivalence, the middle diagram below exists. Gluing the two diagrams, we get the partially filled 3–cube in X (on the right)



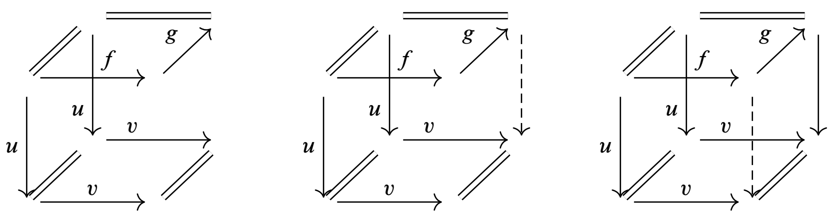
where our conventions for the coordinates in dimensions 2 and 3 are



The map $B\colon \square^2 \to Y$ is represented by the following 2–cube in Y:

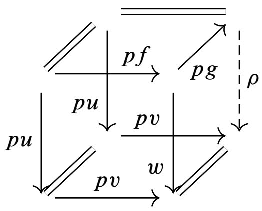


The proof goes in three steps. For the first step, we assume $Y = *$. We complete the above 3–cube cube progressively, as follows:



The top face is full by hypothesis, the left and the bottom faces are given by degeneracies. Because X is a cubical quasicategory, the back face (picture in the middle), then the right face (picture on the right), and finally the whole cube, and hence a fortiori the front face, can be filled. In consequence any special open box in a cubical quasicategory X can be filled by a 2–cube in X.

The second step consists in filling the following 3–cube in Y, where the front, top, left and bottom faces are already filled:



By [Doherty et al. 2024, Lemma 2.6], g is an equivalence because f is and so is pg. Hence the right face is a special open box in the cubical quasicategory Y and thus can be filled by step 1. Then the whole cube is filled because the critical edge associated to the back face is the identity.

For the last step, we resume the filling of the 3–cube of the first step (with the same pictures as above) in X, but now in the general case. The aim is to fill in the front face of the 3–cube in X by a 2–cube A satisfying $pA = B$. Because p is an inner fibration, its back face can be filled by a 2–cube such that the dashed arrow in the picture in the middle is sent to ρ by p. Then the same is true for its right face, so the dashed arrow in the picture on the right is sent to w by p. Finally, the whole cube is filled and sent by p to the 3–cube in Y, and a fortiori its front face A satisfies $pA = B$. □

Proposition 4.2.4 *The functor $N^{\square}$ sends fibrations between fibrant simplicial categories to fibrations between cubical quasicategories.*

Proof Using Proposition 4.1.3, we conclude by adjunction that $N^{\square}(\mathcal{C})$ is a cubical quasicategory if $\mathcal{C}$ is a fibrant simplicial category, and that if $f : \mathcal{C} \to \mathcal{D}$ is a DK-fibration between fibrant simplicial categories, then $N^{\square}(f)$ is an inner fibration between cubical quasicategories. By Theorem 1.2.7, we are left to show that $N^{\square}(f)$ has the right lifting property with respect to the endpoint inclusions $j_0 : \{0\} \to K$ and $j_1 : \{1\} \to K$. These cases being similar, we only treat the first one. Consider a commutative square

$$\begin{array}{ccc} \{0\} & \xrightarrow{\bar{a}} & N^{\square}\mathcal{C} \\ \big\downarrow{\scriptstyle j_0} & & \big\downarrow{\scriptstyle N^{\square} f} \\ K & \longrightarrow & N^{\square}\mathcal{D} \end{array}$$

We shall first lift the middle vertical edge of K. By Lemma 4.2.2, its image in $N^{\square}(\mathcal{D})$ corresponds to some arrow $v \in \mathcal{D}_0(a, b)$, which is an isomorphism in $\pi_0(\mathcal{D})$. The same will have to be true for its image in $N^{\square}(\mathcal{C})$ through the lifting. The object $\bar{a}$ of $\mathcal{C}$ satisfies $f(\bar{a}) = a$. We proceed as follows.

- Since $\pi_0(f)$ is an isofibration of categories, there exists some $\bar{b} \in \mathrm{Ob}(\mathcal{C})$ and some $v' \in \mathcal{C}_0(\bar{a}, \bar{b})$ such that $f(\bar{b}) = b$, $\pi_0(f(v')) = \pi_0(v)$, and $\pi_0(v')$ is an isomorphism in $\pi_0(\mathcal{C})$.
- Since $\mathcal{D}(a, b)$ is a Kan complex, we can find a 1–simplex $\delta \in \mathcal{D}_1(a, b)$ such that $\partial_1 \delta = v$ and $\partial_0 \delta = f(v')$.
- Since $f_{\bar{a},\bar{b}} : \mathcal{C}(\bar{a}, \bar{b}) \to \mathcal{D}(a, b)$ is a Kan fibration, we can lift $\delta \in \mathcal{D}_1(a, b)$ to some $\bar{\delta} \in \mathcal{C}_1(\bar{a}, \bar{b})$ satisfying $\partial_0 \bar{\delta} = v'$.

Then $\bar{v} = \partial_1 \bar{\delta} \in \mathcal{C}_0(\bar{a}, \bar{b})$ meets our goal, ie satisfies $f(\bar{v}) = v$, and $\pi_0(\bar{v}) = \pi_0(v')$ is an isomorphism in $\pi_0(\mathcal{C})$ such that $\bar{v}$, seen as an edge in $N^{\square}(\mathcal{A})$, is an equivalence, by Lemma 4.2.2. Therefore, the two open boxes in

$$\begin{array}{ccccc} 1 & & 0 & = & 0 \\ \| & & {\scriptstyle \bar{v}}\downarrow & & \| \\ 1 & = & 1 & & 0 \end{array}$$

are special. Calling this diagram $\bar{v}$, our lifting problem reduces now to

$$\begin{array}{ccc} K' & \xrightarrow{\bar{v}} & N^{\square}\mathcal{C} \\ \downarrow & & \downarrow {\scriptstyle N^{\square} f} \\ K & \longrightarrow & N^{\square}\mathcal{D} \end{array}$$

where K' is K without its 2–cubes and without its horizontal nondegenerate 1–cubes. This is performed by applying Lemma 4.2.3 to each of the two special boxes above. □

Proposition 4.2.5 *The adjunction $\mathfrak{C}^{\square} \dashv N^{\square}$ is Quillen.*

Proof This follows from Propositions 4.1.1 and 4.2.4, thanks to Joyal's characterisation recalled in Proposition 4.2.1. □

4.3 Quillen equivalence

In order to prove that the Quillen adjunction $\mathfrak{C}^{\square} \dashv N^{\square}$ is a Quillen equivalence, we first compare it with the simplicial rigidification $\mathfrak{C}^{\Delta}$ using the functor Q of Section 1.2 and then use the Quillen equivalences induced by Q and $\mathfrak{C}^{\Delta}$.

Lemma 4.3.1 *There exists a morphism $\phi_n : \mathfrak{C}^{\square}(Q^n) \to \mathfrak{C}^{\Delta}(\Delta^n)$ in $s\mathcal{C}at$, which is a bijection on objects and is natural in $[n] \in \Delta$, ie a natural transformation $\phi : \mathfrak{C}^{\square} \circ Q \Rightarrow \mathfrak{C}^{\Delta} \circ Y$.*

Proof We start by defining a family of morphisms $\psi_n : \mathfrak{C}^{\square}(\square^n) \to \mathfrak{C}^{\Delta}(\Delta^n)$ in $s\mathcal{C}at$. On objects, we set $\psi_n(a) = \sup a$ (cf Lemma 1.2.11). If $a \preccurlyeq b$ ($a \subseteq b$), then $\sup a \leq \sup b$, and we define a map $\psi_n(a,b) : \mathfrak{C}^{\square}(\square^n)(a,b) \to \mathfrak{C}^{\Delta}(\Delta^n)(\sup a, \sup b)$ in $s\mathcal{S}et$, as follows. Since $\mathfrak{C}^{\square}(\square^n)(a,b)$ is the nerve of the poset $\Sigma_{b \setminus a}$ with the weak order $\rightsquigarrow_B$ (see Section 2.3), and $\mathfrak{C}^{\Delta}(\Delta^n)(\sup a, \sup b)$ is the nerve of the poset $\mathcal{P}(]\sup a, \sup b[)$ with the inclusion order, we define this map at the level of the underlying posets. Let $k = d(a,b)$ and $(x_1, \dots, x_k) \in \Sigma_{b \setminus a}$. We set

$$\tilde{\psi}_n(a,b)(x_1, \dots, x_k) = \{x_l \mid x_p < x_l \text{ for all } p < l\} \cap]\sup a, \sup b[.$$

The map $\tilde{\psi}_n(a,b)$ is a morphism of posets. Assume $x_r > x_{r+1}$ for some r. Then

$$x := (x_1, \dots, x_k) \rightsquigarrow_B (x_1, \dots, x_{r+1}, x_r, \dots, x_k) =: y.$$

We write $A(x)=\{x_l \mid x_p < x_l \text{ for all } p<l\}$. Then we observe that $A(x)\setminus\{x_r, x_{r+1}\}=A(y)\setminus\{x_r, x_{r+1}\}$, that $x_{r+1}\notin A(x)$, and that $(x_r \in A(x)) \Rightarrow (x_r \in A(y))$. It follows that $A(x)\subseteq A(y)$, and hence $\tilde{\psi}_n(a,b)(x)\subseteq\tilde{\psi}_n(a,b)(y)$.

We next show that $\tilde{\psi}_n$ preserves the concatenation product. Assume $a \preccurlyeq b \preccurlyeq c$. Let

$$x := (x_1,\dots,x_k)\in\Sigma_{b\setminus a},\quad y := (y_1,\dots,y_{k'})\in\Sigma_{c\setminus b}.$$

We set $z := (x_1,\dots,x_k,y_1,\dots,y_{k'})$. We have to prove that if $\sup b\notin\{\sup a,\sup c\}$ (equivalently $\sup a<\sup b<\sup c$), then

$$\tilde{\psi}_n(a,c)(z)=\tilde{\psi}_n(a,b)(x)\cup\{\sup b\}\cup\tilde{\psi}_n(b,c)(y).$$

We observe that $A(z)$ splits as $A(x)\cup B$, where $B\subseteq A(y)$ and $A(y)\cap\,]\sup b,\sup c[\,\subseteq B$. This settles the left-to-right inclusion, as well as the inclusions $\tilde{\psi}_n(a,b)(x)\subseteq\tilde{\psi}_n(a,c)(z)$ and $\tilde{\psi}_n(b,c)(y)\subseteq\tilde{\psi}_n(a,c)(z)$. Since $\sup b>\sup a$, we have $\sup(b\setminus a)=\sup b$. Thus there exists l such that $x_l=\sup b$ and $x_p\le\sup b$ for all $p\in\{1,\dots,k\}$, and a fortiori $\{\sup b\}\subseteq\tilde{\psi}_n(a,c)(z)$ holds.

In conclusion, setting $\psi_n=N(\tilde{\psi}_n)$, we have shown that $\psi_n\colon\mathfrak{C}^{\square}(\square^n)\to\mathfrak{C}^{\Delta}(\Delta^n)$ is an enriched functor of simplicial categories.

Let us prove that ψ_n factors through the map $\mathfrak{C}^{\square}(\pi_n)\colon\mathfrak{C}^{\square}(\square^n)\to\mathfrak{C}^{\square}(Q^n)$, where π_n is the quotient map of Definition 1.2.10. As $\mathfrak{C}^{\square}$ is cocontinuous, from Definition 1.2.10, the following diagram is a pushout:

$$\begin{array}{ccc}
\mathfrak{C}^{\square}(\square^0\otimes\square^{n-1})\sqcup\mathfrak{C}^{\square}(\square^1\otimes\square^{n-2})\sqcup\dots\sqcup\mathfrak{C}^{\square}(\square^{n-1}\otimes\square^0) & \xrightarrow{(\mathfrak{C}^{\square}(\partial_{1,1}),\mathfrak{C}^{\square}(\partial_{2,1}),\dots,\mathfrak{C}^{\square}(\partial_{n,1}))} & \mathfrak{C}^{\square}(\square^n)\\
\big\downarrow & & \big\downarrow{\scriptstyle\mathfrak{C}^{\square}(\pi_n)}\\
\mathfrak{C}^{\square}(\square^{n-1})\sqcup\mathfrak{C}^{\square}(\square^{n-2})\sqcup\dots\sqcup\mathfrak{C}^{\square}(\square^0) & \longrightarrow & \mathfrak{C}^{\square}(Q^n)
\end{array}$$

So, by universality, all we need to get our factorisation is a commutative square

$$\begin{array}{ccc}
\mathfrak{C}^{\square}(\square^0\otimes\square^{n-1})\sqcup\mathfrak{C}^{\square}(\square^1\otimes\square^{n-2})\sqcup\dots\sqcup\mathfrak{C}^{\square}(\square^{n-1}\otimes\square^0) & \xrightarrow{(\mathfrak{C}^{\square}(\partial_{1,1}),\mathfrak{C}^{\square}(\partial_{2,1}),\dots,\mathfrak{C}^{\square}(\partial_{n,1}))} & \mathfrak{C}^{\square}(\square^n)\\
\big\downarrow & & \big\downarrow{\scriptstyle\psi_n}\\
\mathfrak{C}^{\square}(\square^{n-1})\sqcup\mathfrak{C}^{\square}(\square^{n-2})\sqcup\dots\sqcup\mathfrak{C}^{\square}(\square^0) & \dashrightarrow & \mathfrak{C}^{\Delta}(\Delta^n)
\end{array}$$

ie for all $1\le i\le n$, we want a lift in the diagram

$$\begin{array}{ccccc}
\mathfrak{C}^{\square}(\square^{i-1}\otimes\square^{n-i}) & \xrightarrow{\mathfrak{C}^{\square}(\partial_{i,1})} & \mathfrak{C}^{\square}(\square^n) & \xrightarrow{\psi_n} & \mathfrak{C}^{\Delta}(\Delta^n)\\
\big\downarrow & & & \nearrow{\scriptstyle\gamma_{i,n}} & \\
\mathfrak{C}^{\square}(\square^{n-i}) & & & &
\end{array}$$

We first define a map $\tilde{\gamma}_{i,n}$ between the underlying poset-enriched categories, and then we will set $\gamma_{i,n}=N(\tilde{\gamma}_{i,n})$. Let $a\otimes a'$ be a vertex of $\square^{i-1}\otimes\square^{n-i}$. We note that $(\psi_n\circ\partial_{i,1})(a\otimes a')=\sup(a\cup\{i\}\cup a')$ is independent of a and we set $\tilde{\gamma}_{i,n}(a') = \sup(\{i\} \cup a')$. Let $a \otimes a'$ and $b \otimes b'$ be two objects of $\square^{i-1} \otimes \square^{n-i}$, with $a \preccurlyeq b$ and $a' \preccurlyeq b'$, and $(x_1, \dots, x_k; y_1, \dots, y_{k'})$ be an element of $\Sigma_{b \setminus a} \times \Sigma_{b' \setminus a'}$. We have

$$(\psi_n \circ \partial_{i,1})(a \otimes a'; b \otimes b')(x_1, \dots, x_k; y_1, \dots, y_{k'}) \\ = \{y_l \mid y_p < y_l \text{ for all } p < l\} \cap]\sup(\{i\} \cup a'), \sup(\{i\} \cup b')[,$$

so $\tilde{\gamma}_{i,n}$ is a well-defined morphism of posets, and hence $\gamma_{i,n} = N(\tilde{\gamma}_{i,n})$ provides the required lifting.

As a consequence, there is a well-defined morphism $\phi_n \colon \mathfrak{C}^{\square}(Q^n) \to \mathfrak{C}^{\Delta}(\Delta^n)$ in $s\mathcal{C}at$ for each $[n]$, which is a bijection on objects. It remains to show that it yields a natural transformation $\phi \colon \mathfrak{C}^{\square} \circ Q \Rightarrow \mathfrak{C}^{\Delta} \circ Y$. Namely, given $u \colon [n] \to [m]$ in Δ and denoting the induced map by $u_* \colon Q^n \to Q^m$, we have to prove the commutativity of the diagram

$$\begin{array}{ccc} \mathfrak{C}^{\square}(Q^n) & \xrightarrow{\mathfrak{C}^{\square}(u_*)} & \mathfrak{C}^{\square}(Q^m) \\ {\scriptstyle \phi_n}\downarrow & & \downarrow{\scriptstyle \phi_m} \\ \mathfrak{C}^{\Delta}(\Delta^n) & \xrightarrow{\mathfrak{C}^{\Delta}(u)} & \mathfrak{C}^{\Delta}(\Delta^m) \end{array}$$

Lemma 1.2.11 implies that it is commutative at the level of objects. It is enough to check it for u a face d_j or a degeneracy s_j (left to the reader). □

Remark 4.3.2 Note that $\mathfrak{C}^{\Delta} \circ Y \not\cong \mathfrak{C}^{\square} \circ Q$.

Proposition 4.3.3 *The natural transformation of Lemma 4.3.1 induces a natural DK-equivalence*

$$\phi \colon \mathfrak{C}^{\square} \circ Q \Rightarrow \mathfrak{C}^{\Delta}.$$

Proof We follow closely the proof of [Doherty et al. 2024, Proposition 6.21]. Every simplicial set S is a colimit of representables, and the functors $\mathfrak{C}^{\square} \circ Q$ and $\mathfrak{C}^{\Delta}$ are left adjoint, hence preserve colimits. It follows that we can upgrade the natural transformation of Lemma 4.3.1 as $\phi \colon \mathfrak{C}^{\square} \circ Q \Rightarrow \mathfrak{C}^{\Delta}$, which is componentwise a bijection on objects by Lemma 3.1.1. We prove first that if S is k–skeletal for some k, then ϕ_S is a DK-equivalence. If $k = 0$ or $k = 1$, this is an isomorphism of simplicial categories. Assume it is true for every $k < n$. The functors $\mathfrak{C}^{\square} \circ Q$ and $\mathfrak{C}^{\Delta}$ preserve cofibrations, as well as Joyal weak equivalences (since every object of $s\mathcal{S}et$ is cofibrant). Given any $1 \le i \le n-1$, the cofibration $\Lambda^n_i \to \Delta^n$ is a Joyal weak equivalence of simplicial sets and Λ^n_i is $(n-1)$–skeletal, so that ϕ_{Δ^n} is a DK-equivalence in $s\mathcal{C}at$, by the two-out-of-three property. Given an $(n-1)$–skeletal simplicial set X, let us consider a pushout diagram of the form $\mathcal{X}$:

$$\begin{array}{ccc} \partial\Delta^n & \longrightarrow & X \\ \downarrow & & \downarrow \\ \Delta^n & \longrightarrow & X' \end{array}$$

The left vertical arrow is a cofibration. Applying our functors, we get a cube diagram, where the front face is $(\mathfrak{C}^{\square} \circ Q)(\mathcal{X})$, and the back face is $\mathfrak{C}^{\Delta}(\mathcal{X})$. Both are pushout diagrams, and their left vertical arrow is a cofibration. By the induction hypothesis and by the proof above, the morphisms $\phi_{\partial\Delta^n}$, ϕ_{Δ^n} and ϕ_X

are DK-equivalences. We can thus apply [Hirschhorn 2003, Proposition 15.10.10] and conclude that $\phi_{X'}$ is a DK-equivalence. Since an n–skeletal simplicial set is obtained by transfinite composition of pushouts from its $(n-1)$–skeleton, we obtain that, for any n–skeletal simplicial set X, ϕ_X is a Dwyer–Kan equivalence. Finally, we observe that any simplicial set S is a sequential colimit of cofibrations (the family of inclusions of the n–skeleton into the $(n+1)$–skeleton), preserved by the two functors and thus entailing that ϕ_S is a DK-equivalence, by [Hirschhorn 2003, Proposition 15.10.12]. □

Corollary 4.3.4 *The adjunction $\mathfrak{C}^{\square} \dashv N^{\square}$ is a Quillen equivalence.*

Proof The proposition above implies that the total left derived functor $\mathbb{L}(\mathfrak{C}^{\square} \circ Q)$ is isomorphic to $\mathbb{L}(\mathfrak{C}^{\Delta})$. But $\mathfrak{C}^{\Delta}$ is a Quillen equivalence (Theorem 1.1.3); hence $\mathfrak{C}^{\square} \circ Q$ is also a Quillen equivalence. We conclude by Theorem 1.2.12 and the two-out-of-three property for Quillen equivalences. □

Appendix A Tools in category theory

In this section, we collect some categorical and enriched categorical tools that are needed in the paper.

A.1 Wedge sum and concatenation

Let $\mathcal{C}$ be a category with a distinguished object $*$. Let X be an object of $\mathcal{C}$. A point x in X is a map $x\colon * \to X$. We say also that X is pointed by x. If x is a point in X and y is a point in Y, we define the wedge sum $X \vee Y$ as the pushout (if it exists) of the diagram

$$X \xleftarrow{x} * \xrightarrow{y} Y.$$

Example A.1.1 Taking the category of posets as $\mathcal{C}$, the terminal singleton poset as distinguished object, P a poset with a maximal element ω, Q a poset with a minimal element α, and pointing P and Q by ω and α respectively, the wedge sum $P \vee Q$ is the poset obtained by "placing Q to right of P": $P \vee Q$ as a set is the pushout in the category of sets, and every element of P is less than every element of Q. Note that $P \to P \vee Q$ is an embedding of posets.

We will consider a similar construction in $\mathcal{C}_{*,*} = * \sqcup * \downarrow \mathcal{C}$, the category of double pointed objects in $\mathcal{C}$. We will denote an object $(a,b)\colon * \sqcup * \to X$ in this category by $X_{a,b}$.

Definition A.1.2 Let $X_{a,b}$ and $Y_{u,v}$ be two objects of $\mathcal{C}_{*,*}$.

- The wedge sum $X \vee Y$ of the pointed sets $b\colon * \to X$ and $u\colon * \to Y$ is naturally double pointed by $(a,v)\colon * \sqcup * \to X \vee Y$.
- For $f\colon X_{a,b} \to X'_{a',b'}$ and $g\colon Y_{u,v} \to Y'_{u',v'}$, we denote by $f \vee g\colon (X \vee Y)_{a,v} \to (X' \vee Y')_{a',v'}$ the double pointed map induced by the universal property of the pushout and the natural maps $X' \to X' \vee Y'$ and $Y' \to X' \vee Y'$. It endows $\mathcal{C}_{*,*}$ with a monoidal structure, with unit $*$ (doubled pointed by itself). We call this product the *concatenation product*.

Let X be an object in $\mathcal{C}$ and let u, v and w be points in X. For any maps $f: S_{a,b} \to X_{u,v}$ and $g: T_{a',b'} \to X_{v,w}$ in $\mathcal{C}_{*,*}$, we write $f * g: (S \vee T)_{a,b'} \to X_{u,w}$ for the corresponding structure map out of the pushout.

A.2 P–shaped categories

We introduce the notion of P–shaped category, for P a poset.

Definition A.2.1 A *bounded poset* is a poset P having a least and greatest element denoted respectively by α_P and ω_P.

Remark A.2.2 Given two bounded posets P and Q, the poset $P \vee Q$ (see Example A.1.1), is also bounded, by α_P and ω_Q.

In this section we fix $(\mathcal{V}, \otimes, I)$ a symmetric monoidal category, with initial object denoted by $\varnothing$. We assume that $\varnothing \otimes X \cong \varnothing$ for all objects X in $\mathcal{V}$.

Definition A.2.3 Let P be a poset. A $\mathcal{V}$–enriched category $\mathcal{C}$ is *P–shaped* if

- the set of objects of $\mathcal{C}$ is in bijection with P,
- $\mathcal{C}(p, p) = I$ for all $p \in P$, and
- $\mathcal{C}(p, q) \neq \varnothing$ implies $p \leq q$ for all $p, q \in P$.

It is double pointed by $\alpha_P, \omega_P : * \to \mathcal{C}$, where $*$ denotes the $\mathcal{V}$–enriched category with one object and homset I.

Note that the condition imposed above is implicitly used in this definition. Suppose that $p < q \leq r$. The $\mathcal{V}$–enriched structure implies that there is a composition morphism $\mathcal{C}(r, p) \otimes \mathcal{C}(q, r) \to \mathcal{C}(q, p)$, which by the P–shape axioms is a morphism $\varnothing \otimes \mathcal{C}(q, r) \to \varnothing$, which we take to be the identity up to the identification $\varnothing \otimes \mathcal{C}(q, r) \cong \varnothing$.

Example A.2.4 Any poset P gives rise to a $\mathcal{V}$–enriched category $\widehat{P}$: the objects are the elements of P, and for every p and q in P we have $\widehat{P}(p, q) = I$ if $p \leq q$, and $\widehat{P}(p, q) = \varnothing$ otherwise. Hence $\widehat{P}$ is P–shaped.

The next proposition computes the wedge sum $\mathcal{C} \vee \mathcal{D}$ of a P–shaped $\mathcal{V}$–category $\mathcal{C}$ and a Q–shaped $\mathcal{V}$–category $\mathcal{D}$ along ω_P and α_Q.

Proposition A.2.5 *Let P and Q be two bounded posets. Let $\mathcal{C}$ be a P–shaped $\mathcal{V}$–category and $\mathcal{D}$ be a Q–shaped $\mathcal{V}$–category. The wedge sum $\mathcal{E} = \mathcal{C} \vee \mathcal{D}$ along ω_P and α_Q, in the category of $\mathcal{V}$–categories, exists and is $(P \vee Q)$–shaped. It is described as follows*:

- *the set of objects of* $\mathcal{E}$ *is in bijection with* $P \vee Q$, *so we can identify objects of* $\mathcal{E}$ *with elements of* $P \vee Q$,
- $\mathcal{E}(x,y) = \begin{cases} \mathcal{C}(x,y) & \text{if } x,y \in P, \\ \mathcal{D}(x,y) & \text{if } x,y \in Q, \\ \mathcal{D}(\alpha_Q, y) \otimes \mathcal{C}(x, \omega_P) & \text{if } x \in P \text{ and } y \in Q, \\ \varnothing & \text{otherwise,} \end{cases}$

and the composition is the obvious one.

Proof Note that $\mathcal{C}(\omega_P, \omega_P) = I = \mathcal{D}(\alpha_Q, \alpha_Q)$ implies that the definition above is consistent. Note also that $\mathcal{E}$ is $(P \vee Q)$–shaped. We prove that $\mathcal{E}$ satisfies the required universal property in the category of $\mathcal{V}$–categories, taking $*$ to be the $\mathcal{V}$–category with one object, with homset I. Denote by $\iota_{\mathcal{C}} \colon \mathcal{C} \to \mathcal{E}$ the natural morphism and similarly for $\iota_{\mathcal{D}}$. Given a $\mathcal{V}$–category $\mathcal{F}$ and two $\mathcal{V}$–functors $F \colon \mathcal{C} \to \mathcal{F}$ and $G \colon \mathcal{D} \to \mathcal{F}$ satisfying $F(\omega_P) = G(\alpha_Q)$, we prove that there exists a unique $\mathcal{V}$–functor $H \colon \mathcal{E} \to \mathcal{F}$ such that $H\iota_{\mathcal{C}} = F$ and $H\iota_{\mathcal{D}} = G$. The functor H is clearly uniquely defined on objects, and on most of the morphisms. For $x \in P \setminus \{\omega_P\}$ and $y \in Q \setminus \{\alpha_Q\}$, we (have to) define H as the composite

$$\mathcal{D}(\alpha_Q, y) \otimes \mathcal{C}(x, \omega_P) \xrightarrow{G \otimes F} F(G(\alpha_Q), G(y)) \otimes F(F(x), F(\omega_P)) \xrightarrow{\circ} \mathcal{F}(F(x), G(y)),$$

and we check easily that this defines an enriched functor. □

Corollary A.2.6 *Let* P *and* Q *be two bounded posets. If* $\mathcal{C}$ *is a* P*–shaped poset-category and* $\mathcal{D}$ *is a* Q*–shaped poset-category, then* $N(\mathcal{C}) \vee N(\mathcal{D})$ *is isomorphic to* $N(\mathcal{C} \vee \mathcal{D})$ *as simplicial categories.*

Proof It follows directly from the explicit description of $\mathcal{C} \vee \mathcal{D}$ in Proposition A.2.5 and from the fact that $N(A \times B) \cong N(A) \times N(B)$ for any posets A and B. □

Proposition A.2.7 *Let* $\mathcal{C}$ *be a* P*–shaped simplicial category, with* P *a bounded poset. Let*

$$\varphi \colon \mathcal{C}(\alpha_P, \omega_P) \to Y$$

be a morphism of simplicial sets. Denote by $\mathcal{D}$ *the colimit of the pushout diagram*

$$\Sigma Y \xleftarrow{\Sigma\varphi} \Sigma\mathcal{C}(\alpha_P, \omega_P) \to \mathcal{C},$$

where the right arrow is the counit of the adjunction of Proposition 1.1.1. The simplicial category $\mathcal{D}$ *is* P*–shaped and has simplicial sets of morphisms* $\mathcal{C}(a,b)$ *if* $(a,b) \neq (\alpha_P, \omega_P)$, *and* Y *if* $(a,b) = (\alpha_P, \omega_P)$.

For $a \le b \le c$, *the composition* $\mathcal{D}(b,c) \otimes \mathcal{D}(a,b) \to \mathcal{D}(a,c)$ *is that of* $\mathcal{C}$ *if* $(a,c) \neq (\alpha_P, \omega_P)$ *and is the composition in* $\mathcal{C}$ *followed by* φ *if* $(a,c) = (\alpha_P, \omega_P)$.

Proof Note that since the functor $\mathrm{Ob} \colon s\mathcal{C}at \to \mathcal{S}et$ is cocontinuous, the set of objects of the colimit of the diagram is in bijection with P. Moreover the description of the hom sets of $\mathcal{D}$ shows that the category $\mathcal{D}$ is P–shaped. We check that the proposed simplicial category $\mathcal{D}$ satisfies the universal property of pushouts. Let $\mathcal{E}$ be a simplicial category with $F\colon \mathcal{C} \to \mathcal{E}$ and $g\colon Y \to \mathcal{E}(F(\alpha_P), F(\omega_P))$ such that $g \circ \varphi = F_{\alpha_P,\omega_P}$. There is a unique way to build $H\colon \mathcal{D} \to \mathcal{E}$ making the diagrams commute. It coincides with F on objects as well as on morphisms $\mathcal{D}(a,b) \to \mathcal{E}(F(a), F(b))$ for $(a,b) \neq (\alpha_P, \omega_P)$. In addition, we have $H_{\alpha_P,\omega_P} = g$. It is easy to check that H is a morphism of simplicial categories, using the equation relating g and F. □

Remark A.2.8 Proposition A.2.7 holds more widely for all P–shaped $\mathcal{V}$–categories. All it uses is the adjunction $\Sigma \vdash \mathrm{Hom}$ of Proposition 1.1.1. But the latter (as a mere adjunction) holds in the general setting of bipointed P–shaped $\mathcal{V}$–categories. Moreover, when restricted to [1]–shaped $\mathcal{V}$–categories, the adjunction is an equivalence.

Appendix B Combinatorics

B.1 The category SubNeck(T), with T a necklace

Proposition B.1.1 *The category* SubNeck($\square^n$) *is a poset. It is in bijection with the poset of ordered partitions of* $\{1,\dots,n\}$, *where the order* $A \leq_r B$ *is the refinement inverse order, defined as the reflexive and transitive closure of*

$$(A_1;\dots;A_k) \leq_r (A_1;\dots;A_i \cup A_{i+1};\dots;A_k).$$

In particular, it has a greatest element, the partition with 1 *block* ($\{1,\dots,n\}$), *and the set of minimal elements is naturally in bijection with the symmetric group* Σ_n. *This poset admits all least upper bounds.*

Proof (See also [Ziemiański 2017, Section 7; 2020, Section 7], where elements of SubNeck($\square^n$) are identified with cube chains.) As seen in Lemma 2.1.5, an object $\varphi\colon \square^{n_1} \vee \cdots \square^{n_k} \hookrightarrow \square^n$ is determined

- by a sequence $s_\varphi = (\alpha = a_0 \prec a_1 \prec \cdots \prec a_k = \omega)$, with $a_i \in (\square^n)_0$, $d(a_{i-1}, a_i) = n_i$ and $n_1 + \cdots + n_k = n$, or, equivalently,
- by an ordered partition of $\{1,\dots,n\}$: setting $A_i = a_i \setminus a_{i-1}$, we get $A_\varphi := (A_1;\dots;A_k)$.

We denote the set $\{a_0,\dots,a_k\}$ arising from a sequence s as above by $\bar{s}$. The following easy verifications are left to the reader:

- Given φ and ψ in SubNeck($\square^n$), there is a morphism from φ to ψ if and only if $\bar{s}_\psi \subseteq \bar{s}_\varphi$, and this morphism is unique.
- We have $\bar{s}_\psi \subseteq \bar{s}_\varphi$ if and only if $A_\varphi \leq_r A_\psi$.

In particular, if A is a set of sequences $\{s_1,\dots,s_l\}$ then its least upper bound is the sequence s associated to $\bigcap_{i=1}^{l} \bar{s}_i$. □

The following corollary is a direct consequence of the previous proposition and of Proposition 2.1.2.

Corollary B.1.2 *Let $T = \square^{n_1} \vee \cdots \vee \square^{n_k}$ be a necklace. The category* $\mathrm{SubNeck}(T)$ *is a poset, whose poset relation is denoted by $\leq_r$. It is the product of the categories* $\mathrm{SubNeck}(\square^{n_i})$. *In particular, it has a greatest element and admits all least upper bounds.*

Definition B.1.3 Let $(P, \leq)$ be a poset. A subset A of P is *downward closed* if for all $y \leq x$ in P, $x \in A$ implies $y \in A$. It is *upward closed* if it is downward closed in the opposite poset of P.

Lemma B.1.4 *Let $(P, \leq)$ be a poset and A and B upward closed subsets of P. Then*

$$N(A \cup B) \cong \operatorname{colim}(N(A) \hookleftarrow N(A \cap B) \hookrightarrow N(B)).$$

Proof We claim that

$$\begin{array}{ccc} N(A \cap B) & \longrightarrow & N(A) \\ \downarrow & & \downarrow \\ N(B) & \longrightarrow & N(A \cup B) \end{array}$$

is a pushout diagram in the category $s\mathcal{S}et$. Since colimits in $s\mathcal{S}et$ are computed dimensionwise, it is enough to prove this claim for n–simplices. An element in $N_n(A \cup B)$ is a sequence $x = (x_0 \leq \cdots \leq x_n)$ with $x_i \in A \cup B$ for all i. Since A and B are upward closed subsets of P, if $x_0 \in A$, then $x_i \in A$ for all i, and similarly if $x_0 \in B$. Let X be a simplicial set and $f_A \colon N(A) \to X$ and $f_B \colon N(B) \to X$ be two morphisms of simplicial sets that coincide on $N(A \cap B)$. We define $g \colon N_n(A \cup B) \to X_n$ as $g(x) = f_A(x)$ if $x_0 \in A$ and $g(x) = f_B(x)$ if $x_0 \in B$. It yields a well-defined map of simplicial sets, since if $x_0 \in A \cap B$ then $x_i \in A \cap B$ for all i, and $f_A = f_B$ on $N(A \cap B)$. □

Lemma B.1.5 *Let T be a necklace. If $\mathcal{A} \subseteq \mathrm{SubNeck}(T)$ is downward closed* (*for the order $\leq_r$*), *then the canonical morphism $\operatorname{colim}_{U \in \mathcal{A}} \mathfrak{C}^{\square}_t(U) \to \mathfrak{C}^{\square}_t(T)$ is a monomorphism of simplicial sets.*

Proof We only need to examine the situation of two n–simplices u and v coming from different subnecklaces U and V in $\mathcal{A}$ and whose images in $\mathfrak{C}^{\square}_t(T)$ are identified:

$$\begin{array}{ccc} \Delta^n & \xrightarrow{u} & \mathfrak{C}^{\square}_t(U) \\ {\scriptstyle v}\downarrow & & \downarrow \\ \mathfrak{C}^{\square}_t(V) & \longrightarrow & \mathfrak{C}^{\square}_t(T) \end{array}$$

The composite map $\varphi \colon \Delta^n \to \mathfrak{C}^{\square}_t(T)$ gives a set A_φ of $n+1$ paths of T (by Theorem 3.1.4) with values both in U and V. Let W be the upper bound of A_φ in $\mathrm{SubNeck}(T)$, provided by Corollary B.1.2. Since U and V are upper bounds of A_φ, we have $W \leq_r U$ and $W \leq_r V$ in $\mathrm{SubNeck}(T)$. Hence there is a factorisation in the diagram

$$\begin{array}{ccccc} \Delta^n & & \xrightarrow{\quad\quad u \quad\quad} & & \mathfrak{C}^{\square}_t(U) \\ & \overset{w}{\searrow} & & \nearrow & \\ {\scriptstyle v}\downarrow & & \mathfrak{C}^{\square}_t(W) & & \downarrow \\ & \swarrow & & \searrow & \\ \mathfrak{C}^{\square}_t(V) & & \xrightarrow{\quad\quad\quad\quad} & & \mathfrak{C}^{\square}_t(T) \end{array}$$

Moreover, $W \in \mathcal{A}$ since $\mathcal{A}$ is downward closed. Thus the diagram says that u, v and w are identified in the colimit, which completes the proof. ◻

B.2 On the homotopy type of $\mathrm{SubNeck}(\sqcap^n_{i,\epsilon})$

To simplify the notation in this section, for $n \geq 1$, we will denote by P_n the poset of ordered partitions of $\{1, \dots, n+1\}$, that is, $P_n = \mathrm{SubNeck}(\square^{n+1})$. Similarly, we set $\partial P_n = \mathrm{SubNeck}(\partial\square^{n+1})$ and $\sqcap P_n = \mathrm{SubNeck}(\sqcap^{n+1}_{n+1,0})$ (see Definition 3.3.1). Note that

$$\partial P_n = P_n \setminus (\{1, \dots, n+1\}) \quad \text{and} \quad \sqcap P_n = \partial P_n \setminus (\{1, \dots, n\}; \{n+1\}).$$

The nerve of P_n is contractible, since P_n has a greatest element.

The next proposition is what we need for Proposition 4.1.3.

Proposition B.2.1 *For every $n \geq 2$, the nerve of $\mathrm{SubNeck}(\sqcap^n_{i,\epsilon})$ is contractible.*

Proof We first prove that for $n \geq 1$, the nerve of $\sqcap P_n$ is contractible. For $n = 1$, the poset $\sqcap P_1$ is a singleton, namely the ordered partition $(\{2\}; \{1\})$, hence is contractible. Assume $n \geq 2$. Let us fix some notation:

- For an ordered partition $x = (A_1; \dots; A_k) \in P_n$, and $0 \leq l \leq k$, we set
$$m_l(x) := \#(A_1 \cup \dots \cup A_l),$$
with the convention that $m_0(x) = 0$. Note that $m_k(x) = n+1$.
- For an ordered partition $x = (A_1; \dots; A_k) \in P_n$ with $n+1 \in A_l$, we set
$$\alpha(x) := m_{l-1}(x) \text{ and } \beta(x) := m_l(x).$$
- We have $\partial P_n = T_0 \cup T_1 \cup \dots \cup T_n$ with
$$T_i := \{x \in \partial P_n \mid \alpha(x) \leq i < \beta(x)\}.$$
- For every $0 \leq i \leq n-1$ we set
$$T_{i,i+1} := T_i \cap T_{i+1} = (T_0 \cup \dots \cup T_i) \cap T_{i+1}.$$

We note that $\sqcap P_n = T_0 \cup T_1 \cup \dots T_{n-1} \cup T'_n$ with $T'_n = T_n \cap (\sqcap P_n)$. We also note (by a simple case analysis) that $T_0, \dots, T_{n-1}$ and T'_n are upward closed, allowing us to use Lemma B.1.4 repeatedly and get

$$N(\sqcap P_n) \cong \operatorname{colim} \left(\begin{array}{ccccccccc} N(T_0) & & N(T_1) & \cdots & \cdots & \cdots & N(T_{n-1}) & & N(T'_n) \\ & \nwarrow \quad \nearrow & & \nwarrow & & \nearrow & & \nwarrow \quad \nearrow & \\ & N(T_{0,1}) & & & \cdots & & & N(T_{n-1,n}) & \end{array} \right).$$

We claim that the colimit of this diagram is Kan–Quillen equivalent to its homotopy colimit. Indeed, let D be the underlying category of the diagram, which has objects x_i for $0 \leq i \leq n$ and $y_{i,i+1}$ for $0 \leq i < n$

and morphisms from $y_{i,i+1}$ to x_i and x_{i+1}. We endow D with the following Reedy structure: x_i has degree $2i$ and $y_{i,i+1}$ has degree $2i+1$. Then D has fibrant constants. This follows from [Hirschhorn 2003, Proposition 15.10.2], noting that, for every object α of D, the matching category at D is either empty or the one point category. Moreover, the diagram above is Reedy cofibrant; hence the claim above follows from [Hirschhorn 2003, Proposition 19.9.1]. In consequence, we focus our attention on the homotopy type of $N(T_i)$ for $0 \le i < n$, $N(T'_n)$ and $N(T_{i,i+1})$ for $0 \le i < n$.

Let $\pi\colon P_n \to P_{n-1}$ be the poset morphism removing $(n+1)$ from the ordered partition. Note that $\pi(T_i) \subset \partial P_n$ for every $1 \le i \le n-1$ since

$$\pi^{-1}(\{1,\dots,n\}) = \big\{(\{1,\dots,n+1\}), (\{n+1\};\{1,\dots,n\}), (\{1,\dots,n\};\{n+1\})\big\},$$

and since none of the elements in this set lie in T_i for $1 \le i \le n-1$. Note that, given a partition $x = (A_1;\dots;A_k)$ of P_{n-1} and $0 \le i \le n-1$, there exists a unique $l \in \{0,\dots,k-1\}$ such that $m_l(x) \le i < m_{l+1}(x)$. We leave it to the reader to check the following facts:

- For $0 \le i < n$, the induced map $T_{i,i+1} \xrightarrow{\pi} \partial P_{n-1}$ is an isomorphism of posets with inverse
$$x = (A_1;\dots;A_k) \in \partial P_{n-1} \mapsto (A_1;\dots;A_{l+1} \cup \{n+1\};\dots;A_k),$$
where $m_l(x) \le i < m_{l+1}(x)$.
- For $0 < i < n$, the map $T_i \xrightarrow{\pi} \partial P_{n-1}$ is an adjunction of posets with a section σ given for $x = (A_1;\dots;A_k) \in \partial P_{n-1}$ by
$$\sigma(x) = \begin{cases} (A_1;\dots,A_l;\{n+1\};A_{l+1};\dots;A_k) & \text{if } m_l(x) = i, \\ (A_1;\dots;A_{l+1} \cup \{n+1\};\dots;A_k) & \text{if } m_l(x) < i < m_{l+1}(x). \end{cases}$$
(Indeed, $\pi \circ \sigma = \mathrm{id}$ and $\sigma \circ \pi \le \mathrm{id}$.)
- Similarly, the maps $T_0 \xrightarrow{\pi} P_{n-1}$ and $T'_n \xrightarrow{\pi} \partial P_{n-1}$ are adjunctions of posets with the respective sections
$$(A_1;\dots;A_k) \in P_{n-1} \mapsto (\{n+1\};A_1;\dots;A_k),$$
$$(A_1;\dots;A_k) \in \partial P_{n-1} \mapsto (A_1;\dots;A_k;\{n+1\}).$$

Putting everything together, and using the fact that an adjunction of posets gives rise to a homotopy equivalence, and hence to a Kan–Quillen equivalence between their nerves, we have

$$N(\sqcap P_n) \sim \operatorname{hocolim}\left(\begin{array}{ccccccc} N(P_{n-1}) & & N(\partial P_{n-1}) & \cdots & N(\partial P_{n-1}) & & N(\partial P_{n-1}) \\ & \nwarrow & & & & & \\ & & N(\partial P_{n-1}) \;=\; & \cdots & & = N(\partial P_{n-1}) = & \end{array}\right)$$
$$\sim \operatorname{hocolim}\begin{pmatrix} N(P_{n-1}) \\ \uparrow \\ N(\partial P_{n-1}) \end{pmatrix} \sim \operatorname{colim}\begin{pmatrix} N(P_{n-1}) \\ \uparrow \\ N(\partial P_{n-1}) \end{pmatrix} \sim N(P_{n-1}).$$

Hence the nerve of $\sqcap P_n$ is contractible.

Finally, we have

$$\mathrm{SubNeck}(\sqcap^{n+1}_{i,0}) = \partial P_n \setminus (\{1,\dots,n+1\}\setminus\{i\};\{i\}),$$
$$\mathrm{SubNeck}(\sqcap^{n+1}_{i,1}) = \partial P_n \setminus (\{i\};\{1,\dots,n+1\}\setminus\{i\}),$$

and the proof of the contractibility of these posets is similar to that of $\sqcap P_n$. □

Remark B.2.2 One can find in [Baues 1980] a geometric interpretation linking cellular strings on the cubes and the permutohedra, the original idea being attributed to Milgram. Related results relative to P_n and ∂P_n can be found in, say, [Ziemiański 2017] or in [Ziegler 1995] (where they are put to use to establish connections between cubical sets and higher dimensional automata, through directed path spaces): the geometric realisations of the ordered partition posets are the permutohedra, which are homeomorphic to a ball. Moreover, the pair $(|N(P_n)|, |N(\partial P_n)|)$ is homeomorphic to the pair (D^n, S^{n-1}).

References

[Baues 1980] **H J Baues**, *Geometry of loop spaces and the cobar construction*, Mem. Amer. Math. Soc. 230, Amer. Math. Soc., Providence, RI (1980) MR Zbl

[Bergner 2018] **J E Bergner**, *The homotopy theory of* $(\infty,1)$*–categories*, London Mathematical Society Student Texts 90, Cambridge Univ. Press (2018) MR Zbl

[Björner 1984] **A Björner**, *Orderings of Coxeter groups*, from "Combinatorics and algebra" (C Greene, editor), Contemp. Math. 34, Amer. Math. Soc., Providence, RI (1984) 175–195 MR Zbl

[Boardman and Vogt 1973] **J M Boardman**, **R M Vogt**, *Homotopy invariant algebraic structures on topological spaces*, Lecture Notes in Math. 347, Springer (1973) MR Zbl

[Cisinski 2014] **D-C Cisinski**, *Univalent universes for elegant models of homotopy types*, preprint (2014) arXiv 1406.0058

[Cohen et al. 2018] **C Cohen**, **T Coquand**, **S Huber**, **A Mörtberg**, *Cubical type theory: a constructive interpretation of the univalence axiom*, from "21st international conference on types for proofs and programs" (T Uustalu, editor), Leibniz Int. Proc. Inform. 69, Schloss Dagstuhl, Wadern (2018) art. id. 5 MR Zbl

[Doherty et al. 2024] **B Doherty**, **K Kapulkin**, **Z Lindsey**, **C Sattler**, *Cubical models of* $(\infty,1)$*–categories*, Mem. Amer. Math. Soc. 1484, Amer. Math. Soc., Providence, RI (2024) MR

[Dugger and Spivak 2011a] **D Dugger**, **D I Spivak**, *Rigidification of quasi-categories*, Algebr. Geom. Topol. 11 (2011) 225–261 MR Zbl

[Dugger and Spivak 2011b] **D Dugger**, **D I Spivak**, *Mapping spaces in quasi-categories*, Algebr. Geom. Topol. 11 (2011) 263–325 MR Zbl

[Fajstrup et al. 2016] **L Fajstrup**, **E Goubault**, **E Haucourt**, **S Mimram**, **M Raussen**, *Directed algebraic topology and concurrency*, Springer (2016) MR Zbl

[Gaucher 2008] **P Gaucher**, *Towards a homotopy theory of process algebra*, Homology Homotopy Appl. 10 (2008) 353–388 MR Zbl

[Grandis and Mauri 2003] **M Grandis**, **L Mauri**, *Cubical sets and their site*, Theory Appl. Categ. 11 (2003) 185–211 MR Zbl

[Hirschhorn 2003] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[Joyal 2007] **A Joyal**, *Quasi-categories vs simplicial categories*, unpublished manuscript (2007) Available at `https://www.math.uchicago.edu/~may/IMA/Incoming/Joyal/QvsDJan9(2007).pdf`

[Joyal 2008] **A Joyal**, *The theory of quasi-categories and its applications*, *II*, lecture notes, CRM (2008) Available at `http://mat.uab.cat/~kock/crm/hocat/advanced-course/Quadern45-2.pdf`

[Kapulkin et al. 2019] **K Kapulkin**, **Z Lindsey**, **L Z Wong**, *A co-reflection of cubical sets into simplicial sets with applications to model structures*, New York J. Math. 25 (2019) 627–641 MR Zbl

[Lurie 2009] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR Zbl

[Mac Lane 1998] **S Mac Lane**, *Categories for the working mathematician*, 2nd edition, Graduate Texts in Math. 5, Springer (1998) MR Zbl

[Maltsiniotis 2009] **G Maltsiniotis**, *La catégorie cubique avec connexions est une catégorie test stricte*, Homology Homotopy Appl. 11 (2009) 309–326 MR Zbl

[Pratt 1991] **V Pratt**, *Modeling concurrency with geometry*, from "Proceedings of the 18th ACM SIGPLAN-SIGACT symposium on principles of programming languages", Association for Computing Machinery, New York (1991) 311–322

[Riehl 2014] **E Riehl**, *Categorical homotopy theory*, New Mathematical Monographs 24, Cambridge Univ. Press (2014) MR Zbl

[Serre 1951] **J-P Serre**, *Homologie singulière des espaces fibrés: applications*, Ann. of Math. 54 (1951) 425–505 MR Zbl

[Wong 2020] **L Z Wong**, *Cubes with connections from algebraic weak factorization systems*, presentation slides (2020) Available at `http://sheaves.github.io/slides/cylinders.pdf`

[Ziegler 1995] **G M Ziegler**, *Lectures on polytopes*, Graduate Texts in Math. 152, Springer (1995) MR Zbl

[Ziemiański 2017] **K Ziemiański**, *Spaces of directed paths on pre-cubical sets*, Appl. Algebra Engrg. Comm. Comput. 28 (2017) 497–525 MR Zbl

[Ziemiański 2020] **K Ziemiański**, *Spaces of directed paths on pre-cubical sets, II*, J. Appl. Comput. Topol. 4 (2020) 45–78 MR Zbl

CNRS, IRIF, Université Paris Cité
Paris, France

Université Paris Cité, Institut de Mathématiques de Jussieu–Paris Rive Gauche, CNRS, Sorbonne Université, DMA, ENS-PSL
Paris, France

Department of Mathematics, Stockholm University
Stockholm, Sweden

`curien@irif.fr, muriel.livernet@imj-prg.fr, gabriel.saadia@math.su.se`

Received: 26 January 2023 Revised: 12 June 2023

Algebraic & Geometric Topology **24**:5 (2024) 2889–2932
DOI: 10.2140/agt.2024.24.2889
Published: 19 August 2024

Tautological characteristic classes, I

JAN DYMARA
TADEUSZ JANUSZKIEWICZ

We discuss the formalism of tautological characteristic classes of flat bundles. Applied to $\mathrm{PSL}(2, K)$, it yields the Witt class of Nekovář. Applied to $\mathrm{PGL}_+(2n, K)$, the general linear groups with positive determinant over an arbitrary ordered field, it yields (a generalization of) the Euler class.

20G10; 20J06

Introduction

Take a chain complex C_* and fix the degree k. The identity map $C_k \to C_k$ can be viewed as a cochain of degree k, with values/coefficients in C_k. Usually it is not a cocycle, but we can force it to be, dividing by boundaries, changing coefficients to C_k/B_k. Denote the resulting cocycle by T.

Now suppose that a group G acts on C_*. Clearly T is G–equivariant, in other words it is a cocycle with twisted coefficients. We can force T to be a constant coefficients cocycle simply dividing the coefficients further down to the biggest quotient of the G–module C_k/B_k on which G acts trivially, called coinvariants of G. Denote the resulting image of T by τ.

Besides producing an untwisted cocycle, this construction has an additional crucial advantage: the modules C_k, or even C_k/B_k, are usually very big, while the coinvariants $(C_k/B_k)_G$ are much smaller and sometimes manageable.

It is of interest to go halfway in this procedure: fix a (large) normal subgroup N of G, and take coinvariants $(C_k/B_k)_N$. Then T becomes a (slightly twisted by an action of G/N) cocycle taking values in (sometime still manageable, but bigger) module of N–coinvariants.

One has every reason to expect that this purely algebraic construction has nice functorial properties, and that it carries a significant amount of information about C_* as a G–module. Theorem 1.5 spells out the most natural form of functoriality.

This algebraic construction needs an input. For us such an input comes from a geometry (or, as some will undoubtedly insist, algebra), namely from the complex of geometric configurations. One takes a homogeneous space G/H (for example a projective space over an arbitrary field K) and builds a simplicial complex whose simplices are n–tuples of points "in general position". The notion of general position that we use depends on the situation, and is discussed separately in each case, but the underlying idea is uniform. In all cases the simplicial complexes that we consider have an additional crucial *star property* which is discussed in Section 2. The star property makes the simplicial complex contractible in a strong, geometric sense. The chain complex is just the complex of (alternating or ordered) simplicial chains.

In order to construct characteristic classes of flat G–bundles, we have to address the problem that the G–action on the space of configurations is not free. But this is done in a standard way, by "Borel construction", that we execute on the chain level. We end up with cocycles living in a cochain complex computing group cohomology of G (seen as a discrete group). The star property implies almost immediately that the cohomology class of the cocycle is bounded (see Theorem 3.1). The boundedness of tautological classes is taken with respect to the natural seminorm on the coefficient group. One should keep in mind that this has fairly different overtones from the usual $\mathbb{R}$–coefficients bounded cohomology of Gromov [1982]; compare for example Ghys [1987].

The first interesting case when this construction produces something valuable is that of $\mathrm{PGL}(2, K)$ acting on the projective line. This has been studied by Nekovář, who defined and studied the "Witt class" for $\mathrm{PSL}(2, K)$, with coefficients in the Witt ring of quadratic forms over K. It is an amazing fact that the four-term Witt relation $[a]+[b]=[a+b]+[ab(a+b)]$ is indeed the cocycle relation in the complex of projective point configurations. We review this in detail in Section 7.

The main results of the present paper concern the construction and study of the "Euler class for flat $\mathrm{PGL}(n, K)$–bundles" in the case where K is an arbitrary ordered field and n is even. This class is constructed using the general strategy outlined above. We take the $\mathrm{PGL}(n, K)$–action on the simplicial complex of generic configurations of points in $\mathbb{P}^{n-1}(K)$, the induced action on C_n/B_n, and then we take coinvariants with respect to the group $\mathrm{PGL}_+(n, K)$ of maps with *positive* determinant. (Note that coinvariants with respect to the full projective group are trivial, while coinvariants with respect to $\mathrm{PSL}(n, K)$ are too large for us to handle — for $\mathrm{PGL}_+(n, K)$ we have a nice answer.) The resulting tautological class eu is (an analogue of) the Euler class — for flat $\mathrm{PGL}(n, K)$–bundles. It is twisted by the homomorphism to $\mathbb{Z}/2$ whose kernel consists of maps with positive determinant. The coefficients are $\mathbb{Z}$ for n even and trivial for n odd (see Theorem 8.1).

One can run a parallel construction starting from the $\mathrm{GL}_+(n, K)$–action on the *positive* projective space $\mathbb{P}^{n-1}_+(K)$. The resulting class eu_+ has coefficients in a free abelian group of rank $\lfloor n/2 \rfloor + 1$ (see Theorem 8.1; admittedly, the computation here is somewhat heavy). Consequently, eu_+ can be split into components eu_k that are cohomology classes with $\mathbb{Z}$ coefficients.

We prove several results about the Euler classes eu and eu_+. Theorems 9.1 and 10.1 explain the relation between various components of eu_+. Theorem 11.1 gives a clean formula for the Euler class of a cross product of bundles, while Theorem 12.5 gives a cup product formula for the direct sum. In Section 13 we discuss functoriality. In particular, we relate eu and eu_+ in Theorem 13.1. We also compare the Euler and Witt classes for $\mathrm{PSL}(2, K)$–bundles in Theorem 13.4. Finally, in Theorem 13.6, using the cross-product formula, we show nontriviality of our Euler classes in every even dimension.

Further characteristic classes and more applications are postponed to subsequent papers.

Tautological classes with coefficients in C_k/B_k were defined in a forgotten paper of James Dugundji [1958], where he also proved some results of general nature, like functoriality and universality. The paper was forgotten, probably because the results did not help with actual calculations: modules C_k/B_k are usually very big and unmanageable. We discovered Dugundji's paper when we were already well into our project. Our initial inspiration came from the papers of Nekovář [1990] and Kramer and Tent [2010], where the idea of passing to G–coinvariants is present. With a grain of salt, one may say that the Witt and Maslov classes are constructed in these papers in the tautological way.

Reznikov [1997] noticed that for an ordered field K one has an "Euler class" for $\mathrm{PSL}(2, K)$ with $\mathbb{Z}$ coefficients. In fact, this class is (a multiple of) the image of the Witt class of Nekovář under the signature map from the Witt ring to $\mathbb{Z}$, given by the ordering of K.

The plan of the paper is as follows.

In Part I we discuss the general theory: Section 1 explains definitions and functoriality of tautological classes in a purely algebraic, abstract context; in Section 2 it is shown how actions on simplicial complexes can lead to examples, star-property is recalled, and a method of coefficient calculation for actions on simplicial complexes is described; Section 3 is about (automatic) boundedness of tautological classes; and Section 4 contains a simplicial counterpart of the process of representing classes of flat bundles by pullbacks of invariant forms via sections.

Part II is about GL(2): in Section 6 we discuss various actions of this group with a view towards investigating the corresponding tautological classes; in Section 7 the Witt group appears as the coefficient group coming from the general formalism applied to the homographic action on the projective line, and the tautological Witt class is defined.

In Part III we define Euler classes for the groups $\mathrm{PGL}(n, K)$ and $\mathrm{PGL}_+(n, K)$, where K is an arbitrary ordered field. In Section 8 actions of these groups on $\mathbb{P}^{n-1}(K)$ and on $\mathbb{P}^{n-1}_+(K)$ are used to define tautological Euler classes eu and eu_+; coefficients are calculated, and eu_+ is decomposed into a direct sum of classes eu_k (with coefficients in $\mathbb{Z}$). In Section 9 we establish a general relation between the classes eu_k, and in Section 10 we express all of them in terms of eu_0 in a weak sense using Smillie's argument. In Sections 11 and 12 we show some multiplicativity properties of eu_0. In Section 13 we further investigate relations between various Euler classes (and the Witt class); we also prove that all these classes are nontrivial (for n even).

Acknowledgements

We would like to acknowledge extensive discussions with Linus Kramer that were very helpful in the initial stages of this project.

Both authors were supported by Polish NCN grant UMO-2016/23/B/ST1/01556.

I Generalities

1 Algebraic tautological classes

Chain complexes Let $C_* = (C_n, \partial_n)$ be a chain complex of abelian groups. As usual, we put $Z_n = \ker \partial_n$ (cycles) and $B_n = \operatorname{im} \partial_n$ (boundaries). Let us fix an integer d and consider id_{C_d} as an element of $\mathrm{Hom}(C_d, C_d)$ — the d–cochain group of the complex $\mathrm{Hom}(C_*, C_d)$. This element is usually not a cocycle — yet, if we replace the coefficient group C_d by the quotient C_d/B_d, it becomes one.

Definition 1.1 Let C_* be a chain complex. The tautological cocycle $T^d_{C_*}$ is the d–cycle of the complex $\mathrm{Hom}(C_*, C_d/B_d)$ defined by the quotient map $C_d \to C_d/B_d$. The tautological class $\tau^d_{C_*}$ is the cohomology class of $T^d_{C_*}$ in $H^d(\mathrm{Hom}(C_*, C_d/B_d))$.

The cochain $T^d_{C_*}$ is indeed a cocycle:

$$\delta T^d_{C_*}(c) = T^d_{C_*}(\partial c) = \partial c + B_d = B_d.$$

Notice that $\tau^d_{C_*}$ is functorial, in the following way: Let $f : C_* \to K_*$ be a chain map. Then $f_d : C_d \to K_d$ induces a map $C_d/\ker \partial_d \to K_d/\ker \partial_d$, which in turn induces a map

$$f_*: H^d(\mathrm{Hom}(C_*, C_d/\ker \partial_d)) \to H^d(\mathrm{Hom}(C_*, K_d/\ker \partial_d)).$$

There is also the map $f^*: \mathrm{Hom}(K_*, K_d/\ker \partial_d) \to \mathrm{Hom}(C_*, K_d/\ker \partial_d)$ inducing

$$f^*: H^d(\mathrm{Hom}(K_*, K_d/\ker \partial_d)) \to H^d(\mathrm{Hom}(C_*, K_d/\ker \partial_d)).$$

Clearly, $f^*\tau^d_{K_*} = f_*\tau^d_{C_*}$; indeed, both these classes are represented by the same cocycle

$$C_d \ni c \mapsto f_d(c) + \ker \partial_d \in K_d/\ker \partial_d.$$

***G*–chain complexes** Now suppose that the complex C_* is a G–chain complex, ie it is acted upon by a group G, by chain maps. The group C_d/B_d has the induced G–module structure. The tautological cocycle $T^d_{C_*}: C_d \to C_d/B_d$ is a G–map.

Definition 1.2 Let C_* be a G–chain complex. The tautological class $\tau^d_{C_*,G} \in H^d(\mathrm{Hom}_G(C_*, C_d/B_d))$ (cohomology with twisted coefficients) is the cohomology class of $T^d_{C_*}$.

We have found out that the above class has also been defined and investigated in a forgotten paper of Dugundji [1958].

The G–module C_d/B_d is usually very big. To cut it down in size we will consider its coinvariants group $U_d = (C_d/B_d)_G$ — its largest G–trivial quotient. This group might be either too small to carry information or too big to extract information, yet in some cases it is nontrivial and manageable.

Definition 1.3 Let C_* be a G–chain complex. Let U_d (or $U_d(C_*)$) denote the coinvariants group $(C_d/B_d)_G$. The tautological class $\tau^d_{C_*/G} \in H^d(\mathrm{Hom}_G(C_*, U_d))$ is the cohomology class of $T^d_{C_*/G}$; the cocycle obtained by composing the tautological cocycle $T^d_{C_*}$ with the quotient map $C_d/B_d \to (C_d/B_d)_G$.

Remark (1) The functor of coinvariants is right-exact [Brown 1982, II, Section 2]. Therefore, $U_d = (C_d)_G/(B_d)_G$ (strictly speaking, we divide by the image — not necessarily injective — of $(B_d)_G$ in $(C_d)_G$). Moreover, $\partial\colon C_{d+1} \to C_d$ induces a map $\partial\colon (C_{d+1})_G \to (C_d)_G$, and U_d can also be described as $(C_d)_G/\partial((C_{d+1})_G)$.

(2) If N is a normal subgroup of G then there exists yet another, G/N–twisted tautological class τ in $H^d\big(\mathrm{Hom}_G(C_*, (C_d/B_d)_N)\big)$.

Let us discuss functoriality. Suppose that C_* is a G–complex and that K_* is an H–complex. Assume that $\phi\colon G \to H$ is a homomorphism and that $f\colon C_* \to K_*$ is a ϕ–equivariant chain map. The group $U_d(K_*)$ acquires a G–module structure via ϕ. We have two maps,

$$H^d\big(\mathrm{Hom}_H(K_*, U_d(K_*))\big) \xrightarrow{f^*} H^d\big(\mathrm{Hom}_G(C_*, U_d(K_*))\big) \xleftarrow{f_*} H^d\big(\mathrm{Hom}_G(C_*, U_d(C_*))\big),$$

the right one induced by the f–induced coefficient map $U_d(C_*) \to U_d(K_*)$. As before, it is straightforward to check that $f^*\tau^d_{K_*/H} = f_*\tau^d_{C_*/G}$ — both of these classes are represented by the cocycle $C_d \ni c \mapsto [f_d(c)] \in U_d(K_*)$.

Acyclic G–chain complexes Let us now assume that C_* is an acyclic G–chain complex. By this we mean that

(1) $C_n = 0$ for $n < 0$;

(2) C_* comes equipped with an augmentation map — a G–homomorphism $\epsilon\colon C_0 \to \mathbb{Z}$, where $\mathbb{Z}$ has the trivial G–module structure;

(3) the augmented complex

$$\cdots \to C_n \to C_{n-1} \to \cdots \to C_1 \to C_0 \xrightarrow{\epsilon} \mathbb{Z} \to 0 \to \cdots$$

is exact.

(In other words: C_* is a resolution of the trivial G–module $\mathbb{Z}$.)

The tautological class $\tau^d_{C_*/G}$ can be used to define a cohomology class of the group G, as follows. Let P_* be a projective resolution of the trivial G–module $\mathbb{Z}$. The cohomology groups $H^*(G, U_d)$ are defined as cohomology groups of the complex $\mathrm{Hom}_G(P_*, U_d)$ [Brown 1982, III, Section 1]. There exists a chain map of resolutions $\psi_{C_*}\colon P_* \to C_*$ (respecting augmentations, ie extending by identity on $\mathbb{Z}$ to a chain map of the augmented complexes). Moreover, ψ_{C_*} is unique up to chain homotopy [Brown 1982, I, Lemma 7.4].

Definition 1.4 Let C_* be an acyclic G–chain complex, P_* a projective resolution of the trivial G–module $\mathbb{Z}$, and $\psi_{C_*}\colon P_* \to C_*$ a chain map of resolutions. Let $\psi^*_{C_*}\colon H^d(\mathrm{Hom}_G(C_*, U_d)) \to H^d(G, U_d)$ be the map on cohomology induced by ψ_{C_*}. We define the tautological class

$$\tau^d_{G,C_*} = \psi^*_{C_*}(\tau^d_{C_*/G}) \in H^d(G, U_d).$$

These classes are functorial just as the previous ones:

Theorem 1.5 *Let C_* be an acyclic G–chain complex, K_* an acyclic H–chain complex, $\phi\colon G \to H$ a group homomorphism, and $f : C_* \to K_*$ a ϕ–equivariant chain map. Consider two maps*

$$H^d(H, U_d(K_*)) \xrightarrow{\phi^*} H^d(G, U_d(K_*)) \xleftarrow{f_*} H^d(G, U_d(C_*));$$

the right one induced by the f–induced coefficient map $U_d(C_*) \to U_d(K_*)$. Then

$$\phi^* \tau^d_{H,K_*} = f_* \tau^d_{G,C_*}.$$

Proof Consider the diagram

$$\begin{array}{ccccc} H^d\big(\mathrm{Hom}_H(K_*, U_d(K_*))\big) & \xrightarrow{f^*} & H^d\big(\mathrm{Hom}_G(C_*, U_d(K_*))\big) & \xleftarrow{f_*} & H^d\big(\mathrm{Hom}_G(C_*, U_d(C_*))\big) \\ \downarrow & & \downarrow & & \downarrow \\ H^d(H, U_d(K_*)) & \xrightarrow{\phi^*} & H^d(G, U_d(K_*)) & \xleftarrow{f_*} & H^d(G, U_d(C_*)) \end{array}$$

There are tautological classes $\tau_{K_*/H} \in H^d\big(\mathrm{Hom}_H(K_*, U_d(K_*))\big)$, defined as the class of the tautological cochain $T_{K_*/G}(k) = [k]$, and a similar $\tau_{C_*/G} \in H^d\big(\mathrm{Hom}_G(C_*, U_d(C_*))\big)$. Their images in the group $H^d\big(\mathrm{Hom}_G(C_*, U_d(K_*))\big)$ coincide, since both are clearly equal to the class of T defined by $T(c) = [f(c)]$. The classes τ_{H,K_*} and τ_{G,C_*} are images of $\tau_{K_*/H}$ and $\tau_{C_*/G}$ (respectively) under the vertical maps. Thus, to prove the theorem, we only need to check that the above diagram is commutative.

Commutativity of the right square: The vertical maps are induced by a (unique up to chain homotopy) G–map of chain complexes $P(G)_* \to C_*$. The horizontal maps are induced by the coefficient map f_*. Since these two maps act on different arguments of the Hom functor, they commute.

Commutativity of the left square: That square is the result of applying a cohomology functor to the diagram

$$\begin{array}{ccc} K_* & \xleftarrow{f_*} & C_* \\ \uparrow & & \uparrow \\ P(H)_* & \xleftarrow{\phi_*} & P(G)_* \end{array}$$

The two compositions to compare are G–maps from $P(G)_*$ to the acyclic chain complex K_* (with the G–structure induced via ϕ). Such a map is unique up to chain-homotopy; hence the compositions are chain-homotopic. After passing to cohomology, they become equal. □

Remark 1.6 The procedure applied in Definition 1.4 to the tautological class works in greater generality, for arbitrary coefficient groups and arbitrary classes. In Theorem 4.4 we will need the following version: Let C_* be an acyclic G–chain complex, A a G–module, $T \in Z^d(\mathrm{Hom}_G(C_*, A))$ an A–valued G–invariant d–cocycle, P_* a projective resolution of the trivial G–module $\mathbb{Z}$, and $\psi_{C_*}\colon P_* \to C_*$ a chain map of resolutions. Let $\psi^*_{C_*}\colon \mathrm{Hom}_G(C_*, A) \to \mathrm{Hom}_G(P_*, A)$ be the cochain map induced by ψ_{C_*}. We define the group cohomology class $\tau \in H^d(G, A)$ associated to T by $\tau := [\psi^*_{C_*}(T)]$.

2 Geometric complexes

Our main source of acyclic G–chain complexes is geometry. Suppose that G acts on an acyclic simplicial complex X by simplicial automorphisms. Then the simplicial chain complex C_*X is an acyclic G–chain complex.

Definition 2.1 Let X be an acyclic simplicial G–complex. The definitions of Section 1 applied to the acyclic simplicial G–chain complex C_*X give rise to

- the coefficient group $U_d = U_d(X) := U_d(C_*X)$;
- the tautological cocycle $T^d_{X/G} := T^d_{C_*X/G}$;
- the tautological class $\tau^d_{X/G} := \tau^d_{C_*X/G}$;
- the tautological group cohomology class $\tau^d_{G,X} := \tau^d_{G,C_*X}$.

In our considerations, the G–complexes X will usually arise as restricted configuration complexes of homogeneous G–spaces. We will typically start from a transitive G–action on a space $\mathbb{P}$. We will use $\mathbb{P}$ as the set of vertices of X, and span simplices of X on tuples of elements of $\mathbb{P}$ satisfying some genericity conditions. (A typical example: $G = \mathrm{SL}(2,K)$, $\mathbb{P} = K^2 \setminus \{0\}$, a tuple of vectors spans a simplex if and only if every two of them are linearly independent.) This scheme applies to many algebraic groups over arbitrary infinite fields.

The acyclicity of these restricted configuration complexes is usually the consequence of the star-property defined below.

Definition 2.2 [Kramer and Tent 2010] A simplicial complex X has the star-property if for any finite subcomplex $Y \subseteq X$ there exists a vertex $v \in X^0 \setminus Y^0$ joinable with every simplex of Y (v is joinable with a k–simplex $\sigma = [y_0, \dots, y_k]$ if $v * \sigma = [v, y_0, \dots, y_k]$ is a $(k+1)$–simplex in X).

Fact 2.3 *If X has the star-property, then it is acyclic.*

Proof Let $z = \sum a_\sigma \sigma$ be a cycle in X. Let Y be the union of all simplices σ that appear in z. Let v be a vertex of X witnessing the star-property for Y. Then $z = \partial\big(\sum a_\sigma v * \sigma\big)$. □

For a complex X with the star-property there is another variant of an acyclic chain complex associated to it; the (nondegenerate) ordered chain complex C^o_*X. The group C^o_kX is the free abelian group whose basis is the set of all $(k+1)$–tuples of vertices of X that span k–simplices (in other words, the set of ordered, nondegenerate k–simplices of X). The boundary operator is defined by the usual formula

$$\partial[v_0, \dots, v_k] = \sum_{i=0}^{k} (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_k].$$

By the same argument as in Fact 2.3, the complex C^o_*X is acyclic. (Warning: for finite simplicial complexes the nondegenerate ordered chain complex does not calculate homology correctly, eg the complex $C^o_*(\Delta^1)$ is not acyclic.)

If a simplicial complex is acted upon by a group G, one can use the ordered chain complex to define the coefficient group $U^o_d := (C^o_d X/B^o_d X)_G$, the tautological cocycle $T^d_{C^o_*X/G}$ and the tautological class $\tau^d_{C^o_*X/G}$. (If X has the star-property, one can further define the tautological group cohomology class τ^d_{G,C^o_*X}.) There is a natural epimorphic G–chain map $C^o_*X \to C_*X$; it induces an epimorphism $U^o_d \to U_d$. The group U^o_d is usually insignificantly larger than U_d, as we shall see.

The calculations of the groups U_d and U^o_d are often used in this paper; we now explain how they are done. Let $X^{(n)}$ be the set of nondegenerate ordered n–simplices in a simplicial complex X. Let R_n be a set of representatives of orbits of G on $X^{(n)}$. For any $\sigma \in X^{(n)}$ we denote by σ_R the unique element of R_n that is G–equivalent to σ. For chains we put $\left(\sum a_\sigma \sigma\right)_R = \sum a_\sigma \sigma_R$.

Fact 2.4 *Let X be a simplicial G–complex.*

(a) *The group U^o_d is the quotient of the free abelian group with basis R_d by the subgroup spanned by* $\{(\partial\rho)_R \mid \rho \in R_{d+1}\}$.

(b) *The group U_d is the quotient of the free abelian group with basis R_d by the subgroup spanned by* $\{(\partial\rho)_R \mid \rho \in R_{d+1}\} \cup \{(t\rho)_R - \operatorname{sgn}(t)\rho \mid \rho \in R_d, t \in S_{d+1}\}$. *Moreover, in this description one can change the range of t from the permutation group S_{d+1} to any generating set of this group.*

The proof is based on the formula $U^o_d = (C^o_d X)_G/\partial(C^o_{d+1})_G$ and an analogous formula for U_d. We denote by c_G the image of the chain c in the coinvariants group.

Proof We start with a general remark. Suppose that a group G acts on a set Y. Let $\mathbb{Z}[Y]$ be the free abelian group with basis Y. Then $\mathbb{Z}[Y]$ has a natural G–module structure, and the coinvariants module $\mathbb{Z}[Y]_G$ is the free abelian group with basis Y/G (the orbit space of the G–action on Y). If $R \subseteq Y$ is a set of representatives of G–orbits, then the bijection $R \ni r \mapsto G\cdot r \in Y/G$ induces the natural isomorphism $\mathbb{Z}[R] \to \mathbb{Z}[Y/G] \to \mathbb{Z}[Y]_G$.

Applying this discussion to the G–action on $X^{(n)}$ we see that $(C^o_n X)_G = \mathbb{Z}[X^{(n)}]_G = \mathbb{Z}[X^{(n)}/G] \simeq \mathbb{Z}[R_n]$. This isomorphism $(C^o_n X)_G \to \mathbb{Z}[R_n]$ is clearly given by $c_G \mapsto c_R$. Similarly, $(C^o_{n+1}X)_G$ is isomorphic to $\mathbb{Z}[R_{n+1}]$, which is generated by R_{n+1}. The map $\partial\colon (C^o_{n+1}X)_G \to (C^o_n X)_G$ can be interpreted as the map $\mathbb{Z}[R_{n+1}] \ni c \mapsto (\partial c)_R \in \mathbb{Z}[R_n]$; its image is generated by the images of elements of R_{n+1}, ie by the set $\{(\partial\rho)_R \mid \rho \in R_{n+1}\}$. Part (a) is proved.

For part (b): Let K be the kernel of the epimorphism $C^o_d X \to C_d X$. The group K is generated by $\{t\sigma - (\operatorname{sgn} t)\sigma \mid t \in S_{d+1}, \sigma \in X^{(d)}\}$ (one can change the range of t from S_{d+1} to any generating set of S_{d+1}). Applying the coinvariants functor to the exact sequence $K \to C^o_d X \to C_d X \to 0$ we get the middle row of the following commuting diagram with exact rows and columns:

$$\begin{array}{ccccccc}
 & & (C^o_{d+1}X)_G & \longrightarrow & (C_{d+1}X)_G & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle\partial} & & \downarrow{\scriptstyle\partial} & & \\
K_G & \xrightarrow{\iota_G} & (C^o_dX)_G & \longrightarrow & (C_dX)_G & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \\
 & & U^o_d & \longrightarrow & U_d & & \\
 & & & & \downarrow & & \\
 & & & & 0 & &
\end{array}$$

A diagram chase shows that an element of $(C^o_dX)_G$ that maps to 0 in U_d is a sum of images of elements of K_G and $(C^o_{d+1}X)_G$. Consequently,

$$U_d \simeq (C^o_dX)_G/(\partial(C^o_{d+1}X)_G + \iota_G K_G).$$

Therefore, a presentation of U_d can be obtained from the presentation of U^o_d given in (a) by adjoining extra relations generating $\iota_G K_G$. These extra relations are images of generators of K_G under ι_G, ie are of the form $(t\sigma)_G - (\operatorname{sgn} t)\sigma_G$ ($t \in S_{d+1}$, $\sigma \in X^{(d)}$). Under the isomorphism $(C^o_dX)_G \to \mathbb{Z}[R_d]$ this form maps to $(t\sigma)_R - (\operatorname{sgn} t)\sigma_R$. To finish the proof we will check that $(t\sigma)_R = (t\sigma_R)_R$. We have $\sigma_R = g\sigma$ for some $g \in G$. This implies that $t\sigma_R = g(t\sigma)$, and then $(t\sigma_R)_R = (g(t\sigma))_R = (t\sigma)_R$. □

3 Boundedness

A group cohomology class in $H^d(G,\mathbb{R})$ is called bounded if it can be represented by a bounded cocycle $c\colon S_dBG \to \mathbb{R}$ (or, equivalently, a bounded G–invariant $\mathbb{R}$–valued cocycle on S_dEG). Here S_*BG is the singular chain complex of BG; a cocycle c is bounded if there exists $M > 0$ such that for each singular simplex $\sigma\colon \Delta^d \to BG$ we have $|c(\sigma)| \le M$. Instead of $\mathbb{R}$, one can use other groups with seminorm. In particular, if X is a simplicial G–complex, the coefficient group $U = U_d(X)$ carries a natural seminorm, induced by the ℓ^1–norm on C_dX. Explicitly, for $u \in U$ we consider all chains $\sum \alpha_i\sigma_i \in C_dX$ that represent u, and we declare the infimum of $\sum |\alpha_i|$ over all such chains to be $|u|$.

Theorem 3.1 *Suppose that X is an acyclic simplicial G–complex with the star-property. Then the tautological cohomology class $\tau^d_{G,X} \in H^d(G,U)$ is bounded with respect to the seminorm discussed above.*

Proof We will construct a G–chain map $\Psi_*\colon S_*EG \to C_*X$. For each $n \ge 0$ choose a free basis Σ_n of the free G–module S_nEG. We define Ψ_n inductively. For each $\xi_0 \in \Sigma_0$ we choose a vertex $\Psi_0(\xi_0) \in X^{(0)}$; we extend Ψ_0 to S_0EG by G–equivariance and linearity. Once Ψ_{n-1} is defined, we define Ψ_n on Σ_n as follows. For $\xi_n \in \Sigma_n$ we consider $\Psi_{n-1}(\partial\xi_n) = \sum \sigma_i \in C_{n-1}X$. By the star-property, there exists a vertex $v \in X^{(0)}$ joinable to every σ_i; we put $\Psi_n(\xi_n) = \sum v * \sigma_i$, so as to have $\partial\Psi_n(\xi_n) = \Psi_{n-1}(\partial\xi_n)$. Then we extend Ψ_n to S_nEG by G–equivariance and linearity. A straightforward induction shows that for any singular simplex $\xi_n \in S_nEG$ the chain $\Psi_n(\xi_n)$ is a sum of at most $(n+1)!$ simplices.

The class $\tau^d_{G,X}$ is represented by the cocycle $T^d_{X/G} \circ \Psi_d$. The tautological cocycle $T^d_{X/G}$ has norm at most 1 — it maps a simplex to its class in U_d, and that class has norm ≤ 1 by definition of the seminorm. Therefore, for any singular simplex σ_d in EG,

$$|T^d_{X/G}(\Psi_d(\sigma_d))| \leq (d+1)!. \qquad \square$$

Remark 3.2 There is a different approach to bounded group cohomology, based on the standard homogeneous resolution of the trivial G–module $\mathbb{Z}$ (see [Brown 1982, I, Section 5]). That approach is equivalent to the one used above, as shown in [Gromov 1982, pages 48–49]; for a more detailed account see [Löh 2010, 2.5.5]. In these references real coefficients are used, but the proof works for coefficients in an arbitrary abelian group with seminorm.

4 Characteristic classes

A cohomology class α of a (discrete) group G can serve as a characteristic class of (flat) G–bundles. Suppose that α is obtained from a G–invariant cocycle on an acyclic G–space X as in Remark 1.6. Then it is possible to describe the characteristic class using the cocycle directly, bypassing α (see Theorem 4.4). This section is organized as follows. We start by recalling the connection between group cohomology and characteristic classes. Next, we describe the classical de Rham version of characteristic classes of flat bundles. Then we discuss auxiliary notions and notation and, finally, we state and prove the main statement, Theorem 4.4. (Recall that we consider G with discrete topology, so that all G–bundles are flat — with locally constant transition functions — and BG is $K(G,1)$.)

Let $\alpha \in H^d(G,A) = H^d(BG,A)$ be a cohomology class of a group G. The space BG is the base of a universal principal G–bundle EG. Every principal G–bundle P over a (paracompact) base space B has a classifying map; a map $f_P\colon B \to BG$ such that $f_P^* EG \simeq P$. The map f_P is unique up to homotopy. Notice that we use $f^*\xi$ to denote the pullback of the bundle ξ via the map f, and we also use $f^*\tau$ and f^*T for the pullback of a cohomology class τ or of a cocycle T. Though occasionally confusing, this dual usage is standard practice in bundle theory.

Definition 4.1 The cohomology class $\alpha(P) := f_P^*(\alpha) \in H^d(B,A)$ is functorial in P, and is called the characteristic class (corresponding to α) of the bundle P.

In this definition the G–module A may have nontrivial G–structure. Then the groups $H^d(G,A)$ and $H^d(B,A)$ are cohomology groups with twisted coefficients, ie with coefficients in a flat G–bundle (local system) with fibre A. For $H^d(G,A) = H^d(BG,A)$ the bundle is $EG \times_G A$; for $H^d(B,A)$ we use $P \times_G A$. We have $P \times_G A = f_P^*(EG \times_G A)$, so the coefficient system used over BG pulls back to the one used over B; therefore we get a map $f_P^*\colon H^d(G,A) \to H^d(B,A)$.

In de Rham theory there is a construction of characteristic classes of flat bundles that does not explicitly refer to BG. In fact, it gives an explicit cocycle representative of the characteristic cohomology class in terms of a section. Suppose that $P \to B$ is a principal flat G–bundle over a manifold B, and that $\omega \in \Omega^d(X)$ is a G–invariant closed form on a contractible G–manifold X. To these data we will associate a class in $H^d_{\mathrm{DR}}(B)$. We start by forming the associated bundle $E = P \times_G X$ with fibre X. Then we choose a section $s\colon B \to E$; it exists and is homotopically unique because X is contractible. Now the idea is that a section s of a flat bundle is an ill-defined — G–ambivalent — map from the base to the fibre. The G–ambivalence is countered by the G–invariance of ω, so the pullback of ω by s is well defined. Let us be more precise. Let $\varphi_U\colon E|_U \to U \times X$ be local trivializations of E. Composing φ_U with $\mathrm{pr}_2\colon U \times X \to X$ we get a map $\psi_U\colon E|_U \to X$. The compositions $\psi_U \circ s|_U\colon U \to X$ are locally defined maps; these maps are not compatible. However, due to the G–invariance of ω, the forms $\omega_U = (\psi_U \circ s|_U)^*\omega \in \Omega^d(U)$ are compatible and define a global closed form in $\Omega^d(B)$. Slightly abusing the notation we denote this form by $s^*\omega$. The cohomology class of $s^*\omega$ in $H^d_{\mathrm{DR}}(B)$ is a characteristic class of the bundle P. An alternative description is to define the global form ω^E on E by gluing the compatible collection of forms $\psi_U^*\omega \in \Omega^d(E|_U)$, and then take $s^*\omega^E$ in the standard sense. (See [Morita 2001, Chapter 2] for more information on these classes.)

Let us pass to the simplicial setting. Let $P \to B$ be a principal G–bundle over a Δ–complex B. (For a basic discussion of Δ–complexes see [Hatcher 2002, Section 2.1].) Let $T \in Z^d(\mathrm{Hom}_G(C_*X, A))$ be an A–valued G–invariant simplicial cocycle on an acyclic simplicial G–complex X, and let $\tau \in H^d(G, A)$ be the associated cohomology class (as in Remark 1.6). The characteristic class of P (corresponding to τ) is the cohomology class $\tau(P) \in H^d(B, A)$ (see Definition 4.1). We will use the strategy explained in the de Rham setting and obtain a cochain on B representing $\tau(P)$ (see Theorem 4.4).

To deal with sections in the simplicial context we introduce a special family of trivializations. Let $P \to B$ be a principal G–bundle over a Δ–complex B. Let X be a simplicial G–complex. Let $E = P \times_G X$ be the associated bundle with fibre X. Consider a simplex $\sigma\colon \Delta \to B$, part of the Δ–complex structure. The bundle σ^*P is a flat principal G–bundle over a simplex; hence it has flat sections. Any such flat section $r\colon \Delta \to \sigma^*P$ induces a trivialization of $\sigma^*E \simeq \sigma^*P \times_G X$ — the map

$$\Delta \times X \ni (p, x) \mapsto [r(p), x] \in \sigma^*P \times_G X$$

is an isomorphism, whose inverse $\varphi_{\sigma,r}$ is a trivialization. We put $\psi_{\sigma,r} = \mathrm{pr}_2 \circ \varphi_{\sigma,r}\colon \sigma^*E \to X$. Notice that all possible flat sections of σ^*P are G–related, and that

$$\psi_{\sigma,rg} = g^{-1}\psi_{\sigma,r}. \tag{4-1}$$

Moreover, if σ_i is a face of σ (say $\sigma_i = \sigma|_{\Delta(i)}$, where $\Delta(i) = [e_0, \dots, \hat{e}_i, \dots, e_n]$), then

$$\psi_{\sigma,r}|_{\sigma_i^*E} = \psi_{\sigma_i, r|_{\Delta(i)}}. \tag{4-2}$$

We will now use the maps $\psi_{\sigma,r}$ to define simplicial sections.

Definition 4.2 Let B be a Δ–complex, X a simplicial G–complex, $P \to B$ a principal G–bundle, and $E = P \times_G X$ the associated bundle over B with fibre X. A section $s\colon B \to E$ is called simplicial if for every simplex $\sigma\colon \Delta \to B$ from the Δ–structure of B, and for any $\psi_{\sigma,r}\colon \sigma^*E \to X$ as described above, the composition $\psi_{\sigma,r} \circ s \circ \sigma \colon \Delta \to X$ is an affine map of Δ onto some simplex of the simplicial structure of X — possibly onto a simplex of dimension smaller than $\dim \Delta$ (the composition $s \circ \sigma$ defines a section of $\sigma^* E$ because, for $p \in \Delta$, we have $(\sigma^* E)_p = E_{\sigma(p)}$).

Remark A simplicial section in uniquely determined by its values at the vertices of the base.

A twisted cochain in $C^d(B, A)$ assigns to a simplex $\sigma \colon \Delta \to B$ a value in $(P \times_G A)_{\sigma(e_0)}$ — the fibre of the coefficient bundle over the initial vertex of σ. This value extends to a (unique) flat section of $\sigma^*(P \times_G A) = \sigma^* P \times_G A$. A flat section of that bundle can be described as $[r, a]$, where r is a section of $\sigma^* P$ and $a \in A$. For each $g \in G$ the pair $[rg, g^{-1}a]$ defines the same section; therefore one can also describe sections as continuous (locally constant) G–maps $\sigma^* P \to A$ — or G–maps from the G–torsor of flat sections of $\sigma^* P$ to A.

Definition 4.3 Let B be a Δ–complex, X a simplicial G–complex, $P \to B$ a principal G–bundle, and $E = P \times_G X$ the associated bundle over B with fibre X, s — a simplicial section of E. Consider a simplex $\sigma \colon \Delta \to B$ from the Δ–structure of B and flat sections r of $\sigma^* P$. Then the expression $T(\psi_{\sigma,r} \circ s \circ \sigma)$ is G–equivariant in r (due to (4-1) and the fact that T is a G–map). The formula

$$s^* T(\sigma) = [r, T(\psi_{\sigma,r} \circ s \circ \sigma)]$$

defines the cochain $s^* T \in C^d(B, A)$ (with twisted coefficients).

(The image of the map $\psi_{\sigma,r} \circ s \circ \sigma$ is a simplex in X, on which we put the orientation corresponding under this map to the standard orientation of the standard simplex; we interpret the argument of T as that oriented simplex. If the image of $\psi_{\sigma,r} \circ s \circ \sigma$ has dimension smaller than d, we interpret the argument of T as the zero chain.)

Remark The fact that $s^* T$ is a cocycle will follow from the proof of the next theorem.

Theorem 4.4 *Let T be an A–valued G–invariant cocycle on an acyclic simplicial G–complex X. Let $\tau \in H^d(G, A)$ be the associated group cohomology class (as in Remark 1.6). Let $P \to B$ be a principal (flat) G–bundle over a Δ–complex B. Let $s \colon B \to P \times_G X$ be a simplicial section. Then the class $\tau(P) \in H^d(B, A)$ — the characteristic class of P corresponding to τ — is represented by the simplicial cocycle $s^* T \in Z^d(B, A)$.*

Proof The total space EG of the universal principal G–bundle $EG \to BG$ is contractible (it is also the universal cover of BG). The G–action on EG is free. Therefore, the singular chain complex $S_* EG$ is a projective (in fact, free) resolution of the trivial G–module $\mathbb{Z}$. Moreover, $(S_* EG)_G \simeq S_* BG$. Let $\Psi = \Psi_{C_* X} \colon S_* EG \to C_* X$ be a resolution map from $S_* EG$ to the simplicial chain complex of X. This map induces the map $\Psi^* \colon \mathrm{Hom}_G(C_* X, A) \to \mathrm{Hom}_G(S_* EG, A)$, and

$$\tau = [\Psi^* T] \in H^d(\mathrm{Hom}_G(S_* EG, A)) = H^d(G, A).$$

Let $f = f_P : B \to BG$ the a classifying map of the bundle P, and let $F : P \to EG$ be a G–bundle map covering f. Then $\tau(P) = f^*\tau = [f^*\Psi^*T]$. Let us describe the cocycle $f^*\Psi^*T$ explicitly. This cocycle should assign to any simplex $\sigma : \Delta \to B$ (from the Δ–structure of B) a value in $(P \times_G A)_{\sigma(e_0)}$; as explained in the paragraph preceding Definition 4.3, the choice of that value is equivalent to the choice of a G–map from the G–set of flat sections r of σ^*P to the G–module A. Suppose that $r : \Delta \to \sigma^*P$ is a (flat) section. Then $F \circ r$ is a singular simplex in EG. The map $r \mapsto T(\Psi(F \circ r)) \in A$ is a G–map (since each of F, Ψ and T is a G–map); it defines the value of the cochain $f^*\Psi^*T$ (representing $\tau(P)$) on the simplex σ,

$$f^*\Psi^*T(\sigma) = [r, T(\Psi(F \circ r))]. \tag{4-3}$$

The cochain $f^*\Psi^*T$ depends on several choices:

(1) One can choose the space BG — within the homotopy type.

(2) One can choose $f : B \to BG$ — within the homotopy class.

(3) One can choose Ψ — all resolution maps are possible.

Our strategy is to exploit these choices to ensure that $f^*\Psi^*T = s^*T$.

Let $\Sigma(B)$ be the set of all simplices σ forming the Δ–structure of B.

Lemma 4.5 *One can choose the space BG and the map f such that*

(a) *$f^*EG \simeq P$ (ie f is a classifying map of P);*

(b) *all the maps $f \circ \sigma$ for $\sigma \in \Sigma(B)$ are pairwise distinct.*

Proof For dimension d let m_d be the barycentre of the standard simplex Δ^d. For each $\sigma \in \Sigma(B)$ we put $p_\sigma = \sigma(m_{\dim \sigma})$. Then we choose a collection of pairwise different points $(x_\sigma)_{\sigma \in \Sigma(B)}$ in BG. (If BG is too small for that, we change it by wedging it with a contractible space of sufficiently large cardinality.) Finally, we perform a homotopy of f (inductively over skeleta) to ensure $f(p_\sigma) = x_\sigma$. □

Let $f : B \to BG$ be a classifying map of P satisfying the conditions of Lemma 4.5. Let $F : P \to EG$ be a G–bundle map covering f (the composition of an isomorphism $P \to f^*EG$ with the canonical map $f^*EG \to EG$).

Lemma 4.6 *One can choose the resolution map $\Psi : S_*EG \to C_*X$ such that $f^*\Psi^*T = s^*T$.*

Proof Let us discuss how Ψ may be constructed. For each $n \ge 0$ choose a free basis Σ_n of the free G–module S_nEG. Define Ψ_n inductively. The base case is $\Psi_{-1} = \mathrm{id}_{\mathbb{Z}} : \mathbb{Z} \to \mathbb{Z}$, with $\mathbb{Z}$ connected to the resolutions by the augmentation maps $\partial : S_0EG \to \mathbb{Z}$, $\partial : C_0X \to \mathbb{Z}$. Once Ψ_{n-1} is defined, calculate — for every $\sigma \in \Sigma_n$ — the cycle $\Psi_{n-1}(\partial\sigma)$. Since C_*X is acyclic, this cycle is a boundary of some n–chain; pick one such chain and define it to be $\Psi_n(\sigma)$. A crucial remark is that if, for some $\eta \in \Sigma_n$ and some n–simplex ξ in X, we have $\Psi_{n-1}(\partial\eta) = \partial\xi$, then we may put $\Psi_n(\eta) = \xi$. Once Ψ_n is defined on Σ_n, we extend it to S_nEG by G–equivariance and linearity.

For each $(\sigma\colon \Delta \to B) \in \Sigma(B)$ choose a flat section $r(\sigma)\colon \Delta \to \sigma^* P$. Composing this section with the canonical bundle map $\sigma^* P \to P$, and then with $F\colon P \to EG$, we get a singular simplex $F \circ r(\sigma)$ in EG. We denote this simplex by $\widetilde{f \circ \sigma}$ — it is a lift of $f \circ \sigma$. All the lifts $\widetilde{f \circ \sigma}$ are pairwise G–inequivalent, because all $f \circ \sigma$ are pairwise distinct. Therefore we may choose the free bases Σ_n such that they contain all the lifts $\widetilde{f \circ \sigma}$ (for $\sigma \in \Sigma(B)$). We would like to define

$$\Psi(\widetilde{f \circ \sigma}) = \psi_{\sigma,r(\sigma)} \circ s \circ \sigma. \tag{4-4}$$

To be able to do that we need to check that

$$\Psi(\partial(\widetilde{f \circ \sigma})) = \partial(\psi_{\sigma,r(\sigma)} \circ s \circ \sigma). \tag{4-5}$$

Let $\sigma_i = \sigma|_{\Delta(i)}$, where $\Delta(i) = [e_0, \dots, \hat{e}_i, \dots, e_n]$ with $n = \dim \sigma$. (Strictly speaking, we should also use an extra map identifying $\Delta(i)$ with the standard simplex. We will ignore this in order not to overburden the notation.) We have

$$\partial(\widetilde{f \circ \sigma}) = \sum_{i=0}^{n} (-1)^i (\widetilde{f \circ \sigma})|_{\Delta(i)}. \tag{4-6}$$

Observe that $(\widetilde{f \circ \sigma})|_{\Delta(i)}$ is a lift of $f \circ (\sigma|_{\Delta(i)})$; therefore $(\widetilde{f \circ \sigma})|_{\Delta(i)} = (\widetilde{f \circ \sigma_i}) \cdot g(i)$ for some $g(i) \in G$. By induction on the dimension we know that

$$\Psi((\widetilde{f \circ \sigma_i}) \cdot g(i)) = g(i)^{-1} \Psi(\widetilde{f \circ \sigma_i}) = g(i)^{-1} (\psi_{\sigma_i, r(\sigma_i)} \circ s \circ \sigma_i) = \psi_{\sigma_i, r(\sigma_i) g(i)} \circ s \circ \sigma_i,$$

the last equality following from (4-1). We may finally write

$$\Psi(\partial(\widetilde{f \circ \sigma})) = \sum_{i=0}^{n} (-1)^i \psi_{\sigma_i, r(\sigma_i) g(i)} \circ s \circ \sigma_i. \tag{4-7}$$

On the other hand,

$$\partial(\psi_{\sigma,r(\sigma)} \circ s \circ \sigma) = \sum_{i=0}^{n} (-1)^i \psi_{\sigma,r(\sigma)} \circ s \circ \sigma_i. \tag{4-8}$$

Notice that $s \circ \sigma_i$ is a section of $\sigma_i^* E$ (where $E = P \times_G X$); therefore (4-2) applies and yields

$$\psi_{\sigma,r(\sigma)} \circ s \circ \sigma_i = \psi_{\sigma_i, r(\sigma)|_{\Delta(i)}} \circ s \circ \sigma_i. \tag{4-9}$$

The flat sections $r(\sigma)|_{\Delta(i)}$ and $r(\sigma_i)$ of the G–bundle $\sigma_i^* P$ are G–related: $r(\sigma)|_{\Delta(i)} = r(\sigma_i) \cdot g$ for some $g \in G$. Recall that, for any $\eta \in \Sigma(B)$, the singular simplex $\widetilde{f \circ \eta}$ is the composition of $r(\eta)$ with a G–bundle map $\eta^* P \to EG$. It follows that $(\widetilde{f \circ \sigma})|_{\Delta(i)} = (\widetilde{f \circ \sigma_i}) \cdot g$; therefore $g = g(i)$. Consequently,

$$\psi_{\sigma_i, r(\sigma)|_{\Delta(i)}} \circ s \circ \sigma_i = \psi_{\sigma_i, r(\sigma_i) g(i)} \circ s \circ \sigma_i. \tag{4-10}$$

Putting (4-8), (4-9) and (4-10) together we get

$$\partial(\psi_{\sigma,r(\sigma)} \circ s \circ \sigma) = \sum_{i=0}^{n} (-1)^i \psi_{\sigma_i, r(\sigma_i) g(i)} \circ s \circ \sigma_i. \tag{4-11}$$

Comparing this with (4-7) we obtain (4-5). We may therefore define Ψ so that (4-4) holds for all $\sigma \in \Sigma(B)$. Then, for d–dimensional σ we get

$$\begin{aligned} f^*\Psi^*T(\sigma) &= \bigl[r(\sigma), T\bigl(\Psi(F \circ r(\sigma))\bigr)\bigr] \\ &= [r(\sigma), T(\Psi(\widetilde{f \circ \sigma}))] = [r(\sigma), T(\psi_{\sigma,r(\sigma)} \circ s \circ \sigma)] = s^*T(\sigma). \end{aligned}$$ □

Lemma 4.6 implies $f^*\Psi^*T$ — a cocycle representing $\tau(P)$ — is equal to s^*T; proving Theorem 4.4. □

Remark In our applications of Theorem 4.4 the coefficients will be either untwisted or only mildly twisted (eg a $\mathrm{GL}(2,K)$–module A which is trivial as an $\mathrm{SL}(2,K)$–module).

5 Homological core

Consider an acyclic simplicial G–complex X, the associated coefficient group U_d and the tautological cohomology classes $\tau_{X/G} \in H^d(\mathrm{Hom}_G(C_*X, U_d))$ and $\tau_{G,X} \in H^d(G, U_d)$.

Question Is it possible to represent these tautological classes by cocycles with coefficients in a proper subgroup of U_d?

In general, there is a candidate subgroup. The coefficient group $U_d = (C_dX)_G/\partial(C_{d+1}X)_G$ has a natural homomorphism $\partial\colon U_d \to (C_{d-1}X)_G$ (induced by the usual $\partial\colon C_dX \to C_{d-1}X$).

Definition 5.1 The homological core hU_d of the group U_d is the kernel of the map $\partial\colon U_d \to (C_{d-1}X)_G$.

The following theorem states a weaker property then asked for above, but is quite general.

Theorem 5.2 *Let X be an acyclic simplicial G–complex. Let $\tau_{G,X} \in H^d(G, U_d)$ be the associated tautological class, and let $z \in H_d(BG, \mathbb{Z})$ be a homology class. Then $\langle \tau_{G,X}, z\rangle \in hU_d$.*

Proof The map $\partial\colon U_d \to (C_{d-1}X)_G$ of coefficient groups induces horizontal maps in the commutative diagram

$$\begin{array}{ccc} H^d(\mathrm{Hom}_G(C_*X, U_d)) & \xrightarrow{\partial_*} & H^d\bigl(\mathrm{Hom}_G(C_*X, (C_{d-1}X)_G)\bigr) \\ \downarrow & & \downarrow \\ H^d(G, U_d) & \xrightarrow{\partial_*} & H^d(G, (C_{d-1}X)_G) \end{array}$$

The class $\tau_{X/G} \in H^d(\mathrm{Hom}_G(C_*X, U_d))$ is mapped to 0 by ∂_*. Indeed, the class $\tau_{X/G}$ is represented by the tautological cocycle given by $T_{X/G}(\sigma^d) = [\sigma^d]$. Let $t \in \mathrm{Hom}_G(C_{d-1}, (C_{d-1})_G)$ also be tautological: $t(\sigma^{d-1}) = [\sigma^{d-1}]$. Then

$$(\partial_* T_{X/G})(\sigma^d) = \partial[\sigma^d] = [\partial\sigma^d] = t(\partial\sigma^d) = (\delta t)(\sigma^d),$$

that is, $\partial_* T_{X/G} = \delta t$; hence $\partial_*\tau_{X/G} = 0$.

It now follows from the diagram that $\partial_*\tau_{G,X} = 0$ as well. Therefore, $\partial\langle \tau_{G,X}, z\rangle = \langle \partial_*\tau_{G,X}, z\rangle = 0$. □

II GL(2, K)

In this part we describe some results of Nekovář [1990] from our point of view. This provides a perfect illustration of the general method.

6 Review of possible actions

The first example where the general approach from Part I gives something interesting is $G = \mathrm{GL}(2,K)$. To proceed we need a simplicial action of G on a complex X with desired properties, one of them being high transitivity. Such X can be constructed by taking as the vertex set a homogeneous space G/S for some S and studying the notion of "generic k–tuple". In $\mathrm{GL}(2,K)$ we have the following interesting subgroups.

(1) If $S = \left(\begin{smallmatrix}1 & *\\ 0 & *\end{smallmatrix}\right)$, then $G/S = K^2\setminus\{0\}$. A tuple is generic if it consists of pairwise linearly independent vectors. The action of $\mathrm{GL}(2,K)$ is effective and transitive on generic pairs. We will discuss this case later.

(2) If $S = \left(\begin{smallmatrix}* & 0\\ 0 & *\end{smallmatrix}\right)$, then $G/S = (\mathbb{P}^1(K)\times\mathbb{P}^1(K))\setminus\Delta$ (where Δ is the diagonal). Generic k–tuples are tuples of pairs (p_i, q_i) with all the points p_i and q_j distinct. Here even the action on pairs is not transitive, because the cross-ratio (p_1, q_1, p_2, q_2) is preserved. The action of $\mathrm{GL}(2,K)$ is not effective; it factors through $\mathrm{PGL}(2,K)$.

(3) If $S = \left(\begin{smallmatrix}* & *\\ 0 & *\end{smallmatrix}\right)$, then $G/S = \mathbb{P}^1(K)$. The action factors through $\mathrm{PGL}(2,K)$. Generic k–tuples are tuples of distinct projective points. The action is triply transitive.

In the first two examples our approach yields very big groups of coefficients. We can compute them (and we do so in the $S = \left(\begin{smallmatrix}1 & *\\ 0 & *\end{smallmatrix}\right)$ case), but we cannot say much about them. In the third case transitivity is higher; hence the coefficient group is smaller and easier to understand. We do the computation in detail (following Nekovář). Actually in this case something interesting happens: while $\mathrm{PGL}(2,K)$ acts transitively on triples, the large normal subgroups $\mathrm{PSL}(2,K)$ acts transitively on pairs, while its orbits on triples are indexed by $\dot K/\dot K^2$ — the group of square classes. This gives an untwisted cohomology class for $\mathrm{PSL}(2,K)$ and a (slightly) twisted cohomology class for $\mathrm{PGL}(2,K)$.

Now we proceed with the description of the $S = \left(\begin{smallmatrix}1 & *\\ 0 & *\end{smallmatrix}\right)$ case. We do not discuss the $S = \left(\begin{smallmatrix}* & 0\\ 0 & *\end{smallmatrix}\right)$ case.

GL(2, K)–action on nonzero vectors Consider the $\mathrm{GL}(2,K)$–action on $\mathrm{GL}(2,K)/\left(\begin{smallmatrix}1 & *\\ 0 & *\end{smallmatrix}\right) \simeq K^2\setminus\{0\}$. We declare a tuple of nonzero vectors in K^2 to be generic if its elements are pairwise linearly independent. The complex X has k–simplices spanned on generic $(k+1)$–tuples.

The action of $\mathrm{GL}(2,K)$ is transitive on 1–simplices. However, the action of its large normal subgroup $\mathrm{SL}(2,K)$ does have an invariant: $(v,w)\mapsto \det(v,w)\in\dot K$. Thus, we have a potentially nontrivial 1–

dimensional cocycle, with coefficients in a quotient of $\mathbb{Z}[\dot{K}]$. However, any given generic triple can be normalized by an element of $\mathrm{SL}(2,K)$ to

$$\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\a\end{pmatrix},\begin{pmatrix}-x/a\\y\end{pmatrix}\right) \tag{6-1}$$

with nonzero a, x and y; the three determinants are a, y and x. To get the coefficient group we have to divide $\mathbb{Z}[\dot{K}]$ by relations $\langle a\rangle-\langle x\rangle+\langle y\rangle$, for all nonzero a, x and y. The resulting coefficient group is trivial, hence so is the cocycle.

The action of $\mathrm{GL}(2,K)$ on generic triples of vectors has a complete invariant in $\dot{K}\times\dot{K}\times\dot{K}$: the triple of pairwise determinants. The resulting coefficient group is the quotient of $\mathbb{Z}[\dot{K}\times\dot{K}\times\dot{K}]$ by the relation

$$\left(\frac{be-cd}{a},e,c\right)-\left(\frac{be-cd}{a},d,b\right)+(e,d,a)-(c,b,a)=0$$

(all entries assumed nonzero). We skip the details, as they are not dissimilar to ones in the calculation presented later, and the result is not especially meaningful.

GL(2, K)–action on projective line Consider the $\mathrm{GL}(2,K)$–action on $\mathrm{GL}(2,K)/\left(\begin{smallmatrix}*&*\\0&*\end{smallmatrix}\right)\simeq\mathbb{P}^1(K)$. We declare a $(k+1)$–tuple of projective points generic if they are pairwise distinct; we span k–simplices on such tuples. Thus, the complex X is the (infinite) simplex with vertex set $\mathbb{P}^1(K)$.

The $\mathrm{GL}(2,K)$ (in fact, $\mathrm{PGL}(2,K)$) action on X is transitive on 2–simplices. However, the $\mathrm{SL}(2,K)$–action on 2–simplices has an invariant with values in $\dot{K}/\dot{K}^2$ — the set of square classes. Our procedure will produce a cocycle with constant coefficients (in a quotient group of $\mathbb{Z}[\dot{K}/\dot{K}^2]$) for $\mathrm{PSL}(2,K)$, and with twisted coefficients for $\mathrm{PGL}(2,K)$. We discuss this in detail in the next section.

The cross-ration is a complete invariant of (ordered) 3–simplices, ie of 4–tuples of distinct points in $\mathbb{P}^1(K)$, under the action of $\mathrm{PGL}(2,K)$. Thus, our approach yields a 3–cocycle with coefficients in $\mathbb{Z}[\dot{K}\setminus\{1\}]/I$, where I is the subgroup spanned by

$$\left[\frac{\lambda(\mu-1)}{\mu(\lambda-1)}\right]-\left[\frac{\mu-1}{\lambda-1}\right]+\left[\frac{\mu}{\lambda}\right]-[\mu]+[\lambda]$$

with $\lambda,\mu\in\dot{K}\setminus\{1\}$ and $\lambda\neq\mu$. This is related to the dilogarithm function; we do not pursue it further (but see eg [Bergeron et al. 2014]).

7 Action on triples of points in the projective line

In this section we consider the action of $G=\mathrm{PSL}(2,K)$ on the projective line $\mathbb{P}=\mathbb{P}^1(K)$, for an infinite field K. We define X as the (infinite) simplex with vertex set $\mathbb{P}$; in other words, we span simplices of X on tuples of generic (ie pairwise distinct) points in $\mathbb{P}$. The complex X has the star-property and is contractible. To the induced action of G on X the formalism of Part I (see Definition 2.1) associates the coefficient group U_2, the tautological cocycle T, and a tautological cohomology class of G.

Theorem 7.1 *The group U_2 is isomorphic to $W(K)$, the Witt group of the field K.*

Let us briefly recall the definition of $W(K)$ (for details see [Elman et al. 2008, Chapter I]). The isometry classes of nondegenerate symmetric bilinear forms over K form a semiring, with direct sum as addition and tensor product as multiplication. Passing to the Grothendieck group of the additive structure of this semiring we obtain the Witt–Grothendieck ring $\widehat{W}(K)$. The Witt ring $W(K)$ is the quotient of $\widehat{W}(K)$ by the ideal generated by the hyperbolic plane — the form with matrix $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$. Both rings have explicit presentations in terms of generators and relations; see [Elman et al. 2008, Theorems 4.7 and 4.8].

Proof of Theorem 7.1 We apply Fact 2.4.

Generators We need to find the orbits of the G–action on the set of generic triples of points in $\mathbb{P}$. We denote by $[v]$ the point in $\mathbb{P}$ determined by the vector $v \in K^2$; for $v = \left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$ we shorten $\left[\left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)\right]$ to $\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right]$.

Lemma 7.2 *Every generic triple $([u],[v],[w])$ of points in $\mathbb{P}$ is G–equivalent to a triple of the form*

$$t_\lambda = \left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\\lambda\end{bmatrix}\right), \quad \text{where } \lambda = \det(u,v)\det(v,w)\det(w,u).$$

Triples t_λ and t_μ are equivalent if and only if $\lambda/\mu \in \dot{K}^2$ (the set of squares in $\dot{K}$).

Proof There exists $g \in \mathrm{SL}(2,K)$ such that $gu = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$, and $gv = \left(\begin{smallmatrix}0\\\gamma\end{smallmatrix}\right)$ for $\gamma = \det(u,v)$. Then

$$gw = \begin{pmatrix}\alpha\\\beta\end{pmatrix} = \alpha\begin{pmatrix}1\\\beta/\alpha\end{pmatrix}$$

for some $\alpha, \beta \in \dot{K}$, so

$$g\colon ([u],[v],[w]) \mapsto \left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\\gamma\end{bmatrix}, \begin{bmatrix}\alpha\\\beta\end{bmatrix}\right) = \left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\\lambda\end{bmatrix}\right)$$

for $\lambda = \beta/\alpha$. Notice that

$$\lambda = \frac{\beta}{\alpha} = \begin{vmatrix}1&0\\0&\gamma\end{vmatrix} \cdot \begin{vmatrix}0&\alpha\\\gamma&\beta\end{vmatrix}^{-1} \cdot \begin{vmatrix}\alpha&1\\\beta&0\end{vmatrix} = \det(gu,gv)\det(gv,gw)^{-1}\det(gw,gu) \tag{7-1}$$
$$= \det(u,v)\det(v,w)^{-1}\det(w,u).$$

Notice that the stabilizer of $\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right)$ in $\mathrm{SL}(2,K)$ consists of diagonal matrices of the form $\left(\begin{smallmatrix}\alpha^{-1}&0\\0&\alpha\end{smallmatrix}\right)$. Such a matrix maps t_λ to

$$\left(\begin{bmatrix}\alpha^{-1}\\0\end{bmatrix}, \begin{bmatrix}0\\\alpha\end{bmatrix}, \begin{bmatrix}\alpha^{-1}\\\alpha\lambda\end{bmatrix}\right) = \left(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\\alpha^2\lambda\end{bmatrix}\right) = t_{\alpha^2\lambda}.$$

The last claim of the lemma follows. Finally, the class in $\dot{K}/\dot{K}^2$ of λ given by (7-1) is the same as the class of $\det(u,v)\det(v,w)\det(w,u)$. □

The lemma and Fact 2.4 imply that U_2 is the quotient of $\mathbb{Z}[\dot{K}/\dot{K}^2]$ by two sets of relations (boundary relations and alternation relations). The generator of $\mathbb{Z}[\dot{K}/\dot{K}^2]$ corresponding to t_λ (and the image of this generator in U_2) will be denoted by $[\lambda]$ and called the symbol of the triple.

Alternation relations The transposition of the first two points of a triple maps t_λ to

$$\left(\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}1\\\lambda\end{bmatrix}\right),$$

which can be transformed by $\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$ to

$$\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\-1\end{bmatrix},\begin{bmatrix}\lambda\\-1\end{bmatrix}\right)=\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\-\lambda^{-1}\end{bmatrix}\right).$$

Since $[-\lambda^{-1}]=[-\lambda]$ in $\dot K/\dot K^2$, the resulting relation can be written as $-[\lambda]=[-\lambda]$. Next, consider the transposition of the last two vectors of a triple; this transposition maps t_λ to

$$\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}1\\\lambda\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right),$$

which can be transformed by $\left(\begin{smallmatrix}1&-\lambda^{-1}\\0&1\end{smallmatrix}\right)$ to

$$\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\\lambda\end{bmatrix},\begin{bmatrix}-\lambda^{-1}\\1\end{bmatrix}\right)=\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\-\lambda\end{bmatrix}\right).$$

Again, we get $-[\lambda]=[-\lambda]$.

Boundary relations A generic quadruple of points in $\mathbb{P}$ can be G–transformed to

$$\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\\lambda\end{bmatrix},\begin{bmatrix}1\\\mu\end{bmatrix}\right),$$

where genericity is equivalent to $\lambda,\mu\in\dot K$, $\lambda\neq\mu$. The boundary of the corresponding 3–simplex is the alternating sum of four triangles — triples obtained from the quadruple by omitting one element. We calculate the symbols of those triples:

Omit $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$: $\left[\begin{vmatrix}0&1\\1&\lambda\end{vmatrix}\cdot\begin{vmatrix}1&1\\\lambda&\mu\end{vmatrix}\cdot\begin{vmatrix}1&0\\\mu&1\end{vmatrix}\right]=[(-1)\cdot(\mu-\lambda)\cdot 1]=[\lambda-\mu]$.

Omit $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$: $\left[\begin{vmatrix}1&1\\0&\lambda\end{vmatrix}\cdot\begin{vmatrix}1&1\\\lambda&\mu\end{vmatrix}\cdot\begin{vmatrix}1&1\\\mu&0\end{vmatrix}\right]=[\lambda\cdot(\mu-\lambda)\cdot(-\mu)]=[\lambda\mu(\lambda-\mu)]$.

Omit $\left[\begin{smallmatrix}1\\\lambda\end{smallmatrix}\right]$: $[\mu]$.

Omit $\left[\begin{smallmatrix}1\\\mu\end{smallmatrix}\right]$: $[\lambda]$.

The relation is

$$[\lambda-\mu]-[\lambda\mu(\lambda-\mu)]+[\mu]-[\lambda]=0.$$

Putting $a=\lambda-\mu$ and $b=\mu$, we may rewrite this as

$$[a]+[b]=[a+b]+[ab(a+b)].$$

The relation holds for all $a,b\in\dot K$ that satisfy $a+b\neq 0$. This set of relations, plus the alternation relation $[-a]=-[a]$, gives the classical description of the Witt group $W(K)$; see [Elman et al. 2008, Theorem 4.8]. □

Definition 7.3 The tautological second cohomology class of the group $G = \mathrm{PSL}(2, K)$ with coefficients in $U_2 = W(K)$ associated (as in Definition 2.1) to the action of G on X will be called the Witt class and denoted by $w \in H^2(\mathrm{PSL}(2, K), W(K))$.

Remark 7.4 (1) Let T be the tautological ($W(K)$–valued) cocycle associated to the G–action on X. From the proof of Theorem 7.1 it is useful to extract the following explicit formula for the value of T on a 2–simplex in X determined by a triple of pairwise linearly independent vectors $u, v, w \in K^2$:

$$T([u],[v],[w]) = [\det(u,v)\det(v,w)\det(w,u)]. \tag{7-2}$$

(2) One can see from the proof of Theorem 7.1 that the ordered coefficient group U_2^o is isomorphic to the Witt–Grothendieck group $\widehat{W}(K)$ of the field K.

(3) The space $\mathbb{P} = \mathbb{P}^1(K)$ and the complex X are acted upon by the larger group $\mathrm{PGL}(2, K)$. As a result, the Witt class can be interpreted as a twisted cohomology class of $\mathrm{PGL}(2, K)$. The twisting action of $\mathrm{PGL}(2, K)$ on $W(K)$ is easy to see from the formula for λ in Lemma 7.2; the class of $g \in \mathrm{GL}(2, K)$ acts on the symbol $[\lambda]$ mapping it to $[\det(g)\cdot\lambda]$.

(4) For $K = \mathbb{R}$ we have $W(\mathbb{R}) \simeq \mathbb{Z}$. The isomorphism, called the signature map, maps $[\lambda]$ to $+1$ for $\lambda > 0$ and to -1 for $\lambda < 0$. The pullback of the Witt class to $\mathrm{SL}(2,\mathbb{R})$ is a class in $H^2(\mathrm{SL}(2,\mathbb{R}),\mathbb{Z})$; we will relate it to the usual (topological) Euler class (see Theorem 13.4 and Fact 13.5).

(5) For $K = \mathbb{Q}$ the Witt group has a large torsion part which is a direct summand. Computer calculations (using the computer algebra system FriCAS) indicate that the corresponding part of the Witt class is nontrivial.

It is possible to give an explicit formula for a cocycle representing the Witt class. We use the standard homogeneous resolution (see [Brown 1982, II, Section 3]) to describe group cohomology; $W(K)$–valued 2–cocycles are then represented by functions $G \times G \times G \to W(K)$.

Theorem 7.5 *Let us fix an arbitrary nonzero vector $u \in K^2$. The map*

$$G \times G \times G \ni (g_0, g_1, g_2) \mapsto [\det(\tilde g_0 u, \tilde g_1 u)\det(\tilde g_1 u, \tilde g_2 u)\det(\tilde g_2 u, \tilde g_0 u)] \in W(K) \tag{7-3}$$

is a cocycle representing the Witt class $w \in H^2(\mathrm{PSL}(2, K), W(K))$. (By $\tilde g_i$ we denote an arbitrary lift of $g_i \in \mathrm{PSL}(2, K)$ to $\mathrm{SL}(2, K)$. The senseless symbol $[0]$ is interpreted as 0.)

Proof It is straightforward to check that the maps

$$\Psi_n : (g_0, \dots, g_n) \mapsto \begin{cases} ([\tilde g_0 u], \dots, [\tilde g_n u]) & \text{if the points } [\tilde g_i u] \text{ are pairwise different,} \\ 0 & \text{otherwise,} \end{cases} \tag{7-4}$$

defines a G–chain map from the homogeneous standard resolution of G to C_*X. (The only subtle case is when $[\tilde g_i u] = [\tilde g_j u]$ for exactly one pair of indices i, j. Then $\Psi_{n-1}\partial(g_0, \dots, g_n)$ has two nonzero summands — however, these summands cancel in the alternating chain complex C_*X.) Composing Ψ_2 with the tautological cocycle T given by (7-2) we obtain the statement of the theorem. □

III Euler class for ordered fields

In this part we define and investigate Euler classes for general linear and projective groups over arbitrary ordered fields.

8 Tautological Euler classes: computation of coefficients

Let K be an ordered field. Let $G = \mathrm{GL}(n, K)$; we will also consider the following closely related groups (where we put $\dot{K} = K \setminus \{0\}$ and $K_+ = \{\lambda \in K \mid \lambda > 0\}$):

$$\begin{aligned} PG &:= \mathrm{PGL}(n, K) = G/\{\lambda I \mid \lambda \in \dot{K}\}, \\ P_+G &:= \mathrm{P_+GL}(n, K) = G/\{\lambda I \mid \lambda \in K_+\}, \\ G_+ &:= \mathrm{GL}_+(n, K) = \{g \in G \mid \det g > 0\}, \\ PG_+ &:= \mathrm{PGL}_+(n, K) = G_+/\{\lambda I \in G_+ \mid \lambda \in K\}, \\ P_+G_+ &:= \mathrm{P_+GL_+}(n, K) = G_+/\{\lambda I \mid \lambda \in K_+\}. \end{aligned}$$

The natural maps between these groups are summarized in the diagram

$$\begin{array}{ccccc} G_+ & \longrightarrow & P_+G_+ & \longrightarrow & PG_+ \\ \downarrow & & \downarrow & & \downarrow \\ G & \longrightarrow & P_+G & \longrightarrow & PG \end{array}$$

The defining action of G on K^n induces actions of P_+G on $\mathbb{P}_+$ and of PG on $\mathbb{P}$; here

$$\mathbb{P}_+ := \mathbb{P}_+^{n-1}(K) = (K^n \setminus \{0\})/K_+, \quad \mathbb{P} := \mathbb{P}^{n-1}(K) = (K^n \setminus \{0\})/\dot{K},$$

where the multiplicative groups K_+ and $\dot{K}$ act on K^n by homotheties.

Next we define simplicial complexes X and X_+ by spanning simplices on generic tuples of points in $\mathbb{P}$ and $\mathbb{P}_+$, respectively. We call a tuple $([v_0], \dots, [v_k])$ generic if every subsequence of $(v_0, \dots, v_k)$ of length $\le n$ is linearly independent. Ordered fields are infinite, therefore these complexes have the star-property and are contractible. The complex X is acted upon by PG, and X_+ by P_+G. We restrict these actions to PG_+ and to P_+G_+ and we apply the formalism of Part I. We put

$$U := U_n(X), \quad U_+ = U_n(X_+)$$

(see Definitions 2.1 and 1.3), and we define the Euler classes as tautological classes

$$\text{(8-1)} \qquad \mathrm{eu} := \tau^n_{PG_+,X} \in H^n(PG_+, U), \quad \mathrm{eu}_+ := \tau^n_{P_+G_+,X_+} \in H^n(P_+G_+, U_+).$$

Notice that PG_+ is a normal subgroup of PG. Therefore the group U carries the structure of a PG–module, and the class eu can also be considered as a twisted PG–class. (See the second remark after Definition 1.3.) Similarly, the class eu_+ can be regarded as a twisted P_+G–class.

Our first goal is to compute U and U_+, as abelian groups and as PG– and P_+G–modules.

Theorem 8.1 *Let U (resp. U_+) be the coefficient group associated to the action of* $\mathrm{PGL}_+(n,K)$ *(resp.* $\mathrm{P}_+\mathrm{GL}_+(n,K)$*) on the complex of generic tuples of points in* $\mathbb{P}^{n-1}(K)$ *(resp.* $\mathbb{P}^{n-1}_+(K)$*). Then*

$$U \simeq \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \mathbb{Z} & \text{if } n \text{ is even,} \end{cases} \qquad U_+ \simeq \mathbb{Z}^{\lfloor n/2 \rfloor + 1}.$$

The $\mathrm{PGL}(n,K)$*– and* $\mathrm{P}_+\mathrm{GL}(n,K)$*–structures are given by*

$$[g]\cdot u = \begin{cases} u & \text{if } \det g > 0, \\ -u & \text{if } \det g < 0, \end{cases} \qquad (g \in \mathrm{GL}(n,K)).$$

Proof Both calculations are based on Fact 2.4. We denote by $(e_1,\dots,e_n)$ the standard basis of K^n.

Calculation of U Nondegenerate ordered simplices of X correspond to generic tuples of points in $\mathbb{P}$.

Lemma 8.2 *The action of PG_+ on the set of generic $(n+1)$–tuples of points in $\mathbb{P}$ has one orbit for n odd and two orbits for n even.*

Proof Let $p = (p_1,\dots,p_{n+1})$ be a generic $(n+1)$–tuple of points in $\mathbb{P}$. There is an element $g \in G$ (unique up to scaling) that maps p to the standard tuple $e = \bigl([e_1],\dots,[e_n],\bigl[\sum_{i=1}^n e_i\bigr]\bigr)$. If n is odd then $\det(-g) = -\det(g)$; therefore g may be chosen in G_+. It follows that in this case PG_+ acts transitively on the set of generic $(n+1)$–tuples. For n even, all elements g mapping p to e have determinants of the same sign. This sign is a PG_+–invariant of p that we call the sign of p and denote by $\mathrm{sgn}(p)$. Generic tuples of the same sign are PG_+–equivalent: if $\mathrm{sgn}(p) = \mathrm{sgn}(p')$, $gp = e$ and $gp' = e$, then $g^{-1}g'p' = p$ and $\mathrm{sgn}\det(g^{-1}g') = +1$. □

The case of n odd is now straightforward. The image of any n–simplex of X in $(C_nX)_{PG_+}$ is one and the same generator of that cyclic group. The boundary of an $(n+1)$–simplex of X is an alternating sum of an odd number of n–simplices; hence its image in $(C_nX)_{PG_+}$ is again that generator. It follows that $U = 0$ for n odd.

Suppose now that n is even. Lemma 8.2 and Fact 2.4 imply that U is the quotient of the free abelian group with two generators by two sets of relations (boundary relations and alternation relations). The generators correspond to (representatives of) PG_+–orbits on the set of generic $(n+1)$–tuples of points in $\mathbb{P}$; explicitly, we choose

$$\left([e_1],\dots,[e_n],\left[\sum_{i=1}^n e_i\right]\right) \tag{8-2}$$

and denote it by $[+]$ or $[+1]$, and

$$\left([e_1],\dots,[e_{n-1}],[-e_n],\left[\left(\sum_{i=1}^{n-1} e_i\right) - e_n\right]\right) \tag{8-3}$$

and denote it by $[-]$ or $[-1]$. We call $[+]$ and $[-]$ symbols.

The group $\mathbb{Z}^2$ generated by $[+]$ and $[-]$ is isomorphic to $(C_n^o X)_{PG_+}$. The image in this group of an ordered n–simplex of X corresponding to a generic $(n+1)$–tuple $p = (p_1, \dots, p_{n+1})$ of points in $\mathbb{P}$ is $[\operatorname{sgn}(p)]$. In practice, the sign can be calculated as follows: let $p_i = [v_i]$ for $v_i \in K^n$, and let $v_{n+1} = \sum_{i=1}^n \alpha_i v_i$; then

$$\operatorname{sgn}(p) = \operatorname{sgn}(\det(v_1, \dots, v_n) \cdot \alpha_1 \cdots \alpha_n) \tag{8-4}$$

(this is the sign of the determinant of the matrix mapping $(v_1, \dots, v_{n+1})$ to $\bigl(e_1, \dots, e_n, \sum_{i=1}^n e_i\bigr)$; that matrix is the inverse of the product of the matrix with columns $(v_1, \dots, v_n)$ and the diagonal matrix with diagonal entries $(\alpha_1, \dots, \alpha_n)$).

The alternation relations all reduce to $-[+] = [-]$. Indeed, it is straightforward to check that transposing two neighbouring elements in (8-2) or (8-3) changes the sign of the tuple.

We now discuss the boundary relations. We observe that any ordered nondegenerate $(n+1)$–simplex of X — corresponding to a generic $(n+2)$–tuple of points in $\mathbb{P}$ — can be mapped by an element of PG_+ to

$$\Delta = \biggl([e_1], [e_2], \dots, [e_{n-1}], [se_n], \biggl[\biggl(\sum_{i=1}^{n-1} e_i\biggr) + se_n\biggr], \biggl[\biggl(\sum_{i=1}^{n-1} b_i e_i\biggr) + sb_n e_n\biggr]\biggr).$$

Here $s = \pm 1$ and $b_i \in K$; genericity means that $b_i \neq 0$ and $b_i \neq b_j$. We have

$$\partial\Delta = \sum_{j=1}^{n+2} (-1)^{j-1} [s_j], \tag{8-5}$$

where s_j is the sign of the tuple obtained from Δ by omitting the j^{th} element. We have $s_{n+2} = s$ and $s_{n+1} = s \operatorname{sgn}\bigl(\prod b_i\bigr)$. We claim that, for $j < n+1$,

$$s_j = (-1)^j s \operatorname{sgn}\biggl(b_j \prod_{i \neq j} (b_i - b_j)\biggr). \tag{8-6}$$

Indeed, for $j < n$ we have

$$\operatorname{sgn}\det\biggl(e_1, \dots, \hat{e}_j, \dots, e_{n-1}, se_n, \sum_{i=1}^{n-1} e_i + se_n\biggr) = (-1)^{n-j} s = (-1)^j s,$$

$$\sum_{i=1}^{n-1} b_i e_i + sb_n e_n = b_j \biggl(\sum_{i=1}^{n-1} e_i + se_n\biggr) + \sum_{i \neq j,n} (b_i - b_j) e_i + (b_n - b_j) se_n,$$

while for $j = n$ we have

$$\operatorname{sgn}\det\biggl(e_1, \dots, e_{n-1}, \widehat{se}_n, \sum_{i=1}^{n-1} e_i + se_n\biggr) = s = (-1)^n s,$$

$$\sum_{i=1}^{n-1} b_i e_i + sb_n e_n = b_n \biggl(\sum_{i=1}^{n-1} e_i + se_n\biggr) + \sum_{i=1}^{n-1} (b_i - b_n) e_i.$$

Putting (8-5) and (8-6) together, we get

$$\partial\Delta = \sum_{j=1}^{n} (-1)^{j-1}\Big[(-1)^j s \operatorname{sgn} b_j \prod_{i\neq j} \operatorname{sgn}(b_i - b_j)\Big] + (-1)^n\Big[s \operatorname{sgn}\Big(\prod_i b_i\Big)\Big] - [s]. \tag{8-7}$$

We will show that this relation is trivial, ie that all the symbols cancel. We can assume $s = +1$; the case $s = -1$ will automatically follow. Indeed, changing s flips all the symbols: $[+1] \leftrightarrow [-1]$, and trivially transforms a trivial relation to a trivial relation.

Let us artificially put $b_{n+1} = 0$; then we may rewrite (8-7) more uniformly as

$$\partial\Delta = \sum_{j=1}^{n+1} (-1)^{j-1}\Big[\operatorname{sgn}\Big(\prod_{i=1}^{j-1}(b_j - b_i) \prod_{i=j+1}^{n+1} (b_i - b_j)\Big)\Big] - [+1]. \tag{8-8}$$

Let $\sigma \in S_{n+1}$ be the permutation ordering the numbers (indices) in the same way that the sequence b does: $\sigma(i) < \sigma(k) \iff b_i < b_k$. We put $\operatorname{inv}(j) = \#\{i \mid (i-j)(\sigma(i)-\sigma(j)) < 0\}$ (the number of inversions of σ in which j is involved). Then our relation is

$$\sum_{j=1}^{n+1} (-1)^{j-1}[(-1)^{\operatorname{inv}(j)}] - [+1]. \tag{8-9}$$

Lemma 8.3 $$(-1)^{\operatorname{inv}(j)} = (-1)^{\sigma(j)-j}.$$

Proof If exactly k of the indices smaller than j are mapped by σ to indices larger than $\sigma(j)$, then $\sigma(j) - j + k$ of the ones larger than j have to be mapped to values smaller than $\sigma(j)$. Then

$$\operatorname{inv}(j) = k + \sigma(j) - j + k \equiv \sigma(j) - j \pmod 2. \qquad \square$$

Thus, $\operatorname{inv}(j)$ is odd if and only if j and $\sigma(j)$ differ in parity. Since the number of even j's is equal to the number of even $\sigma(j)$'s, this difference in parity appears equally often in each of the two forms: $(j, \sigma(j)) = (\text{odd}, \text{even})$ and $(j, \sigma(j)) = (\text{even}, \text{odd})$. In (8-9), pairs of the first kind lead to summands $+[-]$, while pairs of the second kind give $-[-]$. Thus, all the appearances of the symbol $[-]$ cancel. It follows that the sum adds up to $[+]$, which is cancelled by the extra term.

We have shown that the boundary relations are trivial. It follows that U is the quotient of $\mathbb{Z}^2$ by the alternation relations, ie $U \simeq \mathbb{Z}$. The PG–structure description follows from the formula

$$\operatorname{sgn}(gp) = \operatorname{sgn}(\det g) \cdot \operatorname{sgn}(p),$$

valid for $g \in G$ and all generic $(n+1)$–tuples p of points in $\mathbb{P}$.

Calculation of U_+ A generic n–tuple of points in $\mathbb{P}_+$ can be lifted to n linearly independent vectors in K^n. The matrix M with columns given by these vectors is well defined up to multiplication on the right by diagonal matrices with positive entries on the diagonal. The sign of $\det M$ is thus an invariant of the tuple (the *sign of the tuple*); it is also a G_+–invariant of the tuple ($\det gM = \det g \cdot \det M = \det M$

for $g \in G_+$). In fact, this sign is the full G_+–invariant: the tuple of vectors is transformed to the standard basis by M^{-1} if $\det M > 0$, and to the basis $(e_1, \dots, e_{n-1}, -e_n)$ by M^{-1} with negated lowest row if $\det M < 0$. We have shown the following statement.

Lemma 8.4 *The action of G_+ on the set of generic n–tuples of points in $\mathbb{P}_+$ has two orbits, detected by the sign of the tuple.*

We now consider the G_+–action on the set of generic $(n+1)$–tuples. The symbol of such a tuple $(p_1, \dots, p_{n+1})$ is defined to be a sequence of $n+1$ signs, $[s; s_1, \dots, s_n]$. Here s is the sign of $(p_1, \dots, p_n)$. To get the other signs, we first lift each p_i to a vector v_i. Then we express v_{n+1} in terms of the other v_i, $v_{n+1} = \sum_{i=1}^{n} a_n v_n$. Finally, $s_i = \operatorname{sgn} a_i$ (genericity implies $a_i \neq 0$). Clearly, the symbol is G_+–invariant.

Lemma 8.5 *Two generic $(n+1)$–tuples of points in $\mathbb{P}_+$ are G_+–equivalent if and only if they have the same symbol.*

Proof Let the tuples be (p_i) and (q_i), with symbol $[s; s_1, \dots, s_n]$. Then a lift of (p_i) is equivalent to

$$\left(e_1, \dots, e_{n-1}, se_n, \sum_{i=1}^{n} a_i e_i\right)$$

for some $a_i \in \dot{K}$, while a lift of (q_i) is equivalent to

$$\left(e_1, \dots, e_{n-1}, se_n, \sum_{i=1}^{n} b_i e_i\right)$$

for some $b_i \in \dot{K}$; moreover, s_i is the (common) sign of a_i and of b_i (for $i < n$; and s_n is the common sign of sa_n and sb_n). These two representing tuples of vectors are projectively related by the diagonal matrix with positive diagonal entries b_i/a_i. □

The following observation describes the G–action on symbols and allows us to determine the P_+G–structure on U_+.

Fact 8.6 *The symbol of a generic $(n+1)$–tuple is G–equivariant: if $g \in G$ and $\det g < 0$, then the tuples (p_i) and (gp_i) have the opposite leading symbol sign s, and coinciding remaining symbol signs.*

It follows from Lemma 8.5 that the group U_+ is the quotient of the free abelian group spanned by symbols by alternation and boundary relations. We first deal with the alternation relations.

Alternation relations The symbol $[s; s_1, \dots, s_n]$ is represented by the tuple

$$\left(e_1, \dots, se_n, \sum_{i=1}^{n-1} s_i e_i + ss_n e_n\right).$$

Suppose that this tuple is permuted; what happens to the symbol? Since permutation commutes with "linear map applied to each element", we may and will assume $s = +1$ — in our arguments, but not in the final statements. We first treat the case of a permutation σ that fixes the last element. Then the symbol of the permuted tuple is $[\operatorname{sgn}\sigma; (s_{\sigma^{-1}(i)})]$. We get the "usual permutation relation"

$$[s; (s_i)] = \operatorname{sgn}\sigma [s \operatorname{sgn}\sigma; (s_{\sigma^{-1}(i)})].$$

Now let us consider the transposition of k and $n+1$. The new leading sign is

$$\det\left(e_1, \dots, e_{k-1}, \sum_{i=1}^{n} s_i e_i, e_{k+1}, \dots, e_n\right) = s_k.$$

We also have

$$e_k = s_k \sum_{i=1}^{n} s_i e_i + \sum_{i \neq k} (-s_k s_i) e_i,$$

so that the total symbol after transposition is

$$[s_k; -s_k s_1, \dots, -s_k s_{k-1}, s_k, -s_k s_{k+1}, \dots, -s_k s_n].$$

The "transposition relation" is thus

$$[s; (s_i)] = -[s s_k; -s_k s_1, \dots, -s_k s_{k-1}, s_k, -s_k s_{k+1}, \dots, -s_k s_n].$$

In words: If the k^{th} sign is $+$, then we can flip all the other signs (except the leading sign); the resulting symbol will be equal to minus the original. If the k^{th} sign is $-$, we get $[s; (s_i)] = -[-s; (s_i)]$.

There is a difference between the cases $n = 2$ and $n > 2$. In the latter case, for any sequence of n signs there exists a stabilizing transposition; therefore, any sequence of n signs can be ordered (put in the form $++\cdots--$) by an even permutation. Let us begin with the case $n > 2$.

The case $n > 2$ As already mentioned, in this case one can use the usual permutation relation to order the nonleading signs of a symbol without changing the leading sign. To shorten the notation, we will use a^+ for $[+; ++\cdots--]$ (a plus signs after the semicolon), and a^- for $[-; ++\cdots--]$ (a plus signs after the semicolon). For example, when $n = 3$, we put $0^+ = [+; ---]$, $2^+ = [+; ++-]$ and $2^- = [-; ++-]$. The transposition relation (with $s_k = +1$) gives $a^\pm = -(n-a+1)^\pm$ (for $a > 0$). Picking $s_k = -1$ in the transposition relation we get $a^+ = -a^-$ — for $a < n$, but $n^+ = -n^-$ also holds, due to $n^+ = -1^+ = 1^- = -n^-$. To summarize:

Lemma 8.7 *Let $n > 2$. Let $A = \{a^+ \mid 0 \le a \le \lfloor n/2 \rfloor\}$. The quotient of the group $(C_n^o X_+)_{P_+ G_+}$ by the set of alternation relations is the free abelian group with generating set A for n even; for n odd it is the direct sum of the free abelian group generated by A and a $\mathbb{Z}/2$ generated by $\left(\frac{n+1}{2}\right)^+$.*

(The extra $\mathbb{Z}/2$–summand appearing for n odd will eventually get killed by the boundary relations.)

The case $n = 2$ There are eight symbols. The usual permutation relation for $\sigma = (12)$ gives

(8-10) $[+; ++] = -[-; ++]$, $[-; -+] = -[+; +-]$, $[-; +-] = -[+; -+]$, $[+; --] = -[-; --]$.

The transposition relation (for $\sigma = (23)$) is $[s; s_1, s_2] = -[ss_2; -s_2 s_1, s_2]$. This gives

(8-11) $-[-;++] = [-;-+], \quad -[+;+-] = [-;+-], \quad -[+;-+] = [+;++], \quad [+;--] = -[-;--].$

We see that all the (signed) symbols appearing in the first three equalities of (8-10) and (8-11) are identified. In particular, $[+,+-] = [+;-+]$, so the $a^{\pm}$ notation still makes sense. Also, the relations $a^+ + a^- = 0$ and $1^+ + 2^+ = 0$ can be read off from the ones displayed above. Therefore the conclusion of Lemma 8.7 holds for $n = 2$.

Boundary relations We will show that they all follow from the alternation relations (with the exception of $\left(\frac{n+1}{2}\right)^+ = 0$). Let us calculate the boundary of an $(n+1)$–simplex of X_+ represented by a generic $(n+2)$–tuple of vectors. Such a tuple of vectors can be transformed by an element of G_+ to

$$\Delta = \left(e_1, \dots, e_{n-1}, se_n, \sum_{i=1}^n s_i e_i, \sum_{i=1}^n s_i b_i e_i\right).$$

The genericity condition (assuming $s = +1$) is that all b_i nonzero and pairwise different. (This will follow from the calculation of $\partial\Delta$.) If $s = -1$, we can transform the tuple by an orientation changing linear map; this will change all leading signs in $\partial\Delta$, and not touch the other signs. Thus, we will assume $s = +1$ — and then double the set of the resulting boundaries by changing the leading signs. If we omit e_j $(i \le n)$ from Δ, then the sign of the determinant of the standardizing matrix is the same as that of

$$\det\left(e_1, \dots, e_{j-1}, e_{j+1}, \dots, e_n, \sum_{i=1}^n s_i e_i\right) = (-1)^{n-j} s_j.$$

The other signs can be read off from

$$\sum_{i=1}^n s_i b_i e_i = \sum_{i \ne j} (s_i b_i - s_i b_j) e_i + b_j \sum_{i=1}^n s_i e_i.$$

The total symbol (for j omitted) is thus $[(-1)^{n-j} s_j; (s_i \operatorname{sgn}(b_i - b_j))_{i \ne j}, \operatorname{sgn} b_j]$.

Omitting the $(n+1)^{\text{st}}$ element gives $[+1; (s_i \operatorname{sgn} b_i)]$.

Omitting the $(n+2)^{\text{nd}}$ element yields $[+1; (s_i)]$. So, finally,

$$\partial\Delta = \sum_{j=1}^n (-1)^{j-1} [(-1)^{n-j} s_j; (s_i \operatorname{sgn}(b_i - b_j))_{i \ne j}, \operatorname{sgn} b_j] + (-1)^n [+1; (s_i \operatorname{sgn} b_i)] + (-1)^{n+1} [+1; (s_i)].$$

Let us rewrite this boundary relation. We put (artificially) $b_{n+1} = 0$ and $s_{n+1} = -1$. Then we have $\operatorname{sgn} b_j = s_{n+1} \operatorname{sgn}(b_{n+1} - b_j)$, $s_i \operatorname{sgn} b_i = s_i \operatorname{sgn}(b_i - b_{n+1})$ and $(-1)^{n-(n+1)} s_{n+1} = +1$, so

$$\begin{aligned}\partial\Delta &= \sum_{j=1}^{n+1} (-1)^{j-1} [(-1)^{n-j} s_j; (s_i \operatorname{sgn}(b_i - b_j))_{i \ne j}] + (-1)^{n+1} [+1; (s_i)] \\ &= \sum_{j=1}^{n+1} (-1)^{n-1} [s_j; (s_i \operatorname{sgn}(b_i - b_j))_{i \ne j}] + (-1)^{n+1} [+1; (s_i)],\end{aligned}$$

where the last equality uses $a^+ = -a^-$. Thus, we need to deduce from our alternation relations (assuming $s_{n+1} = -1$ and $b_{n+1} = 0$) that

$$\sum_{j=1}^{n+1} [s_j; (s_i \operatorname{sgn}(b_i - b_j))_{i \neq j}] + [+1; (s_i)_{i \neq n+1}] = 0.$$

We see that we may change the order of summation to follow the order of the b_i. Indeed, the above summation can be phrased in an index-free way as follows. We have a set of $n+1$ numbers, each with an attached sign (one of these pairs being 0 with $-$). For each element x of the set we form the corresponding symbol k^s, where s is the sign attached to the element x, and k is the number of positive expressions of the form $t(y-x)$, where y runs through our set (and $y \neq x$) and t is the sign attached to y. Finally, there is an extra summand ℓ^+ with ℓ counting all the positive signs.

We may thus renumber the b_i and the s_i (in the same way) so as to have $b_1 > b_2 > \dots > b_n > b_{n+1}$, with an unknown b equal to zero and the corresponding s equal to -1 and omitted in the extra summand $[+1; (s_i)]$. Let us now consider two consecutive summands (numbered j and $j+1$). They differ at most by the leading sign, s_j versus s_{j+1}, and by the j^{th} nonleading sign, $s_{j+1} \operatorname{sgn}(b_{j+1} - b_j) = -s_{j+1}$ versus $s_j \operatorname{sgn}(b_j - b_{j+1}) = s_j$. Substituting all four possible combinations of (s_j, s_{j+1}) we get:

Claim *Two consecutive summands are of one of the forms*

$$(a^\pm, a^\mp), \quad (a^+, (a+1)^+), \quad (a^-, (a-1)^-).$$

Suppose that k of the s_i are positive. Then the extra summand is k^+, while the first and the last one depend on (s_1, s_{n+1}) and are:

s_1	s_{n+1}	first	last
$+$	$+$	$(n-k+1)^+$	$(k-1)^+$
$+$	$-$	$(n-k+1)^+$	k^-
$-$	$+$	$(n-k)^-$	$(k-1)^+$
$-$	$-$	$(n-k)^-$	k^-

We can append the extra summand k^+ to the sum (while keeping the rule of the claim) and get summation starting from $(n-k+1)^+$ or $(n-k)^-$ and ending at k^+. Then we start cancelling consecutive pairs $(a^\pm, a^\mp)$ (except that we do not cancel the first and the last element) until the sequence becomes monotone (possibly except the first or the last pair). If the final monotone sequence runs from $(n-k+1)^+$ to k^+ then the terms pairwise cancel (first with last, second with last-but-one, etc) if n is even, and $\frac{1}{2}(n+1)$ is left if n is odd. If the sequence starts with $(n-k)^-$, we may put an extra pair $((n-k+1)^+, (n-k+1)^-)$ at the beginning of the sequence, to reduce to the former case — except when $k=0$. If $k=0$, we get a sequence running from n^- to 0^+, ie $(n^-, (n-1)^-, \dots, 1^-, 0^-, 0^+)$. The first n terms cancel in the same manner as before, and $0^- + 0^+ = 0$.

Finally, since the set of permutation relations is invariant under the "exponent sign" flip ($a^{\pm} \leftrightarrow a^{\mp}$), the boundary relations obtained from Δ with $s = -1$ are dealt with in the same way.

Fact 8.6 and the relation $a^- = -a^+$ imply the remaining claim (the one describing the P_+G–structure on U_+), completing the proof of Theorem 8.1. □

Remark 8.8 Let T be the tautological (U–valued) n–cocycle associated to the PG_+–action on X, and let T_+ be the tautological (U_+–valued) n–cocycle associated to the P_+G_+–action on X_+. From the proof of Theorem 8.1 it is useful to extract the following explicit description of these cocycles.

(a) Let n be even; then $U \simeq \mathbb{Z}$ is generated by the symbol $[+]$. Suppose that $\sigma = ([v_1], \dots, [v_{n+1}])$ is an n–simplex of X. Then $v_{n+1} = \sum_{i=1}^n \alpha_i v_i$ for some $\alpha_i \in \dot{K}$. We have (see (8-4))

$$T(\sigma) = [\operatorname{sgn}(\det(v_1, \dots, v_n) \cdot \alpha_1 \cdots \alpha_n)].$$

(Recall that $[-] = -[+]$.)

(b) Recall that $U_+ \simeq \mathbb{Z}^{\lfloor n/2 \rfloor + 1}$ with free generating set $A = \{a^+ \mid a = 0, \dots, \lfloor n/2 \rfloor\}$ (see Lemma 8.7). Suppose that $\sigma = ([v_1], \dots, [v_{n+1}])$ is an n–simplex of X_+. Then $v_{n+1} = \sum_{i=1}^n \alpha_i v_i$ for some $\alpha_i \in \dot{K}$. To σ we assign an $(n+1)$–tuple of signs $[s; s_1, \dots, s_n]$, where $s = \operatorname{sgn}\det(v_1, \dots, v_n)$ and $s_i = \operatorname{sgn}\alpha_i$. Next we put $T_+(\sigma) = a^+$ (if $s = +1$ and a of the s_i are $+1$) or $T_+(\sigma) = a^-$ (if $s = -1$ and a of the s_i are $+1$). Finally, we express the symbol in term of the elements of A using the relations $a^+ = -a^-$ and $a^{\pm} = -(n+1-a)^{\pm}$ (for $a > 0$).

Definition 8.9 The splitting of $U_+ = \bigoplus_{k=0}^{\lfloor n/2 \rfloor} \mathbb{Z}k^+$ into cyclic summands generated by the elements k^+ ($0 \le k \le \lfloor n/2 \rfloor$) induces the corresponding splittings of cocycles and cohomology classes:

$$\begin{aligned} T_+ &= \bigoplus T_k, && T_k \in Z^n(\operatorname{Hom}_{P_+G_+}(C_* X_+, \mathbb{Z})); \\ \mathrm{eu}_+ &= \bigoplus \mathrm{eu}_k, && \mathrm{eu}_k \in H^n(P_+G_+, \mathbb{Z}); \\ \tau_+ &= \bigoplus \tau_k, && \tau_k \in H^n(\operatorname{Hom}_{P_+G_+}(C_* X_+, \mathbb{Z})). \end{aligned}$$

(In the last formula, τ_+ (τ_k) is the cohomology class of T_+ (T_k).)

Remark 8.10 Suppose that $K < L$ is a field extension, and that on K and on L there are compatible field orders. Then we have the group embedding $\phi\colon P_+G_+(K) \to P_+G_+(L)$, and the natural ϕ–equivariant simplicial complex embedding $X_+(K) \to X_+(L)$ inducing a coefficient group map $f : U_+(K) \to U_+(L)$. But, in our field-independent description of U_+ (see Theorem 8.1 and Remark 8.8) the map f is represented by the identity. Applying Theorem 1.5 to these data we obtain $\phi^*\mathrm{eu}_+(L) = \mathrm{eu}_+(K)$ — the Euler class eu_+ is stable under ordered field restriction. It follows that all eu_k are also stable. Analogous arguments show the same stability statement for eu.

Remark 8.11 It follows from Theorem 3.1 that the classes eu, eu_+ and eu_k are bounded.

9 Relation between the classes eu_k

The classes eu_k defined in Definition 8.9 are related.

Theorem 9.1
$$\sum_{k=0}^{\lfloor n/2\rfloor}(n-2k+1)\mathrm{eu}_k = 0.$$

Proof We will see that this relation holds already in $H^n(\mathrm{Hom}_{P_+G_+}(C_*X_+,\mathbb{Z}))$ for the classes τ_k. To prove it, we will find a cochain $c \in C^{n-1}(\mathrm{Hom}_{P_+G_+}(C_*X_+,\mathbb{Z}))$ such that

$$\delta c = \sum_{k=0}^{\lfloor n/2\rfloor}(n-2k+1)T_k \tag{9-1}$$

in $C^n(\mathrm{Hom}_{P_+G_+}(C_*X_+,\mathbb{Z}))$. The boundary map

$$\partial\colon (C_nX_+)_{P_+G_+} \to (C_{n-1}X_+)_{P_+G_+}$$

factors as the composition of the projection $(C_nX_+)_{P_+G_+} \to (C_nX_+/B_nX_+)_{P_+G_+} = U_+$ and a map $\partial'\colon U_+ \to (C_{n-1}X_+)_{P_+G_+}$. Each T_k also factors — as the composition of the same projection and the projection T'_k of U_+ on the k^+–summand. Recall that $(C_{n-1}X_+)_{P_+G_+} \cong \mathbb{Z}$ (with generator $[+]$; see Lemma 8.4). We now consider a generator a^+ of U_+ and determine $\partial'(a^+)$. Let

$$v_a = e_1+\cdots+e_a-(e_{a+1}+\cdots+e_n);$$

then $(e_1,\dots,e_n,v_a)$ determines a simplex in X_+ representing a^+. We have

$$\begin{aligned}
\partial'(a^+) &= \partial[e_1,\dots,e_n,v_a]\\
&= \sum_{j=1}^{n}(-1)^{j+1}[e_1,\dots,\hat{e}_j,\dots,e_n,v_a] + (-1)^n[e_1,\dots,e_n]\\
&= \sum_{j=1}^{n}(-1)^{j+1}(-1)^{n-j}[e_1,\dots,v_a,\dots,e_n] + (-1)^n[+]\\
&= \sum_{j=1}^{a}(-1)^{n+1}[+] + \sum_{j=a+1}^{n}(-1)^{n+1}[-] + (-1)^n[+]\\
&= (-1)^n((1-a)[+]-(n-a)[-]) = (-1)^n(n-2a+1)[+].
\end{aligned}$$

Let $c \in C^{n-1}(\mathrm{Hom}_{P_+G_+}(C_*X_+,\mathbb{Z})) = \mathrm{Hom}((C_{n-1}X_+)_{P_+G_+},\mathbb{Z})$ be defined by $c([+]) = (-1)^n$. Then $(c\circ\partial')(a^+) = c(\partial'a^+) = (n-2a+1) = \sum_{k=0}^{\lfloor n/2\rfloor}(n-2k+1)T'_k(a^+)$ holds for each a^+. Formula (9-1) follows. □

10 The Smillie argument

The Smillie argument (see [Gromov 1982, Section 1.3]) can be used to show that the classes eu_k are proportional in a weak sense.

Theorem 10.1 *For any $h \in H_n(BP_+G_+, \mathbb{Z})$ (or $h \in H_n(BP_+G_+, \mathbb{Z}/m)$ for m odd) and any $k \le \lfloor n/2 \rfloor$,*

$$\langle \mathrm{eu}_k, h\rangle = (-1)^k \binom{n+1}{k} \langle \mathrm{eu}_0, h\rangle.$$

If n is odd, then $\langle \mathrm{eu}_k, h\rangle = 0$ for all k.

Proof It is well known that there exists a finite simplicial complex Y, a simplicial cycle $Z \in Z_n(Y, \mathbb{Z})$ (or in $Z_n(Y, \mathbb{Z}/m)$), and a map $f : Y \to BG$, such that $f_*[Z] = h$. Let $P = f^* EP_+G_+$ (the pullback of the universal bundle over BP_+G_+). Then

$$\langle \mathrm{eu}_k, h\rangle = \langle \mathrm{eu}_k, f_*[Z]\rangle = \langle f^* \mathrm{eu}_k, [Z]\rangle = \langle \mathrm{eu}_k(P), [Z]\rangle$$

for each k. We will use Theorem 4.4 to compute $\langle \mathrm{eu}_k(P), [Z]\rangle$. Let $E = P \times_{P_+G_+} \mathbb{P}_+$ be the associated bundle.

Pick a generic section $s : Y^{(0)} \to E$. Genericity means that the values of the section at the vertices of any n–simplex of Y, viewed as points in $\mathbb{P}_+$ via a flat trivialization of E over that simplex, form a generic tuple of points. Such a section can be picked vertex-by-vertex. At a vertex y the genericity conditions mean that a certain finite union of proper projective subspaces of E_y is prohibited; since the ordered field K is infinite, that union does not fill out E_y and a generic choice is possible. Any generic section s determines a simplicial section of the associated X_+–bundle over Y, and then Theorem 4.4 may be applied.

For any function $\epsilon : Y^{(0)} \to \{\pm 1\}$ we can form a modified section $\epsilon s : Y^{(0)} \to E$ defined in the obvious way: if $s(v) = (p, [q])$ (for some $q \in K^n \setminus \{0\}$), then $(\epsilon s)(v) = (p, [\epsilon(v) q])$. Every section ϵs is again generic. Theorem 4.4 gives $\mathrm{eu}_+(P) = [(\epsilon s)^* T_+]$, and coefficient splitting allows us to deduce $\mathrm{eu}_k(P) = [(\epsilon s)^* T_k]$; both formulae hold for all functions ϵ. For a given n–simplex σ of X_+ we will average the expression $\langle (\epsilon s)^* T_k, \sigma\rangle$ over all possible functions ϵ.

Let $\sigma = (v_1, \dots, v_n, v_{n+1})$ be one of the n–simplices of Y. Let us choose a flat section r of P over σ, and let $s(v_i) = [r, [q_i]]$, for $q_i \in K^n \setminus \{0\}$. We denote by $s_*\sigma$ the n–simplex of X_+ given by $([q_1], \dots, [q_n], [q_{n+1}])$. This definition depends on the choice of r, but different choices lead to simplices equivalent under the P_+G_+–action. The expression $\langle T_k, s_*\sigma\rangle$ is well defined and equal to $\langle s^* T_k, \sigma\rangle$. Let $\eta = \operatorname{sgn}\det(q_1, \dots, q_n)$, and let $q_{n+1} = \sum_{i \le n} a_i q_i$. Suppose that exactly ℓ of the coefficients a_i are positive — so that the symbol of $s_*\sigma$ is ℓ^η.

For any function ϵ we have $(\epsilon s)_*\sigma = ([\epsilon_1 q_1], \dots, [\epsilon_n q_n], [\epsilon_{n+1} q_{n+1}])$, where $\epsilon_i = \epsilon(v_i)$. We wish to determine all functions ϵ such that $\langle T_k, (\epsilon s)_*\sigma\rangle \ne 0$. This will happen if and only if the decomposition $\epsilon_{n+1} q_{n+1} = \sum_{i \le n} b_i \epsilon_i q_i$ has either k or $n + 1 - k$ positive coefficients b_i.

Let us first focus on the case of $\epsilon_{n+1} = +1$ and k positive b_i's. We will represent the appropriate functions ϵ in the form $\epsilon'\epsilon''$; the idea is that ϵ' makes all the nonleading signs negative, while ϵ'' changes k of them to $+$. In more detail: $\epsilon'_{n+1} = +1$ and $\epsilon'_i = -\operatorname{sgn} a_i$ for $i \le n$, while ϵ'' is arbitrary with k negative and $n-k$ positive values (plus $\epsilon''_{n+1} = +1$). For such ϵ' and ϵ'' the symbol of $(\epsilon'\epsilon'' s)_*\sigma$ is $k^\pm$, where the exponent is $\prod_{i\le n} \epsilon'_i \prod_{i \le n} \epsilon''_i \cdot \operatorname{sgn}\det(q_1, \dots, q_n) = (-1)^\ell (-1)^k \eta$. There are $\binom{n}{k}$ appropriate functions ϵ.

For $\epsilon_{n+1} = +1$ and $n+1-k$ positive b_i's we get $\binom{n}{n+1-k} = \binom{n}{k-1}$ possibilities, yielding $(n+1-k)^\pm$, with the exponent equal to $(-1)^\ell(-1)^{n+1-k}\eta = -(-1)^n(-1)^\ell(-1)^k\eta$.

If $\epsilon_{n+1} = -1$ the analysis is analogous. The difference is that ϵ' should now be $\epsilon'_i = \operatorname{sgn} a_i$; therefore, the only change is $(-1)^{n-\ell}$ instead of $(-1)^\ell$ in the final exponent sign formulae.

Putting these together, we get (with $N = \#Y^{(0)}$)

$$\begin{aligned}\Big\langle T_k, \sum_\epsilon (\epsilon s)_*\sigma \Big\rangle &= \Big\langle T_k, 2^{N-(n+1)} \Big(\binom{n}{k} k^{(-1)^\ell(-1)^k\eta} + \binom{n}{k-1}(n+1-k)^{-(-1)^n(-1)^\ell(-1)^k\eta} \\ &\qquad + \binom{n}{k} k^{(-1)^{n-\ell}(-1)^k\eta} + \binom{n}{k-1}(n+1-k)^{-(-1)^n(-1)^{n-\ell}(-1)^k\eta}\Big)\Big\rangle \\ &= 2^{N-(n+1)}\Big\langle T_k, \Big(\binom{n}{k}((-1)^\ell(-1)^k\eta + (-1)^{n-\ell}(-1)^k\eta)k^+ \\ &\qquad - \binom{n}{k-1}((-1)^n(-1)^\ell(-1)^k\eta + (-1)^n(-1)^{n-\ell}(-1)^k\eta)(n+1-k)^+\Big)\Big\rangle \\ &= 2^{N-(n+1)}\Big\langle T_k, \Big(\binom{n}{k}(1+(-1)^n)(-1)^\ell(-1)^k\eta k^+ \\ &\qquad - \binom{n}{k-1}(1+(-1)^n)(-1)^n(-1)^\ell(-1)^k\eta(n+1-k)^+\Big)\Big\rangle.\end{aligned}$$

For n odd, this is zero. Thus, we assume n even; then the coefficients in the above expression add up to

$$2^{N-n}\Big(\binom{n}{k} + \binom{n}{k-1}\Big)(-1)^k(-1)^\ell\eta = 2^{N-n}\binom{n+1}{k}(-1)^k(-1)^\ell\eta.$$

Similarly, if $n+1-\ell$ of the coefficients a_i are positive, we get

$$2^{N-n}\binom{n+1}{k}(-1)^k(-1)^{n+1-\ell}\eta.$$

In both cases, the result can be expressed as

$$2^{N-n}\binom{n+1}{k}(-1)^k\langle(-1)^\ell T_\ell, s_*\sigma\rangle.$$

Since on any $s_*\sigma$ exactly one of T_ℓ is nonzero, we can rewrite this formula as

$$2^{N-n}\binom{n+1}{k}(-1)^k\sum_\ell\langle(-1)^\ell T_\ell, s_*\sigma\rangle,$$

with summation over $\ell \le n/2$. Let us summarize:

$$\left\langle \sum_{\epsilon} (\epsilon s)^* T_k, \sigma \right\rangle = 2^{N-n} \binom{n+1}{k} (-1)^k \sum_{\ell} \langle (-1)^\ell s^* T_\ell, \sigma \rangle.$$

It follows that

$$\sum_{\epsilon} (\epsilon s)^* T_k = 2^{N-n} \binom{n+1}{k} (-1)^k \sum_{\ell} (-1)^\ell s^* T_\ell.$$

Now recall that, by Theorem 4.4, each $(\epsilon s)^* T_k$ is a cocycle representing the cohomology class $\mathrm{eu}_k(P)$. Therefore

$$2^N \mathrm{eu}_k(P) = 2^{N-n} \binom{n+1}{k} (-1)^k \sum_{\ell} (-1)^\ell \mathrm{eu}_\ell(P).$$

Comparing this formula for $k = 0$ and for any other value of k we get the following lemma (which may be regarded as a variant of Theorem 10.1).

Lemma 10.2 *Let P be a flat principal P_+G_+–bundle over a finite simplicial complex Y that has N vertices. Then*

$$2^N \mathrm{eu}_k(P) = 2^N (-1)^k \binom{n+1}{k} \mathrm{eu}_0(P).$$

Evaluating both sides of the formula from the lemma on $[Z]$ concludes the proof of Theorem 10.1. □

Corollary 10.3 *Let P be a $\mathrm{P}_+\mathrm{GL}_+(n, K)$–bundle over an even-dimensional manifold M^n. Then any triangulation of M has at least $2^n |\langle \mathrm{eu}_0(P), [M] \rangle|$ simplices of dimension n.*

Proof Pick a generic section s, over the given triangulation, of the associated bundle with fibre $\mathbb{P}_+$. Then, by Theorem 10.1, $|\langle s^* T_k, [M] \rangle| = |\langle \mathrm{eu}_k(P), [M] \rangle| = \binom{n+1}{k} |\langle \mathrm{eu}_0(P), [M] \rangle|$. Since $|\langle s^* T_k, \sigma \rangle| \le 1$ and the supports of the cocycles $s^* T_k$ are pairwise disjoint, the number of n–simplices of the triangulation is at least

$$\sum_{k=0}^{\lfloor n/2 \rfloor} |\langle s^* T_k, [M] \rangle| = |\langle \mathrm{eu}_0(P), [M] \rangle| \cdot \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n+1}{k} = 2^n |\langle \mathrm{eu}_0(P), [M] \rangle|. \qquad \square$$

11 Cross product of Euler classes

It will be convenient to put $[n] = \{0, 1, \dots, n\}$. We will use groups $\mathrm{GL}_+(n, K)$ for varying n; therefore we denote U_+ by $U_{n,+}$ in this and the next section.

Theorem 11.1 *Let E and E' be $\mathrm{GL}_+(n, K)$– and $\mathrm{GL}_+(k, K)$–bundles over simplicial complexes X and X' respectively. Let $E \times E'$ be the product bundle over $X \times X'$. For any simplicial cycles $z \in Z_n(X, \mathbb{Z})$ and $z' \in Z_k(X', \mathbb{Z})$,*

$$\langle \mathrm{eu}_0(E), z \rangle \cdot \langle \mathrm{eu}_0(E'), z' \rangle = \langle \mathrm{eu}_0(E \times E'), z \times z' \rangle. \tag{11-1}$$

Proof We first explain the general strategy of the proof. We may and do assume that $X = \mathrm{supp}\, z$ and $X' = \mathrm{supp}\, z'$. We triangulate $X \times X'$ subdividing each product of simplices $\sigma \times \sigma'$ in a standard way (to be recalled later). It is also convenient to treat E, E' and $E\times E'$ not as principal bundles, but as vector bundles; eg the fibre E_x will be an n–dimensional vector space over K. We pick generic sections s of E and s' of E', and combine them to a section S of $E\times E'$. To ensure genericity of S we impose stronger than usual, weird-looking genericity conditions on s and on s'. The section s induces a simplicial section s_+ of the associated $X_{n,+}$–bundle, where $X_{n,+}$ is the complex of generic tuples of points in $\mathbb{P}^{n-1}_+(K)$. Then, by Theorem 4.4, we get cocycles $s_+^*T_+$ and $s_+^*T_0$ representing $\mathrm{eu}_+(E)$ and $\mathrm{eu}_0(E)$. For each n–simplex σ in X we have $\langle s_+^*T_+,\sigma\rangle = k^{\pm}$ for some k; if $k=0$ then $\langle s_+^*T_0,\sigma\rangle=\pm1$, otherwise $\langle s_+^*T_0,\sigma\rangle=0$. Similarly, we have cocycles $s_+'^*T_0$ and $S_+^*T_0$ representing $\mathrm{eu}_0(E')$ and $\mathrm{eu}_0(E\times E')$. Suppose that $z=\sum_\sigma n_\sigma\sigma$ and $z=\sum_{\sigma'} n_{\sigma'}\sigma'$; then $z\times z'=\sum_{\sigma,\sigma'} n_\sigma n_{\sigma'}\cdot\sigma\times\sigma'$, where $\sigma\times\sigma'$ denotes the chain representing the standard subdivision of the product of simplices. We have

$$\langle \mathrm{eu}_0(E),z\rangle=\sum_\sigma n_\sigma\langle s_+^*T_0,\sigma\rangle,\quad \langle \mathrm{eu}_0(E'),z'\rangle=\sum_{\sigma'} n_{\sigma'}\langle s_+'^*T_0,\sigma'\rangle,$$

$$\langle \mathrm{eu}_0(E\times E'),z\times z'\rangle=\sum_{\sigma,\sigma'} n_\sigma n_{\sigma'}\langle S_+^*T_0,\sigma\times\sigma'\rangle.$$

Thus, to establish the theorem it is enough to show that

$$\langle s_+^*T_0,\sigma\rangle\cdot\langle s_+'^*T_0,\sigma'\rangle=\langle S_+^*T_0,\sigma\times\sigma'\rangle. \tag{11-2}$$

We do this step-by-step. In Corollary 11.4 we show that if the left-hand side of (11-2) is zero, then so is the right-hand side. In Corollary 11.5 we show that if the left-hand side is nonzero, then in the chain $\sigma\times\sigma'$ there is a unique summand (unique $(n+k)$–simplex) on which $S_+^*T_0$ evaluates to ±1. Finally, in Lemma 11.6 we check that the sign of that evaluation is consistent with (11-2).

We proceed to the details. First we pick a generic section s of E over $X^{(0)}$. The genericity condition is as follows. For each ℓ–simplex $\sigma=(x_0,\dots,x_\ell)$ of X ($\ell<n$), the vectors $(s(x_0),\dots,s(x_\ell))$ are linearly independent. For each n–simplex $\sigma=(x_0,\dots,x_n)$, if $\sum_{i=0}^n\alpha_i s(x_i)=0$ is a nontrivial linear relation (projectively unique, because of the previous condition), then $\sum_{i=0}^n\alpha_i\neq0$. To make sense of these conditions we choose a (flat) trivialization of E over σ.

This kind of section can be chosen vertex-by-vertex. Let $X^{(0)}=(x_1,x_2,\dots,x_N)$. First, we choose any nonzero $s(x_1)\in E_{x_1}$. When $s(x_1),\dots,s(x_{i-1})$ have been chosen, we choose (flat) trivializations over all simplices with vertex x_i. If $\sigma=(x_i,y_1,\dots,y_\ell)$ is an ℓ–simplex of X ($\ell<n$) such that $s(y_1),\dots,s(y_\ell)$ have already been chosen, we use the trivialization of E over σ to transport all $s(y_j)$ to E_{x_i}. There, these vectors span a linear subspace E_i^σ of dimension $\ell<n$. We have to ensure $s(x_i)\notin E_i^\sigma$ in order to fulfill the first genericity condition for σ.

Next, for each simplex $\sigma=(x_i,y_1,\dots,y_n)$ with $s(y_1),\dots,s(y_n)$ already chosen we pick a (flat) trivialization of E over σ and use it to transport the $s(y_j)$ to vectors $s_j^\sigma\in E_{x_i}$. Then we form an affine subspace

$$E_i^\sigma=\{\alpha_1 s_1^\sigma+\dots+\alpha_n s_n^\sigma\mid \alpha_1+\dots+\alpha_n=1\}.$$

We have to choose $s(x_i)$ outside of this subspace in order to fulfill the second genericity condition for σ.

A linear space over an ordered (hence infinite) field is not a union of finitely many proper affine subspaces. Therefore, $s(x_i)$ can be suitably chosen. By induction, there exists a generic section s of E over $X^{(0)}$.

With the section s we associate a collection of scalars $\mathscr{A}$. For each n–simplex $\sigma = (x_0, \dots, x_n)$ of X let $\sum_{i=0}^n \alpha_i s(x_i) = 0$ be the linear relation with $\sum_{i=0}^n \alpha_i = 1$ (in some trivialization of E over σ). For every proper nonempty subset of $[n]$ we sum the corresponding α_i's. The set $\mathscr{A}$ is the collection of all such sums (over all n–simplices).

Now, analogously, we choose a generic section s' of E' over $X'^{(0)}$. It has its own collection of scalars $\mathscr{A}'$. We want $\mathscr{A}$ and $\mathscr{A}'$ to be disjoint. To this end, we perform the above section-choosing procedure for E' with supplementary restrictions. Suppose that we are at step i, choosing $s'(x'_i)$. There is a collection of proper affine subspaces in $E'_{x'_i}$ that we need to avoid; we now describe an additional finite collection, that will enforce our extra "joint genericity" condition. Let $\sigma' = (x'_i, y'_1, \dots, y'_k)$ be a k–simplex of X', such that $s'(y'_j)$ are already chosen, and let $s_j^{\sigma'}$ be $s'(y'_j)$ transported to $E'_{x'_i}$ via a chosen trivialization of E' over σ'. For any generic (in the previous sense) $s'_i = s'(x'_i)$ there is a unique relation $\beta_0 s'_i + \sum_{j=1}^k \beta_j s_j^{\sigma'} = 0$ satisfying $\sum_{j=0}^k \beta_j = 1$. Pick an $\alpha \in \mathscr{A}$ and a proper nonempty $J \subset [k]$; we want to ensure that $\sum_{j \in J} \beta_j \neq \alpha$. Let us express this as a restriction for the possible position of s'_i. Suppose that $s'_i = \sum_{j=1}^k \gamma_j s_j^{\sigma'}$, and that $\sum_{j \in J} \beta_j = \alpha$. Let us express β_j in terms of the γ_j. By the original genericity requirement we know that $\Gamma := -1 + \sum_{j=1}^k \gamma_j \neq 0$; therefore

$$-\frac{1}{\Gamma} s'_j + \sum_{i=1}^k \frac{\gamma_i}{\Gamma} s_j^{\sigma'} = 0,$$

so $\beta_0 = -1/\Gamma$ and $\beta_j = \gamma_j/\Gamma$. Thus, the condition $\sum_{j \in J} \beta_j = \alpha$ can be rewritten in terms of the γ_j (putting $\gamma_0 = -1$): $\sum_{j \in J} \gamma_j/\Gamma = \alpha$, or $\sum_{j \in J} \gamma_j = \alpha\Gamma$, or finally

$$\sum_{j=0}^k (\alpha - \delta_J(j))\gamma_j = 0.$$

Since J is proper and nonempty, regardless of the value of α the set of vectors $\sum_{j=1}^k \gamma_j s_j^{\sigma'}$ for γ_j satisfying this condition forms a proper affine subspace of $E'_{x'_i}$. (The two special suspect cases, $J = \{0\}$ with $\alpha = 0$ and $J = \{1, \dots, k\}$ with $\alpha = 1$, are easily seen to be impossible.) Thus, the extra genericity conditions produce a new finite collection of proper affine subspaces to avoid, so that it is possible to fulfill them.

Assume then that we have chosen jointly generic (in the above sense) sections — s of E and s' of E'. We now form a generic section S of $E \times E'$ over $(X \times X')^{(0)}$ by $S(x, x') = (s(x), s'(x'))$. To claim genericity, we need to describe the (standard) triangulation of $X \times X'$. We choose some total orders on $X^{(0)}$ and on $X'^{(0)}$, and order each simplex of $X^{(n)}$ and of $X'^{(k)}$ accordingly. Let $\sigma = (x_0, \dots, x_n) \in X^{(n)}$, and let $\sigma' = (x'_0, \dots, x'_k) \in X'^{(k)}$. Let $x_{(i,j)} = (x_i, x'_j)$. We form the $n \times k$ integer grid — with vertex set $[n] \times [k]$ and edges connecting pairs that differ on exactly one coordinate and exactly by 1. Shortest paths from $(0, 0)$ to (n, k) will be called *admissible*. ("Shortest" is equivalent to "going right or up at each step".) For each admissible path $\gamma\colon [n+k] \to [n] \times [k]$ we span an $(n{+}k)$–simplex σ_γ in $\sigma \times \sigma'$ on the vertices $(x_{\gamma(j)} \mid j \in [n+k])$. It is well known that the set of all such σ_γ triangulates $\sigma \times \sigma'$ (see [Gelfand and Manin 2003, I.1.5]).

We will call an $(n{+}1)$–tuple of vectors in an n–dimensional vector space *linearly generic*, if every n of them are linearly independent.

Lemma 11.2 *Vectors $(v_0, \dots, v_n)$ are linearly generic if and only if there is a projectively unique linear relation $\sum_{i=0}^n \alpha_i v_i = 0$, and the coefficients in this relation are all nonzero.*

Proof ($\Leftarrow$) If some n of the v_i's were linearly dependent, a nontrivial linear relation between them could be extended — by adding 0 times the remaining vector — to a nontrivial relation with coefficient 0. This is a contradiction.

($\Rightarrow$) For dimensional reasons, there is a nontrivial linear relation between the v_i's; if some of its coefficients were 0, it would give linear dependence of a proper subset of the v_i's. If the relation was not projectively unique, one could form a linear combination of two nonproportional relations and obtain a nontrivial relation with coefficient 0. □

Observe that for a linearly generic tuple $(v_0, \dots, v_n)$, the class $[(v_0, \dots, v_n)]$ in $U_{n,+}$ is $0^{\pm}$ if and only if all the coefficients in the linear relation $\sum_{i=0}^n \alpha_i v_i = 0$ are of the same sign.

Now we will tackle the question of genericity of the section S (of $E \times E'$ over $(X \times X')^{(0)}$). Let $\sigma = (x_0, \dots, x_n) \in X^{(n)}$ and $\sigma' = (x'_0, \dots, x'_k) \in X'^{(k)}$. Using trivializations of E over σ and of E' over σ', we identify each E_{x_i} with the same vector space V ($\cong K^n$), and each $E'_{x'_j}$ with W ($\cong K^k$). Thus, we put $v_i = s(x_i) \in V$, $w_j = s'(x'_j) \in W$, $V_{(i,j)} = (v_i, w_j) = S(x_{(i,j)}) \in V \oplus W$. We would like to show that for each admissible path γ the vectors $(V_{\gamma(j)} \mid j \in [n+k])$ are linearly generic. There are unique scalars α_i and β_j such that

$$\sum_{i=0}^n \alpha_i v_i = 0, \quad \sum_{i=0}^n \alpha_i = 1, \quad \sum_{j=0}^k \beta_j w_j = 0, \quad \sum_{j=0}^k \beta_j = 1.$$

Let $A_u = \sum_{i=0}^u \alpha_i$ and $B_s = \sum_{j=0}^s \beta_j$. For a given path γ, let us arrange these two sequences into one:

$$C_j = \begin{cases} A_i & \text{if } \gamma(j) = (i, *) \text{ and } \gamma(j+1) = (i+1, *), \\ B_i & \text{if } \gamma(j) = (*, i) \text{ and } \gamma(j+1) = (*, i+1). \end{cases}$$

We put $C_{n+k} = A_n = B_k = 1$ and $C_{-1} = 0$.

Lemma 11.3 *There is a projectively unique linear relation between the vectors $(V_{\gamma(j)} \mid j \in [n+k])$. If we require that the sum of coefficients be 1, this relation is*

$$\sum_{j=0}^{n+k} (C_j - C_{j-1}) V_{\gamma(j)} = 0.$$

If all A_u and B_s are distinct, all coefficients of this relation are nonzero.

Proof First, let us prove the formula. Projecting onto the first factor we get

$$\sum_{i=0}^{n}\Bigl(\sum_{j\in\Gamma_i}(C_j-C_{j-1})\Bigr)v_i=0, \tag{11-3}$$

where $\Gamma_i=\{j\in[n+k]\mid\gamma(j)=(i,*)\}$ (similarly, we put $\Gamma^i=\{j\in[n+k]\mid\gamma(j)=(*,i)\}$). We have $\Gamma_i=\{u,u+1,\dots,u+\ell\}$ for some integers u, ℓ. Therefore

$$\sum_{j\in\Gamma_i}(C_j-C_{j-1})=C_{u+\ell}-C_{u-1}=A_i-A_{i-1}=\alpha_i.$$

Consequently, (11-3) becomes $\sum_{i=0}^{n}\alpha_i v_i=0$. Similarly, the projection onto W is 0. Thus, the relation stated in the lemma holds.

Suppose now that $\sum_{j=0}^{n+k}\Xi_j V_{\gamma(j)}=0$ is a linear relation with $\sum_{j=0}^{n+k}\Xi_j=1$. Projecting onto the first factor, we get $\sum_{i=0}^{n}\xi_i v_i=0$, where $\xi_i=\sum_{j\in\Gamma_i}\Xi_j$. Since $\sum_{i=0}^{n}\xi_i=\sum_{j=0}^{n+k}\Xi_j=1$ and the vectors v_i are linearly generic, we know that $\xi_i=\alpha_i$. Thus,

$$\sum_{j\in\Gamma_i}\Xi_j=\alpha_i.$$

Similarly,

$$\sum_{j\in\Gamma^i}\Xi_j=\beta_i.$$

Since each j is the largest element of exactly one set Γ_i or Γ^i, these equations recursively and uniquely determine all the Ξ_j.

Consequently, a linear relation between the $V_{\gamma(j)}$ with nonzero sum of coefficients is projectively unique. This implies that there is no nontrivial relation with sum of coefficients 0 — if it existed, it could be added to the one with sum of coefficients 1, contradicting the uniqueness of the latter.

The last claim of the lemma follows directly from the formula. □

Corollary 11.4 *Suppose that the class $[(v_0,\dots,v_n)]$ in $U_{n,+}$, or the class $[(w_0,\dots,w_k)]$ in $U_{k,+}$, is not $0^{\pm}$. Then, for every admissible path γ, the class $[(V_{\gamma(j)}\mid j\in[n+k])]$ in $U_{n+k,+}$ is not $0^{\pm}$.*

Proof The assumption can be interpreted as $\alpha_i<0$ for some i, or $\beta_j<0$ for some j. In each case one of the sequences (A_u) or (B_s) is not increasing; therefore, independent of γ, the sequence (C_j) is not increasing. Consequently, the relation between the $V_{\gamma(j)}$ (as in the lemma) cannot have all positive coefficients, while it does have some since their sum is 1. Hence the claim. □

Corollary 11.5 *Suppose that the class $[(v_0,\dots,v_n)]=0^{\pm}$ in $U_{n,+}$, and the class $[(w_0,\dots,w_k)]=0^{\pm}$ in $U_{k,+}$. Then there is a unique admissible path γ such that $[(V_{\gamma(j)}\mid j\in[n+k])]=0^{\pm}$ in $U_{n+k,+}$.*

Proof The sequences (A_u) and (B_s) are increasing. There is a unique γ such that (C_j) is increasing as well — then $[(V_{\gamma(j)}\mid j\in[n+k])]=0^{\pm}$. For other γ we conclude as in the previous corollary. □

Now we know that (11-2) holds up to sign. To finish the proof it remains to work out the relation between the signs that appear in the exponents in Corollary 11.5, and to check that this relation is consistent with (11-2). The cycle $z \times z'$ contains a triangulated version of the product $\sigma \times \sigma'$, for $\sigma \in \operatorname{supp} z$ and $\sigma' \in \operatorname{supp}(z')$, in the form $\sum_\gamma \operatorname{sgn}(\gamma)\sigma_\gamma$. The summation is over all admissible γ. The sign $\operatorname{sgn}(\gamma)$ equals $(-1)^{A(\gamma)}$, where $A(\gamma)$ is the area of the part of the grid that lies under the image of γ. In particular, if γ goes along the lower edge and the right-hand edge of the grid, the sign is $+1$. If we change γ by moving one $\gamma(j)$ to the opposite vertex of a 1×1 square — and get an admissible γ' — then $\operatorname{sgn}(\gamma') = -\operatorname{sgn}(\gamma)$.

Lemma 11.6 *Suppose that* $[(v_0, \dots, v_n)] = 0^s$ *and* $[(w_0, \dots, w_k)] = 0^{s'}$. *Then*

$$[(V_{\gamma(j)} \mid j \in [n+k])] = \operatorname{sgn}(\gamma) \cdot 0^{ss'}$$

for the γ *from the previous corollary.*

Proof We choose orientations of the bundles E and E'; we get induced orientations of V, W and $V \oplus W$. With respect to some positively oriented bases of V and W we have $\operatorname{sgn}\det(v_0, \dots, v_{n-1}) = s$ and $\operatorname{sgn}\det(w_0, \dots, w_{k-1}) = s'$. We will show that for every admissible γ the sign formula

$$\operatorname{sgn}\det_B(V_{\gamma(j)} \mid j \in [n+k-1]) = \operatorname{sgn}(\gamma)$$

holds, where the determinant is calculated with respect to the basis

$$B = ((v_0, 0), \dots, (v_{n-1}, 0), (0, w_0), \dots, (0, w_{k-1})).$$

(This claim implies the lemma.)

First, for the γ with $A(\gamma) = 0$, the determinant is

$$\begin{vmatrix} 1 & 0 & \cdots & 0 & a_0 & a_0 & \cdots & a_0 \\ 0 & 1 & \cdots & 0 & a_1 & a_1 & \cdots & a_1 \\ & \vdots & & & \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 & a_{n-1} & a_{n-1} & \cdots & a_{n-1} \\ 1 & 1 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ & \vdots & & & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{vmatrix},$$

where all a_i are negative ($v_n = \sum_{i=0}^{n-1} a_i v_i$). To calculate it, we use lower rows to cancel all the a_i except the ones in the $(n+1)^{\text{st}}$ column. Then we use the left columns to cancel all the remaining a_i — this increases the $(n+1, n+1)$–entry. The result is now lower-triangular and positive on the diagonal.

Now let us consider the change of the determinant as $\gamma(j)$ moves across a 1×1 square. This changes one column. That column, and the neighbouring ones, are

$$(\dots, (v_i, w_j), (v_{i+1}, w_j), (v_{i+1}, w_{j+1}), \dots) \leftrightarrow (\dots, (v_i, w_j), (v_i, w_{j+1}), (v_{i+1}, w_{j+1}), \dots).$$

The change, up to sign, can be performed by two column operations:

$$-(v_i, w_{j+1}) = (v_{i+1}, w_j) - (v_i, w_j) - (v_{i+1}, w_{j+1}).$$

Since every admissible γ can be obtained by such operations from the one with $A(\gamma) = 0$, the sign formula holds for all admissible paths. □

This completes the proof of Theorem 11.1. □

12 Cup product of Euler classes

Let E be a (flat) $\mathrm{GL}_+(n, K)$–bundle over a simplicial complex X. We will often trivialize this bundle over simplices of X; to facilitate the use of such trivializations we introduce the following convention. Let $\sigma = (x_0, \dots, x_\ell)$ be a simplex of X. We put $E_\sigma := E_{x_0}$, and we use any (flat) trivialization of E over σ to isomorphically identify all the other E_{x_i} with E_σ. Thus, if $s\colon X^{(0)} \to E$ is a section, we write $s(x_0), \dots, s(x_\ell) \in E_\sigma$.

Definition 12.1 A section $s\colon X^{(0)} \to E$ is called positive, if for every simplex $\sigma = (x_0, \dots, x_\ell)$ of X there is a functional $\phi_\sigma \in E_\sigma^*$ such that $\phi_\sigma(s(x_i)) > 0$ for $i = 0, \dots, \ell$.

If a $\mathrm{GL}_+(n, K)$–bundle E over X admits a generic positive section s, then $\langle \mathrm{eu}_0(E), z\rangle = 0$ for every cycle $z \in Z_n(X, \mathbb{Z})$. Indeed, for every simplex $\sigma \in X^{(n)}$ we have $s_*\sigma \neq 0^\pm$ in $U_{n,+}$, since the values of s at the vertices of σ do not admit a linear relation with all positive coefficients — by positivity of s. It turns out that (over a cycle) every positive section can be perturbed to a generic positive section.

Lemma 12.2 *If a $\mathrm{GL}_+(n, K)$–bundle E over a finite simplicial complex X admits a positive section, then it admits a generic positive section.*

Proof Let s be a positive section, as witnessed by functionals $\phi_\sigma \in E_\sigma^*$ ($\sigma \in X^{(n)}$). We will construct, vertex-by-vertex, a new generic section s', positive with respect to the same collection of functionals. We order the vertices of X, and we start with $s'(x_0) = s(x_0)$. Suppose that $s'(x_\ell)$ have already been chosen for $\ell < i$. Put $V = E_{x_i}$. When choosing $s'(x_i)$ in V, in order to ensure genericity, we need to avoid a finite collection of affine hyperplanes, say defined by equations $(\psi_j(v) = \alpha_j)_{j \in J}$ (where $\psi_j \in V^*$, $\alpha_j \in K$). Also, for each n–simplex σ with vertex x_i, we need to ensure that $\phi_\sigma(s'(x_i)) > 0$ (we identify E_σ with V). Let $w \in V$ be such that $\psi_j(w) \neq 0$ for all $j \in J$; such w exists, since V is not the union of finitely many hyperplanes $(\ker \psi_j)_{j \in J}$. We will find suitable $s'(x_i)$ in the form $v(\alpha) := s(x_i) + \alpha w$, for some scalar α. First, observe that the equation $\psi_j(v(\beta)) = \alpha_j$ has a unique solution $\beta_j = \bigl(\alpha_j - \psi_j(s(x_i))\bigr)/\psi_j(w)$. Let $B := \min\{\beta_j \mid \beta_j > 0\}$. The condition $\phi_\sigma(v(\beta)) > 0$, ie $\phi_\sigma(s(x_i)) + \beta\phi_\sigma(w) > 0$, is equivalent to $\beta > -\phi_\sigma(s(x_i))/\phi_\sigma(w)$ (if $\phi_\sigma(w) > 0$) or to $\beta < -\phi_\sigma(s(x_i))/\phi_\sigma(w)$ (if $\phi_\sigma(w) < 0$). We know that $\beta = 0$ satisfies all these inequalities. Therefore, the scalar $M := \min\{-\phi_\sigma(s(x_i))/\phi_\sigma(w) \mid \phi_\sigma(w) < 0\}$ is positive. We put $\alpha := \frac{1}{4}\min(B, M)$ and $s'(x_i) = v(\alpha)$. □

Corollary 12.3 *Let E and E' be $\mathrm{GL}_+(n,K)$– and $\mathrm{GL}_+(k,K)$–bundles over simplicial complexes X and X', respectively. For any simplicial cycles $z \in Z_{n-\ell}(X,\mathbb{Z})$ and $z' \in Z_{k+\ell}(X',\mathbb{Z})$, where $\ell > 0$,*

$$\langle \mathrm{eu}_0(E \times E'), z \times z' \rangle = 0.$$

Proof We may and do assume that $X = \operatorname{supp} z$ and $X' = \operatorname{supp} z'$. Let s be a generic section of E. For dimensional reasons, the values of s at the vertices of any simplex σ of X are linearly independent; therefore, a functional ϕ_σ can be chosen that evaluates to 1 on each of them. Thus, s is positive. Now define $S\colon (X \times X')^{(0)} \to E \times E'$ by $S(x,x') = (s(x),0)$. Then, for any simplices $\sigma \in X^{(n-\ell)}$ and $\sigma' \in X'^{(k+\ell)}$, and any $(n+k)$–dimensional simplex σ_γ in the standard triangulation of $\sigma \times \sigma'$, we may put $\phi_{\sigma_\gamma} = \phi_\sigma \circ \pi_E$. Then, for every vertex (x,x') of σ_γ we have

$$\phi_{\sigma_\gamma}(S(x,x')) = \phi_\sigma\bigl(\pi_E(s(x),0)\bigr) = \phi_\sigma(s(x)) > 0.$$

Therefore, S is a positive section of $E \times E'$ over $\operatorname{supp} z \times z'$. By the lemma above, there exists a generic positive section, and that implies the asserted vanishing. □

Corollary 12.4 *Let E and E' be $\mathrm{GL}_+(n,K)$– and $\mathrm{GL}_+(k,K)$–bundles over simplicial complexes X and X' respectively. For any simplicial cycle $Z \in Z_{n+k}(X \times X',\mathbb{Z})$,*

$$\langle \mathrm{eu}_0(E) \times \mathrm{eu}_0(E'), Z \rangle = \langle \mathrm{eu}_0(E \times E'), Z \rangle.$$

Proof Indeed, by Künneth's formula, an integer multiple of Z is homologous to a combination of cycles of the form $z \times z'$; for the latter, the formula holds either by the previous corollary, or by Theorem 11.1. □

Theorem 12.5 *Let E and E' be $\mathrm{GL}_+(n,K)$– and $\mathrm{GL}_+(k,K)$–bundles over a simplicial complex X. For any simplicial cycle $z \in Z_{n+k}(X,\mathbb{Z})$,*

$$\langle \mathrm{eu}_0(E) \cup \mathrm{eu}_0(E'), z \rangle = \langle \mathrm{eu}_0(E \oplus E'), z \rangle.$$

Proof Let $\Delta\colon X \to X \times X$ be the diagonal map. Then

$$\begin{aligned}\langle \mathrm{eu}_0(E) \cup \mathrm{eu}_0(E'), z \rangle &= \langle \Delta^*(\mathrm{eu}_0(E) \times \mathrm{eu}_0(E')), [z] \rangle \\ &= \langle \mathrm{eu}_0(E) \times \mathrm{eu}_0(E'), \Delta_*[z] \rangle = \langle \mathrm{eu}_0(E \times E'), \Delta_*[z] \rangle \\ &= \langle \Delta^* \mathrm{eu}_0(E \times E'), [z] \rangle = \langle \mathrm{eu}_0(\Delta^*(E \times E')), [z] \rangle \\ &= \langle \mathrm{eu}_0(E \oplus E'), z \rangle.\end{aligned}$$

□

13 Comparison of Euler and Witt classes

We use the functoriality theorem (Theorem 1.5) to compare various tautological classes that we have constructed. We begin with eu and eu_+.

Euler classes We assume n is even. There is a natural map $\mathbb{P}_+ \to \mathbb{P}$; it induces a simplicial (nondegenerate) map $f: X_+ \to X$. The groups P_+G_+ and PG_+ acting on X_+ and X (respectively) are also related by the natural projection homomorphism $\phi\colon P_+G_+ \to PG_+$. The map f is ϕ–equivariant, and induces a coefficient group map $f: U_+ \to U$. Theorem 1.5 applies and gives the diagram

$$H^n(PG_+, U) \xrightarrow{\phi^*} H^n(P_+G_+, U) \xleftarrow{f_*} H^n(P_+G_+, U_+).$$

Recall that $U \simeq \mathbb{Z}$ and $U_+ \simeq \mathbb{Z}^{(n/2)+1}$. The map $f\colon U_+ \to U$ can be described explicitly using Remark 8.8. The generator a^+ of U_+ is represented by the simplex $([e_1], \dots, [e_n], [v_a])$, where

$$v_a = e_1 + \cdots + e_a - (e_{a+1} + \cdots + e_n).$$

The image of this simplex in X determines in U the symbol $[\operatorname{sgn}(\det(e_1, \dots, e_n) \cdot (-1)^{n-a})] = [(-1)^a]$. Therefore, $f(a^+) = [(-1)^a] = (-1)^a[+]$. It follows that the induced map on cohomology,

$$f_*\colon H^n(P_+G_+, U_+) \to H^n(PG_+, U),$$

maps $\mathrm{eu}_+ = \bigoplus_a \mathrm{eu}_a$ to $\sum_a (-1)^a \mathrm{eu}_a$. Theorem 1.5 implies the following result.

Theorem 13.1 *Let $\phi\colon P_+G_+ \to PG_+$ be the natural projection homomorphism. Then*

$$\phi^*\mathrm{eu} = \sum_{a=0}^{n/2} (-1)^a \mathrm{eu}_a.$$

A (flat) P_+G_+–bundle P over Y determines a PG_+–bundle P' over Y. As is usual in such cases, we put $\mathrm{eu}(P) := \mathrm{eu}(P') \in H^n(Y, \mathbb{Z})$.

Corollary 13.2 *Let P be a (flat) $\mathrm{P_+GL_+}(n, K)$–bundle over an oriented closed n–manifold M. Then*

$$\langle \mathrm{eu}(P), [M] \rangle = 2^n \langle \mathrm{eu}_0(P), [M] \rangle.$$

Proof Using Theorems 10.1 and 13.1, we calculate

$$\begin{aligned}
\langle \mathrm{eu}(P), [M] \rangle &= \Big\langle \sum_{k=0}^{n/2} (-1)^k \mathrm{eu}_k(P), [M] \Big\rangle \\
&= \sum_{k=0}^{n/2} (-1)^k \langle \mathrm{eu}_k(P), [M] \rangle \\
&= \sum_{k=0}^{n/2} (-1)^k (-1)^k \binom{n+1}{k} \langle \mathrm{eu}_0(P), [M] \rangle = 2^n \langle \mathrm{eu}_0(P), [M] \rangle.
\end{aligned}$$ □

Remark 13.3 For $n = 2$, Theorem 9.1 gives $3\mathrm{eu}_0 + \mathrm{eu}_1 = 0$. Theorem 13.1 now implies $\phi^*\mathrm{eu} = 4\mathrm{eu}_0$, ie in this case Corollary 13.2 can be strengthened to equality in $H^2(\mathrm{P_+GL_+}(2, K), \mathbb{Z})$ — there is no need to evaluate on cycles.

Witt class In Section 7 we discussed the action of $\mathrm{PSL}(2, K)$ on $\mathbb{P}^1$, on the associated complex X, and the resulting Witt class $w \in H^2(\mathrm{PSL}(2, K), W(K))$. In Section 8 we considered the action of $\mathrm{PGL_+}(2, K)$

on the same spaces, and the resulting cohomology class $\mathrm{eu} \in H^2(\mathrm{PGL}_+(2,K),\mathbb{Z})$. Theorem 1.5 may be applied to the identity map $\iota\colon X \to X$ and the injection homomorphism $\phi\colon \mathrm{PSL}(2,K) \to \mathrm{PGL}_+(2,K)$. Before stating the result we compute the coefficient map $\iota\colon W(K) \to \mathbb{Z}$. The symbol $[\lambda]$ is represented by the triple $t_\lambda = \left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\\lambda\end{smallmatrix}\right]\right)$. To find the symbol of t_λ in $U_2(X,\mathrm{PGL}_+(2,K))$ we write

$$\begin{pmatrix}1\\\lambda\end{pmatrix} = 1\cdot\begin{pmatrix}1\\0\end{pmatrix} + \lambda\cdot\begin{pmatrix}0\\1\end{pmatrix}.$$

Then, using Remark 8.8, we get

$$\left[\mathrm{sgn}\left(\begin{vmatrix}1&0\\0&1\end{vmatrix}\cdot 1\cdot\lambda\right)\right] = [\mathrm{sgn}(\lambda)].$$

Therefore, the map $\iota\colon W(K) \to \mathbb{Z}$ is just the signature map σ, given by $\sigma([\lambda]) = \mathrm{sgn}(\lambda)$. The diagram is

$$H^2(\mathrm{PGL}_+(2,K),\mathbb{Z}) \xrightarrow{\phi^*} H^2(\mathrm{PSL}(2,K),\mathbb{Z}) \xleftarrow{\sigma_*} H^2(\mathrm{PSL}(2,K),W(K))$$

and the theorem is as follows.

Theorem 13.4 *Let $\phi\colon \mathrm{PSL}(2,K) \to \mathrm{PGL}_+(2,K)$ be the standard inclusion. Then*

$$\phi^*\mathrm{eu} = \sigma_* w.$$

Furthermore, the pullback of this class to $\mathrm{SL}(2,K)$ is equal to $4\mathrm{eu}_0$.

The last claim of the theorem follows from Remark 13.3.

Nonvanishing Consider a flat vector $\mathrm{SL}(2,\mathbb{R})$–bundle E over a closed oriented surface Σ. The (classical, topological) Euler class $\mathrm{eu}_t(E)$ of E (more precisely, the Euler number $\langle \mathrm{eu}_t(E),[\Sigma]\rangle$) can be computed as the signed number of zeroes of a generic section of E; generic means transversal to the zero section. Consider now a triangulation Y of Σ. Let $s\colon Y^{(0)} \to E$ be a generic section over the set of vertices of Y. Here genericity means that for every 2–simplex σ of Y the values of s at the vertices of σ are pairwise linearly independent (as usual, we compare them using a flat trivialization of E over σ). The section s can be affinely extended to each simplex of Y. Together, these extensions define a generic section of E over Σ in the previous, classical sense. Moreover, the zeroes of this extended section occur exactly in simplices σ on which s^*T_0 (the cocycle representing $\mathrm{eu}_0(E)$; see Remark 8.8 and Definition 8.9) is nonzero, and the sign of the zero in σ is equal to $s^*T_0(\sigma)$. These arguments prove the following statement.

Fact 13.5 *Let E be a flat $\mathrm{SL}(2,\mathbb{R})$–bundle over a closed surface Σ. Then*

$$\langle \mathrm{eu}_0(E),[\Sigma]\rangle = \langle \mathrm{eu}_t(E),[\Sigma]\rangle.$$

We will now prove that all the Euler classes constructed in this paper are nonzero (for n even).

Theorem 13.6 *Let K be an ordered field and let n be even. Then the Euler classes eu, eu_+ and all eu_k are nonzero.*

Proof Recall that an ordered field contains $\mathbb{Q}$ as a subfield, and the order restricted to $\mathbb{Q}$ is standard. Due to field restriction stability of our classes (see Remark 8.10) it is enough to show the theorem for $K = \mathbb{Q}$.

Assume first that $n = 2$. Recall that over a closed oriented surface Σ of genus $g \geq 2$ there are flat vector $\mathrm{SL}(2, \mathbb{R})$–bundles E with nontrivial Euler class eu_t (see [Milnor and Stasheff 1974, Appendix C]). Moreover, Takeuchi proved that $\mathrm{SL}(2, \mathbb{Q})$ can be used as the structure group of such bundles (see [Takeuchi 1971]); let us call such examples (flat $\mathrm{SL}(2, \mathbb{Q})$–bundles with nontrivial eu_t) Takeuchi bundles. Fact 13.5 implies that the Euler class eu_0 is nonzero for Takeuchi bundles. Theorem 10.1 and Corollary 13.2 imply that also eu_1 and eu are nontrivial for them.

For larger even $n = 2k$ we consider the Cartesian product Y of k copies of Σ, and over Y the product bundle $E^{\times k}$ of k–copies of a Takeuchi bundle E. Then Theorem 11.1 shows that

$$\langle \mathrm{eu}_0(E^{\times k}), [Y] \rangle = \langle \mathrm{eu}_0(E), [\Sigma] \rangle^k \neq 0.$$

Again, it follows from Theorem 10.1 and Corollary 13.2 that all eu_k as well as eu are nontrivial on $E^{\times k}$. □

References

[Bergeron et al. 2014] **N Bergeron**, **E Falbel**, **A Guilloux**, *Tetrahedra of flags, volume and homology of* SL(3), Geom. Topol. 18 (2014) 1911–1971 MR Zbl

[Brown 1982] **K S Brown**, *Cohomology of groups*, Graduate Texts in Mathematics 87, Springer (1982) MR Zbl

[Dugundji 1958] **J Dugundji**, *Cohomology of equivariant maps*, Trans. Amer. Math. Soc. 89 (1958) 408–420 MR Zbl

[Elman et al. 2008] **R Elman**, **N Karpenko**, **A Merkurjev**, *The algebraic and geometric theory of quadratic forms*, American Mathematical Society Colloquium Publications 56, Amer. Math. Soc., Providence, RI (2008) MR Zbl

[Gelfand and Manin 2003] **S I Gelfand**, **Y I Manin**, *Methods of homological algebra*, 2nd edition, Springer (2003) MR Zbl

[Ghys 1987] **E Ghys**, *Groupes d'homéomorphismes du cercle et cohomologie bornée*, from "The Lefschetz centennial conference, III" (A Verjovsky, editor), Contemp. Math. 58, Amer. Math. Soc., Providence, RI (1987) 81–106 MR Zbl

[Gromov 1982] **M Gromov**, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. 56 (1982) 5–99 MR Zbl

[Hatcher 2002] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[Kramer and Tent 2010] **L Kramer**, **K Tent**, *A Maslov cocycle for unitary groups*, Proc. Lond. Math. Soc. 100 (2010) 91–115 MR Zbl

[Löh 2010] **C Löh**, *Group cohomology & bounded cohomology*: *an introduction for topologists*, book project (2010) Available at `https://loeh.app.uni-regensburg.de/teaching/topologie3_ws0910/prelim.pdf`

[Milnor and Stasheff 1974] **J W Milnor**, **J D Stasheff**, *Characteristic classes*, Annals of Mathematics Studies 76, Princeton Univ. Press (1974) MR Zbl

[Morita 2001] **S Morita**, *Geometry of characteristic classes*, Translations of Mathematical Monographs 199, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[Nekovář 1990] **Y Nekovář**, *Maslov index and Clifford algebras*, Funct. Anal. Appl. 24 (1990) 196–204 MR Zbl

[Reznikov 1997] **A Reznikov**, *Euler class and free generation*, preprint (1997) arXiv dg-ga/9709009

[Takeuchi 1971] **K Takeuchi**, *Fuchsian groups contained in* $\mathrm{SL}_2(\mathbb{Q})$, J. Math. Soc. Japan 23 (1971) 82–94 MR Zbl

Instytut Matematyczny, Uniwersytet Wrocławski
Wrocław, Poland

Instytut Matematyczny PAN
Wrocław, Poland

jan.dymara@uwr.edu.pl, tjan@impan.pl

Received: 25 February 2023 Revised: 30 June 2023

Algebraic & Geometric Topology **24**:5 (2024) 2933–2956
DOI: 10.2140/agt.2024.24.2933
Published: 19 August 2024

Homotopy types of suspended 4–manifolds

PENGCHENG LI

Given a closed, smooth, connected, orientable 4–manifold M whose integral homology groups can have 2–torsion, we determine the homotopy decomposition of the double suspension $\Sigma^2 M$ as wedge sums of some elementary $\mathbf{A}_3^3$–complexes which are 2–connected finite complexes of dimension at most 6. Furthermore, we utilize the Postnikov square (or equivalently Pontryagin square) to find sufficient conditions for the homotopy decompositions of $\Sigma^2 M$ to desuspend to that of ΣM.

55P15, 55P40, 57N65

1 Introduction

Recently, research on the homotopy properties of manifolds has emerged in two directions. The first direction is the loop homotopy of manifolds, which can be traced back to Beben and Wu's work [6] in 2011. After them, many people made efforts to promote the development of this project, such as Beben, Theriault and Huang [4; 5; 15]. On the other hand, as exhibited by So and Theriault [19], the suspension homotopy of manifolds has rich applications in some important objects of geometry and physics, such as gauge groups and current groups. Hereafter, this research direction has been widely studied, for instance in Huang [11; 12; 13], Cutler and So [8] and Huang and Li [14].

This paper contributes to further research on the suspension homotopy of manifolds. In the above related literature, due to some intractable obstructions, the authors usually avoid handling 2–torsions of the integral homology groups of the manifolds. For example, So and Theriault [19] required the 4–manifolds are 2–torsion-free in integral homology, Huang [13] restricts to 6–manifolds with integral homology groups containing no 2– or 3–torsions, while Cutler and So [8] and Huang and Li [14] respectively studied the suspension homotopy of simply connected 6–manifolds and 7–manifolds after localization away from 2.

In this paper we developed new techniques and tools in homotopy theory to obtain *complete* classification of the homotopy types of suspended 4–manifolds which can have 2–torsion in homology. For instance, we successfully apply certain homotopy properties of some $\mathbf{A}_n^3$*–complexes* (defined below) to obtain the homotopy decompositions of $\Sigma^2 M$. Moreover, the Postnikov squaring operation (1-1) and the Pontryagin squaring operation (1-2) appear to be powerful in the characterizations of the homotopy type of ΣM; see Section 5.

To make sense of the introduction, we need the following notions and notation. Let G be an abelian group and let n be a positive integer. Denote by $H_n(X;G)$ (resp. $H^n(X;G)$) the n^{th} (singular) homology (resp. cohomology) group of X with coefficients in G, and denote by $P^n(G)$ the n–dimensional Peterson space (see Neisendorfer [18]) which admits a unique nontrivial reduced integral cohomology group G in dimension n. In particular, for integers $n,k \geq 2$, we denote by $\mathbb{Z}/k = \mathbb{Z}/k\mathbb{Z}$ the group of integers modulo k. Recall the Peterson spaces have the cell structure

$$P^n(k) = P^n(\mathbb{Z}/k) = S^{n-1} \cup_k e^n,$$

which admits the obvious inclusion i_{n-1} of the bottom sphere S^{n-1} into $P^n(k)$ and the pinch map q_n onto S^n. For each $n \geq 3$, there is a generator $\tilde{\eta}_r \in \pi_{n+1}(P^n(2^r))$ satisfying the formula

$$q_n \tilde{\eta}_r \simeq \eta_n;$$

see Lemma 2.1, where $\eta_n\colon S^{n+1} \to S^n$ is the iterated suspensions of the Hopf map $\eta\colon S^3 \to S^2$. For a homomorphism $\phi\colon G \to G'$ of groups, $\ker(\phi)$ and $\operatorname{im}(\phi)$ denote the kernel and the image subgroups of ϕ, respectively.

A finite CW–complex X is called an $\boldsymbol{A}_n^k$*–complex* if X is $(n-1)$–connected and has dimension at most $n+k$. It is well known that elementary (or called indecomposable) $\boldsymbol{A}_n^1$–complexes consist of spheres S^n, S^{n+1} and the Moore spaces $P^{n+1}(p^r)$ with p odd primes and $r \geq 1$. One may consult Zhu, Li and Pan [24; 16; 23; 25] and Baues and Hennes [3] for more homotopy theory of such complexes. We need the following elementary $\boldsymbol{A}_n^3$–complexes with $n \geq 3$ and $r,s \geq 1$:

$$\begin{aligned}
C_\eta^{n+2} &= S^n \cup_\eta \boldsymbol{C}S^{n+1} = \Sigma^{n-2}\mathbb{C}P^2, & C_r^{n+2} &= P^{n+1}(2^r) \cup_{i_n\eta} \boldsymbol{C}S^{n+1},\\
C^{n+2,s} &= S^n \cup_{\eta q_{n+1}} \boldsymbol{C}P^{n+1}(2^s), & C_r^{n+2,s} &= P^{n+1}(2^r) \cup_{i_n\eta q_{n+1}} \boldsymbol{C}P^{n+1}(2^s),\\
A^{n+3}(\eta^2) &= S^n \cup_{\eta^2} \boldsymbol{C}S^{n+2}, & A^{n+3}(\tilde{\eta}_r) &= P^{n+1}(2^r) \cup_{\tilde{\eta}_r} \boldsymbol{C}S^{n+2},\\
A^{n+3}(2^r\eta^2) &= P^{n+1}(2^r) \cup_{i_n\eta^2} \boldsymbol{C}S^{n+2}.
\end{aligned}$$

Here the first four $\boldsymbol{A}_n^2$–complexes are the *elementary Chang complexes* (due to Chang [7]), and the last two spaces are the only two $\boldsymbol{A}_n^3$–complexes with the homology groups

$$H_n \cong \mathbb{Z}/2^r, \qquad H_{n+3} = H_0 \cong \mathbb{Z}, \qquad H_i = 0 \quad \text{for } i \neq 0, n, n+3.$$

Compare [2, Theorem 10.3.1]. Note that all of the above $\boldsymbol{A}_n^3$–complexes desuspend: they can be defined for $n \geq 2$.

To deal with 2–torsions in $H_*(M;\mathbb{Z})$, we shall employ the following cohomology operations. Let X be a connected CW–complex. For each $r \geq 1$, there are unstable cohomology operations: the *Postnikov square*

$$\mathfrak{P}_0\colon H^1(X;\mathbb{Z}/2^r) \to H^3(X;\mathbb{Z}/2^{r+1}) \tag{1-1}$$

and the *Pontryagin square*

$$\mathfrak{P}_1\colon H^2(X;\mathbb{Z}/2^r) \to H^4(X;\mathbb{Z}/2^{r+1}). \tag{1-2}$$

These two operations were carefully studied by Whitehead [21; 22]. Given a cohomology operation C which maps $H^r(X,A;G_1)\to H^s(X,A;G_2)$ for each pair (X,A), the *suspension operation* $S(C)$ is the composition (see [20, Section 3])

$$H^{r-1}(Y;G_1)\xrightarrow{\sigma}H^r(\Sigma Y;G_1)\xrightarrow{C}H^s(\Sigma Y;G_2)\xrightarrow{\sigma^{-1}}H^{s-1}(Y;G_2),$$

where σ is the suspension isomorphism. Note that $\mathfrak{P}_0$ is the suspension operation of $\mathfrak{P}_1$:

$$\sigma\mathfrak{P}_0=\mathfrak{P}_1\sigma; \tag{1-3}$$

see [20, Theorem I(i)]. The Adem relations

$$\mathrm{Sq}^3=\mathrm{Sq}^1\,\mathrm{Sq}^2 \quad\text{and}\quad \mathrm{Sq}^3\,\mathrm{Sq}^1+\mathrm{Sq}^2\,\mathrm{Sq}^2=0$$

yield the secondary operation Θ_n based on the relation $\varphi_n\theta_n=0$ with

$$\begin{aligned}\theta_n&=\begin{pmatrix}\mathrm{Sq}^2\,\mathrm{Sq}^1\\ \mathrm{Sq}^2\end{pmatrix}\colon K_n\to K_{n+3}\times K_{n+2},\\ \varphi_n&=(\mathrm{Sq}^1,\mathrm{Sq}^2)\colon K_{n+3}\times K_{n+2}\to K_{n+4},\end{aligned} \tag{1-4}$$

where $n\ge 1$, $K_m=K_m(\mathbb{Z}/2)$ denotes the Eilenberg–Mac Lane space of type $(\mathbb{Z}/2,m)$. For a space X, the secondary operation $\Theta_n\colon S_n(X)\to T_n(X)$ is the induced homomorphism with

$$\begin{aligned}S_n(X)&=\ker(\theta_n)_\sharp=\ker(\mathrm{Sq}^2)\cap\ker(\mathrm{Sq}^2\,\mathrm{Sq}^1),\\ T_n(X)&=\mathrm{coker}(\Omega\varphi_n)_\sharp=H^{n+3}(X;\mathbb{Z}/2)/\mathrm{im}(\mathrm{Sq}^1+\mathrm{Sq}^2).\end{aligned}$$

The secondary operation Θ_n detects the map $\eta^2=\eta_n\eta^{n+1}\colon S^{n+2}\to S^n$; see [9, page 96] or Lemma 2.7. For each $r\ge 1$, the higher-order Bockstein operations

$$\beta_r\colon H^*(X;\mathbb{Z}/2)\dashrightarrow H^{*+1}(X;\mathbb{Z}/2) \tag{1-5}$$

are inductively defined by setting β_1 as the usual Bockstein homomorphism associated to the short exact sequence

$$0\to\mathbb{Z}/2\to\mathbb{Z}/4\to\mathbb{Z}/2\to 0;$$

for $r\ge 2$, β_r is defined on the intersection of $\ker(\beta_i)$, $i<r$, and takes values in the quotient by the $\mathrm{im}(\beta_i)$, $i<r$. This is also indicated by the dashed arrow in (1-5). See [9, Section 5.2] for more details. Note that the higher Bocksteins β_r and the sequence $\Theta=\{\Theta_n\}_{n\ge 1}$ are both *stable* (cf [9, 4.2.2]):

$$\Omega\beta_r=\beta_r,\quad \Omega\Theta_{n+1}=\Theta_n.$$

Let M be a closed, smooth, connected, orientable 4–manifold. By Poincaré duality and the universal coefficient theorem for cohomology, the homology groups $H_*(M;\mathbb{Z})$ are given by Table 1, where m,d are nonnegative integers, and T is a finitely generated torsion abelian group. Denote the 2–primary component of T by

$$T_2=\bigoplus_{j=1}^{n}\mathbb{Z}/2^{r_j}.$$

Now we are prepared to state our first main theorem.

i	$0,4$	1	2	3	≥ 5
$H_i(M;\mathbb{Z})$	$\mathbb{Z}$	$\mathbb{Z}^m \oplus T$	$\mathbb{Z}^d \oplus T$	$\mathbb{Z}^m$	0

Table 1: $H_*(M;\mathbb{Z})$.

Theorem 1.1 *Let M be a closed, smooth, connected, orientable 4–manifold with integral homology $H_*(M;\mathbb{Z})$ given by Table 1.*

(1) *Suppose that M is spin, then $\Sigma^2 M$ has two possible homotopy types:*

(a) *If $\Theta(H^1(M;\mathbb{Z}/2)) = 0$, then there is a homotopy equivalence*

$$\Sigma^2 M \simeq \Big(\bigvee_{i=1}^{m} (S^3 \vee S^5)\Big) \vee \Big(\bigvee_{i=1}^{d} S^4\Big) \vee P^4(T) \vee P^5(T) \vee S^6.$$

(b) *If $\Theta(H^1(M;\mathbb{Z}/2)) \neq 0$, then*

$$\Sigma^2 M \simeq \Big(\bigvee_{i=1}^{m} (S^3 \vee S^5)\Big) \vee \Big(\bigvee_{i=1}^{d} S^4\Big) \vee P^4\Big(\frac{T}{\mathbb{Z}/2^{r_{j_0}}}\Big) \vee P^5(T) \vee A^6(2^{r_{j_0}}\eta^2),$$

where j_0 is the maximum of the indices $j \leq n$ such that

$$\Theta(x) \neq 0 \quad \text{and} \quad \beta_{r_j}(x) \neq 0 \quad \text{for } x \in H^1(M;\mathbb{Z}/2).$$

(2) *Suppose that M is nonspin. Then the suspension $\Sigma^i M$ has the following possible homotopy types:*

(a) *If for any $u \in H^4(\Sigma^2 M;\mathbb{Z}/2)$ with $\mathrm{Sq}^2(u) \neq 0$ and any $v \in \ker(\mathrm{Sq}^2)$ it holds that*

$$\beta_r(u+v) = 0 \quad \text{and} \quad u+v \notin \mathrm{im}(\beta_s) \quad \text{for all } r,s \geq 1,$$

then there is a homotopy equivalence

$$\Sigma^2 M \simeq \Big(\bigvee_{i=1}^{m} (S^3 \vee S^5)\Big) \vee \Big(\bigvee_{i=1}^{d-1} S^4\Big) \vee P^4(T) \vee P^5(T) \vee C^6_\eta.$$

(b) *Suppose that for any $u \in H^2(M;\mathbb{Z}/2)$ with $\mathrm{Sq}^2(u) \neq 0$ and any $v \in \ker(\mathrm{Sq}^2)$, it holds that*

$$u+v \notin \mathrm{im}(\beta_s) \quad \text{for all } s \geq 1,$$

while there exist $u' \in H^2(M;\mathbb{Z}/2)$ with $\mathrm{Sq}^2(u') \neq 0$ and $v' \in \ker(\mathrm{Sq}^2)$ such that

$$\beta_r(u'+v') \neq 0 \quad \text{for some } r \geq 1.$$

Then there is a homotopy equivalence

$$\Sigma^2 M \simeq \bigvee_{i=1}^{m} (S^3 \vee S^5) \vee \bigvee_{i=1}^{d} S^4 \vee P^4(T) \vee P^5\Big(\frac{T}{\mathbb{Z}/2^{r_{j_1}}}\Big) \vee C^6_{r_{j_1}},$$

where j_1 is the maximum of the indices $j \leq n$ such that

$$\mathrm{Sq}^2(u') \neq 0 \quad \text{and} \quad \beta_r(u'+v') \neq 0.$$

(c) *Suppose that there exist* $u \in H^2(M;\mathbb{Z}/2)$ *with* $\mathrm{Sq}^2(u) \neq 0$ *and* $v \in \ker(\mathrm{Sq}^2)$ *such that* $u+v \in \mathrm{im}(\beta_r)$ *for some* r.

(i) *If* $\Theta(H^1(M;\mathbb{Z}/2)) = 0$, *then there is a homotopy equivalence*

$$\Sigma^2 M \simeq \Big(\bigvee_{i=1}^m (S^3 \vee S^5)\Big) \vee \Big(\bigvee_{i=1}^d S^4\Big) \vee P^5(T) \vee P^4\Big(\frac{T}{\mathbb{Z}/2^{r_{j_2}}}\Big) \vee A^6(\tilde{\eta}_{r_{j_2}}),$$

where j_2 *is the minimum of the indices* $j \leq n$ *such that* $u+v \in \mathrm{im}(\beta_{r_j})$.

(ii) *If* $\Theta(H^1(M;\mathbb{Z}/2)) \neq 0$ *and* $T_2 \cong \mathbb{Z}/2^{r_{j_2}}$, *then there is a homotopy equivalence*

$$\Sigma^2 M \simeq \Big(\bigvee_{i=1}^m (S^3 \vee S^5)\Big) \vee \Big(\bigvee_{i=1}^d S^4\Big) \vee P^5(T) \vee P^4\Big(\frac{T}{\mathbb{Z}/2^{r_{j_2}}}\Big) \vee A^6_\varepsilon(\tilde{\eta}_{r_{j_2}}),$$

where $A^6_\varepsilon(\tilde{\eta}_{r_{j_2}})$ *is the homotopy cofiber of* $\tilde{\eta}_{r_{j_2}} + \varepsilon \cdot i_3\eta^2$ *with* $\varepsilon \in \{0,1\}$.

(iii) *If* $\Theta(H^1(M;\mathbb{Z}/2)) \neq 0$ *and* $n \geq 2$ (ie T_2 has at least 2 direct summands), *then there is a homotopy equivalence*

$$\Sigma^3 M \simeq$$
$$\Big(\bigvee_{i=1}^m (S^4 \vee S^6)\Big) \vee \Big(\bigvee_{i=1}^d S^5\Big) \vee P^6(T) \vee A^7_\varepsilon(\tilde{\eta}_{r_{j_2}}) \vee P^5\Big(\frac{T}{\mathbb{Z}/2^{r_{j_2}} \oplus \mathbb{Z}/2^{r_{j_0'}}}\Big) \vee A^7(2^{r_{j_0'}}\eta^2),$$

where $A^7_\varepsilon(\tilde{\eta}_{r_{j_2}}) = \Sigma A^6_\varepsilon(\tilde{\eta}_{r_{j_2}})$, *the index* j_2 *the minimum of the indices* $j \leq n$ *such that* $u+v \in \mathrm{im}(\beta_{r_j})$, *and* j_0' *is the maximum of the indices* $j \leq n$ *with* $j \neq j_2$ *such that*

$$\Theta(x) \neq 0 \quad \text{and} \quad \beta_{r_j}(x) \neq 0 \quad \text{for all } x \in H^3(C_\varphi;\mathbb{Z}/2).$$

From the above complete discussion we see that when M is nonspin, the nontriviality of the secondary operation Θ on $H^1(M;\mathbb{Z}/2)$ only affects case when $u+v \in \mathrm{im}(\beta_r)$ for some r. In the last case (iii) we made one more suspension to cancel the possible nontrivial Whitehead products in k'–invariant of the homology decomposition of the $\Sigma^2 M$.

We also study the homotopy type of the suspension ΣM in terms of the Postnikov square $\mathfrak{P}_0$ (or equivalently the Pontryagin square $\mathfrak{P}_1$).

Theorem 1.2 *Let* M *be a closed, smooth, connected, orientable* 4*–manifold with* $H_*(M;\mathbb{Z})$ *given by Table 1. If the Postnikov square*

$$\mathfrak{P}_0\colon H^1(M;\mathbb{Z}/2^{r_j}) \to H^3(M;\mathbb{Z}/2^{r_j+1})$$

is trivial for each $j = 1,2,\dots,n$, *then the desuspensions of the homotopy decompositions of* $\Sigma^2 M$ *in Theorem 1.1 yield the homotopy decompositions of* ΣM.

If $H_*(M;\mathbb{Z})$ contains no 2–torsion (ie $T_2 = 0$), then the homotopy decomposition $\Sigma M \simeq \bigvee_{i=1}^m S^2 \vee \Sigma W$ (4-1) implies that the Pontryagin square

$$\mathfrak{P}_1\colon H^1(\Sigma M;\mathbb{Z}/2^{r_j}) \to H^3(\Sigma M;\mathbb{Z}/2^{r_j+1})$$

is trivial, hence so is $\mathfrak{P}_0$ by (1-3). Hence Theorem 1.2 extends So and Theriault's results [19, Theorem 1.1].

However, the author didn't find any other 4–manifolds M satisfying conditions in Theorem 1.2. This is also why we arrange the above theorem after Theorem 1.1.

The paper is organized as follows. In Section 2 we review some homotopy theory of partial elementary A_n^3–complexes and list some technical lemmas about the Pontryagin or Steenrod square operations. Section 3 introduces the main analysis methods adopted in this paper, including a useful criterion to determine the homotopy type of suspensions and the matrix method to determine the homotopy type of homotopy cofibers of certain maps. Section 4 simply analyses the homology decomposition of the suspension ΣM. In Section 5 we utilize the methods developed in Section 3 to give a detailed discussion on the homotopy decompositions of our suspended four-manifolds. At the end, we prove Theorems 1.1 and 1.2, respectively.

Acknowledgements The author would like to thank Jianzhong Pan for some helpful discussion on Proposition 5.2. The author was partially supported by the National Natural Science Foundation of China grant 12101290.

2 Some technical lemmas

In this section we recall some homotopy groups of mod 2^r Moore spaces and prove some lemmas about the Pontryagin or Steenrod square operations.

Throughout, all spaces $X, Y, \dots$ are based connected CW–complexes, and $[X, Y]$ is the set of based homotopy classes of based maps from X to Y. We identify a map f with its homotopy class in notation. For composable maps g and f, denote by gf or $g \circ f$ the composition of g with f. Unless otherwise specified, CX denotes the reduced mapping cone of a space X, and C_f denotes the homotopy cofiber of a given map $f\colon X \to Y$. For a cyclic group G, we write $G\langle x\rangle$ to mean x is a generator of G.

2.1 Some homotopy theory of mod 2^r Moore spaces

Let $n, k \geq 2$. There is a homotopy cofibration for the mod k Moore space $P^n(k)$:

$$S^{n-1} \xrightarrow{k} S^{n-1} \xrightarrow{i_{n-1}} P^n(k) \xrightarrow{q_n} S^n,$$

where i_{n-1} and q_n are the canonical inclusion and projection, respectively. Recall that if 2 doesn't divide k, then

$$\pi_n(P^n(k)) = \pi_{n+1}(P^n(k)) = 0 \quad \text{for all } n \geq 3.$$

For each $r, s \geq 1$, let $\rho_r\colon \mathbb{Z} \to \mathbb{Z}/2^r$ be the reduction mod 2^r with $1_r = \rho_r(1)$, let $\chi_s^r\colon \mathbb{Z}/2^r \to \mathbb{Z}/2^s$ be the homomorphism given by

$$\chi_s^r(1_r) = \begin{cases} 1_s & \text{if } r \geq s, \\ 2^{s-r} 1_s & \text{if } r < s. \end{cases} \tag{2-1}$$

For each $n \geq 3$, there exists a map (with n omitted in notation)

$$B(\chi_s^r)\colon P^{n+1}(2^r) \to P^{n+1}(2^s)$$

such that

$$H_n(B(\chi_s^r)) = \chi_s^r \quad \text{and} \quad \Sigma B(\chi_s^r) = B(\chi_s^r).$$

Moreover, $B(\chi_s^r)$ satisfies the relation formulas (cf [3])

$$B(\chi_s^r) i_n = \begin{cases} i_n & \text{if } r \geq s, \\ 2^{s-r} i_n & \text{if } r \leq s, \end{cases} \quad \text{and} \quad q_{n+1} B(\chi_s^r) = \begin{cases} 2^{r-s} q_{n+1} & \text{if } r \geq s, \\ q_{n+1} & \text{if } r \leq s. \end{cases} \tag{2-2}$$

Note that a multiple $t\alpha$ (written also as $t \cdot \alpha$) of an element $\alpha \in \pi_k(X)$ coincides with the composite $\alpha \circ t$.

Lemma 2.1 *Let $r \geq 1$ and $n \geq 3$ be integers.*

(1) $\pi_{n-1}(P^n(2^r)) \cong \mathbb{Z}/2^r \langle i_{n-1} \rangle$.

(2) $\pi_3(P^3(2^r)) \cong \mathbb{Z}/2^{r+1}\langle i_2\eta\rangle$, $\pi_{n+1}(P^{n+1}(2^r)) \cong \mathbb{Z}/2\langle i_n\eta\rangle$.

(3) *There are isomorphisms*

$$\pi_{n+1}(P^n(2^r)) \cong \begin{cases} \mathbb{Z}/4\langle \tilde{\eta}_1 \rangle & \text{if } r = 1, \\ \mathbb{Z}/2\langle \tilde{\eta}_r \rangle \oplus \mathbb{Z}/2\langle i_{n-1}\eta^2 \rangle & \text{if } r \geq 2, \end{cases}$$

where $\tilde{\eta}_r$ satisfies the formulas

$$\tilde{\eta}_r = B(\chi_r^1)\tilde{\eta}_1, \quad q_n \tilde{\eta}_r = \eta, \quad 2\tilde{\eta}_1 = i_{n-1}\eta^2. \tag{2-3}$$

(4) *Dually, there are isomorphisms*

$$\pi^n(P^{n+2}(2^r)) \cong \begin{cases} \mathbb{Z}/4\langle \bar{\eta}_1 \rangle & \text{if } r = 1, \\ \mathbb{Z}/2\langle \bar{\eta}_r \rangle \oplus \mathbb{Z}/2\langle \eta^2 q_{n+2} \rangle & \text{if } r \geq 2, \end{cases}$$

where $\bar{\eta}_r$ satisfies the formula

$$\bar{\eta}_r i_{n+1} = \eta_n, \quad 2\bar{\eta}_1 = \eta^2 q_{n+2}.$$

Proof (1) The isomorphism holds by the Hurewicz theorem.

(2) By [2, bottom of page 19, top of page 20], it holds that

$$\pi_n(P^n(2^r)) \cong \begin{cases} \Gamma(\mathbb{Z}/2^r) \cong \mathbb{Z}/2^{r+1} & \text{if } n = 3, \\ \mathbb{Z}/2^r \otimes \mathbb{Z}/2 \cong \mathbb{Z}/2 & \text{if } n \geq 4. \end{cases}$$

Here $\Gamma(\mathbb{Z}/2^r)$ is the Whitehead quadratic group; see [2] or [21]. The composite $i_{n-1}\eta$ is clearly a generator of $\pi_n(P^n(2^r))$.

(3) By [2, Proposition 11.1.12], $\pi_4(P^3(2^r))$ is isomorphic to the stable homotopy group $\pi_4^s(P^3(2^r))$, whose generators and the relations (2-3) refer to [3].

(4) The isomorphisms and the relation formulas follow by (3) under the Spanier–Whitehead duality:

$$\pi^n(P^{n+2}(2^r)) \cong \pi_{n+2}(P^{n+1}(2^r)).$$

□

For simplicity we still denote $\widetilde{\eta}_r \colon S^{n+1} \to P^n(2^r)$ the iterated suspensions of the generator $\widetilde{\eta}_r$ of $\pi_4(P^3(2^r))$. Combining (2-2) and (2-3), we have:

Corollary 2.2 *Let $r, s \geq 1$. There hold relations*

$$B(\chi_s^r)\widetilde{\eta}_r = \begin{cases} \widetilde{\eta}_s & \text{if } s \geq r, \\ 2^{r-s}\widetilde{\eta}_s & \text{if } s \leq r. \end{cases}$$

2.2 Whitehead's quadratic functor

Recall the *Whitehead quadratic functor*

$$\Gamma \colon \mathbf{Ab} \to \mathbf{Ab}$$

on the category **Ab** of abelian groups [21; 1]. The functor Γ is characterized by the following property: a function $\varphi \colon G \to G'$ between abelian groups is called *quadratic* if $\varphi(x) = \varphi(-x)$ and the function $G \times G \to G'$ with $(x, y) \mapsto \varphi(x+y) - \varphi(x) - \varphi(y)$ is bilinear. For each abelian group G, there is a *universal quadratic function*

$$\gamma = \gamma_G \colon G \to \Gamma(G)$$

such that for any quadratic function $\varphi \colon G \to G'$, there is a unique homomorphism $\varphi^\square \colon \Gamma(G) \to G'$ such that $\varphi = \varphi^\square \circ \gamma$. It follows that for a homomorphism $\phi \colon G \to G'$, there is a unique induced homomorphism $\Gamma(\phi) \colon \Gamma(G) \to \Gamma(G')$ such that $\Gamma(\phi) \circ \gamma_G = \gamma_{G'} \circ \phi$. The universal quadratic function $\gamma = \gamma_G$ induces the bilinear pairing

$$[1,1] \colon G \otimes G \to \Gamma(G), \quad [1,1](x, y) = \gamma(x+y) - \gamma(x) - \gamma(y). \tag{2-4}$$

Lemma 2.3 (cf [2]) *Let G be an abelian group and let $n \geq 0$.*

1. *For the cyclic group $G = \mathbb{Z}/n$ we have*

$$\Gamma(\mathbb{Z}/n) \cong \mathbb{Z}/(n^2, 2n),$$

where $\mathbb{Z}/0 = \mathbb{Z}$ and $(n^2, 2n)$ is the greatest common divisor. The group is generated by $\gamma(1_n)$ with $1_n = 1 + n\mathbb{Z}$.

2. *For any $x \in G$, there holds $\gamma(nx) = n^2\gamma(x)$.*

2.3 Squaring operations

For an abelian group G, the Pontryagin square

$$\mathfrak{P}_1 \colon H^2(X; G) \to H^4(X; \Gamma(G))$$

is a *quadratic* function with respect to the cup product $\smile$:

$$\mathfrak{P}_1(-x) = \mathfrak{P}_1(x), \quad \mathfrak{P}_1(nx) = n^2\mathfrak{P}_1(x), \quad \mathfrak{P}_1(x+y) = \mathfrak{P}_1(x) + \mathfrak{P}_1(y) + [1,1]_*(x \smile y), \tag{2-5}$$

where $[1,1]_*$ is induced by the coefficient homomorphism (2-4). The Pontryagin square is natural with respect to maps $X \to Y$ between spaces and with respect to homomorphisms $G \to G'$ between groups.

Let X be an A_2^2–complex and let

$$C_4(X) \xrightarrow{d} C_3(X) \xrightarrow{d} C_2(X)$$

be its cellular chain complex. Represent a cohomology class $x \in H^2(X;G)$ by a cocycle $\hat{x}\colon C_2(X) \to G$, which induces a unique homomorphism

$$\tilde{x}\colon H_2(X) = C_2(X)/dC_3(X) \to G,$$

and therefore a unique homomorphism

$$\Gamma(\tilde{x})\colon \Gamma(H_2(X)) \to \Gamma(G).$$

By the universal coefficient theorem, there is an isomorphism

$$\mu\colon H^2(X;H_2(X)) \xrightarrow{\cong} \mathrm{Hom}(H_2(X), H_2(X)).$$

Let $\iota_2 \in H^2(X;H_2(X))$ be given such that $\mu(\iota_2)$ is the identity on $H_2(X)$. By [1, Chapter I] we know that the Pontryagin square

$$\mathfrak{P}_1\colon H^2(X;G) \to H^4(X;\Gamma(G))$$

is completely determined by the *Pontryagin element*

$$\mathfrak{P}_1(\iota_2) \in H^4(X;\Gamma(H_2(X)))$$

in the sense that there holds an formula

$$\mathfrak{P}_1(x) = \Gamma(\tilde{x})_*(\mathfrak{P}_1(\iota_2)),$$

where $\Gamma(\tilde{x})_*$ is induced by the coefficient homomorphism.

Let $C_r(t\eta)$ be the homotopy cofiber of $t \cdot i_2\eta\colon S^3 \to P^3(2^r)$, where $r \geq 1$ and $t \in \mathbb{Z}/2^{r+1}$. Note that $C_r(t\eta)$ is an A_2^2–polyhedron and has the *A_2^2–form*

$$f = (t\eta, 2^r)\colon S^3 \vee S^2 \to S^2, \tag{2-6}$$

ie $C_r(t\eta)$ is the homotopy cofiber of the attaching map f between spheres.

Lemma 2.4 *Let $t \in \mathbb{Z}/2^{r+1}$ and $r \geq 1$. The Pontryagin square*

$$\mathfrak{P}_1\colon H^2(C_r(t\eta);\mathbb{Z}/2^r) \to H^4(C_r(t\eta);\mathbb{Z}/2^{r+1})$$

is trivial if and only if $t = 0$.

Proof Let $\iota_2 \in H^2(C_r(t\eta);\mathbb{Z}/2^r)$ be the generator which corresponds to the identity on $H_2(C_r(t\eta))$. By [1, Chapter I, Proposition 7.6] and the A_2^2–form (2-6), the Pontryagin element $\mathfrak{P}_1(\iota_2)$ is represented by the cocycle

$$t \cdot \Gamma(\rho_r)\gamma = \Gamma(\rho_r)(t\gamma)\colon \mathbb{Z} \xrightarrow{t\gamma} \Gamma(\mathbb{Z}) \xrightarrow{\Gamma(\rho_r)} \Gamma(\mathbb{Z}/2^r).$$

Note that $\Gamma(\rho_r)\gamma = \gamma\rho_r$ represents a generator of $H^4(C_r(t\eta);\Gamma(\mathbb{Z}/2^r))$ by the universal coefficient theorem. Then it follows by Lemma 2.3 that $\mathfrak{P}_1 = 0$ if and only if $t = 0$. □

Recall that the Steenrod square

$$\mathrm{Sq}^2\colon H^n(-;\mathbb{Z}/2)\to H^{n+2}(-;\mathbb{Z}/2)$$

is a stable cohomology operation such that $\mathrm{Sq}^2(x)=x^2$ for any cohomology class x of dimension 2; see [10, Section 4.L].

Lemma 2.5 (cf [24]) *For any $n\geq 3$, the Steenrod square*

$$\mathrm{Sq}^2\colon H^n(C;\mathbb{Z}/2)\to H^{n+2}(C;\mathbb{Z}/2)$$

is an isomorphism for each (elementary) Chang complex C.

Lemma 2.6 *For each $n\geq 2$ and $r\geq 1$, the Steenrod square*

$$\mathrm{Sq}^2\colon H^{n+1}(A^{n+3}(\tilde{\eta}_r);\mathbb{Z}/2)\to H^{n+3}(A^{n+3}(\tilde{\eta}_r);\mathbb{Z}/2)$$

is an isomorphism.

Proof By (2-3) there is a homotopy commutative diagram of homotopy cofibrations (in which rows and columns are all homotopy cofibrations):

$$\begin{array}{ccccc}
* & \longrightarrow & S^n & = & S^n \\
\downarrow & & \downarrow{\scriptstyle i_n} & & \downarrow \\
S^{n+2} & \xrightarrow{\tilde{\eta}_r} & P^{n+1}(2^r) & \longrightarrow & A^{n+3}(\tilde{\eta}_r) \\
\| & & \downarrow{\scriptstyle q_{n+1}} & & \downarrow{\scriptstyle d} \\
S^{n+2} & \xrightarrow{\eta} & S^{n+1} & \longrightarrow & C_\eta^{n+3}
\end{array}$$

It follows that $d^*\colon H^k(C_\eta^{n+3};\mathbb{Z}/2)\to H^k(A^{n+3}(\tilde{\eta}_r);\mathbb{Z}/2)$ is an isomorphism for $k=n+1,n+3$. The isomorphism in the lemma then follows by Lemma 2.5 and the commutative square

$$\begin{array}{ccc}
H^{n+1}(C_\eta^{n+3};\mathbb{Z}/2) & \xrightarrow[\cong]{\mathrm{Sq}^2} & H^{n+3}(C_\eta^{n+3};\mathbb{Z}/2) \\
\cong\downarrow{\scriptstyle d^*} & & \cong\downarrow{\scriptstyle d^*} \\
H^{n+1}(A^{n+3}(\tilde{\eta}_r);\mathbb{Z}/2) & \xrightarrow{\mathrm{Sq}^2} & H^{n+3}(A^{n+3}(\tilde{\eta}_r);\mathbb{Z}/2)
\end{array}$$

□

2.4 Higher-order cohomology operations

Recall the secondary cohomology operations

$$\Theta_n\colon S_n(X)\to T_n(X) \tag{2-7}$$

based on the relation $\varphi_n\theta_n=0$ of (1-4), where

$$S_n(X)=\ker(\theta_n)_\sharp=\ker(\mathrm{Sq}^2)\cap\ker(\mathrm{Sq}^2\,\mathrm{Sq}^1),$$
$$T_n(X)=\operatorname{coker}(\Omega\varphi_n)_\sharp=H^{n+3}(X;\mathbb{Z}/2)/\operatorname{im}(\mathrm{Sq}^1+\mathrm{Sq}^2).$$

Lemma 2.7 *Let $n \geq 2$ and $r \geq 1$. For $X = A^{n+3}(\eta^2)$ or $A^{n+3}(2^r\eta^2)$, the secondary operation Θ_n acts nontrivially on $H^n(X;\mathbb{Z}/2)$; that is,*

$$0 \neq \Theta_n \colon H^n(X;\mathbb{Z}/2) \to H^{n+3}(X;\mathbb{Z}/2).$$

Proof For $X = A^{n+3}(\eta^2)$ or $A^{n+3}(2^r\eta)$, we compute that

$$S_n(X) = H^n(X;\mathbb{Z}/2) \cong \mathbb{Z}/2 \quad \text{and} \quad T_n(X) = H^{n+3}(X;\mathbb{Z}/2) \cong \mathbb{Z}/2.$$

The proof of $\Theta_n \neq 0$ for $X = A^{n+3}(\eta^2)$ refers to [9, page 96]. There is a homotopy cofibration

$$S^n \xrightarrow{i_n \circ 2^r} A^{n+3}(\eta^2) \xrightarrow{j} A^{n+3}(2^r\eta),$$

which induces the commutative square

$$\begin{array}{ccc} H^n(A^{n+3}(2^r\eta);\mathbb{Z}/2) & \xrightarrow{\Theta_n} & H^{n+3}(A^{n+3}(2^r\eta);\mathbb{Z}/2) \\ \cong \big\downarrow j^* & & \cong \big\downarrow j^* \\ H^n(A^{n+3}(\eta^2);\mathbb{Z}/2) & \xrightarrow{\Theta_n \neq 0} & H^{n+3}(A^{n+3}(\eta^2);\mathbb{Z}/2) \end{array}$$

Thus $\Theta \neq 0$ for $X = A^{n+3}(2^r\eta)$. □

The higher-order Bocksteins (1-5)

$$\beta_r \colon H^n(X;\mathbb{Z}/2) \dashrightarrow H^{n+1}(X;\mathbb{Z}/2)$$

are helpful to detect torsion elements of $H_*(X;\mathbb{Z})$ or $H^*(X;\mathbb{Z})$.

Lemma 2.8 (cf [17, pages 173 and 61]) *The following statements hold*:

(1) *The higher Bockstein β_r detects the degree 2^r map on S^n; in other words, for each $r \geq 1$, there is exactly one nontrivial higher Bockstein*

$$\beta_r \colon H^{n-1}(P^n(2^r);\mathbb{Z}/2) \to H^n(P^n(2^r);\mathbb{Z}/2).$$

(2) *For each $r \geq 1$, elements of $H^*(X;\mathbb{Z}/2)$ coming from free integral homology class lie in* $\ker(\beta_r)$ *and not in* $\operatorname{im}(\beta_r)$.

(3) *If $z \in H^{n+1}(X;\mathbb{Z})$ generates a direct summand $\mathbb{Z}/2^r$ for some r, then there exist generators $z' \in H^n(X;\mathbb{Z}/2)$ and $z'' \in H^{n+1}(X;\mathbb{Z}/2)$ such that*

$$\beta_r(z') = z'' \qquad \text{and} \qquad \beta_i(z') = \beta_i(z'') = 0 \quad \text{for } i < r.$$

3 Analysis methods

In this section we list some auxiliary lemmas that simplify the proof arguments in the next section. These lemmas appear to be applicable to other similar problems as well, so we leave them in a separate section.

We say that a map $f: X \to Y$ is *homologically trivial* if the induced homomorphism $f_*: H_i(X) \to H_i(Y)$ is trivial for each i.

Lemma 3.1 [10, Theorem 4H.3] *Suppose that X is a simply connected space of dimension N. Write $H_i = H_i(X)$. Then there is a sequence $X_2 \subseteq X_3 \subseteq \cdots \subseteq X_m$ of subcomplexes X_j of X such that*

1. $i_*: H_j(X_n) \cong H_j(X)$ *for* $j \le n$ *and* $H_j(X_n) = 0$ *for* $j > n$,
2. $X_2 = M_2(H_2)$ *and* $X_N = X$,
3. *there is a principal homotopy cofibration*

$$M_n(H_{n+1}) \xrightarrow{k_n} X_n \xrightarrow{i_n} X_{n+1} \to M_{n+1}(H_{n+1})$$

with k_n homologically trivial.

Note that we have the canonical inclusions $X^n \subseteq X_n \subseteq X^{n+1}$, where X^k denotes the k–skeleton of X. The map k_n above is called the n^{th} *k'–invariant*, and plays a key role in the homology decomposition of X. For instance, k_n is null-homotopic if and only if $X_n \simeq X_{n-1} \vee M_n(H_n X)$.

Lemma 3.2 *Let $f: \bigvee_{i=1}^m A_i \to \bigvee_{j=1}^n B_j$ be a map which induces trivial homomorphism in cohomology groups with coefficients in abelian groups G and G'. Let*

$$f_\jmath = p_\jmath \circ f \quad \text{and} \quad f_{\imath,\jmath} = f_\jmath \circ i_\imath = p_\jmath \circ f \circ i_\imath,$$

where $i_\imath: A_\imath \to \bigvee_{i=1}^m A_i$ and $p_\jmath: \bigvee_{j=1}^n B_j \to B_\jmath$ are respectively the canonical inclusion and projection, with $1 \le \imath \le m$ and $1 \le \jmath \le n$.

1. *If $H^*(C_f; G)$ contains no nontrivial cup products, then so do $H^*(C_{f_\jmath}; G)$ and $H^*(C_{f_{\imath,\jmath}}; G)$ for all $\imath$ and $\jmath$.*
2. *If the cohomology operation $\mathbb{O}: H^k(C_f; G) \to H^l(C_f; G')$ is trivial, then so are the operations*

$$\mathbb{O}_\jmath: H^k(C_{f_\jmath}; G) \to H^l(C_{f_\jmath}; G') \quad \text{and} \quad \mathbb{O}_{\imath\jmath}: H^k(C_{f_{\imath\jmath}}; G) \to H^l(C_{f_{\imath\jmath}}; G').$$

where $\mathbb{O}_\jmath$ and $\mathbb{O}_{\imath\jmath}$ are the cohomology operation of the same type as $\mathbb{O}$.

Proof (1) The statement (1) is due to [19, Lemma 4.2].

(2) By the proof of [19, Lemma 4.2], for any integer $k \ge 0$ and any coefficient group G there are monomorphisms

$$d_\jmath^*: H^k(C_{f_\jmath}; G) \to H^k(C_f; G)$$

and epimorphisms

$$d_{\imath\jmath}^*: H^k(C_{f_\jmath}; G) \to H^k(C_{f_{\imath\jmath}^*}; G).$$

Consider the commutative diagrams

$$\begin{array}{ccccc}
H^k(C_f;G) & \xleftarrow{d_\jmath^*} & H^k(C_{f_\jmath};G) & \xrightarrow{d_{\imath,\jmath}^*} & H^k(C_{f_{\imath,\jmath}};G) \\
\downarrow{\scriptstyle \mathbb{C}} & & \downarrow{\scriptstyle \mathbb{C}_\jmath} & & \downarrow{\scriptstyle \mathbb{C}_{\imath,\jmath}} \\
H^l(C_f;G') & \xleftarrow{d_\jmath^*} & H^l(C_{f_\jmath};G') & \xrightarrow{d_{\imath,\jmath}^*} & H^k(C_{f_{\imath,\jmath}};G')
\end{array}$$

It follows that $\mathbb{C}_\jmath$ is the restriction of $\mathbb{C}$, and $\mathbb{C}_{\imath\jmath}$ is induced by $\mathbb{C}_\jmath$. Thus if $\mathbb{C}$ is trivial, so are $\mathbb{C}_\jmath$ and $\mathbb{C}_{\imath,\jmath}$. □

The following lemma is useful to determine the homotopy type of a suspension; see [14, Lemma 6.4] or [19, Lemma 5.6].

Lemma 3.3 *Let $S \xrightarrow{f} \left(\bigvee_{i=1}^n A_i\right) \vee B \xrightarrow{g} \Sigma C$ be a homotopy cofibration of simply connected CW–complexes. For $j = 1, \dots, n$, let $p_j : \bigvee_i A_i \to A_j$ be the canonical projection onto the wedge summand A_j. Suppose that each composition*

$$f_j : S \xrightarrow{f} \bigvee_i A_i \xrightarrow{p_j} A_j$$

is null-homotopic. Then there is a homotopy equivalence

$$\Sigma C \simeq \bigvee_{i=1}^n A_i \vee D,$$

where D is the homotopy cofiber of the composition $S \xrightarrow{f} \left(\bigvee_i A_i\right) \vee B \xrightarrow{q_B} B$, with q_B the obvious projection.

Let $X = \Sigma X'$ and $Y_i = \Sigma Y_i'$ be suspensions, for $i = 1, 2, \dots, n$. Let

$$i_l : Y_l \to \bigvee_{j=i}^n Y_i \quad \text{and} \quad p_k : \bigvee_{i=1}^n Y_i \to Y_k$$

be, respectively, the canonical inclusions and projections, for $1 \le k, l \le n$. By the Hilton–Milnor theorem, we may write a map $f : X \to \bigvee_{i=1}^n Y_i$ as

$$f = \sum_{k=1}^n i_k \circ f_k + \theta,$$

where $f_k = p_k \circ f : X \to Y_k$ and θ satisfies $\Sigma\theta = 0$. The first part $\sum_{k=1}^n i_k \circ f_k$ is usually represented by a vector

$$u_f = (f_1, f_2, \dots, f_n)^t.$$

We say that f is completely determined by its components f_k if $\theta = 0$; in this case, write $f = u_f$. Let $h = \sum_{k,l} i_l h_{lk} p_k$ be a self-map of $\bigvee_{i=1}^n Y_i$ which is completely determined by its components $h_{kl} = p_k \circ h \circ i_l : Y_l \to Y_k$. Write

$$M_h := (h_{kl})_{n\times n} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n1} & \cdots & h_{nn} \end{pmatrix}.$$

Then the composition law $h(f+g) \simeq hf + hg$ implies that the product

$$M_h(f_1, f_2, \dots, f_n)^t$$

given by the matrix multiplication represents the composite $h \circ f$. Two maps $f = u_f$ and $g = u_g$ are called *equivalent*, and we write

$$(f_1, f_2, \dots, f_n)^t \sim (g_1, g_2, \dots, g_n)^t,$$

if there is a self-homotopy equivalence h of $\bigvee_{i=1}^n Y_i$, which can be represented by the matrix M_h, such that

$$M_h(f_1, f_2, \dots, f_n)^t \simeq (g_1, g_2, \dots, g_n)^t.$$

Recall that the above matrix multiplication refers to elementary row operations in matrix theory; and note that the homotopy cofibers of the maps $f = u_f$ and $g = u_g$ are homotopy equivalent if f and g are equivalent.

The following lemma serves as an example of the above matrix method.

Lemma 3.4 *Define X by the homotopy cofibration*

$$S^4 \xrightarrow{(f_1, f_2, \dots, f_n)^t} \bigvee_{j=1}^n V_j \longrightarrow X,$$

where $f_j \colon S^4 \to V_j$ for $j = 1, \dots, n$.

(1) *If $V_j = S^3$ for $j = 1, 2, \dots, n$ and $f_{j_0} = \eta$ for some j_0, then there is a homotopy equivalence*

$$X \simeq C_\eta^5 \vee \bigvee_{j \neq j_0} S^3.$$

(2) *If $V_j = P^4(2^{r_j})$ for $j = 1, 2, \dots, n$, and $f_j = i_3\eta$ for some j, then there is a homotopy equivalence*

$$X \simeq C_{r_{j_1}}^5 \vee \bigvee_{j \neq j_1} P^4(2^{r_j}),$$

where $j_1 = \max\{1 \le j \le n \mid f_j = i_3\eta\}$.

Proof (1) If there is a unique $f_{j_0} = \eta$, the statement clearly holds. We may assume that $f_1 = \eta$ and $f_i = \varepsilon_i \cdot \eta$, with $\varepsilon_i \in \{0, 1\}$. Then

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\varepsilon_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\varepsilon_n & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \eta \\ \varepsilon_2 \cdot \eta \\ \vdots \\ \varepsilon_n \cdot \eta \end{pmatrix} \simeq \begin{pmatrix} \eta \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It follows that there exists a self-homotopy equivalence e_S of $\bigvee_{j=1}^m S^3$ such that

$$e_S f \sim (\eta, 0, \dots, 0)^t,$$

and hence there is a homotopy equivalence

$$X = C_f \simeq C_{e_S f} \simeq C^5_\eta \vee \bigvee_{j=2}^{m} S^3.$$

(2) The statement clearly holds if there is a unique j such that $f_j = i_3\eta$. Let j_1 be defined in the lemma. If there is an index j_2 such that

$$f_{j_2} = i_3\eta \in \pi_4(P^4(2^{r_{j_2}})),$$

then the matrix multiplication

$$\begin{pmatrix} 1_P & 0 \\ -B(\chi^r_s) & 1_P \end{pmatrix} \begin{pmatrix} i_3\eta \\ i_3\eta \end{pmatrix} \simeq \begin{pmatrix} i_3\eta \\ 0 \end{pmatrix}$$

implies $(f_{j_1}, f_{j_2})^t \sim (f_{j_1}, 0)^t$. By induction, it follows that there exists a self-homotopy equivalence e_P of $\bigvee_{j=1}^m P^4(2^{r_j})$ such that

$$e_P \circ (f_1, f_2, \dots, f_n)^t \simeq (0, \dots, 0, i_3\eta, 0, \dots, 0)^t,$$

where $i_3\eta$ in the latter vector lies in the j_1^{th} position. Thus we have a homotopy equivalence

$$X = C_f \simeq C^5_{r_{j_1}} \vee \bigvee_{j \neq j_1} P^4(2^{r_j}). \qquad \square$$

4 Homology decomposition of ΣM

By Table 1 and [19, Lemma 5.1], there is a homotopy equivalence

$$\Sigma M \simeq \Big(\bigvee_{i=1}^{m} S^2\Big) \vee \Sigma W, \tag{4-1}$$

where W is a CW–complex with integral homology given by Table 2. By Lemma 3.1 and Table 2, there are homotopy cofibrations

$$\bigvee_{i=1}^{d} S^2 \vee P^3(T) \xrightarrow{k_3} P^3(T) \to W_3, \quad \bigvee_{i=1}^{m} S^3 \xrightarrow{k_4} W_3 \to W_4, \quad S^4 \xrightarrow{k_5} W_4 \to \Sigma W, \tag{4-2}$$

where k_3, k_4, k_5 are homologically trivial maps. Let $T_2 = \bigoplus_{j=1}^n \mathbb{Z}/2^{r_j}$ be the 2–primary component of T and write $T = T_2 \oplus T_{\neq 2}$. For each $k \geq 3$, there are homotopy equivalences (cf [18])

$$P^k(T) \simeq P^k(T_2) \oplus P^k(T_{\neq 2}) \simeq \Big(\bigvee_{j=1}^{n} P^k(2^{r_j})\Big) \vee P^k(T_{\neq 2}).$$

i	0, 4	1	2	3	≥ 5
$H_i(W)$	$\mathbb{Z}$	T	$\mathbb{Z}^d \oplus T$	$\mathbb{Z}^m$	0

Table 2: $H_*(W; \mathbb{Z})$.

Lemma 4.1 *There is a homotopy equivalence*

$$W_3 \simeq \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^3(T) \vee P^4(T).$$

Proof By (4-2), there is a homotopy cofibration

$$\Big(\bigvee_{i=1}^{d} S^2\Big) \vee P^3(T) \xrightarrow{f} P^3(T) \to W_3,$$

where f is a homologically trivial map with its two components of the types

$$f_1^S : \Big(\bigvee_{i=1}^{d} S^2\Big) \hookrightarrow \Big(\bigvee_{i=1}^{d} S^2\Big) \vee P^3(T) \xrightarrow{f} P^3(T) \quad \text{and} \quad f_2^T : P^3(T) \hookrightarrow \Big(\bigvee_{i=1}^{d} S^2\Big) \vee P^3(T) \xrightarrow{f} P^3(T).$$

Here the arrows $\hookrightarrow$ denote the obvious inclusions. Clearly f_1^S and f_2^T are both homologically trivial. Set $T = \bigoplus_{k=1}^{l} p_k^{r_k}$ with p_k primes. Then the Hurewicz isomorphism $\pi_2(P^3(T)) \cong H_2(P^3(T))$ implies that both f_1^S and the composite

$$S_T = \bigvee_{k=1}^{l} S^2 \xrightarrow{j} P^3(T) \xrightarrow{f_2^T} P^3(T)$$

are null-homotopic, where j is the canonical inclusion. Let

$$m_T = \bigvee_k p_k^{r_k} : S_T \to S_T$$

be the attaching map of $P^3(T)$. There is a homotopy commutative diagram of homotopy cofibrations

$$\begin{array}{ccccc}
S_T & \longrightarrow & * & \longrightarrow & \Sigma S_T \\
\downarrow{\scriptstyle m_T} & & \downarrow & & \downarrow{\scriptstyle i_2 \circ (\Sigma m_T)} \\
S_T & \xrightarrow{0} & P^3(T) & \longrightarrow & P^3(T) \vee \Sigma S_T \\
\downarrow{\scriptstyle i} & & \| & & \downarrow \\
P^3(T) & \xrightarrow{f_2^T} & P^3(T) & \longrightarrow & C_{f_2^T}
\end{array}$$

in which rows are columns are homotopy cofibrations. It follows that

$$C_{f_2^T} \simeq P^3(T) \vee P^4(T),$$

and hence there is a homotopy equivalence

$$W_3 \simeq \bigvee_{i=1}^{d} S^3 \vee C_{f_2^T} \simeq \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^3(T) \vee P^4(T).$$ □

Lemma 4.2 *There is a homotopy equivalence*

$$W_4 \simeq \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^4(T) \vee C_{g_2}$$

for some homologically trivial map $g_2 : \bigvee_{i=1}^{m} S^3 \to P^3(T)$. *Moreover, there is a homotopy equivalence*

$$\Sigma W_4 \simeq \Big(\bigvee_{i=1}^{d} S^4\Big) \vee P^4(T) \vee P^5(T) \vee \bigvee_{i=1}^{m} S^5.$$

Proof By (4-2) and Lemma 4.1, there is a homotopy cofibration

$$\bigvee_{i=1}^{m} S^3 \xrightarrow{g} \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^3(T) \vee P^4(T) \to W_4,$$

with g a homologically trivial map. The map g is determined by the following components:

$$g_1 \colon S^3 \to \bigvee_{i=1}^{m} S^3 \xrightarrow{g} \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^3(T) \vee P^4(T) \to \bigvee_{i=1}^{d} S^3 \to S^3,$$
$$g_2 \colon S^3 \to \bigvee_{i=1}^{m} S^3 \xrightarrow{g} \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^3(T) \vee P^4(T) \to P^3(T),$$
$$g_3 \colon S^3 \to \bigvee_{i=1}^{m} S^3 \xrightarrow{g} \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^3(T) \vee P^4(T) \to P^4(T).$$

Here the unlabeled maps are the obvious inclusions and projections. The maps g_1, g_2, g_3 are all homologically trivial. The Hurewicz theorem implies that both g_1 and g_3 are null-homotopic. Then by Lemma 3.3 we get the first statement.

To prove the second homotopy equivalence, it suffices to show that if $f : S^4 \to P^4(T)$ is homologically trivial, then f is null-homotopic. Consider the following homologically trivial components of f:

$$f_1 \colon S^4 \xrightarrow{f} P^4(T) \to P^4(T_{\neq 2}),$$
$$f_2^j : S^4 \xrightarrow{f} P^4(T) \to P^4(T_2) \to P^4(2^{r_j}) \quad \text{for } j = 1, 2, \dots, n.$$

The map f_1 is clearly null-homotopic, because $\pi_4(P^4(p^r)) = 0$ for odd primes p. Observe that $W_4 = \Sigma W^4$ is a suspension, the Steenrod square Sq^2 acts trivially on $H^2(W_4; \mathbb{Z}/2)$. By Lemma 3.2(2), Sq^2 acts trivially on $H^3(C_{f_2^j}; \mathbb{Z}/2)$. Since $\pi_4(P^4(2^{r_j})) \cong \mathbb{Z}/2\langle i_3\eta\rangle$ (Lemma 2.1), we may set

$$f_2^j = \varepsilon_j \cdot i_3\eta, \quad \text{where } \varepsilon_j \in \mathbb{Z}/2.$$

Note that the homotopy cofiber of $i_3\eta \in \pi_4(P^4(2^{r_j}))$ is the Chang complex $C^5_{r_j}$, by Lemma 2.5 we then get that $\varepsilon_j = 0$, or equivalently f_2^j is null-homotopic for each $j = 1, 2, \dots, n$. Thus f is null-homotopic, by Lemma 3.3. □

5 Proofs of Theorems 1.1 and 1.2

By (4-2) and Lemma 4.2, there is a homotopy cofibration

$$S^5 \xrightarrow{h} \Big(\bigvee_{i=1}^{d} S^4\Big) \vee P^4(T) \vee P^5(T) \vee \bigvee_{i=1}^{m} S^5 \to \Sigma^2 W,$$

where h a homologically trivial map, $T \cong T_2 \oplus T_{\neq 2}$ with $T_2 \cong \bigoplus_{j=1}^{n} \mathbb{Z}/2^{r_j}$. Since $\pi^5(P^4(p^r)) = \pi_5(P^5(p^s)) = 0$ for any odd primes p, Lemma 3.3 indicates that there is a homotopy equivalence

$$\Sigma^2 W \simeq P^4(T_{\neq 2}) \vee P^5(T_{\neq 2}) \vee \Big(\bigvee_{i=1}^{m} S^5\Big) \vee C_\varphi, \tag{5-1}$$

where $\varphi\colon S^5 \to \big(\bigvee_{i=1}^d S^4\big) \vee P^4(T_2) \vee P^5(T_2)$ is a homologically trivial map. The map φ has the following three types of components:

$$\varphi_1\colon S^5 \xrightarrow{\varphi} \Big(\bigvee_{i=1}^d S^4\Big) \vee P^4(T_2) \vee P^5(T_2) \to S^4,$$
$$\varphi_2^j\colon S^5 \xrightarrow{\varphi} \Big(\bigvee_{i=1}^d S^4\Big) \vee P^4(T_2) \vee P^5(T_2) \to P^4(T_2) \to P^4(2^{r_j}),$$
$$\varphi_3^j\colon S^5 \xrightarrow{\varphi} \Big(\bigvee_{i=1}^d S^4\Big) \vee P^4(T_2) \vee P^5(T_2) \to P^5(T_2) \to P^5(2^{r_j}),$$

where $j = 1, 2, \dots, n$ and the unlabeled maps are the obvious projections.

Proposition 5.1 *If* $\mathrm{Sq}^2(H^4(\Sigma^2 W;\mathbb{Z}/2)) = 0$, *then the homotopy type of* $\Sigma^2 W$ *is determined by the secondary operation* Θ *of equation* (2-7) *and the higher Bockstein* β_r. *Explicitly, if* $\Theta(H^3(C_\varphi;\mathbb{Z}/2)) = 0$, *then there is a homotopy equivalence*

$$C_\varphi \simeq \Big(\bigvee_{i=1}^d S^4\Big) \vee P^4(T_2) \vee P^5(T_2) \vee S^6.$$

Otherwise we have

$$C_\varphi \simeq \Big(\bigvee_{i=1}^d S^4\Big) \vee P^4\Big(\frac{T_2}{\mathbb{Z}/2^{r_{j_0}}}\Big) \vee P^5(T_2) \vee A^6(2^{r_{j_0}}\eta^2),$$

where j_0 *is the maximum of the indices* j *satisfying*

$$\Theta(x) \neq 0 \quad \text{and} \quad \beta_{r_j}(x) \neq 0 \quad \text{for all } x \in H^3(C_\varphi;\mathbb{Z}/2).$$

Proof By assumption and (5-1), Sq^2 acts trivially on $H^4(C_\varphi;\mathbb{Z}/2)$, and hence so does Sq^2 on $H^4(C_{\varphi_1};\mathbb{Z}/2)$ and $H^4(C_{\varphi_k^j};\mathbb{Z}/2)$ for each $k = 2, 3$ and $j = 1, 2, \dots, n$, by Lemma 3.2(2). It follows by Lemmas 2.5 and 2.6 that φ_1 and φ_3^i are null-homotopic, and

$$\varphi_2^j = y_j \cdot i_3\eta^2 \quad \text{for all } y_j \in \mathbb{Z}/2 \text{ and } j = 1, 2, \dots, n.$$

By Lemma 2.7, the coefficients y_j can be detected by the secondary operation Θ. There are possibly many such indices j; however, similar arguments to those in the proof of Lemma 3.4 show that there exists a homotopy equivalence e of $P^4(T_2)$ such that

$$e(\varphi_2^1, \varphi_2^2, \dots, \varphi_2^n)^t \simeq (0, \dots, 0, \varphi_2^{j_0} = i_3\eta^2, 0, \dots, 0)^t,$$

with j_0 described in the proposition. The proof is then completed by applying Lemma 3.3. □

Proposition 5.2 *Suppose that* $\mathrm{Sq}^2(H^4(\Sigma^2 W;\mathbb{Z}/2)) \neq 0$. *Then the homotopy types of* C_φ *or* ΣC_φ *can be characterized as follows*:

(1) *Suppose that for any* $u \in H^4(\Sigma^2 M;\mathbb{Z}/2)$ *with* $\mathrm{Sq}^2(u)\neq 0$ *and any* $v\in\ker(\mathrm{Sq}^2)$, *it holds that*

$$\beta_r(u+v)=0 \quad \textit{and} \quad u+v\notin \mathrm{im}(\beta_s) \quad \textit{for all } r,s\geq 1.$$

Then there is a homotopy equivalence

$$C_\varphi \simeq \Big(\bigvee_{i=1}^{d-1} S^4\Big)\vee P^5(T_2)\vee P^4(T_2)\vee C^6_\eta.$$

(2) *Suppose that for any* $u \in H^4(\Sigma^2 M;\mathbb{Z}/2)$ *with* $\mathrm{Sq}^2(u)\neq 0$ *and any* $v\in\ker(\mathrm{Sq}^2)$, *it holds that*

$$u+v\notin \mathrm{im}(\beta_s) \quad \textit{for all } s\geq 1,$$

while there exist $u' \in H^4(\Sigma^2 M;\mathbb{Z}/2)$ *with* $\mathrm{Sq}^2(u')\neq 0$ *and* $v'\in\ker(\mathrm{Sq}^2)$ *such that*

$$\beta_r(u'+v')\neq 0 \quad \textit{for some } r\geq 1.$$

Then there is a homotopy equivalence

$$C_\varphi \simeq \Big(\bigvee_{i=1}^{d} S^4\Big)\vee P^5\Big(\frac{T_2}{\mathbb{Z}/2^{r_{j_1}}}\Big)\vee P^4(T_2)\vee C^6_{r_{j_1}},$$

with j_1 *the maximum of indices* j *such that*

$$\mathrm{Sq}^2(u')\neq 0 \quad \textit{and} \quad \beta_{r_{j_1}}(u'+v')\neq 0.$$

(3) *Suppose that there exist* $u \in H^4(\Sigma^2 M;\mathbb{Z}/2)$ *with* $\mathrm{Sq}^2(u)\neq 0$ *and* $v\in\ker(\mathrm{Sq}^2)$ *such that*

$$u+v\in \mathrm{im}(\beta_r) \quad \textit{for some } r\geq 1.$$

(a) *If* $\Theta(H^3(C_\varphi;\mathbb{Z}/2))=0$, *then there is a homotopy equivalence*

$$C_\varphi \simeq \Big(\bigvee_{i=1}^{d} S^4\Big)\vee P^5(T_2)\vee P^4\Big(\frac{T_2}{\mathbb{Z}/2^{r_{j_2}}}\Big)\vee A^6(\tilde{\eta}_{r_{j_2}}),$$

with j_2 *the minimum of the indices* j *such that* $u+v\in\mathrm{im}(\beta_{r_j})$.

(b) *If* $\Theta(H^3(C_\varphi;\mathbb{Z}/2))\neq 0$ *and* $T_2\cong\mathbb{Z}/2^{r_{j_2}}$, *then*

$$C_\varphi \simeq \Big(\bigvee_{i=1}^{d} S^4\Big)\vee P^5(T_2)\vee P^4\Big(\frac{T_2}{\mathbb{Z}/2^{r_{j_2}}}\Big)\vee A^6_\varepsilon(\tilde{\eta}_{r_{j_2}}),$$

where $A^6_\varepsilon(\tilde{\eta}_{r_{j_2}})$ *is the homotopy cofiber of* $\tilde{\eta}_{r_{j_2}}+\varepsilon\cdot i_3\eta^2$ *with* $\varepsilon\in\{0,1\}$.

(c) *If* $\Theta(H^3(C_\varphi;\mathbb{Z}/2))\neq 0$ *and* T_2 *has at least two direct summands, then*

$$\Sigma C_\varphi \simeq \Big(\bigvee_{i=1}^{d} S^5\Big)\vee P^6(T_2)\vee P^5\Big(\frac{T_2}{\mathbb{Z}/2^{r_{j_2}}\oplus\mathbb{Z}/2^{r_{j_0'}}}\Big)\vee A^7_\varepsilon(\tilde{\eta}_{r_{j_2}})\vee A^7(2^{r_{j_0'}}\eta^2),$$

where $A^7_\varepsilon(\tilde{\eta}_{r_{j_2}})=\Sigma A^6_\varepsilon(\tilde{\eta}_{r_{j_2}})$, *the index* j_2 *the minimum of the indices* $j\leq n$ *such that* $u+v\in\mathrm{im}(\beta_{r_j})$, *and* j_0' *is the maximum of the indices* $j\leq n$, $j\neq j_2$ *such that*

$$\Theta(x)\neq 0 \quad \textit{and} \quad \beta_{r_j}(x)\neq 0 \quad \textit{for all } x\in H^3(C_\varphi;\mathbb{Z}/2).$$

Proof By the Hilton–Milnor theorem and Lemmas 2.1 and 2.7, we can put

$$\varphi=\sum_{i=1}^{d}x_i\cdot\eta+\sum_{j=1}^{n}y_j\cdot i_4\eta+\sum_{k=1}^{n}z_k\cdot\tilde{\eta}_{r_k}+\sum_{l=1}^{n}w_l\cdot i_3\eta^2+\theta, \tag{5-2}$$

where θ is a linear combination of Whitehead products in $\pi_5(P^4(T_2))$.

By Lemmas 2.5 and 2.6, the condition $\mathrm{Sq}^2(H^4(\Sigma^2W;\mathbb{Z}/2))\neq 0$ enforces that at least one of these coefficients x_i, y_j, z_k is nonzero.

(1) Under the conditions in (1), we deduce from Lemma 2.8 that u comes from a free integral homology class. It follows that

$$y_j=z_k=0 \qquad\text{and}\qquad x_i=1 \quad\text{for some } i$$

in the expression (5-2). By Lemma 3.4(1), we may assume that there is exactly one index i such that $x_i=1$. Consider the homotopy type of the map

$$\begin{pmatrix}\eta\\ i_3\eta^2\end{pmatrix}: S^5\to S^4\vee P^4(2^{r_{j_0}}).$$

Clearly we have an equivalence

$$\begin{pmatrix}\eta\\ i_3\eta^2\end{pmatrix}\sim\begin{pmatrix}\eta\\ 0\end{pmatrix}.$$

After composing with self-homotopy equivalences of $\left(\bigvee_{i=1}^{d}S^4\right)\vee P^4(T_2)\vee P^5(T_2)$, we may assume that the above x_i is the unique nonzero coefficient, which completes the proof of the homotopy equivalence in (1) by Lemma 3.3.

(2) The arguments are similar to (1). The conditions (2) imply that

$$x_i=z_k=0 \qquad\text{and}\qquad y_j=1 \quad\text{for some } j,$$

while Lemma 3.4(2) guarantees that we may assume that there is exactly one such j, which is equal to j_1, as described in the proposition. By Lemma 2.1(4), we have

$$(i_3\overline{\eta}_{r_{j_1}})(i_4\eta)=i_3(\overline{\eta}_{r_{j_1}}i_4)\eta=i_3\eta^2,$$

hence

$$\begin{pmatrix}i_4\eta\\ i_3\eta^2\end{pmatrix}\sim\begin{pmatrix}i_4\eta\\ 0\end{pmatrix}.$$

Thus we may assume that all coefficients $w_l=0$ and the homotopy equivalence in (2) then follows.

(3) The conditions (3) imply that

$$z_k\equiv 1 \pmod 2 \quad\text{for some } k.$$

By Lemma 3.4, we may firstly assume that

$$x_1=\epsilon\in\mathbb{Z}/2,\qquad x_i=0\quad\text{for } i>1,$$
$$y_{j_0}=\varepsilon\in\mathbb{Z}/2,\qquad y_j=0\quad\text{for } j\neq j_0.$$

Note that $(\epsilon, \varepsilon) \neq (1,1)$, because

$$\begin{pmatrix}\eta\\ i_4\eta\end{pmatrix} \sim \begin{pmatrix}\eta\\ 0\end{pmatrix} : S^5 \to S^4 \vee P^5(2^r).$$

By the relation $q_4\tilde{\eta}_{r_k} = \eta$ in (2-3), we have

$$\begin{pmatrix}\eta\\ \tilde{\eta}_{r_k}\end{pmatrix} \sim \begin{pmatrix}0\\ \tilde{\eta}_{r_k}\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}i_4\eta\\ \tilde{\eta}_{r_k}\end{pmatrix} \sim \begin{pmatrix}0\\ \tilde{\eta}_{r_k}\end{pmatrix}.$$

It follows that $z_k \equiv 1 \pmod 2$ implies that $\epsilon = \varepsilon = 0$. By Corollary 2.2 we have

$$\begin{pmatrix}1 & 0\\ -B(\chi^r_s) & 1\end{pmatrix}\begin{pmatrix}\tilde{\eta}_r\\ \tilde{\eta}_s\end{pmatrix} \simeq \begin{pmatrix}\tilde{\eta}_r\\ 0\end{pmatrix} \quad \text{for } r \leq s.$$

Thus up to homotopy we may assume that $x_i = y_j = 0$ and there exists exactly one $z_{k_0} \equiv 1 \pmod 2$ with $k_0 = j_2$ as described in the proposition.

If $\Theta(H^3(C_\varphi;\mathbb{Z}/2)) = 0$, then Lemma 2.7 implies that $w_l = 0$ for all l. The first homotopy equivalence in (3) then follows by Lemma 3.3.

If $\Theta(H^3(C_\varphi;\mathbb{Z}/2)) \neq 0$, by Lemma 2.7 we have $w_l \neq 0$ for at least one $l \leq n$. It reduces to considering the homotopy type of the homotopy cofiber of the component

$$\varphi_2\colon S^5 \to P^4(T_2).$$

Note that when composing a self-homotopy equivalence of $P^4(T_2)$ to get z_{j_2}, it happens that $w_{j_2} = 0$ or $w_{j_2} = 1$.

If $n = 1$ and $T_2 \cong \mathbb{Z}/2^{r_{j_2}}$, the second homotopy equivalence in (3) then follows by Lemma 3.3. If $n \geq 2$, there are indices $l \neq j_2$, then similar arguments to that in the proof of Proposition 5.1 show that there is an equivalence

$$(i_3\eta^2, \dots, i_3\eta^2, \dots)^t \sim (0, \dots, 0, i_3\eta^2, 0, \dots, 0)^t,$$

where the unique $i_3\eta^2$ appears at the maximal j_0' among indices $l \leq n$, with $l \neq j_2$, such that

$$\Theta(x) \neq 0 \quad \text{and} \quad \beta_{r_{j_0'}}(x) \neq 0 \quad \text{for all } x \in H^3(C_\varphi;\mathbb{Z}/2).$$

Thus we get a homotopy equivalence $C_{\varphi_2} \simeq C_{\varphi_2'+\theta}$, where

$$\varphi_2' = (0, \dots, 0, \tilde{\eta}_{r_{j_2}} + \varepsilon \cdot i_3\eta^2, 0, \dots, 0, i_3\eta^2, 0, \dots, 0)^t.$$

After one suspension the possible Whitehead product θ becomes trivial. Thus we get

$$\Sigma C_{\varphi_2'} \simeq P^5\left(\frac{T_2}{\mathbb{Z}/2^{r_{j_2}} \oplus \mathbb{Z}/2^{j_0'}}\right) \vee A^7_\varepsilon(\tilde{\eta}_{r_{j_2}}) \vee A^7(2^{r_{j_0'}}\eta^2),$$

and therefore $\Sigma C_\varphi \simeq \left(\bigvee_{i=1}^d S^5\right) \vee P^6(T_2) \vee \Sigma C_{\varphi_2'}$. □

Proof of Theorem 1.1 It is well known that a closed, smooth, connected, orientable 4–manifold M is spin if and only if the Steenrod square Sq^2 acts trivially on $H^2(M;\mathbb{Z}/2)$. The homotopy types of $\Sigma^2 M$ in Theorem 1.1 then are obtained by (4-1), (5-1) and Propositions 5.1 and 5.2. □

Next, we give a proof of Theorem 1.2. By (1-3), there hold equivalence relations

$$\mathfrak{P}_0(H^1(M;\mathbb{Z}/2^r)) = 0 \iff \mathfrak{P}_1(H^2(\Sigma M;\mathbb{Z}/2^r)) = 0.$$

Lemma 5.3 *If the Pontryagin square $\mathfrak{P}_1$ acts trivially on $H^2(\Sigma M;\mathbb{Z}/2^r)$, then so does $\mathfrak{P}_1$ on $H^2(W_4;\mathbb{Z}/2^r)$.*

Proof By Lemma 3.1 and the universal coefficient theorem for cohomology, the canonical inclusion $i\colon W_4 \to \Sigma W$ induces isomorphisms

$$i^*\colon H^2(\Sigma W;\mathbb{Z}/2^r) \xrightarrow{\cong} H^2(W_4;\mathbb{Z}/2^r) \quad \text{and} \quad i^*\colon H^4(\Sigma W;\mathbb{Z}/2^{r+1}) \xrightarrow{\cong} H^2(W_4;\mathbb{Z}/2^{r+1}).$$

If $\mathfrak{P}_1$ acts trivially on $H^2(\Sigma M;\mathbb{Z}/2^r)$, then so does $\mathfrak{P}_1$ on $H^2(\Sigma W;\mathbb{Z}/2^r)$, by (4-1). The commutative diagram

$$\begin{array}{ccc} H^2(\Sigma W;\mathbb{Z}/2^r) & \xrightarrow{\mathfrak{P}_1=0} & H^4(\Sigma W;\mathbb{Z}/2^{r+1}) \\ \cong\big\downarrow i^* & & \cong\big\downarrow i^* \\ H^2(W_4;\mathbb{Z}/2^r) & \xrightarrow{\mathfrak{P}_1} & H^4(W_4;\mathbb{Z}/2^{r+1}) \end{array}$$

then implies $\mathfrak{P}_1 = 0$ on the second row. □

Lemma 5.4 *If the Pontryagin square $\mathfrak{P}_1$ acts trivially on $H^2(\Sigma M;\mathbb{Z}/2^{r_j})$ for each $j = 1, 2, \ldots, n$, then there is a homotopy equivalence*

$$W_4 \simeq \Big(\bigvee_{i=1}^{d} S^3\Big) \vee \Big(\bigvee_{i=1}^{m} S^4\Big) \vee P^3(T) \vee P^4(T).$$

Proof By Lemma 4.2 there is a homotopy equivalence

$$W_4 \simeq \Big(\bigvee_{i=1}^{d} S^3\Big) \vee P^4(T) \vee C_{g_2}$$

for some homologically trivial map $g_2\colon \bigvee_{i=1}^{m} S^3 \to P^3(T)$. It suffices to show the homologically trivial component

$$g_2\colon S^3 \to P^3(T)$$

is null-homotopic. By Lemma 3.3, it suffices to show that the components

$$g_2^j\colon S^3 \xrightarrow{g_2} P^3(T_2) \xrightarrow{p_j} P^3(2^{r_j})$$

are null-homotopic for each $j = 1, 2, \ldots, n$.

Since $\pi_3(P^3(2^r))\cong\mathbb{Z}/2^{r+1}$, for all $j=1,2,\dots,n$ we may set

$$g_2^j=t_j\cdot i_2\eta$$

for some $t_j\in\mathbb{Z}/2^{r_j+1}$. The assumption and Lemma 5.3 imply that the Pontryagin square

$$\mathfrak{P}_1\colon H^2(W_4;\mathbb{Z}/2^{r_j})\to H^4(W_4;\mathbb{Z}/2^{r_j+1})$$

is trivial. By the universal coefficient theorem for cohomology, g_2^j induces trivial homomorphism in mod 2^{r_j} or mod 2^{r_j+1} cohomology, and hence by Lemma 3.2(2), the Pontryagin square

$$\mathfrak{P}_1\colon H^2(C_{g_2^j};\mathbb{Z}/2^{r_j})\to H^4(C_{g_2^j};\mathbb{Z}/2^{r_j+1})$$

is trivial for each j. Then it follows by Lemma 2.4 that $t_j=0$, or equivalently g_2^j is null-homotopic for each $j=1,2,\dots,n$. □

Proof of Theorem 1.2 By Lemma 5.4 and (4-2), there is a homotopy cofibration

$$S^4\xrightarrow{k_5}W_4\simeq\Big(\bigvee_{i=1}^d S^3\Big)\vee\Big(\bigvee_{i=1}^m S^4\Big)\vee P^3(T)\vee P^4(T)\to\Sigma W,$$

with k_5 homologically trivial. Since $\pi_4(P^3(p^r))=\pi_4(P^4(p^r))=0$, Lemma 3.3 implies that there is a homotopy equivalence

$$\Sigma W\simeq\Big(\bigvee_{i=1}^m S^4\Big)\vee P^3(T_{\neq 2})\vee P^4(T_{\neq 2})\vee C_\phi,$$

where $\phi\colon S^4\to\big(\bigvee_{i=1}^d S^3\big)\vee P^3(T_2)\vee P^4(T_2)$ is a homologically trivial map. Compare (5-1). The discussion on the homotopy type of ΣW is totally parallel to that of $\Sigma^2 W$ in the proofs of Propositions 5.1 and 5.2. The proof is then completed by (4-1). □

References

[1] **H J Baues**, *Combinatorial homotopy and 4–dimensional complexes*, de Gruyter Expos. Math. 2, de Gruyter, Berlin (1991) MR Zbl

[2] **H J Baues**, *Homotopy type and homology*, Oxford Univ. Press (1996) MR Zbl

[3] **H J Baues**, **M Hennes**, *The homotopy classification of $(n-1)$–connected $(n+3)$–dimensional polyhedra,* $n\geq 4$, Topology 30 (1991) 373–408 MR Zbl

[4] **P Beben**, **S Theriault**, *The loop space homotopy type of simply-connected four-manifolds and their generalizations*, Adv. Math. 262 (2014) 213–238 MR Zbl

[5] **P Beben**, **S Theriault**, *Homotopy groups of highly connected Poincaré duality complexes*, Doc. Math. 27 (2022) 183–211 MR Zbl

[6] **P Beben**, **J Wu**, *The homotopy type of a Poincaré duality complex after looping*, Proc. Edinb. Math. Soc. 58 (2015) 581–616 MR Zbl

[7] **S-C Chang**, *Homotopy invariants and continuous mappings*, Proc. Roy. Soc. Lond. Ser. A 202 (1950) 253–263 MR Zbl

[8] **T Cutler**, **T So**, *The homotopy type of a once-suspended* 6*–manifold and its applications*, Topology Appl. 318 (2022) art. id. 108213 MR Zbl

[9] **J R Harper**, *Secondary cohomology operations*, Graduate Studies in Math. 49, Amer. Math. Soc., Providence, RI (2002) MR Zbl

[10] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[11] **R Huang**, *Homotopy of gauge groups over non-simply connected five-dimensional manifolds*, Sci. China Math. 64 (2021) 1061–1092 MR Zbl

[12] **R Huang**, *Homotopy of gauge groups over high-dimensional manifolds*, Proc. Roy. Soc. Edinburgh Sect. A 152 (2022) 182–208 MR Zbl

[13] **R Huang**, *Suspension homotopy of* 6*–manifolds*, Algebr. Geom. Topol. 23 (2023) 439–460 MR Zbl

[14] **R Huang**, **P Li**, *Suspension homotopy of simply connected* 7*–manifolds*, preprint (2022) arXiv 2208.13145

[15] **R Huang**, **S Theriault**, *Loop space decompositions of* $(2n-2)$*–connected* $(4n-1)$*–dimensional Poincaré duality complexes*, Res. Math. Sci. 9 (2022) art. id. 53 MR Zbl

[16] **P Li**, *Homotopy classification of maps between* $\boldsymbol{A}_n^2$*–complexes*, preprint (2020) arXiv 2008.03049

[17] **R E Mosher**, **M C Tangora**, *Cohomology operations and applications in homotopy theory*, Harper & Row, New York (1968) MR Zbl

[18] **J Neisendorfer**, *Algebraic methods in unstable homotopy theory*, New Math. Monogr. 12, Cambridge Univ. Press (2010) MR Zbl

[19] **T So**, **S Theriault**, *The suspension of a* 4*–manifold and its applications* (2019) arXiv 1909.11129 To appear in Israel J. Math.

[20] **E Thomas**, *The suspension of the generalized Pontrjagin cohomology operations*, Pacific J. Math. 9 (1959) 897–911 MR Zbl

[21] **J H C Whitehead**, *A certain exact sequence*, Ann. of Math. 52 (1950) 51–110 MR Zbl

[22] **J H C Whitehead**, *On the theory of obstructions*, Ann. of Math. 54 (1951) 68–84 MR Zbl

[23] **Z Zhu**, **P Li**, **J Pan**, *Periodic problem on homotopy groups of Chang complexes* $C_r^{n+2,r}$, Homology Homotopy Appl. 21 (2019) 363–375 MR Zbl

[24] **Z Zhu**, **J Pan**, *The decomposability of a smash product of* $\boldsymbol{A}_n^2$*–complexes*, Homology Homotopy Appl. 19 (2017) 293–318 MR Zbl

[25] **Z Zhu**, **J Pan**, *The local hyperbolicity of* $\boldsymbol{A}_n^2$*–complexes*, Homology Homotopy Appl. 23 (2021) 367–386 MR Zbl

Department of Mathematics, School of Sciences, Great Bay University
Dongguan, Guangdong, China

lipcaty@outlook.com

https://lipcaty.github.io

Received: 8 March 2023 Revised: 9 May 2023

Algebraic & Geometric Topology **24**:5 (2024) 2957–2970
DOI: 10.2140/agt.2024.24.2957
Published: 19 August 2024

The braid indices of the reverse parallel links of alternating knots

YUANAN DIAO
HUGH MORTON

The braid indices of most links remain unknown as there is no known universal method for determining the braid index of an arbitrary knot. This is also the case for alternating knots. We show that if K is an alternating knot, then the braid index of any reverse parallel link of K can be precisely determined. Specifically, if D is a reduced diagram of K, $v_+(D)$ (resp. $v_-(D)$) is the number of regions in the checkerboard shading of D for which all crossings are positive (resp. negative) and $w(D)$ is the writhe of D, then the braid index of a reverse parallel link of K with framing f, denoted by $\mathbb{K}_f$, is given by the precise formula

$$\boldsymbol{b}(\mathbb{K}_f) = \begin{cases} c(D)+2+a(D)-f & \text{if } f < a(D), \\ c(D)+2 & \text{if } a(D) \le f \le b(D), \\ c(D)+2-b(D)+f & \text{if } f > b(D), \end{cases}$$

where $a(D) = -v_-(D) + w(D)$ and $b(D) = v_+(D) + w(D)$.

57K10, 57K31

1 Introduction

The determination of the braid index of a knot or a link is known to be a challenging problem. To date there is no known method that can be used to determine the precise braid index of an arbitrarily given knot/link. This is also the case when we restrict ourselves to alternating knots and links, although the braid indices of many alternating knots and links can now be determined. For example, all 2–bridge links and all alternating Montesinos links; see Diao, Ernst, Hetyei and Liu [6] and Murasugi [13]. However, we prove a somewhat surprising result: the braid index of any *reverse parallel* link of an alternating knot can be precisely determined. Furthermore, the formula can be derived easily from any reduced diagram of the alternating knot.

Here we study the *reverse parallel* links of alternating knots. A reverse parallel link of a knot consists of the two boundary components of an annulus A embedded in S^3 with the said knot being one of the components and such that the two components are assigned opposite orientations. Let K and K' be the two components of a reverse parallel link induced by an annulus A. Following the convention that has been used in the literature (such as by Nutt [15] and Rudolph [16]), we shall call the linking number f between K and K' when they are assigned parallel orientations the *framing* of K. We note that a reverse

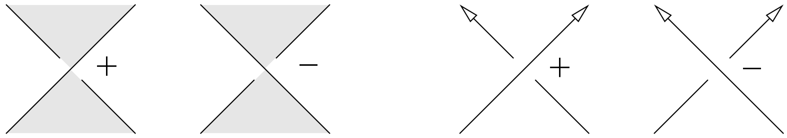


Figure 1: Left: the crossing with respect to a checkerboard shading. Right: the crossing sign with respect to the orientation of the knot.

parallel link of K with framing f is denoted by $K *_f A$ in [15] and by $\mathrm{Bd}_A(K, f)$ in [16]. The framing is independent of the orientation of K, and the ambient isotopy class of A in S^3 depends only on K and the framing. Therefore, the reverse parallel links of K are characterized by the framing f. Since our results (and proofs) only depend on the framing, not the actual annulus A, we shall introduce a new notation $\mathbb{K}_f$ for the reverse parallel link of K with framing f. Keep in mind that the framing f is the linking number of the two components of $\mathbb{K}_f$ with parallel orientations, and hence the linking number of $\mathbb{K}_f$ itself is $-f$.

For a given knot diagram D with a checkerboard shading, a crossing can be assigned a $+$ or a $-$ sign relative to this shading, as shown on the left side of Figure 1. This is not to be confused with the crossing sign with respect to the orientation of the knot which is used in the definition of the writhe of D, as shown on the right side of Figure 1.

Now let K be an alternating knot with a reduced diagram D. It is known that in such a case crossings of D are all positive with respect to one checkerboard shading of D and are all negative with respect to the other checkerboard shading of D. Furthermore, if we let $v_+(D)$ be the number of shaded regions in the shading with respect to which all crossings are positive, and $v_-(D)$ be the number of shaded regions in the complementary shading with respect to which all crossings are negative, then $v_+(D)+v_-(D)-2=c(D)$ where $c(D)$ is the number of crossings in D; see Kauffman [9]. From D we can also obtain its so-called blackboard reverse parallel annulus (resp. framing), which provides a good reference for other choices of annuli (resp. framings) as the other choices come from this one by adding either right-handed or left-handed twists. If the writhe of D is $w(D)$, then the framing of the blackboard reverse parallel is also $w(D)$. If k right-handed (resp. left-handed) twists are added between the two components, then the resulting reverse parallel has framing $w(D)+k$ (resp. $w(D)-k$). See Figure 2 for an illustration.

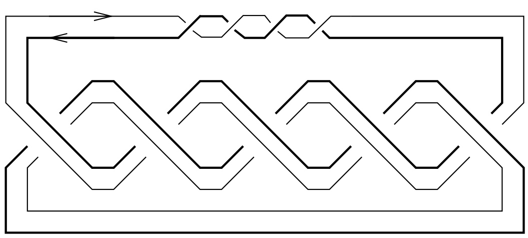


Figure 2: The blackboard reverse parallel of the $(2,5)$ torus knot with two left-handed twists added. The framing of the resulting reverse parallel link (with the added twists) is thus $5+(-2)=3$.

Our main result is the following theorem:

Theorem 1.1 *Let K be an alternating knot and D a reduced diagram of K. Let $c(D)$, $w(D)$, $v_+(D)$ and $v_-(D)$ be as defined above. Then the braid index of $\mathbb{K}_f$, denoted by $\boldsymbol{b}(\mathbb{K}_f)$, is given by the formula*

$$\boldsymbol{b}(\mathbb{K}_f) = \begin{cases} c(D)+2+a(D)-f & \text{if } f < a(D), \\ c(D)+2 & \text{if } a(D) \le f \le b(D), \\ c(D)+2-b(D)+f & \text{if } f > b(D), \end{cases} \tag{1-1}$$

where $a(D) = -v_-(D) + w(D)$ and $b(D) = v_+(D) + w(D)$.

We can summarize Theorem 1.1 pictorially in terms of the blackboard reverse parallel of D:

- The blackboard reverse parallel has braid index $c(D) + 2$.
- The braid index remains $c(D) + 2$ after adding up to $v_+(D)$ right-handed twists, or up to $v_-(D)$ left-handed twists.
- Each further right or left-handed twist increases the braid index by 1.

So for example, since $v_-(D) = 2$ and $v_+(D) = 5$ for the $(2, 5)$ torus knot, the braid index for the reverse parallel shown in Figure 2 is $c(D) + 2 = 7$. Adding one further left-handed twist would increase the braid index to 8, while we would still have braid index 7 after adding up to 5 right-handed twists to the blackboard parallel.

We shall establish (1-1) by proving that the right side expression is both a lower bound and an upper bound for the $\boldsymbol{b}(\mathbb{K}_f)$. The lower bound is obtained by the Morton–Franks–Williams inequality, while the upper bound is established by direct construction.

2 The lower bound

In this section, we shall prove the following theorem:

Theorem 2.1 *Let $\mathbb{K}_f$ be the reverse parallel link of an alternating knot K with framing f and D a reduced diagram of K. Then*

$$\boldsymbol{b}(\mathbb{K}_f) \ge \begin{cases} c(D)+2+a(D)-f & \text{if } f < a(D), \\ c(D)+2 & \text{if } a(D) \le f \le b(D), \\ c(D)+2-b(D)+f & \text{if } f > b(D), \end{cases} \tag{2-1}$$

where $a(D) = -v_-(D) + w(D)$ and $b(D) = v_+(D) + w(D)$.

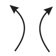
2.1 The Homfly and Kauffman polynomials

Before proving this theorem we note some properties of the Homfly and Kauffman polynomials of a link L.

The *Homfly polynomial* $P_L(v, z) \in \mathbb{Z}[v^{\pm 1}, z^{\pm 1}]$ of an oriented link L is determined by the skein relations

$$v^{-1} P_{L^+} - v P_{L^-} = z P_{L^0},$$

where $L^{\pm}$ and L^0 differ only near one crossing as shown below, and takes the value 1 on the unknot:

$L^+ =$  , $L^- =$  , $L^0 =$  .

The *Kauffman polynomial* $F_L(a,z) \in \mathbb{Z}[a^{\pm 1}, z^{\pm 1}]$ for an unoriented link L is defined in [10]. Again it takes the value 1 on the unknot.

When an extra distant unknotted component O is adjoined to the link L to make $L \sqcup O$, each polynomial changes in the following simple way:

$$P_{L\sqcup O}(v,z) = \frac{v^{-1}-v}{z} P_L(v,z), \quad F_{L\sqcup O}(a,z) = \Big(\frac{a+a^{-1}}{z} - 1\Big) F_L(a,z).$$

Define the *extended Homfly polynomial* EP by

$$\mathrm{EP}_L(v,z) = \frac{v^{-1}-v}{z} P_L(v,z) = P_{L\sqcup O}(v,z) \tag{2-2}$$

and the *extended Kauffman polynomial* EF by

$$\mathrm{EF}_L(a,z) = \Big(\frac{a+a^{-1}}{z} - 1\Big) F_L(a,z) = F_{L\sqcup O}(a,z). \tag{2-3}$$

Remark This extended normalization is often used in the context of quantum invariants, where it allows for more natural specializations of the knot polynomials. It is also more useful in that context to use the Dubrovnik variant of the Kauffman polynomial in place of F.

By plugging in $L = \phi$ on both sides of (2-2) and (2-3), the extended polynomials can be thought of as taking the value 1 on the empty link ϕ.

2.2 Bounds from the Homfly and Kauffman polynomials

The Morton–Franks–Williams inequality [7; 11] gives a lower bound for the braid index $\boldsymbol{b}(L)$ of the link L in terms of the v–spread of the Homfly polynomial $P_L(v,z)$ or its extended version. Explicitly

$$\boldsymbol{b}(L) \ge 1 + \tfrac{1}{2}\operatorname{spr}_v P_L(v,z) = \tfrac{1}{2}\operatorname{spr}_v \mathrm{EP}_L(v,z). \tag{2-4}$$

The a–spread of the Kauffman polynomial is shown by Morton and Beltrami [12] to give a bound for the arc index $\alpha(L)$. Explicitly this is

$$\operatorname{spr}_a F_L(a,z) \le \alpha(L) - 2.$$

Bae and Park [1] showed that the arc index $\alpha(L)$ is bounded above by $c(L)+2$, that is, $\alpha(L) \le c(L)+2$. Combining these results shows that

$$\operatorname{spr}_a F_L(a,z) \le c(L). \tag{2-5}$$

2.3 A congruence result

Rudolph [16] relates the Kauffman polynomial of a link L with the Homfly polynomial of the reverse parallels of L.

Notation For Laurent polynomials $A = \sum a_{i,j} v^i z^j$ and $B = \sum b_{i,j} v^i z^j \in \mathbb{Z}[v^{\pm 1}, z^{\pm 1}]$ we write $A \cong_{\mathbb{Z}_2} B$ when $a_{i,j} \cong b_{i,j} \bmod 2$ for all i and j.

In the case of a knot K, Rudolph's theorem for the reverse parallel $\mathbb{K}_f$ can then be stated very cleanly in terms of the extended polynomials.

Theorem 2.2 [16, congruence theorem] $\mathrm{EP}_{\mathbb{K}_f}(v,z)-1 \cong_{\mathbb{Z}_2} v^{-2f}\,\mathrm{EF}_K(v^{-2},z^2)$.

2.4 Alternating knots

We can apply these bounds to the case of alternating knots, starting from observations of Cromwell [3] about their Kauffman polynomial.

For any knot K with a diagram D, write the Kauffman polynomial $F_K(a,z)$ of K as

$$F_K(a,z)=a^{-w(D)}\sum_{i,j} a_{i,j}a^i z^j. \tag{2-6}$$

In this form the coefficients $a_{i,j}$ are only nonzero in the range $|i|+j \le c(D)$.

Cromwell extends work of Thistlethwaite [17] to identify two nonzero coefficients $a_{i,j}$ which realize the maximum possible a–spread $c(D)$ for $F_K(a,z)$ in the case of an alternating knot K with reduced diagram D.

Theorem 2.3 [3] *Let K be an alternating knot and D a reduced diagram of K. Then $a_{i,j}=1$ in the two cases $i=1-v_+(D)$, $j=c(D)+i$ and $i=v_-(D)-1$, $j=c(D)-i$.*

Corollary 2.4 *We have $\mathrm{spr}_a\, F_K(a,z)=c(D)$, and $a_{i,j}=0$ in (2-6) unless $1-v_+(D)\le i\le v_-(D)-1$.*

Proof By Theorem 2.3 $\mathrm{spr}_a\, F_K(a,z)\ge v_-(D)-1-(1-v_+(D))=c(D)$, while $\mathrm{spr}_a\, F_K(a,z)\le c(D)$ by (2-5). □

Now set

$$B_D(a,z)=a^{w(D)}\,\mathrm{EF}_K(a,z)=\Big(\frac{a+a^{-1}}{z}-1\Big)\sum_{i,j} a_{i,j}a^i z^j. \tag{2-7}$$

Then $\mathrm{spr}_a\, B_D(a,z)=\mathrm{spr}_a\, F_K(a,z)+2=c(D)+2$. Furthermore, if we write

$$B_D(a,z)=\sum_{i,j} b_{i,j}a^i z^j, \tag{2-8}$$

then $b_{i,j}=0$ unless $-v_+(D)\le i\le v_-(D)$ by Corollary 2.4.

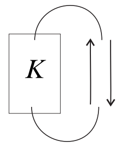
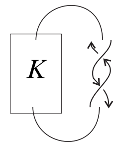
The two critical monomials $a^{-v_+(D)}z^{c(D)-v_+(D)}$ and $a^{v_-(D)}z^{c(D)-v_-(D)}$ in $B_D(a,z)$, which correspond to $i=-v_+(D)$ and $i=v_-(D)$, respectively, both have coefficient $b_{i,j}=1$, by Theorem 2.3. We will use these critical monomials in finding a lower bound for the v–spread of the extended Homfly polynomial of the reverse parallels of D.

Theorem 2.5 gives a simple formula to calculate the extended Homfly polynomial of $\mathbb{K}_{k+f}$ in terms of the polynomial of $\mathbb{K}_k$.

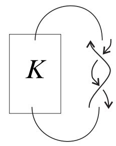
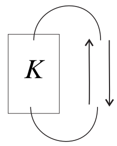
Theorem 2.5 *For any f and k we have*

$$v^{2f}(\mathrm{EP}_{\mathbb{K}_{k+f}}(v,z)-1)=\mathrm{EP}_{\mathbb{K}_k}(v,z)-1.$$

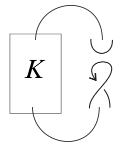
Proof While this is in effect shown by Rudolph [16, Proposition 2(5)] it is easy to give a direct skein theory proof. It is enough to prove it in the case $f = 1$. Now $\mathbb{K}_{k+1}$ is given from $\mathbb{K}_k$ by adding one extra twist in the annulus, as shown:

$\mathbb{K}_k =$  , $\mathbb{K}_{k+1} =$  .

With the reverse parallel orientation on the strings, apply the Homfly skein relation at one of the crossings in the diagram for $\mathbb{K}_{k+1}$. Since this is a negative crossing, $\mathbb{K}_{k+1}$ plays the role of L^-. Switching the crossing gives

$L^+ =$  $=$  $= \mathbb{K}_k$,

while the smoothed diagram

$L^0 =$ 

is simply an unknotted curve.

The skein relation, in the form

$$\mathrm{EP}_{L^+} = vz\,\mathrm{EP}_{L^0} + v^2\,\mathrm{EP}_{L^-},$$

then gives

$$\mathrm{EP}_{\mathbb{K}_k} = vz\frac{v^{-1}-v}{z} + v^2\,\mathrm{EP}_{\mathbb{K}_{k+1}} = 1 - v^2 + v^2\,\mathrm{EP}_{\mathbb{K}_{k+1}}.$$

Thus

$$v^2(\mathrm{EP}_{\mathbb{K}_{k+1}} - 1) = \mathrm{EP}_{\mathbb{K}_k} - 1. \qquad \square$$

We can now specify a lower bound for the v–spread of the extended Homfly polynomial of the parallels $\mathbb{K}_{w(D)+f}$ as f varies.

Theorem 2.6 *Let K be an alternating knot with reduced diagram D. The framed reverse parallel $\mathbb{K}_{w(D)+f}$ has the following lower bound for the v–spread of its extended Homfly polynomial*:

$$\mathrm{spr}_v\,\mathrm{EP}_{\mathbb{K}_{w(D)+f}}(v,z) \geq \begin{cases} 2(v_+(D) - f) & \text{if } f < -v_-(D), \\ 2(v_+(D) + v_-(D)) & \text{if } -v_-(D) \leq f \leq v_+(D), \\ 2(f + v_-(D)) & \text{if } f > v_+(D). \end{cases}$$

Proof Since K is an alternating knot with reduced diagram D, Theorem 2.2 shows that

$$B_D(v^{-2}, z^2) = v^{-2w(D)}\,\mathrm{EF}_K(v^{-2}, z^2) \cong_{\mathbb{Z}_2} \mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z) - 1. \tag{2-9}$$

In $B_D(v^{-2}, z^2) = \sum b_{i,j} v^{-2i} z^{2j}$ there are two critical monomials $v^{-2i}z^{2j}$, one with $i = -v_+(D)$ and $j = c(D) - v_+(D)$, and the other with $i = v_-(D)$ and $j = c(D) - v_-(D)$, where $b_{i,j} = 1$. By (2-9) there are two corresponding critical monomials $v^{-2i}z^{2j}$ in $\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z) - 1$ whose coefficients are congruent to $b_{i,j}$, and hence are odd. One term has v–degree $-2v_-(D)$ and the other has v–degree $2v_+(D)$.

By Theorem 2.5 we have

$$v^{2f}\,\mathrm{EP}_{\mathbb{K}_{w(D)+f}}(v,z) = (\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z)-1)+v^{2f}.$$

The v–spread of $\mathrm{EP}_{\mathbb{K}_{w(D)+f}}(v,z)$ is the same as the v–spread of $(\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z)-1)+v^{2f}$. In this Laurent polynomial consider the appearance of the two critical monomials along with the monomial v^{2f}. Unless one of the two critical monomials $v^{2v_+(D)}z^{2c(D)-2v_+(D)}$ and $v^{-2v_-(D)}z^{2c(D)-2v_-(D)}$ in $B_D(v^{-2},z^2)$ is v^{2f} they will each still have odd coefficients, and the v–spread will be at least $2(v_+(D)+v_-(D))$.

If $f<-v_-(D)$ or $f>v_+(D)$ the monomial v^{2f} has even coefficient in $\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z)-1$ since it has coefficient 0 in $B_D(v^{-2},z^2)$. In this range of f it then has nonzero coefficient in $(\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z)-1)+v^{2f}$. This gives the lower bound $2(v_+(D)-f)$ when $f<-v_-(D)$, and $2(v_-(D)+f)$ when $f>v^+(D)$ for $\mathrm{spr}_v\,\mathrm{EP}_{\mathbb{K}_{w(D)+f}}(v,z)$.

To complete the proof of Theorem 2.6 it remains to deal with the cases where v^{2f} is one of the two critical monomials $v^{2v_+(D)}z^{2c(D)-2v_+(D)}$ and $v^{-2v_-(D)}z^{2c(D)-2v_-(D)}$ in $B_D(v^{-2},z^2)$. In the first case this means that $f=v_+(D)$ and $0=c(D)-v_+(D)$. Then $f=c(D)=v_+(D)=n$ and D is the reduced diagram of the $(2,n)$ torus knot. In the other case $-f=c(D)=v_-(D)=n$. Hence D is the reduced diagram of the $(2,-n)$ torus knot.

In the $(2,n)$ case we need to show that the coefficient of v^{2n} in $(\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z)-1)+v^{2n}$ is nonzero. In Theorem 2.7 we show that this coefficient is 2 by showing that v^{2n} has coefficient 1 in $\mathrm{EP}_{\mathbb{K}_{w(D)}}(v,z)$, where $\mathbb{K}_{w(D)}$ is the blackboard reverse parallel of D.

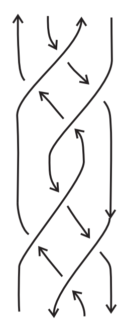
The $(2,-n)$ case follows directly by considering the polynomial of the mirror image. □

The detailed calculation for the special case of the $(2,n)$ torus knot will now be shown.

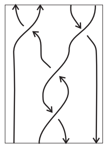
Theorem 2.7 *The blackboard reverse parallel $\mathbb{K}_n$ of the $(2,n)$ torus knot K satisfies*

$$\mathrm{EP}_{\mathbb{K}_n}(v,z) = v^{2n} + \sum_{i<2n,j} a_{i,j}v^iz^j.$$

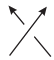
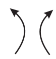
Proof We can draw a diagram of $\mathbb{K}_n$ as the closure of a 4–strand tangle with two upward and two downward strings, as shown:



It is more convenient to place the upward pair of strings on the left at the top and bottom, and write $\mathbb{K}_n$ as the closure of the tangle T^n, where

$T =$  .

We use the skein relations in the form

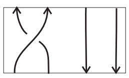
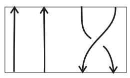
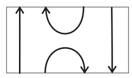
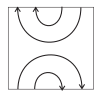
v^{-1}  $- v$  $= z$ 

to write the closure of T^n as a linear combination of the closures of simpler tangles.

Notation The 4–strand tangle U *evaluates to* the extended Homfly polynomial of its closure, which we write as $\mathrm{ev}(U) \in \mathbb{Z}[v^{\pm 1}, z^{\pm 1}]$.

Remark Evaluation is linear on tangles and respects the skein relations. It is a sort of trace function in that $\mathrm{ev}(AB) = \mathrm{ev}(BA)$.

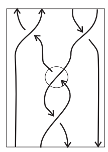
Our first step is to expand T as a combination of the tangles

$\sigma_1 =$  , $\sigma_3 =$  , $h =$  and $H =$  ,

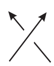
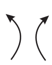
and their products when placed one above the other.

Remark By using the skein relations we are in effect working in a version of the mixed Hecke algebra $H_{2,2}(v, z)$ spanned by tangles with two upward and two downward strings [8].

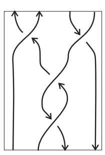
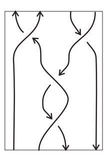
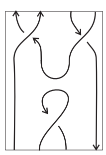
The crossing circled here in

$T =$ 

is a negative crossing, so we can use the skein relation at this crossing in the form

 $= v^{-2}$  $- v^{-1} z$  .

Then

$T =$  $= v^{-2}$  $- v^{-1} z$  $= v^{-2}\sigma_1\sigma_3 - v^{-1} z \sigma_1\sigma_3 h = C + C\tau,$

where for convenience we set $C = c_1 c_3 = (v^{-1}\sigma_1)(v^{-1}\sigma_3)$ and $\tau = (-zv)h$.

Then $T^n = (C + C\tau)^n$. Now C and τ do not commute, so we write

$$T^n = C^n + (C\tau)^n + \sum_{0<k<n} C^{r_1}\tau C^{r_2}\tau \cdots C^{r_k}\tau C^r, \tag{2-10}$$

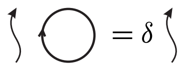
where $r_i \geq 1$, $r \geq 0$ and $r + \sum r_i = n$.

We can estimate the contribution of these terms to the evaluation of T^n.

- The evaluation of C^n only contributes terms up to v–degree 4, by Proposition 2.9.
- The terms in the large sum with weight k in τ evaluate to terms of v–degree at most $2k$. Without changing the evaluation we can assume that $r = 0$, since we can cycle C^r from the end to the beginning of the product and amalgamate it with C^{r_1}. The contribution of these terms with $k < n$ to the evaluation of T^n is shown in Proposition 2.10 to have degree no more than $2k$ (and thus at most $2n-2$) in v.
- The most important contribution comes from the evaluation of $(C\tau)^n$, which gives v^{2n}, and no other terms with v–degree $2n$ or larger, as stated in Proposition 2.8.

Before making detailed calculations we note some useful properties, which can be quickly checked diagrammatically:

$$\sigma_1 H = \sigma_3 H, \qquad H\sigma_1 = H\sigma_3, \qquad H = h\sigma_1\sigma_3^{-1}h,$$

 where $\delta = \dfrac{v^{-1}-v}{z}$, $\qquad h^2 = \delta h, \qquad h\sigma_1 h = h.$

Here are some consequences for our use of $c_1 = v^{-1}\sigma_1$, $c_3 = v^{-1}\sigma_3$, $C = c_1c_3$ and $\tau = (-zv)h$, which follow algebraically:

- $c_1 = c_1^{-1} + zI$ and $c_3 = c_3^{-1} + zI$ (the skein relation, where I stands for the identity tangle).
- $\tau c_1 c_3^{-1}\tau = (-zv)^2 h c_1 c_3^{-1} h = (zv)^2 H$.
- $\tau^2 = (-zv)\delta\tau = (v^2-1)\tau$.
- $\tau c_1 \tau = v^{-1}(-zv)^2 h\sigma_1 h = -z\tau$.
- $\tau C \tau = \tau(c_1c_3^{-1} + zc_1)\tau = (zv)^2 H - z^2\tau$.

Proposition 2.8 *The extended polynomial of the closure of $(C\tau)^n$ is v^{2n} plus lower terms in v for $n > 1$, and $1 - v^{-2}$ when $n = 1$.*

Proof When $n = 1$ we have $C\tau = (-zv^{-1})\sigma_1\sigma_3 h$. Now $\sigma_1\sigma_3 h$ closes to a single unknotted curve, so $C\tau$ evaluates to $-zv^{-1}\delta = 1 - v^{-2}$.

For $n > 1$ write

$$(C\tau)^n = C(\tau C\tau)(C\tau)^{n-2} = (zv)^2 CH(C\tau)^{n-2} - z^2 C\tau(C\tau)^{n-2}.$$

The evaluation of the second term has v–degree at most $2n-2$, by induction on n, so any monomials of larger v–degree must come from the first term.

Now $Hh = \delta H$ and $H\sigma_1 h = H$. We can then write

$$HC\tau = H(c_1c_3^{-1} + zc_1)\tau = H\tau + zHc_1\tau = (-zv)(\delta + zv^{-1})H = (v^2 - 1 - z^2)H.$$

So the first term expands to

$$(zv)^2 CH(C\tau)^{n-2} = (zv)^2(v^2 - 1 - z^2)^{n-2} CH.$$

Now $CH = c_1c_3^{-1}H + zc_1H = H + zv^{-1}\sigma_1 H$. The closure of H is two disjoint unknotted curves, and $\sigma_1 H$ closes to one unknotted curve. These evaluate to δ^2 and δ, respectively. The first term then evaluates to

$$(v^2-1-z^2)^{n-2}(\delta^2(-zv)^2 - z^2(-zv\delta)) = (v^2-1-z^2)^{n-1}(v^2-1).$$

This contributes a single term v^{2n} and no further terms of v–degree larger than $2n-2$. □

The skein relation, in the form $c_1^2 = I + zc_1$, allows us to write c_1^r recursively as a linear combination of c_1 and the identity tangle I,

$$c_1^r = a_r(z)I + b_r(z)c_1,$$

with coefficients which are polynomials in z only. Similarly

$$c_3^r = a_r(z)I + b_r(z)c_3.$$

We can then expand C^r as a linear combination of I, c_1, c_3 and c_1c_3, with coefficients in $\mathbb{Z}[z]$. Explicitly

$$C^r = (a_r I + b_r c_1)(a_r I + b_r c_3).$$

Proposition 2.9 *The term C^n in the expansion of T^n provides terms of degree at most 4 in v, in the evaluation.*

Proof We have

$$C^n = a_n^2 I + a_n b_n(c_1 + c_3) + b_n^2 c_1 c_3 = a_n^2 I + a_n b_n v^{-1}(\sigma_1 + \sigma_3) + b_n^2 v^{-2}\sigma_1\sigma_3.$$

Now I closes to four unknotted curves evaluating to δ^4, σ_1 and σ_3 close to three unknotted curves evaluating to δ^3, and $\sigma_1\sigma_3$ closes to two unknotted curves evaluating to δ^2. The term C^n then contributes $a_n^2\delta^4 + 2a_nb_nv^{-1}\delta^3 + b_n^2v^{-2}\delta^2$ to the evaluation. Since $\delta = (v^{-1}-v)z^{-1}$, and a_n and b_n depend only on z, these terms have v–degree at most 4. □

To complete our proof of Theorem 2.7 we show that the evaluation of the remaining terms in (2-10) has v–degree at most $2n-2$:

Proposition 2.10 *The evaluation of*

$$C^{r_1}\tau \cdots C^{r_i}\tau \cdots C^{r_k}\tau$$

with $r_i \geq 1$ has terms of degree at most $2k$ in v.

Proof We proceed by induction on the number of exponents r_i for which $r_i > 1$.

When $r_i = 1$ for all i this follows from Proposition 2.8.

Otherwise we can cycle the terms in the product without changing its evaluation, and arrange that $r_k = r > 1$. Then

$$\tau C^r \tau = a_r^2\tau^2 + a_r b_r \tau(c_1+c_3)\tau + b_r^2\tau C\tau = a_r^2(v^2-1)\tau - 2za_rb_r\tau + b_r^2\tau C\tau.$$

So

$$C^{r_1}\tau\cdots C^{r_k}\tau = (a_r^2(v^2-1) - 2za_rb_r)C^{r_1}\tau\cdots C^{r_{k-1}}\tau + b_r^2 C^{r_1}\tau\cdots C^{r_{k-1}}\tau C\tau.$$

These expressions both have one fewer term C^{r_i} for which $r_i > 1$, so by our induction hypothesis the evaluation of $C^{r_1}\tau\cdots C^{r_{k-1}}\tau C\tau$ has terms of degree at most $2k$ in v while $C^{r_1}\tau\cdots C^{r_{k-1}}\tau$ has terms of degree at most $2k-2$. With the coefficient $a_r^2(v^2-1)-2za_rb_r$ adding 2, in this case all terms in the final evaluation have degree at most $2k$ in v. This establishes the proposition. □

Now all the terms in (2-10) have been dealt with, and Theorem 2.7 for the evaluation of the reverse blackboard parallel of the $(2,n)$ torus knot follows. □

The proof of Theorem 2.6 is then complete. We can now prove Theorem 2.1, which was the goal of this section.

Proof of Theorem 2.1 Using the Morton–Franks–Williams bound (2-4) in Theorem 2.6 immediately gives the lower bound for the braid index of $\mathbb{K}_{w(D)+f}$ as

$$\boldsymbol{b}(\mathbb{K}_{w(D)+f}) \geq \begin{cases} v_+(D)-f & \text{if } f < -v_-(D), \\ v_+(D)+v_-(D) & \text{if } -v_-(D) \leq f \leq v_+(D), \\ f+v_-(D) & \text{if } f > v_+(D). \end{cases}$$

Replacing f by $f-w(D)$ then gives

$$\boldsymbol{b}(\mathbb{K}_{f}) \geq \begin{cases} v_+(D)-f+w(D) & \text{if } f-w(D) < -v_-(D), \\ v_+(D)+v_-(D) & \text{if } -v_-(D) \leq f-w(D) \leq v_+(D), \\ f-w(D)+v_-(D) & \text{if } f-w(D) > v_+(D). \end{cases}$$

Now $v_+(D)+v_-(D)=c(D)+2$, so after setting $a(D)=w(D)-v_-(D)$ and $b(D)=w(D)+v_+(D)$ this lower bound becomes

$$\boldsymbol{b}(\mathbb{K}_{f}) \geq \begin{cases} c(D)+2+a(D)-f & \text{if } f < a(D), \\ c(D)+2 & \text{if } a(D) \leq f \leq b(D), \\ c(D)+2-b(D)+f & \text{if } f > b(D), \end{cases}$$

which is the formula (2-1) claimed in Theorem 2.1 □

3 The upper bound

In this section, we shall prove the following theorem, which provides us the desired upper bound for the braid index of $\mathbb{K}_f$.

Theorem 3.1 *If $\mathbb{K}_f$ is a reverse parallel link of an alternating knot K with framing f and D is a reduced diagram of K, then*

$$\boldsymbol{b}(\mathbb{K}_{f}) \leq \begin{cases} c(D)+2+a(D)-f & \text{if } f < a(D), \\ c(D)+2 & \text{if } a(D) \leq f \leq b(D), \\ c(D)+2-b(D)+f & \text{if } f > b(D), \end{cases} \tag{3-1}$$

where $a(D)=-v_-(D)+w(D)$ and $b(D)=v_+(D)+w(D)$.

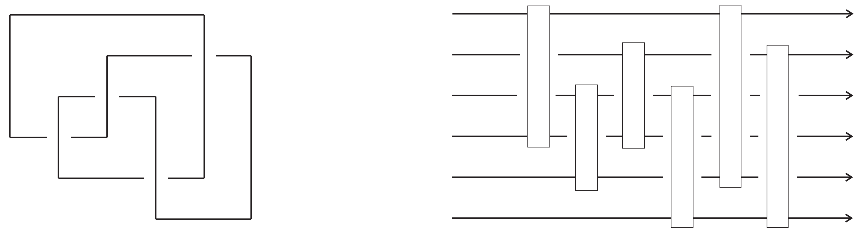


Figure 3: Left: a grid diagram for the figure-eight knot with $\alpha = 6$ arcs. Right: the resulting braid template.

Proof It suffices to show that braid presentations of $\mathbb{K}_f$ can be constructed with the number of strings given in the theorem.

Nutt [15] constructs reverse string satellites of a knot K using an arc presentation for K with α arcs. In its simplest form, Nutt's construction gives an α–string braid presentation of the reverse parallels $\mathbb{K}_f$ for $\alpha + 1$ consecutive values of f.

An arc presentation of K on α arcs provides a grid diagram for K made up of α vertical arcs joined by α horizontal arcs such that the horizontal arcs can only pass under the vertical arcs. Convert this to a braid template by the following procedure. First we extend each horizontal arc (from the points where it is connected to vertical arcs) left and right. If an extended arc runs into a vertical arc, then we make it pass under the vertical arc. Notice that each vertical arc now ends in a $\top$ at its top and in a $\perp$ at its bottom. We then "thicken" each vertical arc slightly to create an empty 2–braid box as shown in Figure 3. In this construction, the horizontal arcs at a $\top$ or $\perp$ of a vertical arc will meet the vertical sides of the corresponding braid box, while any intermediate horizontal arcs pass entirely underneath the braid box.

We can now place a single positive or negative crossing in each braid box, with strings running from left to right, connecting the horizontal arcs that meet the boundary of the box. The other horizontal arcs either do not pass through this crossing, or will pass under it. This gives an α–string closed braid which is a link with two components. Furthermore, one can verify that this closed braid is ambient isotopic to a link within a tubular neighborhood of the grid diagram such that each component runs parallel to K but with opposite orientations. That is, the resulting closed braid presents $\mathbb{K}_f$ for some framing f.

Write $f = a$ for the framing which arises when all the crossings used in the braid boxes are positive. If k of the boxes are filled with negative crossings instead, then we have framing $f = a + k$, so with all crossings negative we have $f = a + \alpha = b$, say.

In the case that $f < a$ or $f > b$, we can present $\mathbb{K}_f$ as a closed braid on $\alpha + \tau$ strings for $f = a - \tau$ or $f = b + \tau$ ($\tau \geq 1$) by adding τ extra suitably chosen arcs to the grid diagram. This gives us an upper bound for $\boldsymbol{b}(\mathbb{K}_f)$,

$$\boldsymbol{b}(\mathbb{K}_f) \leq \begin{cases} \alpha + a - f & \text{if } f < a, \\ \alpha & \text{if } a \leq f \leq b, \\ \alpha + f - b & \text{if } f > b, \end{cases} \tag{3-2}$$

with $b = a + \alpha$.

An alternating knot K has an arc representation with $c(D)+2$ arcs [1], so we can set $\alpha = c(D)+2$ in (3-2). It remains to show that $a = a(D)$ and $b = b(D)$ in this case.

By our lower bound calculation (2-1), $\mathbb{K}_f$ cannot be presented as a braid on $\alpha = c(D)+2$ strings if $f < a(D)$ or $f > b(D)$. Now $\mathbb{K}_a$ and $\mathbb{K}_b$ are each presented by an α–string braid, so $a \geq a(D)$ and $b \leq b(D)$. Thus $b = c(D)+2+a \geq c(D)+2+a(D) = b(D)$, giving

$$b = b(D), a = a(D)$$

in (3-2). □

We end our paper with the following remarks:

Remark If one desires to use the linking number l of $\mathbb{K}_f$ in the formulation of $\boldsymbol{b}(\mathbb{K}_f)$ instead of the framing f, then the formulation can be easily obtained by substituting f by $-l$ in (1-1). Specifically, (1-1) becomes

$$\boldsymbol{b}(\mathbb{K}_f) = \begin{cases} c(D)+2+a'(D)-l & \text{if } l < a'(D), \\ c(D)+2 & \text{if } a'(D) \leq l \leq b'(D), \\ c(D)+2-b'(D)+l & \text{if } l > b'(D), \end{cases} \tag{3-3}$$

where $a'(D) = -v_+(D) - w(D)$, $b'(D) = v_-(D) - w(D)$ and $l = -f$ is the linking number of $\mathbb{K}_f$. The corresponding formulation of (2-1) matches the one given in [5]. We need to point out that the lower bound formula derived in [5] uses a graph-theoretic approach on the Seifert graphs of D and $\mathbb{K}_f$ constructed from the blackboard reverse parallel of D. However, that approach only works for the special alternating knots, namely those alternating knots which admit a reduced alternating diagram in which the crossings are either all positive or all negative.

Remark The general question of finding the braid index for a satellite of a knot K with some form of reverse string pattern has been considered by Birman and Menasco [2]. Our reverse parallels, along with Whitehead doubles, are the simplest such satellites. Nutt [15] draws on [2] to give lower bounds for the braid index in terms of the arc index of K, as well as the upper bounds which we have used. Coupled with the later work of Bae and Park [1], this would provide our result without the use of Rudolph's congruence.

Some descriptions given by [2] were later found to be incomplete, with Ka Yi Ng [14] providing details of the missing cases. Nutt's lower bound argument needs the analysis in [2] which shows that the arc index of K is a lower bound for the braid index of any reverse string satellite of K. We have not been able to confirm how well the arc index analysis in the original paper extends to Ng's extra cases.

Remark Theorem 1.1 allows us to settle the long-standing conjecture that the ropelength of an alternating knot K is at least proportional to its crossing number. This statement is a consequence of [4, Theorem 3.1] and the fact that the ropelength of K is bounded below by a (fixed) constant multiple of the ropelength of $\mathbb{K}_f$ for some f. The more general case of an alternating link with two or more components remains open.

References

[1] **Y Bae**, **C-Y Park**, *An upper bound of arc index of links*, Math. Proc. Cambridge Philos. Soc. 129 (2000) 491–500 MR Zbl

[2] **J S Birman**, **W W Menasco**, *Special positions for essential tori in link complements*, Topology 33 (1994) 525–556 MR Zbl

[3] **P R Cromwell**, *Arc presentations of knots and links*, from "Knot theory" (V F R Jones, J Kania-Bartoszyńska, J H Przytycki, P Traczyk, V G Turaev, editors), Banach Center Publ. 42, Polish Acad. Sci. Inst. Math., Warsaw (1998) 57–64 MR Zbl

[4] **Y Diao**, *Braid index bounds ropelength from below*, J. Knot Theory Ramifications 29 (2020) art. id. 2050019 MR Zbl

[5] **Y Diao**, *Anti-parallel links as boundaries of knotted ribbons*, preprint (2021) arXiv 2011.06200

[6] **Y Diao**, **C Ernst**, **G Hetyei**, **P Liu**, *A diagrammatic approach for determining the braid index of alternating links*, J. Knot Theory Ramifications 30 (2021) art. id. 2150035 MR Zbl

[7] **J Franks**, **R F Williams**, *Braids and the Jones polynomial*, Trans. Amer. Math. Soc. 303 (1987) 97–108 MR Zbl

[8] **R J Hadji**, **H R Morton**, *A basis for the full Homfly skein of the annulus*, Math. Proc. Cambridge Philos. Soc. 141 (2006) 81–100 MR Zbl

[9] **L H Kauffman**, *State models and the Jones polynomial*, Topology 26 (1987) 395–407 MR Zbl

[10] **L H Kauffman**, *An invariant of regular isotopy*, Trans. Amer. Math. Soc. 318 (1990) 417–471 MR Zbl

[11] **H R Morton**, *Seifert circles and knot polynomials*, Math. Proc. Cambridge Philos. Soc. 99 (1986) 107–109 MR Zbl

[12] **H R Morton**, **E Beltrami**, *Arc index and the Kauffman polynomial*, Math. Proc. Cambridge Philos. Soc. 123 (1998) 41–48 MR Zbl

[13] **K Murasugi**, *On the braid index of alternating links*, Trans. Amer. Math. Soc. 326 (1991) 237–260 MR Zbl

[14] **K Y Ng**, *Essential tori in link complements*, J. Knot Theory Ramifications 7 (1998) 205–216 MR Zbl

[15] **I J Nutt**, *Embedding knots and links in an open book, III: On the braid index of satellite links*, Math. Proc. Cambridge Philos. Soc. 126 (1999) 77–98 MR Zbl

[16] **L Rudolph**, *A congruence between link polynomials*, Math. Proc. Cambridge Philos. Soc. 107 (1990) 319–327 MR Zbl

[17] **M B Thistlethwaite**, *On the Kauffman polynomial of an adequate link*, Invent. Math. 93 (1988) 285–296 MR Zbl

Department of Mathematics, University of North Carolina Charlotte
Charlotte, NC, United States

Department of Mathematical Sciences, University of Liverpool
Liverpool, United Kingdom

ydiao@charlotte.edu, morton@liverpool.ac.uk

Received: 4 April 2023 Revised: 8 August 2023

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 24 Issue 5 (pages 2389–2970) 2024