

*AG
T*

*Algebraic & Geometric
Topology*

Volume 23 (2023)

Issue 9 (pages 3909–4400)

ALGEBRAIC & GEOMETRIC TOPOLOGY

msp.org/agt

EDITORS

PRINCIPAL ACADEMIC EDITORS

John Etnyre
etnyre@math.gatech.edu
Georgia Institute of Technology

Kathryn Hess
kathryn.hess@epfl.ch
École Polytechnique Fédérale de Lausanne

BOARD OF EDITORS

Julie Bergner	University of Virginia jeb2md@eservices.virginia.edu	Robert Lipshitz	University of Oregon lipshitz@uoregon.edu
Steven Boyer	Université du Québec à Montréal cohf@math.rochester.edu	Norihiko Minami	Nagoya Institute of Technology nori@nitech.ac.jp
Tara E Brendle	University of Glasgow tara.brendle@glasgow.ac.uk	Andrés Navas	Universidad de Santiago de Chile andres.navas@usach.cl
Indira Chatterji	CNRS & Univ. Côte d'Azur (Nice) indira.chatterji@math.cnrs.fr	Thomas Nikolaus	University of Münster nikolaus@uni-muenster.de
Alexander Dranishnikov	University of Florida dranish@math.ufl.edu	Robert Oliver	Université Paris 13 bobol@math.univ-paris13.fr
Tobias Ekholm	Uppsala University, Sweden tobias.ekholm@math.uu.se	Jessica S Purcell	Monash University jessica.purcell@monash.edu
Mario Eudave-Muñoz	Univ. Nacional Autónoma de México mario@matem.unam.mx	Birgit Richter	Universität Hamburg birgit.richter@uni-hamburg.de
David Futer	Temple University dfuter@temple.edu	Jérôme Scherer	École Polytech. Féd. de Lausanne jerome.scherer@epfl.ch
John Greenlees	University of Warwick john.greenlees@warwick.ac.uk	Vesna Stojanoska	Univ. of Illinois at Urbana-Champaign vesna@illinois.edu
Ian Hambleton	McMaster University ian@math.mcmaster.ca	Zoltán Szabó	Princeton University szabo@math.princeton.edu
Matthew Hedden	Michigan State University mhedden@math.msu.edu	Maggy Tomova	University of Iowa maggy-tomova@uiowa.edu
Hans-Werner Henn	Université Louis Pasteur henn@math.u-strasbg.fr	Nathalie Wahl	University of Copenhagen wahl@math.ku.dk
Daniel Isaksen	Wayne State University isaksen@math.wayne.edu	Chris Wendl	Humboldt-Universität zu Berlin wendl@math.hu-berlin.de
Thomas Koberda	University of Virginia thomas.koberda@virginia.edu	Daniel T Wise	McGill University, Canada daniel.wise@mcgill.ca
Christine Lescop	Université Joseph Fourier lescop@ujf-grenoble.fr		

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

The subscription price for 2023 is US \$650/year for the electronic version, and \$940/year (+ \$70, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by [Mathematical Reviews](#), [Zentralblatt MATH](#), [Current Mathematical Publications](#) and the [Science Citation Index](#).

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

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

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2023 Mathematical Sciences Publishers

Two-dimensional extended homotopy field theories

KÜRŞAT SÖZER

We give another definition of 2–dimensional extended homotopy field theories (EHFTs) with aspherical targets and classify them. When the target of EHFT is chosen to be a $K(G, 1)$ –space, we classify EHFTs taking values in the symmetric monoidal bicategory of algebras, bimodules, and bimodule maps by certain Frobenius G –algebras called quasibiangular G –algebras. As an application, for any discrete group G , we verify a special case of the $(G \times \mathrm{SO}(2))$ –structured cobordism hypothesis due to Lurie.

57R56

1. Introduction	3909
2. The 2–dimensional X –cobordism bicategory	3914
3. The G –planar decompositions	3921
4. The classification of 2–dimensional extended X –HFTs	3944
5. Extended unoriented X –HFTs and their classifications	3974
Appendix. Unbiased semistrict symmetric monoidal 2–categories	3985
References	3994

1 Introduction

Extended topological field theories (ETFTs) are generalizations of topological field theories (usually called TQFTs or TFTs) to manifolds with corners and higher categories; see Freed [7], Lawrence [14] and J Baez and J Dolan [2]. A different generalization of TFTs is obtained by considering manifolds equipped with principal G –bundles. When G is a discrete group, such a generalization was introduced by Turaev [27], who called them homotopy (quantum) field theories (HFTs). These theories are defined

by applying axioms of TFTs to manifolds and cobordisms endowed with maps to a fixed target space. In this paper, we combine ETFTs and HFTs in dimension 2. More precisely, we define 2–dimensional extended homotopy field theories (EHFTs) with aspherical targets and classify them.

1.1 Main results

To define a 2–dimensional EHFT with target $X \simeq K(G, 1)$, we introduce an X –cobordism bicategory $X\text{Bord}_2$. The objects of $X\text{Bord}_2$ are compact oriented 0–dimensional manifolds and the 1–morphisms are oriented cobordisms between such manifolds equipped with homotopy classes of maps to X . The 2–morphisms of $X\text{Bord}_2$ are equivalence classes of pairs (S, P) , where S is a certain type of oriented surface with corners and P is a homotopy class of a map from S to X . The equivalence relation is given by diffeomorphisms respecting P and restricting to the identity on the boundary. The disjoint union operation turns $X\text{Bord}_2$ into a symmetric monoidal bicategory and a 2–dimensional EHFT with target X (extended X –HFT) is defined as a symmetric monoidal 2–functor from $X\text{Bord}_2$ to any other symmetric monoidal bicategory.

For a given symmetric monoidal bicategory \mathcal{C} , our classification of \mathcal{C} –valued 2–dimensional extended X –HFTs comprises two steps. Firstly, we define certain combinatorial diagrams in $I = [0, 1]$, I^2 , and I^3 , called G –linear, G –planar, and G –spatial diagrams, respectively. These diagrams generalize the ones of Schommer-Pries [22] and they possess the same information as the morphisms of $X\text{Bord}_2$. As the second step, we define a symmetric monoidal bicategory XB^{PD} whose 1– and 2–morphisms are defined using these diagrams. This bicategory is equivalent to $X\text{Bord}_2$ and has a convenient description in terms of generators and relations. Figures 25 and 26 show the corresponding list of generators and relations for $X\text{Bord}_2$. Then, the classification of 2–dimensional extended X –HFTs reduces to an application of the cofibrancy theorem [22], which is a coherence theorem for symmetric monoidal 2–functors explained below.

For a computadic symmetric monoidal bicategory $F(\mathbb{P})$ constructed from a list of generators and relations \mathbb{P} , let $\text{SymMon}(F(\mathbb{P}), \mathcal{C})$ denote the bicategory of symmetric monoidal 2–functors, transformations, and modifications. The cofibrancy theorem gives an equivalence $\text{SymMon}(F(\mathbb{P}), \mathcal{C}) \simeq \mathbb{P}(\mathcal{C})$ of bicategories, where $\mathbb{P}(\mathcal{C})$ is the bicategory, called \mathbb{P} –data in \mathcal{C} , whose objects are assignments of generators in \mathbb{P} to the objects, 1–morphisms, and 2–morphisms of \mathcal{C} subject to relations (see Section 4.2). Applying this theorem to the list of generators and relations $\mathbb{X}\mathbb{P}$ of XB^{PD} , along with the equivalence $X\text{Bord}_2 \simeq \text{XB}^{\text{PD}}$, gives the following classification theorem:

Theorem 4.10 For any symmetric monoidal bicategory \mathcal{C} , there is an equivalence of bicategories

$$\text{SymMon}(X\text{Bord}_2, \mathcal{C}) \simeq \mathbb{X}\mathbb{P}(\mathcal{C}).$$

Next, we consider a specific target bicategory $\text{Alg}_{\mathbb{k}}^2$ of \mathbb{k} -algebras, bimodules, and bimodule maps for a commutative ring \mathbb{k} with unit. The following notions are the main ingredients of our result on $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFTs. For a discrete group G with identity element e , a strongly graded G -algebra is a G -graded associative \mathbb{k} -algebra $A = \bigoplus_{g \in G} A_g$ with unity such that $A_g A_{g'} = A_{gg'}$ for all $g, g' \in G$. The opposite G -algebra of A is $A^{\text{op}} = \bigoplus_{g \in G} A_{g^{-1}}$, where the order of multiplication is reversed.

A Frobenius G -algebra is a pair (A, η) , where $A = \bigoplus_{g \in G} A_g$ is a G -algebra such that each A_g is a finitely generated projective \mathbb{k} -module, and $\eta: A \otimes A \rightarrow \mathbb{k}$ is a nondegenerate bilinear form satisfying $\eta(ab, c) = \eta(a, bc)$ for any $a, b, c \in A$. A quasibiangular G -algebra is a strongly graded Frobenius G -algebra (A, η) in which the identity component A_e is separable and η satisfies certain conditions (see Section 4.3). We also need G -graded Morita contexts between G -algebras, which were introduced by Boisen [3]. We recall their definition and introduce a notion of compatibility with Frobenius structures in Section 4.3.

Theorem 4.19 Let \mathbb{k} be a commutative ring and X be a CW-complex which is a $K(G, 1)$ -space for a discrete group G . Then any $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFT $Z: X\text{Bord}_2 \rightarrow \text{Alg}_{\mathbb{k}}^2$ whose precomposition $X\text{B}^{\text{PD}} \xrightarrow{\simeq} X\text{Bord}_2 \xrightarrow{Z} \text{Alg}_{\mathbb{k}}^2$ gives a strict symmetric monoidal 2-functor determines a triple (A, B, ζ) , where A and B are quasibiangular G -algebras and ζ is a compatible G -graded Morita context between A and B^{op} . Conversely, for any such triple (A, B, ζ) there exists an $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFT.

This generalizes Schommer-Pries' classification of $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended TFTs [22] in terms of separable symmetric Frobenius algebras. Theorem 4.10 suggests studying the bicategory $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ to understand $\text{SymMon}(X\text{Bord}_2, \text{Alg}_{\mathbb{k}}^2)$, more specifically to answer the question of which triples yield equivalent extended X -HFTs. This study leads us to define a bicategory, Frob^G , which has quasibiangular G -algebras as objects, compatible G -graded Morita contexts as 1-morphisms, and equivalences of such Morita contexts as 2-morphisms (see Section 4.3).

Theorem 4.24 Under the assumptions of [Theorem 4.19](#), there is an equivalence of bicategories

$$\text{SymMon}(X\text{Bord}_2, \text{Alg}_{\mathbb{k}}^2) \simeq \text{Frob}^G.$$

On the level of objects, this equivalence maps a triple (A, B, ζ) to A . Consequently, [Theorem 4.24](#) implies the following corollary:

Corollary 4.25 Under the assumptions of [Theorem 4.19](#), two triples (A_1, B_1, ζ_1) and (A_2, B_2, ζ_2) produce equivalent 2–dimensional extended X –HFTs if and only if there exists a compatible G –graded Morita context between A_1 and A_2 .

A different approach to the classification of 2–dimensional EHFTs with $K(G, 1)$ targets is given by the $(G \times \text{SO}(2))$ –structured cobordism hypothesis due to J Lurie [\[15\]](#). This hypothesis states a classification of such EHFTs in terms of homotopy $(G \times \text{SO}(2))$ –fixed points (see [Section 4.5](#)). Davidovich [\[6\]](#) computed these fixed points in $\text{Alg}_{\mathbb{k}}^2$ when \mathbb{k} is an algebraically closed field of characteristic zero. By comparing [Theorem 4.24](#) with Davidovich’s results, we verify a special case of the $(G \times \text{SO}(2))$ –structured cobordism hypothesis as follows:

Corollary 4.27 For any discrete group G and any algebraically closed field \mathbb{k} of characteristic zero, the $(G \times \text{SO}(2))$ –structured cobordism hypothesis for $\text{Alg}_{\mathbb{k}}^2$ –valued oriented EHFTs with target $X \simeq K(G, 1)$ holds true.

In the definition of $X\text{Bord}_2$, we use oriented manifolds. By using unoriented manifolds, we define the unoriented X –cobordism bicategory $X\text{Bord}_2^{\text{un}}$ and provide a list of generators and relations. Then, parallel to the oriented case, we classify 2–dimensional extended unoriented HFTs and verify a special case of the $(G \times O(2))$ –structured cobordism hypothesis.

1.2 Related works

Our main reference is Schommer-Pries’ thesis [\[22\]](#) on the classification of 2–dimensional extended TFTs. In addition to detailed classification of oriented and unoriented extended TFTs, Schommer-Pries sketched the classification of 2–dimensional structured extended TFTs (see [\[22, Section 3.5\]](#)). In this approach the structured cobordism bicategory is defined using topological stacks. In particular, a stack corresponding to principal G –bundles and orientation structures provides an alternative formulation for extended $(G \times \text{SO}(2))$ –structured TFTs, or equivalently 2–dimensional extended HFTs with $K(G, 1)$ targets.

When defining 2–dimensional extended TFTs, Schommer-Pries [22] defined the cobordism bicategory in all dimensions, not only in dimension 2. Schweigert and Woike [23] extended this cobordism bicategory to the setting of homotopy field theories and defined extended HFTs in all dimensions. Similarly, Müller and Szabo [18] constructed a geometric cobordism bicategory $\text{Cob}_{n,n-1,n-2}^{\mathcal{F}}$, where \mathcal{F} is a general stack which encodes the arbitrary background fields in the corresponding quantum field theory (see also Müller [17]). In their work [19], when the background fields \mathcal{F} are chosen for 2–dimensional Dijkgraaf–Witten theories, namely principal bundles with finite structure group G and orientations, the symmetric monoidal bicategory $\text{Cob}_{2,1,0}^{\mathcal{F}}$ is equivalent to $X\text{Bord}_2$, described above. Using the n –dimensional version $\text{Cob}_{n,n-1,n-2}^{\mathcal{F}}$ of this bicategory, they defined an n –dimensional extended homotopy field theory for all $n \geq 2$. Moreover, Müller and Woike [20] constructed an n –dimensional extended HFT from a flat $(n-1)$ –gerbe on a target space represented by a $U(1)$ –valued singular cocycle (see also [17] and Bunke, Turner and Willerton [4]). This construction was generalized to unoriented extended HFTs by Young [29].

A state sum approach to both 2–dimensional extended TFTs and HFTs was taken by Davidovich [6]. As mentioned above, Davidovich [6] also classified $\text{Alg}_{\mathbb{k}}^2$ –valued 2–dimensional extended $(G \times \text{SO}(2))$ –structured TFTs following the cobordism hypothesis.

Conventions Throughout the paper, G is a discrete group with identity element e and the target space is a pointed aspherical CW–complex (X, x) with $\pi_1(X, x) = G$. All manifolds are assumed to be smooth and all algebras are unital. By a closed manifold we mean a compact manifold without boundary. For smooth manifolds M and N , the space of smooth maps $C^\infty(M, N)$ is provided with the Whitney C^∞ –topology. For subsets $K \subset M$ and $L \subset N$, the notation $[(M, K), (N, L)]$ stands for the set of relative homotopy classes of maps between pairs.

Acknowledgments I would like to thank my advisor Vladimir Turaev for introducing this problem to me and his support throughout this project. I would also like to thank Noah Snyder for fruitful and enlightening discussions on extended field theories and the cobordism hypothesis. I am grateful to Patrick Chu for helpful discussions and to Alexis Virelizier for his comments on the earlier version of this paper. I would like to thank the referee for valuable comments and suggestions. This work was supported by NSF grant DMS-1664358.

2 The 2–dimensional X –cobordism bicategory

In his study of HFTs, Turaev [27] introduced notions of X –manifold and X –cobordism using pointed manifolds, where X is a connected CW–complex with a specified point $x \in X$. In this paper, X is always a $K(G, 1)$ –space. In this case, a *pointed manifold* is a manifold with a basepoint on each connected component. We denote the set of basepoints of a pointed manifold M by bp_M . An n –dimensional X –manifold is a pair (M, g) consisting of a closed pointed n –manifold M and a homotopy class $g \in [(M, \text{bp}_M), (X, x)]$, called the *characteristic map*. An X –cobordism between X –manifolds (M, g) and (M', g') is a pair (W, P) consisting of a cobordism W between M and M' and a homotopy class $P \in [(W, \text{bp}_M \cup \text{bp}_{M'}), (X, x)]$ restricting to g and g' on the corresponding boundary components.

Given an aspherical space X , a 2–dimensional extended X –HFT is a symmetric monoidal 2–functor from the 2–dimensional X –cobordism bicategory $X\text{Bord}_2$ to another symmetric monoidal bicategory. Therefore, this bicategory plays a key role in the definition of 2–dimensional extended X –HFTs.

There are existing definitions of structured or equivariant cobordism bicategories related to $X\text{Bord}_2$. These include the homotopy bicategory of the $(\infty, 2)$ –category of cobordisms with $(G \times \text{SO}(2))$ –structures (see [15; 5]), the structured cobordism bicategory $\text{Bord}_2^{\mathcal{F}}$ introduced in [22] with a topological stack \mathcal{F} corresponding to principal G –bundles and orientation structures, the G –equivariant cobordism bicategory $G\text{–Cob}(2, 1, 0)$ introduced in [23], and the structured cobordism bicategory $\text{Cob}_{2,1,0}^{\mathcal{F}}$ introduced in [18] with an appropriate choice of a stack \mathcal{F} (see [19; 17]). It can be shown that these symmetric monoidal bicategories are equivalent to $X\text{Bord}_2$ and hence the corresponding extended HFTs with aspherical targets are equivalent.

We start this section with a descriptive definition of $X\text{Bord}_2$ to motivate the types of X –manifolds and structures on them, which form the 1– and 2–morphisms of this bicategory. Then we provide the complete definition of $X\text{Bord}_2$.

2.1 Surfaces with corners

Roughly, the objects of $X\text{Bord}_2$ are compact oriented 0–manifolds, 1–morphisms are 1–dimensional X –cobordisms, and 2–morphisms are X –homeomorphism classes of 2–dimensional X –cobordisms between those X –cobordisms. This hints that the underlying manifold of a 2–morphism must be a surface with corners. Recall that a surface with corners M is a 2–dimensional topological manifold whose coordinate

charts are of the form $\varphi: U \rightarrow \mathbb{R}_+^2$, where $U \subset M$ is open and $\mathbb{R}_+^2 = [0, \infty) \times [0, \infty)$. Compatibility of charts is given by diffeomorphisms; that is, two charts (U, φ) and (U', φ') with $U \cap U' \neq \emptyset$ are compatible if the composition $\varphi' \circ \varphi^{-1}: \varphi(U \cap U') \rightarrow \varphi'(U \cap U')$ is a diffeomorphism. Here, when $\varphi(U \cap U') \cap \partial \mathbb{R}_+^2 \neq \emptyset$, the map $\varphi' \circ \varphi^{-1}$ is a diffeomorphism if it is a restriction of a diffeomorphism defined on an open set containing $\varphi(U \cap U')$.

The composition of 2-morphisms in the X -cobordism bicategory is given by gluing surfaces with corners along their common boundaries. Recall that the first step of gluing construction is to choose collar neighborhoods. Nevertheless, not every surface with corners admits collar neighborhoods. Following [22], we use $\langle 2 \rangle$ -surfaces, which are special cases of $\langle n \rangle$ -manifolds defined in [13]. These surfaces admit collar neighborhoods (see [13]) and they are a certain type of surfaces with faces.

A surface with faces M is a surface with corners such that any point $m \in M$ belongs to $\text{index}(m)$ different connected faces. Here, the index of a point m is the number of zeros in $\varphi(m) \in \mathbb{R}_+^2$, where (U, φ) is a chart with $m \in U$ and a connected face of a surface with corners M is the closure of a component of $\{m \in M \mid \text{index}(m) = 1\}$. A face is a disjoint union of connected faces.

Definition 2.1 A $\langle 2 \rangle$ -surface is a 2-dimensional compact manifold with faces S equipped with two submanifolds with faces $\partial_h S$ and $\partial_v S$, called the *horizontal* and *vertical faces*, respectively, such that $\partial S = \partial_h S \cup \partial_v S$ and $\partial_h S \cap \partial_v S$ is either empty or a face of both. A $\langle 2 \rangle$ -surface S is *pointed* if it is equipped with a finite set $R \subset \partial S$ such that $\partial_h S \cap \partial_v S \subset R$, $\partial_v S \cap R = \partial_h S \cap \partial_v S$, and every connected component of $\partial_h S$ contains at least two elements of R .

Definition 2.2 A $\langle 2 \rangle$ - X -surface is a triple (S, R, P) , where (S, R) is a pointed oriented $\langle 2 \rangle$ -surface and $P \in [(S, R), (X, x)]$ is a homotopy class. A $\langle 2 \rangle$ - X -surface (S, R, P) is said to be of *cobordism type* if $\partial_v S$ is diffeomorphic to a product X -manifold with a constant characteristic map, ie $(\partial_v S, P|_{\partial_v S}) \cong (M \times I, P|_{M \times I})$, where $I = [0, 1]$, $(M, P|_M)$ is a 0-dimensional X -manifold, and the restriction of $P|_{M \times I} \in [(M \times I, \partial(M \times I)), (X, x)]$ to each connected component is the constant homotopy class.

Figure 1 shows an example of a cobordism type $\langle 2 \rangle$ - X -surface (S, R, P) , where we encode the data of relative homotopy class P by arrows and G -labels are determined uniquely by P and arrows. Observe that the horizontal boundary of a $\langle 2 \rangle$ - X -surface (S', R', P') is not a 1-dimensional X -cobordism if $R' \neq \partial(\partial_h S')$. Since we regard

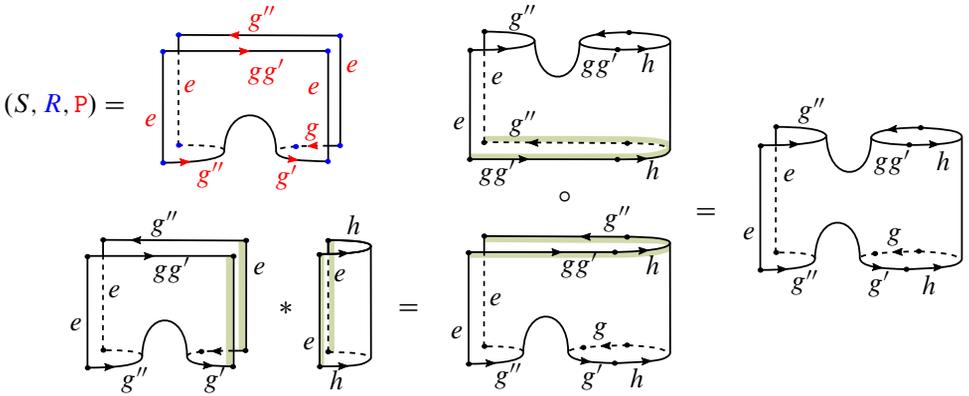


Figure 1: Examples of cobordism type $\langle 2 \rangle$ - X -surfaces and their compositions.

2-morphisms of $X\text{Bord}_2$ to be X -cobordisms between 1-morphisms, this observation implies that 1-morphisms are more general than 1-dimensional X -cobordisms in that there are possibly extra points in the interior of the underlying 1-dimensional compact manifold.

Definition 2.3 A 1-dimensional marked X -manifold is a triple (M, T, g) , where M is an oriented compact 1-manifold, $T \subseteq M$ is a finite set with $\partial M \subset T$ such that each connected component of M contains at least two elements of T , and $g \in [(M, T), (X, x)]$.

According to the arguments above, the bicategory $X\text{Bord}_2$ is expected to have compact oriented 0-manifolds as objects, 1-dimensional marked X -manifolds as 1-morphisms, and cobordism type $\langle 2 \rangle$ - X -surfaces as 2-morphisms. In this case, for a given cobordism type $\langle 2 \rangle$ - X -surface, its source and target 1-morphisms are certain components of the horizontal boundary. However, the composition of morphisms is a delicate issue, especially the composition of 1-morphisms.

When we glue two manifolds along their common boundary, the smooth structure on the resulting topological manifold depends on the choice of collar neighborhoods (see [16]). Equivalently, different choices of collars give different smooth structures. However, different choices give diffeomorphic smooth manifolds and diffeomorphisms are noncanonical. Therefore, gluing operation on smooth manifolds is not well defined on the nose, but up to a noncanonical diffeomorphism.

The same results continue to hold for $\langle 2 \rangle$ - X -surfaces. Let (S, R, P) be a $\langle 2 \rangle$ - X -surface and $(N, N \cap R, P|_N)$ be a face with the inclusion map $\iota: (N, N \cap R) \hookrightarrow (S, R)$. For a

collar neighborhood $U_N \subset S$ of N , a *collar* is a diffeomorphism $\Psi_N: U_N \rightarrow N \times \mathbb{R}_+$ (see Figure 1). The following proposition implies that 2-morphisms of $X\text{Bord}_2$ must be (relative) diffeomorphism classes of cobordism type $\langle 2 \rangle$ - X -surfaces in order to have well-defined horizontal and vertical compositions of 2-morphisms:

Proposition 2.4 *Let (S, R, P) and (S', R', P') be cobordism type $\langle 2 \rangle$ - X -surfaces with faces and (N, T, g) be a 1-dimensional marked X -manifold together with inclusions $\iota: (N, T, g) \hookrightarrow (S, R, P)$ and $\iota': (N, T, g) \hookrightarrow (S', R', P')$ realizing (N, T, g) as a face of both $\langle 2 \rangle$ - X -surfaces. Then $S \cup_N S'$ is a topological manifold with boundary. If in addition we are given collars $\Psi_+: N \times \mathbb{R}_+ \rightarrow S$ and $\Psi_-: N \times \mathbb{R}_+ \rightarrow S'$, then there exists a canonical smooth structure on $S \cup_N S'$ which is compatible with the smooth structures on S and S' . Moreover, different choices of collars produce noncanonically diffeomorphic cobordism type $\langle 2 \rangle$ -surfaces.*

The proof of this theorem follows from the proofs of Proposition 3.1 and Theorem 3.3 in [22]. Note that gluing $\langle 2 \rangle$ - X -surfaces vertically along their common horizontal boundary components does not yield a cobordism type $\langle 2 \rangle$ - X -surface. One needs to choose a diffeomorphism $I \cup I \cong I$ and omit the points on the faces through which $\langle 2 \rangle$ - X -surfaces are glued. Figure 1 shows examples of vertical and horizontal compositions of $\langle 2 \rangle$ - X -surfaces, denoted by \circ and $*$, respectively. Following [22], we solve the problem of composition of 1-morphisms by equipping manifolds with germs of neighborhoods. The notion of a germ of neighborhoods was made precise by Schommer-Pries (see Section 3.2.3 in [22]) using halations which are formulated as maps of pro-manifolds.

2.2 Pro- X -manifolds and X -halations

Recall that a directed set is a tuple (D, \leq) , where D is a nonempty set and \leq is a reflexive and transitive binary relation such that, for any $x, y \in D$, there exists $z \in D$ with $x \leq z$ and $y \leq z$. We think of a directed set (D, \leq) as a category \mathcal{D} whose objects are elements of D and morphisms are given by the relation \leq . Let Man^X be the category of smooth X -manifolds and smooth pointed maps commuting with characteristic maps. A *pro- X -manifold* is a pair (\mathcal{D}, A) , where \mathcal{D} is a directed set and $A: \mathcal{D} \rightarrow \text{Man}^X$ is a functor.

Two directed sets play an important role in describing germs of neighborhoods of X -manifolds: the first one is trivial one, $\mathcal{D}_\bullet = \{\bullet\}$, and the second one is associated to an embedding of X -manifolds as follows. Let (M, g) and (N, h) be X -manifolds

possibly with boundary or corners, and let $\iota: (M, \mathfrak{g}) \hookrightarrow (N, \mathfrak{h})$ be an embedding of X -manifolds, ie $\iota(\text{bp}_M) = \text{bp}_N$ and $\mathfrak{g} = \mathfrak{h} \circ [\iota]$ as elements of $[(M, \text{bp}_M), (X, x)]$. Then the directed set \mathcal{D}_N is given by codimension zero closed X -submanifolds of N containing $\iota(M)$ and the relation is inclusion.

For a given X -manifold (M, \mathfrak{g}) , we denote the pro- X -manifolds corresponding to these directed sets with (M, \mathfrak{g}) and $(\widehat{M} \subset N, \widehat{\mathfrak{g}})$, respectively. There is an obvious inclusion $\iota_M: (M, \mathfrak{g}) \hookrightarrow (\widehat{M} \subset N, \widehat{\mathfrak{g}})$ of pro- X -manifolds and an X -halation is an inclusion of pro- X -manifolds isomorphic to ι_M . Here morphisms of pro- X -manifolds are the morphisms in Man^X of the corresponding limits and colimits of the diagrams. More precisely, the set of morphisms between two pro- X -manifolds discussed above is

$$\text{Hom}_{\text{pro-Man}^X}(\mathcal{D}_\bullet, \mathcal{D}_N) = \lim_p \text{colim}_q \text{Hom}_{\text{Man}^X}(\mathcal{D}_\bullet(p), \mathcal{D}_N(q)),$$

where the limit and colimit are taken in the category of sets. The codimension of an X -halation is the codimension of the embedding. We denote an X -halation by $(M, \widehat{M}, \widehat{\mathfrak{g}})$ and call an X -manifold equipped with an X -halation an X -haloed manifold. A map between X -haloed manifolds $(A, \widehat{A}, \widehat{\mathfrak{a}})$ and $(B, \widehat{B}, \widehat{\mathfrak{b}})$ is a pair of pro- X -manifold morphisms $A \rightarrow B$ and $\widehat{A} \rightarrow \widehat{B}$ such that the diagram involving the inclusions $A \hookrightarrow \widehat{A}$ and $B \hookrightarrow \widehat{B}$ commutes. The category of pro-objects in a category \mathcal{C} is generally defined using cofiltered diagrams instead of directed sets. Here we use the results in [22] to simplify arguments and refer reader to [22, Sections 3.2.1 and 3.2.2] for a more detailed exposition on halations.

The solution to the problem of composition of 1-morphisms is to use compatible X -haloed manifolds. The compatibility is given by choosing an orientation for the normal bundle of the embedding which defines X -halation. Such an X -halation is called *cooriented*. Now assume that $(M_0, T_0, \mathfrak{g}_0)$ and $(M_1, T_1, \mathfrak{g}_1)$ are 1-dimensional marked X -manifolds equipped with cooriented codimension one X -halations $(\widehat{M}_0 \subset N_0, \widehat{\mathfrak{g}}_0)$ and $(\widehat{M}_1 \subset N_1, \widehat{\mathfrak{g}}_1)$, respectively. Using the tubular neighborhood theorem and the coorientations of the source and target objects, we write these X -halations as $(M_i \subset M_i \times \mathbb{R} \cup_{\partial M_i \times \mathbb{R} \times \{0\}} \partial M_i \times (\mathbb{R} \times \mathbb{R}_+))$ for $i = 0, 1$. Using [22, Lemma 3.25], we refine the index of both pro- X -manifolds to natural numbers \mathbb{N} as

$$\begin{aligned} i &\mapsto M_0 \times \{0\} \cup_{\partial_t M_0} \partial_t M_0 \times \{0\} \times \left(-\frac{1}{i}, 0\right] \cup_{\partial_s M_0} \partial_s M_0 \times \{0\} \times \left[0, \frac{1}{i}\right), \\ i &\mapsto M_1 \times \{0\} \cup_{\partial_s M_1} \partial_s M_1 \times \{0\} \times \left[0, \frac{1}{i}\right) \cup_{\partial_t M_1} \partial_t M_1 \cup \{0\} \times \left(-\frac{1}{i}, 0\right], \\ i &\mapsto Y \times \left(-\frac{1}{i}, \frac{1}{i}\right), \end{aligned}$$

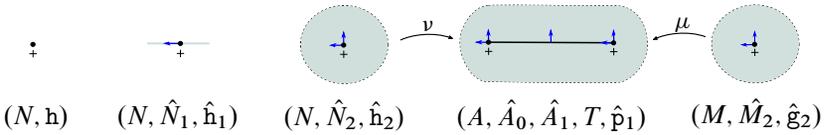


Figure 2: Cooriented X -halations and an X -haloed 1-cobordism.

where ∂_s and ∂_t denote the source and target boundary components and $Y \cong \partial_t M_0 \cong \partial_s M_1$. For each $i \in \mathbb{N}$, the pushout exists and it is the smooth manifold $M_1 \cup_Y M_0$ whose smooth structure is determined by the embedding of $Y \times (-1/i, 1/i)$ and by the smooth structures of M_0 and M_1 . The pushout as a cooriented codimension one X -haloed manifold, ie as a 1-morphism of $X\text{Bord}_2$, exists by the results in [11] on commutativity of finite colimits and (cofiltered) limits. Here note that the colimit is taken over the pushout diagram $K = \bullet \leftarrow \bullet \rightarrow \bullet$, which is clearly finite (see [22, Section 3.2.3] for details).

2.3 The X -cobordism bicategory $X\text{Bord}_2$

We are now equipped with the necessary information to define the 2-dimensional X -cobordism bicategory.

Definition 2.5 The 2-dimensional X -cobordism bicategory $X\text{Bord}_2$ is as follows:

(1) The objects are triples $\{(M, \mathfrak{g}), (M, \hat{M}_1, \hat{\mathfrak{g}}_1), (M, \hat{M}_2, \hat{\mathfrak{g}}_2)\}$, where (M, \mathfrak{g}) is a compact oriented 0-manifold, and $(M, \hat{M}_1, \hat{\mathfrak{g}}_1)$ and $(M, \hat{M}_2, \hat{\mathfrak{g}}_2)$ are cooriented codimension one and codimension two X -halations, respectively, with inclusions $(M, \mathfrak{g}) \hookrightarrow (M, \hat{M}_1, \hat{\mathfrak{g}}_1) \hookrightarrow (M, \hat{M}_2, \hat{\mathfrak{g}}_2)$. For brevity we denote such an object as $(M, \hat{M}_1, \hat{M}_2, \hat{\mathfrak{g}}_2)$.

(2) The 1-morphisms are X -haloed 1-dimensional X -cobordisms; an X -haloed 1-dimensional X -cobordism $(A, \hat{A}_0, \hat{A}_1, T, \hat{p}_1)$ from $(M, \hat{M}_1, \hat{M}_2, \hat{\mathfrak{g}}_2)$ to $(N, \hat{N}_1, \hat{N}_2, \hat{\mathfrak{h}}_2)$ consists of

- a 1-dimensional marked X -manifold (A, T, \mathfrak{p}) ,
- a codimension zero X -halation $(A, \hat{A}_0, \hat{\mathfrak{p}}_0)$ and a cooriented codimension one X -halation $(A, \hat{A}_1, \hat{\mathfrak{p}}_1)$ with inclusions $(A, \mathfrak{p}) \hookrightarrow (A, \hat{A}_0, \hat{\mathfrak{p}}_0) \hookrightarrow (A, \hat{A}_1, \hat{\mathfrak{p}}_1)$,
- a decomposition of the boundary of (A, T, \mathfrak{p}) as $\partial A = \partial_{\text{in}} A \amalg \partial_{\text{out}} A$ with isomorphisms of X -halations preserving coorientations (see Figure 2)

$$\begin{aligned}
 (M, \hat{M}_1, \hat{M}_2, \hat{\mathfrak{g}}_2) &\xrightarrow{\mu} (\partial_{\text{in}} A, \hat{A}_0|_{\partial_{\text{in}}}, \hat{A}_1|_{\partial_{\text{in}}}, \hat{\mathfrak{p}}_1), \\
 (N, \hat{N}_1, \hat{N}_2, \hat{\mathfrak{h}}_2) &\xrightarrow{\nu} (\partial_{\text{out}} A, \hat{A}_0|_{\partial_{\text{out}}}, \hat{A}_1|_{\partial_{\text{out}}}, \hat{\mathfrak{p}}_1),
 \end{aligned}$$

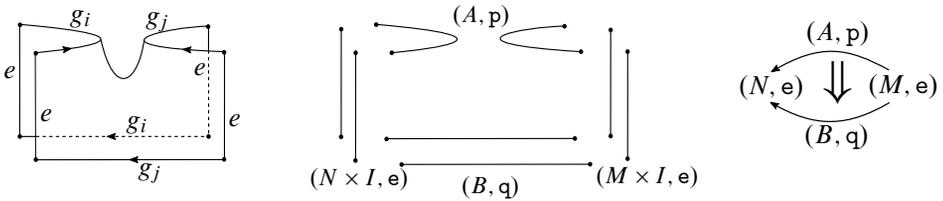


Figure 3: An example of decomposition of a 2-morphism in $X \text{Bord}_2$.

where $\widehat{A}_0|_{\partial_{\text{in}}}$ is cooriented by an inward pointing normal vector and $\widehat{A}_0|_{\partial_{\text{out}}}$ is cooriented by an outward pointing normal vector.

(3) The 2-morphisms are isomorphism classes of X -haloed 2-dimensional X -cobordisms; an X -haloed 2-dimensional X -cobordism $(S, \widehat{S}, R, \widehat{F})$ from $(A, \widehat{A}_0, \widehat{A}_1, T, \widehat{p}_1)$ to $(B, \widehat{B}_0, \widehat{B}_1, Q, \widehat{q}_1)$ consists of a cobordism type $\langle 2 \rangle$ - X -surface (S, R, F) together with a codimension zero X -halation $(S, \widehat{S}, \widehat{F})$ and isomorphisms of X -halations (see Figure 3)

$$(A, \widehat{A}_1, \widehat{p}_1) \amalg (B, \widehat{B}_1, \widehat{q}_1) \xrightarrow{\theta} (\partial_h S, \widehat{S}|_{\partial_h S}, \widehat{F}|_{\partial_h S}),$$

$$(M \times I, \widehat{M \times I^2}, \widehat{e}) \amalg (N \times I, \widehat{N \times I^2}, \widehat{e}) \xrightarrow{\eta} (\partial_v S, \widehat{S}|_{\partial_v S}, \widehat{F}|_{\partial_v S}),$$

where $(A, \widehat{A}_1, \widehat{p}_1)$ is cooriented by an inward pointing normal vector and $(B, \widehat{B}_1, \widehat{q}_1)$ is cooriented by an outward pointing normal vector. The X -halations of $M \times I$ and $N \times I$ are induced by their embeddings into $M \times I^2$ and $N \times I^2$ with constant homotopy class \widehat{e} . Coorientations are given by an inward pointing normal vector for $(M \times I, \widehat{M \times I^2}, \widehat{e})$ and an outward pointing normal vector for $(N \times I, \widehat{N \times I^2}, \widehat{e})$. For such an X -haloed 2-cobordism we have the notation $\theta(A) = \partial_{h,\text{in}} S$, $\theta(B) = \partial_{h,\text{out}} S$, $\eta(M \times I) = \partial_{v,\text{in}} S$, and $\eta(N \times I) = \partial_{v,\text{out}} S$.

Two X -haloed 2-cobordisms $(S_0, \widehat{S}_0, R_0, \widehat{F}_0)$ and $(S_1, \widehat{S}_1, R_1, \widehat{F}_1)$ are isomorphic if there is an isomorphism of X -halations $\xi: (S_0, \widehat{S}_0, \widehat{F}_0) \rightarrow (S_1, \widehat{S}_1, \widehat{F}_1)$ which restricts to isomorphisms

$$(\partial_{h,\text{in}} S_0, (\widehat{S}_0)|_{\partial_{h,\text{in}} S_0}, (\widehat{F}_0)|_{\partial_{h,\text{in}} S_0}) \rightarrow (\partial_{h,\text{in}} S_1, (\widehat{S}_1)|_{\partial_{h,\text{in}} S_1}, (\widehat{F}_1)|_{\partial_{h,\text{in}} S_1}),$$

$$(\partial_{h,\text{out}} S_0, (\widehat{S}_0)|_{\partial_{h,\text{out}} S_0}, (\widehat{F}_0)|_{\partial_{h,\text{out}} S_0}) \rightarrow (\partial_{h,\text{out}} S_1, (\widehat{S}_1)|_{\partial_{h,\text{out}} S_1}, (\widehat{F}_1)|_{\partial_{h,\text{out}} S_1}),$$

$$(\partial_{v,\text{in}} S_0, (\widehat{S}_0)|_{\partial_{v,\text{in}} S_0}, (\widehat{F}_0)|_{\partial_{v,\text{in}} S_0}) \rightarrow (\partial_{v,\text{in}} S_1, (\widehat{S}_1)|_{\partial_{v,\text{in}} S_1}, (\widehat{F}_1)|_{\partial_{v,\text{in}} S_1}),$$

$$(\partial_{v,\text{out}} S_0, (\widehat{S}_0)|_{\partial_{v,\text{out}} S_0}, (\widehat{F}_0)|_{\partial_{v,\text{out}} S_0}) \rightarrow (\partial_{v,\text{out}} S_1, (\widehat{S}_1)|_{\partial_{v,\text{out}} S_1}, (\widehat{F}_1)|_{\partial_{v,\text{out}} S_1})$$

such that $\xi|_{\partial S_0}$ is identity, $\xi \circ \eta = \eta'$ and $\xi \circ \theta = \theta'$, where θ' and η' are isomorphisms of cooriented X -halations corresponding to the decomposition $\partial S_1 = \partial_h S_1 \amalg \partial_v S_1$.

Lemma 2.6 *The bicategory $X\text{Bord}_2$ is a symmetric monoidal bicategory under disjoint union.*

We skip the proof, which is given in [25] using a method developed by Shulman [24]. Recall that two major goals of this paper are to define 2–dimensional extended homotopy field theories and classify them. The following definition addresses to the first one:

Definition 2.7 Let \mathcal{C} be a symmetric monoidal bicategory. A \mathcal{C} –valued 2–dimensional extended homotopy field theory with target X is a symmetric monoidal 2–functor from $X\text{Bord}_2$ to \mathcal{C} .

3 The G –planar decompositions

3.1 G –linear diagrams

Linear diagrams, introduced by Schommer-Pries [22], represent 1–dimensional compact manifolds equipped with a Morse function to $[0, 1]$. Briefly speaking, a linear diagram is a triple formed by the set of critical values of a Morse function on a compact 1–manifold, an open cover of $[0, 1]$, and combinatorial data describing preimages of a Morse function on open sets. By labeling critical values with cup or cap instead of their indices (see Figure 4) the first ingredient of a linear diagram is defined as follows:

Definition 3.1 [22] A 1–dimensional graphic Ψ is a finite subset of $(0, 1)$ where each point is labeled with either cup or cap.

For a given 1–dimensional graphic Ψ , an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ of $[0, 1]$ having at most double intersections is said to be Ψ –compatible if each U_α contains at most one element from Ψ and double intersections are disjoint from Ψ . It is not hard to find such open covers and the second ingredient of a linear diagram is defined as follows:

Definition 3.2 [22] Let Ψ be a 1–dimensional graphic. A chambering set Γ for Ψ is a set of isolated points in $(0, 1)$ disjoint from μ . Chambers of Γ are the connected components of $[0, 1] \setminus (\Gamma \cup \mu)$. A chambering set Γ is said to be subordinate to an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ of $[0, 1]$ if each chamber is a subset of at least one U_α .

Example 3.3 Figure 5 shows an example of a 1–dimensional oriented compact manifold M equipped with a Morse function $f: (M, \partial M) \rightarrow ([0, 1], \{0, 1\})$. The

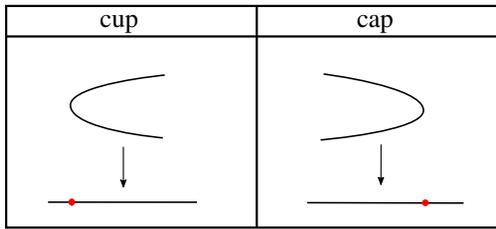


Figure 4: Singularities of a Morse function on a 1-manifold and their images in \mathbb{R} .

critical values of f define a 1-dimensional graphic Ψ (see Figure 4). An open cover $\mathcal{U} = \{U_i\}_{i=1}^4$ of $[0, 1]$ is a Ψ -compatible open cover and turquoise points form a chambering set subordinate to \mathcal{U} .

For an oriented compact 1-manifold M , a Morse function of the form $f : (M, \partial M) \rightarrow ([0, 1], \{0, 1\})$ whose critical values are distinct and lie in $(0, 1)$ is called a *generic map*. Let Ψ be a 1-dimensional graphic induced from a pair (M, f) of an oriented compact 1-manifold equipped with a generic map. Let Γ be a chambering set subordinate to a Ψ -compatible open cover \mathcal{U} . Since f is a Morse function and chambers are disjoint from μ , the preimage $f^{-1}(V)$ of a chamber V consists of a disjoint union of arcs (possibly empty), each mapping diffeomorphically onto V under f . A *trivialization* of V is an identification of $f^{-1}(V)$ with $\mathbb{N}_{\leq N} \times V$ for some $N \in \mathbb{N}$, where $\mathbb{N}_{\leq N} = \{a \in \mathbb{N} \mid 0 < a \leq N\}$ if $f^{-1}(V)$ is nonempty and an identification with the empty set otherwise. In this case, each $\{i\} \times V$ is called a *sheet* and each sheet is equipped with an orientation.

Trivializations of two neighboring chambers have the same number of sheets if chambers are separated by a point in Γ . If a point in μ separates chambers, then, by the Morse

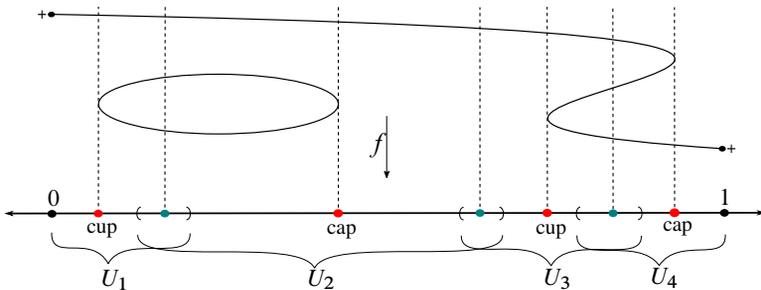


Figure 5: Induced 1-dimensional graphic and a chambering set.

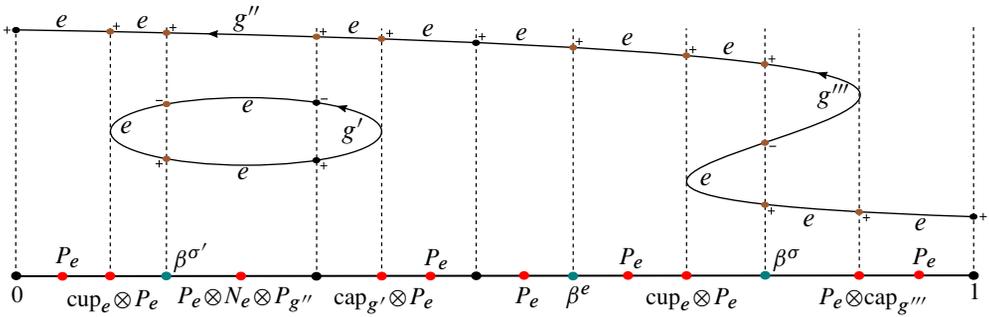


Figure 7: Example of a G -linear diagram without sheet data.

point is given as follows. We assign P_{g_1} for such a sheet with $+$ boundary points and g_1 -label, and assign N_{g_2} for a such a sheet with $-$ boundary points and g_2 -label. Then the label of the added point is given by $A_1 \otimes A_2 \otimes \cdots \otimes A_n$, where A_i is the assigned label of the i^{th} sheet according to the trivialization of the chamber for $i = 1, \dots, n$. Similarly, we modify labels cup and cap according to assignments to sheets which are in the same trivialization with these singularities (see Figure 7). Lastly, on sheet data trivialization of sheets involves labeling each sheet with a group element with an arrow as described above and each (unlabeled) point is lifted to boundary of a sheet. We denote this modified linear diagram using the extra data of the 1-dimensional marked X -manifold with $(\Psi^G, \Gamma^G, \mathcal{S}^G)$ and call it a G -linear diagram.

It is clear that any G -linear diagram $(\Psi^G, \Gamma^G, \mathcal{S}^G)$ produces a pair $((M', T', g'), f')$. For any such pair, by choosing a compatible chambering set, we obtain a new pair. These two pairs are related by the following notion. An X -homeomorphism between (marked) X -manifolds is a pointed orientation-preserving diffeomorphism commuting with the characteristic maps. An X -homeomorphism between such pairs is called over $[0, 1]$ if it commutes with the fixed generic maps.

Proposition 3.5 *Let $(\Psi^G, \Gamma^G, \mathcal{S}^G)$ be a G -linear diagram induced from a pair $((M, T, g), f)$ of a 1-dimensional oriented marked X -manifold and a generic map $f : (M, \partial M) \rightarrow ([0, 1], \{0, 1\})$, and a chambering set Γ for a Ψ -compatible open cover \mathcal{U} of $[0, 1]$. If the pair $((M', T', g'), f')$ is constructed from $(\Psi^G, \Gamma^G, \mathcal{S}^G)$, then there exists an X -homeomorphism $F : M \rightarrow M'$ over $[0, 1]$.*

Proof The diffeomorphism F maps inverse images of chambers to corresponding trivializations. Since corresponding connected components have the same G -labels

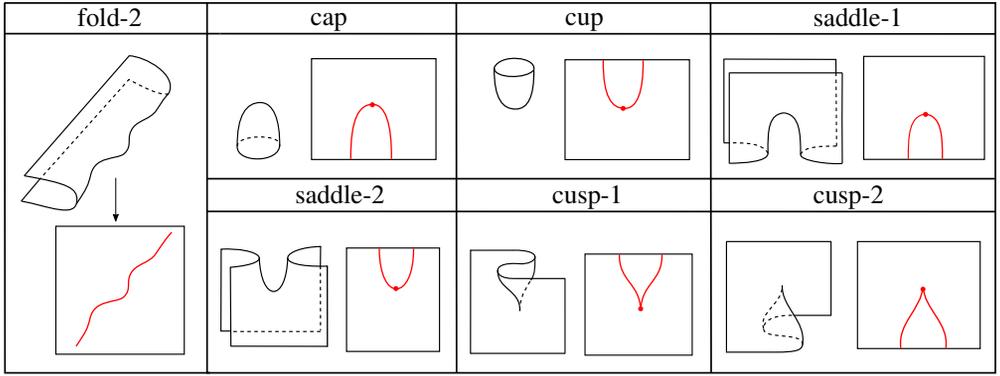


Figure 8: Singularities of Schommer-Pries stratification and their graphics in \mathbb{R}^2 .

and both f and $f' \circ F$ restrict to the same map on $f^{-1}(V)$ for any chamber V , F is an X -homeomorphism over $[0, 1]$. □

3.2 G -planar diagrams

Planar diagrams, introduced by Schommer-Pries [22], represent cobordism type $\langle 2 \rangle$ -surfaces equipped with a generic map to $I^2 = [0, 1] \times [0, 1]$. Here generic maps refer to Schommer-Pries' stratification of jet spaces described below. Parallel to linear diagrams, a planar diagram consists of a graphic of a generic map, an open cover of I^2 , and a combinatorial data describing preimages of a generic map on open sets.

In his classification of 2-dimensional extended TFTs, Schommer-Pries [22] studied maps from cobordism type $\langle 2 \rangle$ -surfaces to I^2 and refined the Thom-Boardman stratification of jet spaces. Figure 8 shows the singularities of Schommer-Pries' stratification in normal coordinates and their graphics in I^2 . Here, by a *graphic* we mean the image of a singularity under a generic map. In this context, by a generic map we mean a map whose jet sections are transversal to each stratum. In Figure 8, generic maps are projections to the page. The numbers on singularity names indicate their indices. By an index of a singularity, we mean a symmetry of either a singularity or its graphic. For example, fold-1 is obtained from fold-2 by changing the folding direction. Similarly, cap, cup, saddle-1, and saddle-2 are different indices of the Morse singularity. Observe that a cusp singularity has four indices.

For a given cobordism type $\langle 2 \rangle$ -surface Σ , a generic map for Schommer-Pries stratification has the form

$$f : (\Sigma, \partial_v \Sigma, \partial_h \Sigma) \rightarrow (I^2, \partial I \times I, I \times \partial I).$$

Transversality theorems (see [22; 8]) imply that the set of generic maps is dense in $C^\infty((\Sigma, \partial_v \Sigma, \partial_h \Sigma), (I^2, \partial I \times I, I \times \partial I))$. The properties of Schommer-Pries' stratification are listed in the following definition. In particular, the graphic of a generic map for this stratification is a 2-dimensional graphic, which is defined as follows:

Definition 3.6 [22, Definition 1.29] A 2-dimensional graphic $\Phi = (\eta, \mu)$ is a diagram in I^2 consisting of a finite number of embedded labeled curves (η) and a finite number of labeled points (μ) satisfying the following conditions:

- (i) Elements of η can only have transversal intersections and no three elements intersect at a point. Each element of η is labeled with either fold-1 or fold-2.
- (ii) Elements of η are disjoint from $\partial I \times I$ and intersect transversely with $I \times \partial I$. Labeling each of these intersection points on $I \times \partial I$ with cup for fold-1 labeled curves and with cap for fold-2 labeled curves produces 1-dimensional graphics on $I \times \{0\}$ and $I \times \{1\}$.
- (iii) Projections of elements of η to the last coordinate of I^2 are local diffeomorphisms.
- (iv) Elements of μ are isolated and disjoint from $\partial(I \times I)$. Each element is labeled with one of cup, cap, saddle-1, saddle-2 or cusp- i for $i = 1, 2, 3, 4$.
- (v) Each element in μ has a neighborhood in which two elements of η form one of cup, cap, saddle-1, saddle-2 or cusp- i graphic for $i = 1, 2, 3, 4$ (see Figure 8).

We want to extend this definition to cobordism type $\langle 2 \rangle$ - X -surfaces so that a 2-dimensional graphic additionally contains the X -manifold data. First we consider $\langle 2 \rangle$ - X -surfaces whose underlying manifolds are singularities of Schommer-Pries' stratification in normal coordinates (see Figure 8). Figure 9 shows their graphics with the X -manifold data. Note that we abbreviate fold- i label to F_i , saddle- i to S_i , and cusp- i to C_i for $i = 1, 2$. Also observe that fold-1 and fold-2 singularities are paths of cup and cap singularities in the previous section. For this reason, henceforth, on any G -linear diagram we replace cup and cap labels with F_1 and F_2 labels, respectively. This implies that the restriction of each diagram in Figure 9 to $I \times \partial I$ yields two partial G -linear diagrams. Later we complete them to G -linear diagrams by adding chambering sets and sheet data.

Compared to graphics of singularities in Figure 8, there are additional arcs connecting graphics of Morse¹ and cusp singularities to the (red) points of the boundary G -linear

¹Similar to the graphics of saddles, one can add arcs to the graphics of cup and cap and label them with \emptyset .

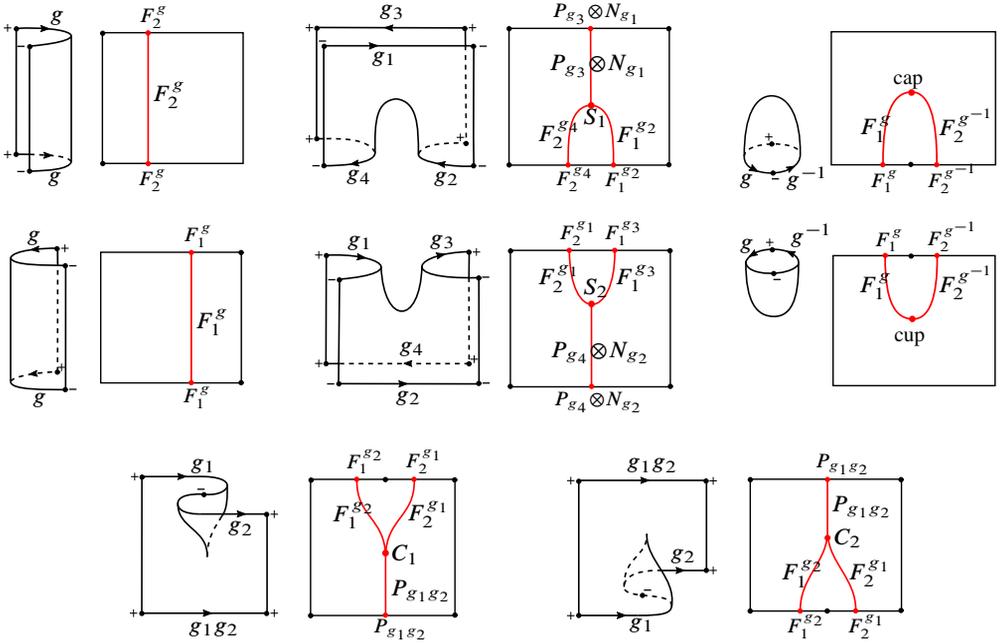


Figure 9: Graphics of singularities with X -manifold data.

diagram. The reason behind the addition of these arcs is the connection between these diagrams and string diagrams, which is the content of [Theorem 4.5](#). We call $\langle 2 \rangle$ - X -surfaces in [Figure 9](#) *elementary $\langle 2 \rangle$ - X -surfaces* since they are building blocks of cobordism type $\langle 2 \rangle$ - X -surfaces under horizontal and vertical gluing operations. However, this list is not complete. The rest of the elementary $\langle 2 \rangle$ - X -surfaces and their graphics are given in [Figure 10](#).

We now know the extra data on 2-dimensional graphics of elementary $\langle 2 \rangle$ - X -surfaces. Using these modified diagrams, for any generic map on a cobordism type $\langle 2 \rangle$ - X -surface (Σ, R, P) we add the X -manifold data (R, P) to the graphic of the generic map in two steps. First we decompose (Σ, R, P) into horizontal and vertical compositions of elementary $\langle 2 \rangle$ - X -surfaces. This is always possible by the nature of Schommer-Pries' stratification and above arguments. Using P we choose G -labels on each elementary $\langle 2 \rangle$ - X -surface in the decomposition. We then consider the modified diagrams of these $\langle 2 \rangle$ - X -surfaces in I^2 , as described above. [Figure 11](#) shows an example of this process where the generic map is projection to the page. For a given 2-dimensional graphic $\Phi = (\eta, \mu)$, we denote the union of η and arcs encoding the X -manifold data by η^G and, similarly, μ^G denotes the union of μ and additional labeled points. We call such a

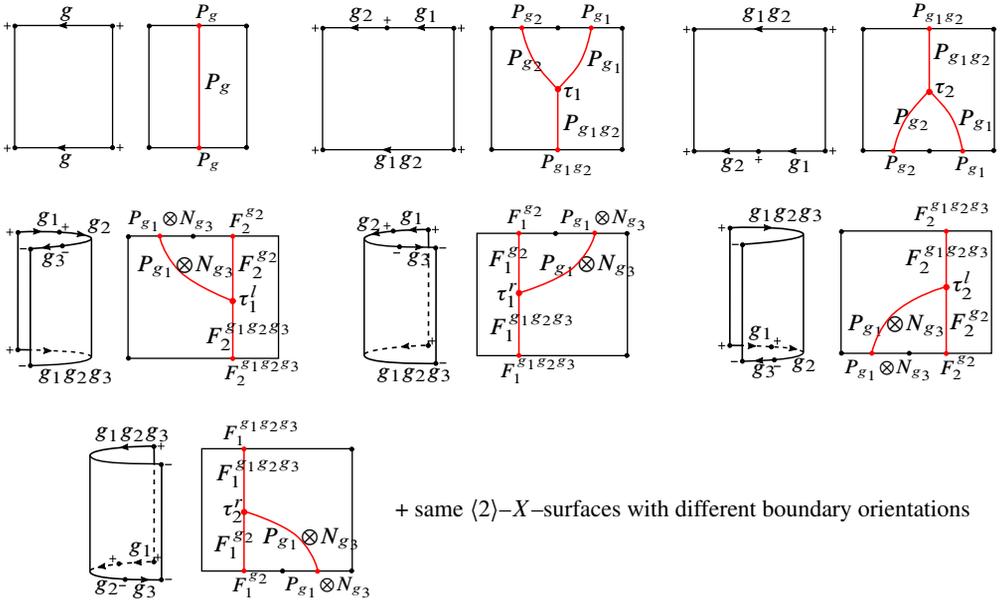


Figure 10: The remaining elementary $\langle 2 \rangle$ - X -surfaces and their graphics.

2-dimensional graphic Φ equipped with X -manifold data a 2-dimensional G -graphic and denote it by $\Phi^G = (\eta^G, \mu^G)$.

Let $\Phi^G = (\eta^G, \mu^G)$ be a 2-dimensional G -graphic; an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ of I^2 with at most triple intersections is said to be Φ -compatible [22] if each triple intersection is disjoint from μ , each double intersection is disjoint from $\eta \cup \mu$ or contains a single element from η , and the open covers $\{U_\alpha \cap (I \times \{i\})\}_{\alpha \in J}$ of $I \times \{i\}$ for $i = 0, 1$ are

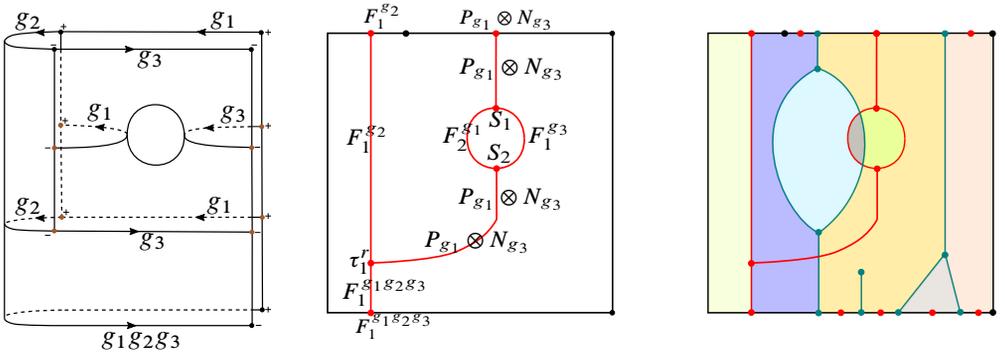


Figure 11: Adding X -manifold data to a graphic and an example of a chambering graph.

compatible with corresponding 1-dimensional graphics obtained from Φ^G . Knowing that \mathbb{R}^2 has covering dimension two and sets η and μ are finite, it is not hard to find Φ -compatible open covers for a given graphic Φ .

Definition 3.7 [22, Definition 1.42] Let $\Phi^G = (\eta^G, \mu^G)$ be a 2-dimensional G -graphic. A *chambering graph* Γ for Φ^G is a smoothly embedded graph in I^2 satisfying the following conditions. Vertices of Γ are disjoint from elements of Φ^G and have degree either one or three. Edges of Γ are disjoint from $\partial I \times I$ and transverse to Φ^G and $I \times \partial I$. Furthermore, projection of each edge to the last coordinate is a local diffeomorphism and around each trivalent vertex one of the edges projects to the opposite side of the projection of the other two edges with respect to the image of the vertex.

Definition 3.8 Let Γ be a chambering graph for $\Phi^G = (\eta^G, \mu^G)$. *Chambers* of Γ are the connected components of $I^2 \setminus (\Gamma \cup \eta \cup \mu)$. A chambering graph Γ is said to be subordinate to an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ of I^2 if each chamber is a subset of at least one U_α with $\alpha \in J$ and the chambering sets $\Gamma \cap (I \times \{i\})$ are compatible with the restricted open covers $\{U_\alpha \cap (I \times \{i\})\}_{\alpha \in J}$ for $i = 0, 1$.

Example 3.9 Figure 11 shows an example of a chambering graph Γ where each colored region is a chamber. Note that the chambering graph leads to new red points on $I \times \partial I$ forming two partial G -linear graphs. In this example all new points are labeled with $P_e \otimes N_e$.

Proposition 3.10 Let Φ^G be a 2-dimensional G -graphic in I^2 and let $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ be a Φ -compatible open cover of I^2 . Then there exists a chambering graph Γ for Φ^G subordinate to \mathcal{U} .

The 2-dimensional graphic version of this proposition was proven in Proposition 1.46 of [22]. This version follows from that only using transversality arguments. From now on we assume that all chambering graphs are subordinate to some compatible open cover.

Next, we recall sheet data associated to a pair (Φ, Γ) . We know sheet data on the components of $I \times \partial I$ from the previous section. For the other boundary component $\partial I \times I$, sheet data is similar and indeed simpler since vertical boundary components are all identical. Therefore, we consider the open subsets of chambers by removing

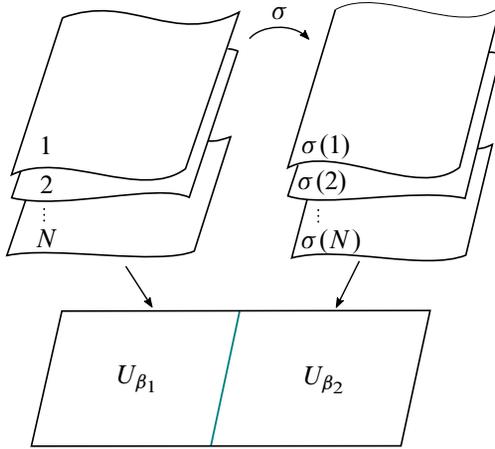


Figure 12

boundary components. That is, for any chamber U_β which intersects with $\partial(I \times I)$ we consider $U'_\beta = U_\beta - (U_\beta \cap \partial(I \times I))$. Since f is generic, the preimage $f^{-1}(U'_\beta)$ consists of a disjoint union of open sets (possibly empty) each mapping diffeomorphically onto U'_β . A *trivialization* of U'_β is an identification of $f^{-1}(U'_\beta)$ with $\mathbb{N}_{\leq N} \times U'_\beta$ for some $N \in \mathbb{N}$ if $f^{-1}(U'_\beta)$ is nonempty and an identification with the empty set otherwise. In this case, each $\{i\} \times U'_\beta$ is called a *sheet* and each sheet is oriented. By requiring the same trivializations on U'_β and ∂U_β , we extend these identifications to $\mathbb{N}_{\leq N} \times U_\beta$.

Similar to the 1-dimensional case, trivializations of two neighboring chambers have the same number of sheets if chambers are separated by an edge of Γ (see Figure 12). If an element in η separates chambers then the number of sheets differ by two because it is a fold graphic (see Figure 8). *Sheet data* \mathcal{S} [22] for a pair (Φ, Γ) consists of a

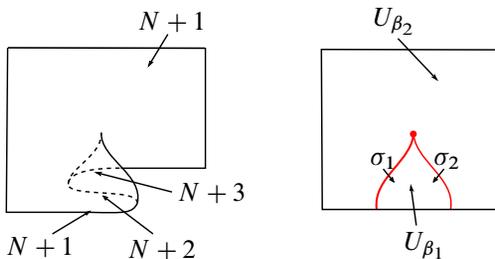


Figure 13

trivialization of each chamber and an injection or a permutation between trivializations of neighboring chambers preserving orientations and describing how sheets are glued (see Figure 12). The gluing description of sheets requires certain conditions on permutations and injections. For example, if three chambers are separated by edges of a trivalent vertex of Γ , then the circular composition of permutations must be the identity. Also, the permutation corresponding to the edge of a univalent vertex must be the trivial permutation. The only nontrivial sheet data is that of a cusp graphic, which we briefly describe. Consider the cusp-2 labeled point in Figure 13. Let $\mathbb{N}_{\leq N+3} \times U_{\beta_1}$ and $\mathbb{N}_{\leq N+1} \times U_{\beta_2}$ be the trivializations such that the sheets $\bigcup_{i=N+1}^{N+3} i \times U_{\beta_1}$ and $(N+1) \times U_{\beta_2}$ belong to a cusp singularity as shown in Figure 13. In this case, restriction of injections to the cusp singularity gives $\sigma_1(N+1) = N+1$ and $\sigma_2(N+1) = N+3$.

Definition 3.11 [22, Definition 1.48] A *planar diagram* is a triple $(\Phi, \Gamma, \mathcal{S})$ consisting of a 2-dimensional graphic Φ , a chambering graph Γ for Φ subordinate to a Φ -compatible open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ of I^2 , and a sheet data \mathcal{S} associated to the pair (Φ, Γ) .

Any planar diagram $(\Phi, \Gamma, \mathcal{S})$ produces a cobordism type $\langle 2 \rangle$ -surface Σ with a generic map $f: (\Sigma, \partial_v \Sigma, \partial_h \Sigma) \rightarrow (I^2, \partial I \times I, I \times \partial I)$. In the case of a tuple (Φ^G, Γ) , the associated sheet data can be improved to produce a $\langle 2 \rangle$ - X -surface (Σ, R, P) . We call such sheet data carrying X -manifold data to sheets *G-sheet data* and denote it by \mathcal{S}^G . Then, generalizing a planar diagram, we define a *G-planar diagram* as a triple $(\Phi^G, \Gamma, \mathcal{S}^G)$. As an extension of G -linear diagrams, we label edges of a chambering graph with β^σ , where σ is the permutation coming from sheet data. If both sheets are trivialized by the empty set, then the separating edge is labeled with $\beta^!$. We also label vertices of chambering graph as follows. For a fixed trivalent vertex, if two of the edges direct upward, the vertex is labeled with $X^{\sigma, \sigma'}$, and if two of the edges direct downward then it is labeled with $(X^{\sigma, \sigma'})^{-1}$. Here σ and σ' are the permutations corresponding to the sheet data of these edges. A univalent vertex is labeled with X^e if its edge directs upward and it is labeled with $(X^e)^{-1}$ if its edge directs downward. We also label intersections of edges of the chambering graph and 2-dimensional G -graphic. For such an intersection, if an edge of the chambering graph is labeled with β^σ and an arc of the 2-dimensional G -graphic is labeled with A , then the intersection is labeled with β_A^σ .

Example 3.12 Figure 14 shows an example of a cobordism type $\langle 2 \rangle$ - X -surface (Σ, R, P) and its G -planar diagram with respect to the projection map. We denote the

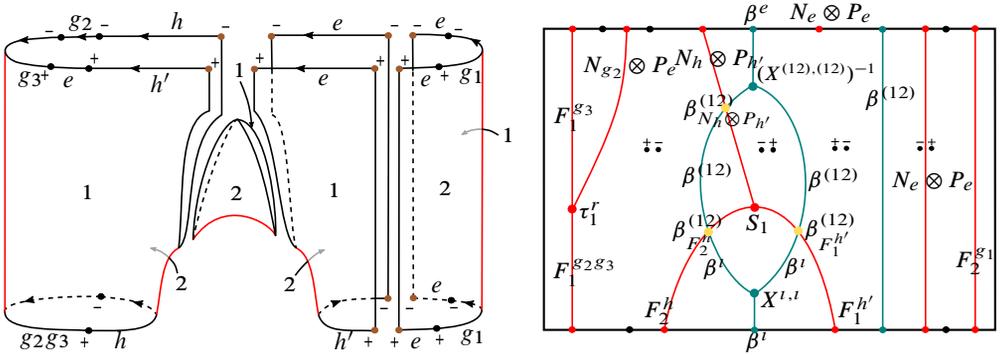


Figure 14: Example of a cobordism type $\langle 2 \rangle$ - X -surface and its G -planar diagram.

trivializations of sheets with numbers on Σ . Correspondingly, we encode this data on the G -planar diagram by labeling chambers with ordered signed points. Here signs come from the sign of the points on the corner of the corresponding sheet.

Let (Σ, R, P) and (Σ', R', P') be $\langle 2 \rangle$ - X -surfaces endowed with generic maps f and f' , respectively. An X -homeomorphism $F : (\Sigma, R, P) \rightarrow (\Sigma', R', P')$ is said to be *over I^2* if it commutes with the fixed generic maps, ie $f' \circ F = f$.

Proposition 3.13 *Let $f : (\Sigma, \partial_v \Sigma, \partial_h \Sigma) \rightarrow (I^2, \partial I \times I, I \times \partial I)$ be a generic map on a cobordism type $\langle 2 \rangle$ - X -surface (Σ, R, P) inducing a 2-dimensional graphic Φ^G . Let Γ be a chambering graph for Φ^G subordinate to a Φ -compatible open cover giving a G -planar diagram $(\Phi^G, \Gamma, \mathcal{S}^G)$. If the pair $((\Sigma', R', P'), f')$ is constructed from $(\Phi^G, \Gamma, \mathcal{S}^G)$, then there exists an X -homeomorphism $F : \Sigma \rightarrow \Sigma'$ over I^2 .*

Proof The diffeomorphism $F : \Sigma \rightarrow \Sigma'$ maps inverse images of chambers to the corresponding trivializations. Since $F(R) = R'$, $[P' \circ F] = P$, and both f and $f' \circ F$ restrict to the same map on $f^{-1}(U_\beta)$ for any chamber U_β , F is an X -homeomorphism over I^2 . □

3.3 G -spatial diagrams

Schommer-Pries [22] introduced spatial diagrams to identify planar diagrams which produce diffeomorphic cobordism type $\langle 2 \rangle$ -surfaces. We extend them to G -spatial diagrams which identify those G -planar diagrams producing X -homeomorphic cobordism type $\langle 2 \rangle$ - X -surfaces. Then, using these identifications, we define an equivalence relation among G -planar diagrams, and prove the G -planar decomposition theorem.

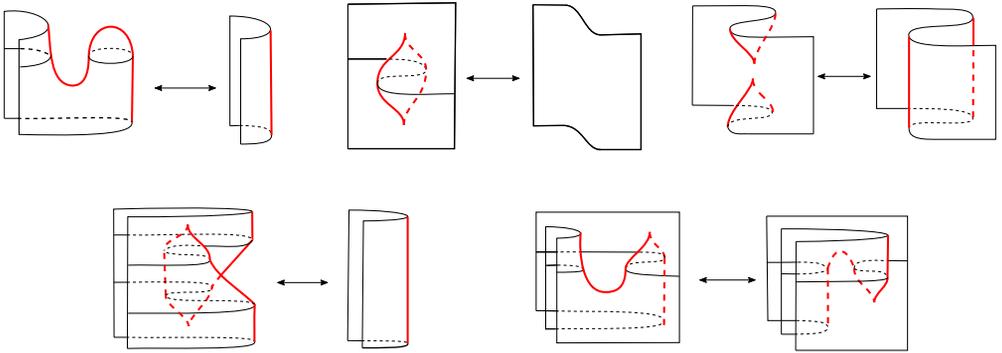


Figure 15: Movie moves coming from codimension 3 singularities.

Different generic maps on a fixed cobordism type $\langle 2 \rangle$ -surface yield different graphics just as different Morse functions yield different critical values. In the latter case, Cerf theory relates different sets of critical values in terms of isotopies or birth and death of critical values. Similarly, Schommer-Pries [22] related different graphics obtained from different generic maps in terms of isotopies and certain local moves of graphics, called *movie moves* (see Figure 15). These movie moves are obtained from the singularities of certain stratification of jet spaces

$$J^r((\Sigma \times I, \partial_h \Sigma \times I, \partial_v \Sigma \times I), (I^2 \times I, I \times \partial I \times I, \partial I \times I \times I))$$

for a cobordism type $\langle 2 \rangle$ -surface Σ . Note that in general the map F is not of the form $F(x, t) = (f_t(x), t)$. For our purposes, we consider the subspace of $J^r(\Sigma \times I, I^2 \times I)$ consisting of paths of functions.

Figure 16 shows the graphics of singularities for the Schommer-Pries stratification in normal coordinates. Observe the relation between movie moves in Figure 15 and the horizontal boundary components of the new graphics. The remaining movie moves coming from this stratification are shown in Figure 17. The properties of this stratification are listed in the following definition. In particular, the graphic of a generic map for this stratification is a 3-dimensional graphic, which is defined as follows:

Definition 3.14 [22, Definition 1.30] A 3-dimensional graphic $\Delta = (\delta, \eta, \mu)$ is a diagram in $I^2 \times I$ consisting of a finite number of embedded compact labeled surfaces (δ), a finite number of embedded labeled curves (η), and a finite number of embedded labeled points (μ) satisfying the following conditions:

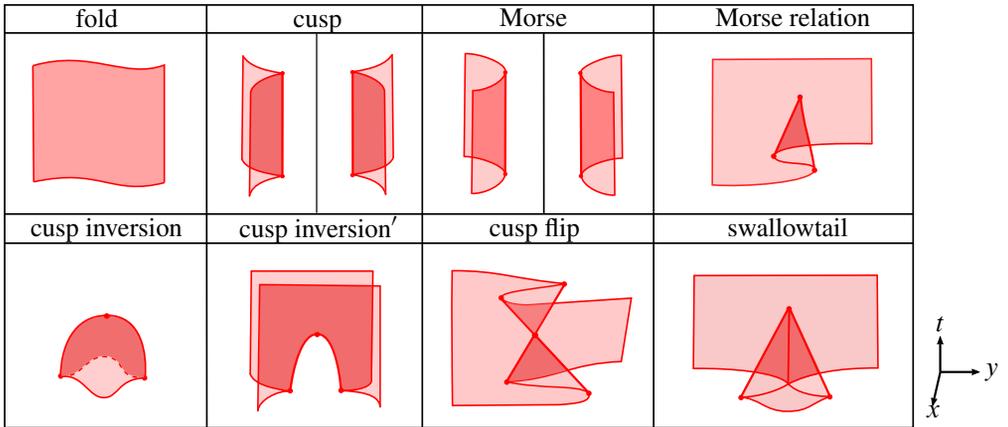


Figure 16: Graphics of the singularities in $I^2 \times I$.

- (i) Projections of elements of δ to the last two coordinates are local diffeomorphisms. Elements of δ intersect with $I \times \{i\} \times I$ along vertical line segments $\{(x, i, s)\}_{s \in [0,1]}$ for some $x \in (0, 1)$ and for $i = 0, 1$. Each element of δ is labeled with either fold-1 or fold-2.
- (ii) Projections of elements of η to the last coordinate are local diffeomorphisms. Every element of η has a neighborhood in which two elements of δ form either Morse or cusp graphics (see Figure 16). Each element of η is labeled with either Morse- i or cusp- i , where $i = 1, 2, 3, 4$ indicates the indices.²
- (iii) Each element of μ has a neighborhood in which some elements of δ and η form one of the graphics Morse relation- i , cusp inversion- j , cusp inversion'- j , cusp flip- j , or swallowtail- i , where $i = 1, 2, 3, 4$ and $j = 1, 2$ indicate graphics of different indices.
- (iv) Elements of μ are labeled with one of the singularities Morse relation- i , cusp inversion- j , cusp inversion'- j , cusp flip- j or swallowtail- j , where $1 \leq i \leq 8$ and $j = 1, 2, 3, 4$ indicate the indices.
- (v) The restriction of the graphic to the components of $I^2 \times \partial I$ gives 2-dimensional graphics.
- (vi) Elements of δ , η , and μ are disjoint from $\partial I \times I^2$. They are transversal with respect to each other and to $I^2 \times \partial I$. Moreover, when two surfaces intersect

²Morse singularities are paths of cap, cup, saddle-1 and saddle-2 singularities.

along an arc, there can only be finitely many points on the arc with tangent space lying in $\langle \partial_x, \partial_y \rangle$, where (x, y, t) is the coordinate for $I^2 \times I$.

Let Σ be a cobordism type $\langle 2 \rangle$ -surface and

$$F : (\Sigma \times I, \partial_h \Sigma \times I, \partial_v \Sigma \times I) \rightarrow (I^2 \times I, I \times \partial I \times I, \partial I \times I \times I)$$

be a generic map. We know that the restriction of F to the boundary components $\Sigma \times \{0\}$ and $\Sigma \times \{1\}$ gives two generic maps on Σ . The converse is also true. That is, for any given two generic maps $f_1, f_2 : (\Sigma, \partial_v \Sigma, \partial_h \Sigma) \rightarrow (I^2, \partial I \times I, I \times \partial I)$, there exists a generic map F on $\Sigma \times I$ with $F|_{\Sigma \times \{0\}} = f_1$ and $F|_{\Sigma \times \{1\}} = f_2$. Therefore, 3-dimensional graphics are designed to identify 2-dimensional graphics obtained from generic maps on $\langle 2 \rangle$ -surfaces which are diffeomorphic relative to boundary. Using 2-dimensional G -graphics, we extend this result to $\langle 2 \rangle$ - X -surfaces which are X -homeomorphic relative to their boundary. This is useful since, in the cobordism bicategory $X\text{Bord}_2$, the 2-morphisms are X -homeomorphism classes of cobordism type $\langle 2 \rangle$ - X -surfaces relative to their boundary.

Movie moves are local relations on 2-dimensional graphics generating the identification of 2-dimensional graphics induced from different generic maps. Figures 18 and 19 show some of the generalized movie moves relating 2-dimensional G -graphics. The remaining generalized movie moves involve singularities with different indices, orientations, and decomposition into elementary $\langle 2 \rangle$ - X -surfaces. Similarly, movie moves given in Figure 17 are generalized and their possible versions (different indices, orientations

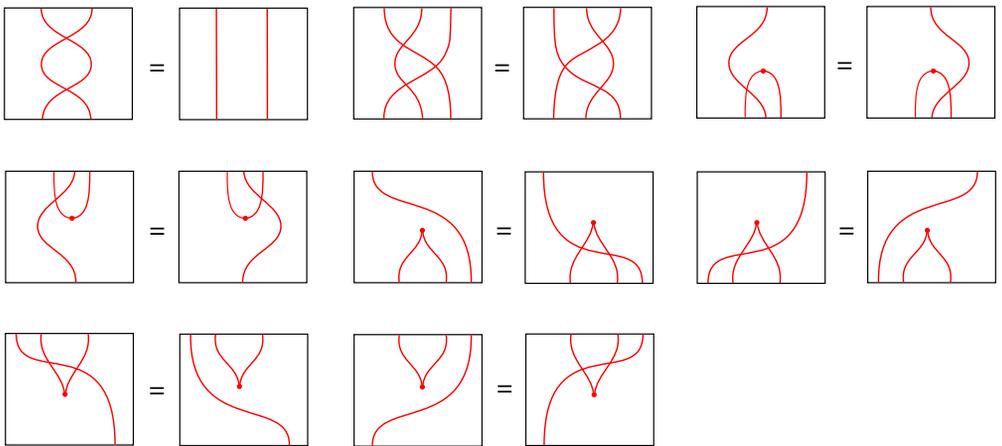


Figure 17: Movie moves coming from intersection of codimension 1 and 2 singularities.

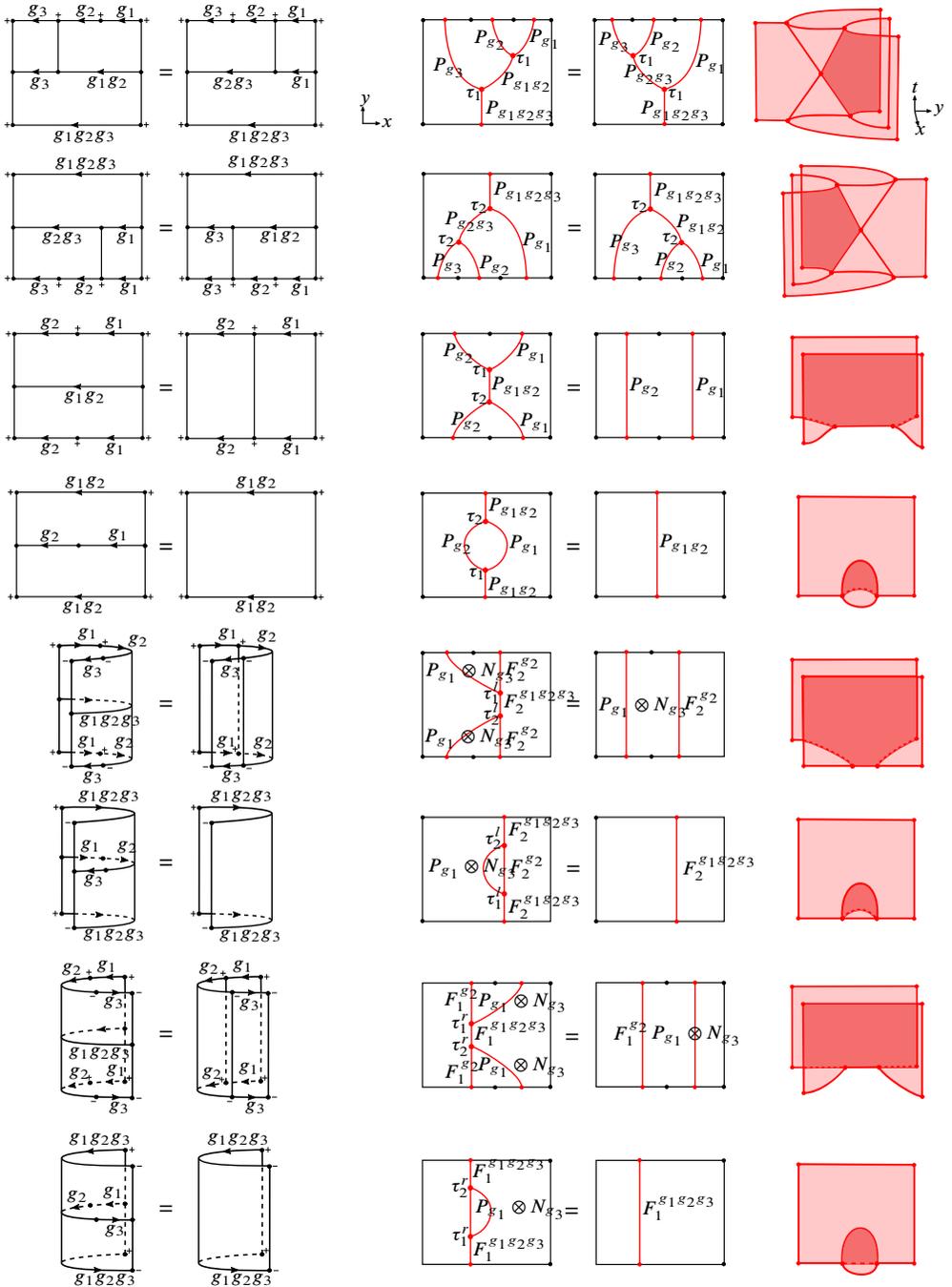


Figure 19: Some of the generalized movie moves obtained by gluing elementary (2)-X-surfaces.

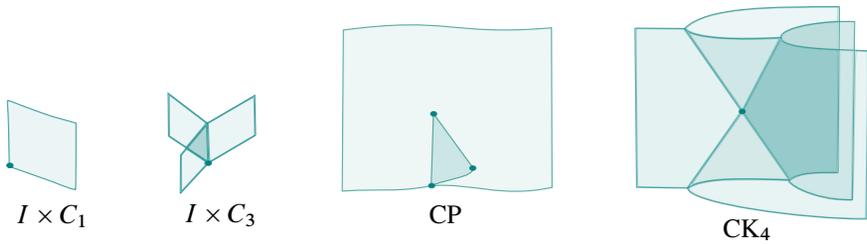


Figure 20: Local models for a chambering foam.

and decompositions) form generalized movie moves. From now on, whenever we refer to these figures we mean the complete list of movie moves.

Similar to the previous two sections, we generalize a 3–dimensional graphic to a G –graphic by adding (labeled) surfaces, arcs, and points. For every new movie move, the corresponding graphic is shown in Figure 19, right. Then we define a 3–dimensional G –graphic $\Delta^G = (\delta^G, \eta^G, \mu^G)$ as a 3–dimensional graphic $\Delta = (\delta, \eta, \mu)$ along with a 2–dimensional locally conical stratified space of compact type which is transverse to the graphic and whose local models are given in Figure 19. The notion of locally conical stratified space is also the main ingredient of Definition 3.15 below. Its definition can be found in [22, Definition 1.41]. The reader can think of a locally conical stratified space as a space constructed from given local models just as a manifold is built locally from disks.

Let $\Delta^G = (\delta^G, \eta^G, \mu^G)$ be a 3–dimensional G –graphic. An open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ of I^3 with at most 4–fold intersections is said to be Δ –compatible [22] if each 4–fold intersection is disjoint from $\delta \cup \eta \cup \mu$, each 3–fold intersection is disjoint from $\mu \cup \eta$ and contains at most a single component of surfaces in δ , each double intersection is disjoint from points in μ , and the open covers $\{U_\alpha \cap (I^2 \times \{i\})\}_{\alpha \in J}$ of $I^2 \times \{i\}$ for $i = 0, 1$ are compatible with the corresponding 2–dimensional graphics obtained from Δ^G . Since I^3 has covering dimension 3 and there are only finitely many elements in δ, η and μ , Δ –compatible open covers exist.

Definition 3.15 [22, Definition 1.43] Let $\Delta^G = (\delta^G, \eta^G, \mu^G)$ be a 3–dimensional G –graphic. A chambering foam Γ for Δ^G is a smooth embedding of a 2–dimensional locally conical stratified space Γ of compact type into $I \times (0, 1) \times I$ with the following properties. The space Γ is locally conical with respect to the system of local models $I^2, I \times C_1, I \times C_3, CP$, and CK_4 shown in Figure 20. Vertices are disjoint from Δ^G

and lie in the interior. Edges can only intersect with a surface from δ^G . Faces can only intersect with surfaces from δ^G and arcs from η^G . All intersections are transversal and Γ additionally satisfies the following conditions:

- (I) The projection $p: \Gamma \rightarrow I \times I$ to the last two coordinates has no singularity and projection of faces to the last coordinate has no singularity.
- (II) For every $t \in I$ satisfying that $(I^2 \times \{t\}) \cap \mu^G = \emptyset$, t is not a critical value of projection $\text{pr}: \Gamma \rightarrow I$ to the last coordinate, and $(I^2 \times \{t\}) \cap \Gamma$ does not include a vertex of Γ , the graph $(I^2 \times \{t\}) \cap \Gamma$ forms a chambering graph for the 2-dimensional G -graphic $\Delta^G \cap (I^2 \times \{t\})$.
- (III) The projection of each one of four edges in the CK_4 -model connecting at the cone point to the last coordinate is a local diffeomorphism. Additionally, at least one of them must map to downward of the cone point and at least one of them must map to upward of the cone point.
- (IV) The projection of the two edges in the CP -model connecting at the cone point to the last coordinate maps both edges to the same direction with respect to the image of the cone point.

Definition 3.16 Let $\Delta^G = (\delta^G, \eta^G, \mu^G)$ be a 3-dimensional G -graphic and let Γ be a chambering foam for Δ^G . Chambers of Γ are the connected components of $I^2 \times I \setminus (\Gamma \cup \delta \cup \eta \cup \mu)$. A chambering foam Γ is said to be *subordinate to an open cover* $\mathcal{O} = \{O_\alpha\}_{\alpha \in J}$ of $I^2 \times I$ if each chamber is a subset of at least one O_α with $\alpha \in J$ and the chambering graphs $\Gamma \cap (I^2 \times \{i\})$ are compatible with the restricted open cover $\{O_\alpha \cap I^2 \times \{i\}\}_{\alpha \in J}$ for $i = 0, 1$.

Lemma 3.17 Let Γ be a chambering foam for a 3-dimensional G -graphic Δ^G inducing 2-dimensional G -graphics and chambering graphs (Φ_0^G, Γ_0) and (Φ_1^G, Γ_1) on $I^2 \times \{0\}$ and $I^2 \times \{1\}$, respectively. Let $\mathcal{O} = \{O_\alpha\}_{\alpha \in J}$ be a Δ -compatible open cover of I^3 such that Γ_i is subordinate to $\mathcal{O}_i = \mathcal{O}|_{I^2 \times \{i\}}$ for $i = 0, 1$. Then there exists a chambering foam Γ' for Δ^G subordinate to \mathcal{O} whose restrictions to $I^2 \times \{0\}$ and $I^2 \times \{1\}$ yield Γ_0 and Γ_1 , respectively.

In [22], the corresponding statement for a 3-dimensional graphic was proven (see [22, Corollary 1.47]). In the case of nontransversal intersections with new elements encoding X -manifold data, Γ can be slightly modified to make all intersections transversal while being compatible with \mathcal{O} .

Just as the movie moves in Figures 18 and 19 generate relations locally between two G -planar diagrams on the boundary of a G -spatial diagram, there are movie moves describing local relations between two chambering graphs on the boundary of a compatible chambering foam. These moves along with corresponding chambering foams are shown in Figure 21. Moreover, there are movie moves coming from an intersection of a 3-dimensional G -graphic with a chambering foam. These local relations are shown in Figure 22, in which the labels are omitted.

Next, we describe the sheet data. Let $\Delta = (\delta, \eta, \mu)$ be a 3-dimensional graphic induced from a generic map $F: (\Sigma \times I, \partial_h \Sigma \times I, \partial_v \Sigma \times I) \rightarrow (I^2 \times I, I \times \partial I \times I, \partial I \times I \times I)$, where Σ is a cobordism type $\langle 2 \rangle$ -surface. Let Γ be a chambering foam subordinate to a Δ -compatible open cover. Similar to the previous section, sheet data associated to (Δ, Γ) extends the sheet data of planar diagrams on faces $I^2 \times \{0, 1\}$. Since F is generic, the preimage $F^{-1}(O_\beta)$ of an open chamber consists of a disjoint union of open sets, each mapping diffeomorphically onto O_β . If a chamber O_β is not open then we consider $O'_\beta = O_\beta \setminus (O_\beta \cap \partial(I^2 \times I))$. Then, a *trivialization* of a chamber is the identification of $F^{-1}(O_\beta)$ with $\mathbb{N}_{\leq N} \times O_\beta$ if $F^{-1}(O_\beta)$ is nonempty and an identification with the empty set otherwise. Each $\{i\} \times O_\beta$ is called a *sheet* and each sheet is oriented. For every chamber O_β which is not open, we extend identifications to $\mathbb{N}_{\leq N} \times O_\beta$ by requiring the same trivializations on O'_β and $O_\beta \cap \partial(I^2 \times I)$ coming from sheet data of planar diagrams.

Trivializations of two neighboring chambers have the same number of sheets if chambers are separated by a 2-dimensional stratum of Γ . If an element in δ separates chambers, then the number of sheets differs by two. Sheet data \mathcal{S} [22] for a pair (Δ, Γ) consists of a trivialization of each chamber and an injection or a permutation between trivializations of neighboring chambers preserving orientations and describing how sheets are glued.

The gluing description of sheets requires the following conditions on permutations and injections. In the local models $I \times C_3$, CP , and CK_4 , circular compositions of three or four permutations must be the identity. Since fold, cusp, and Morse graphics are paths of the corresponding graphics in the previous section, their sheet data do not change. According to properties of multijet stratification, transversal double and triple fold intersections are possible. There are four chambers for the double and eight for the triple fold intersection. In both cases, different compositions of injections starting from the chamber with the least number of sheets and ending at the chamber with the maximum number of sheets must be the same. The sheet data for the intersection of fold and Morse graphics is the same as double fold intersection and the sheet data

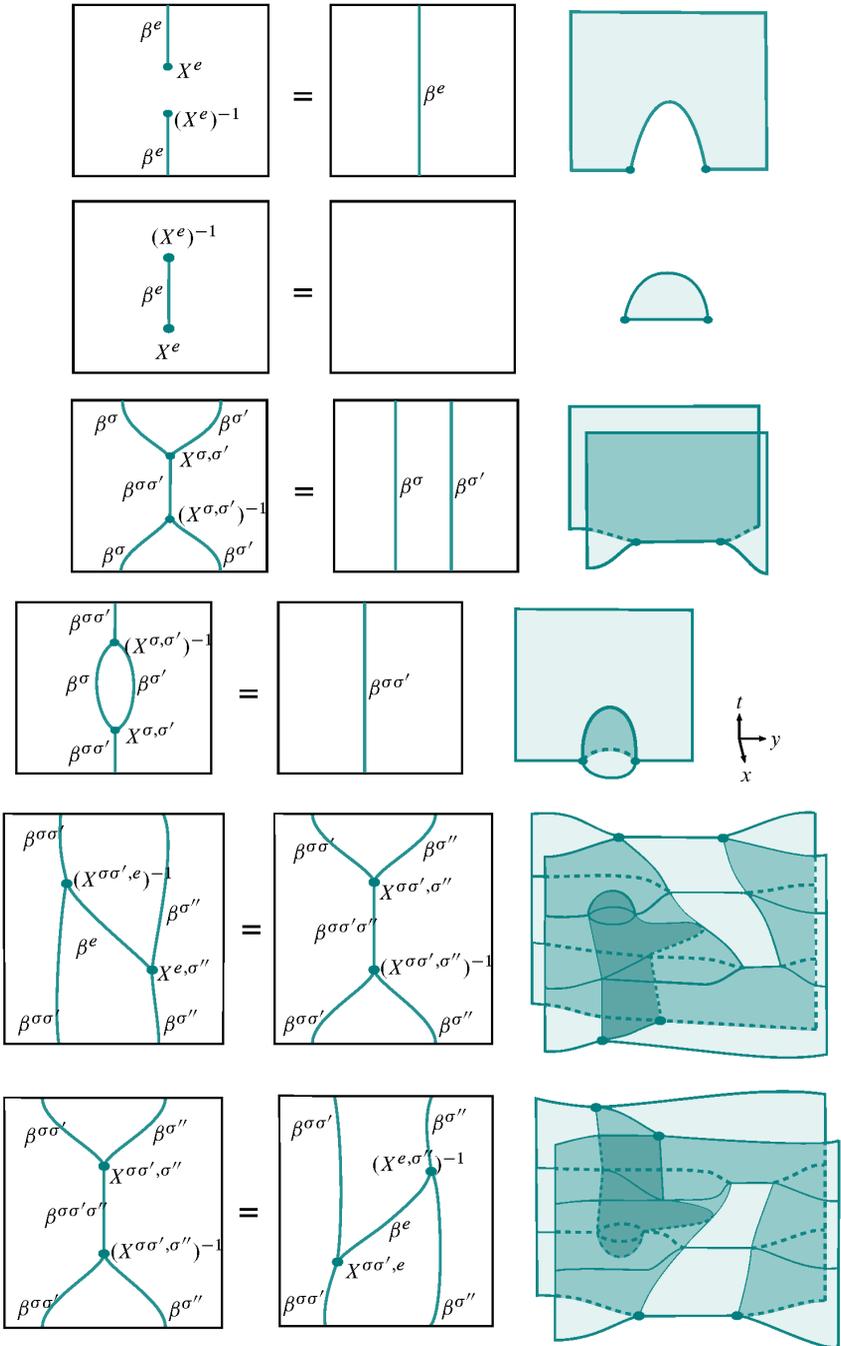


Figure 21: Movie moves for chambering graphs and the corresponding chambering foams.

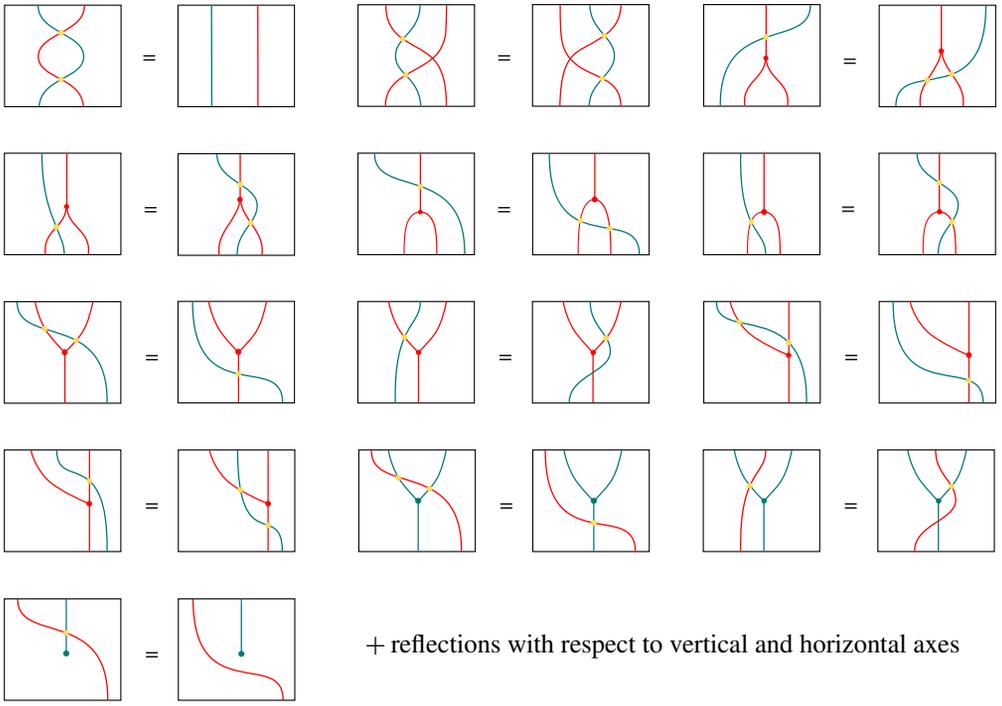


Figure 22: Generating relations (\mathcal{XR}) from movie moves of graphics and chambering graphs.

for intersection of fold and cusp graphics follows from the sheet data of cusp graphic. Sheet data of Morse relation, cusp inversion, cusp inversion', and cusp flip graphics can be interpreted from the corresponding movie moves (see Figure 15). For the details of these sheet data, see [22, Section 1.5.2].

We briefly describe the sheet data of swallowtail-1 graphic shown in Figure 23, where two (blue and green) out of three fold singularities form a double fold crossing. Let $\mathbb{N}_{\leq N} \times U_{\beta_1}$, $\mathbb{N}_{\leq N+2} \times U_{\beta_2}$, and $\mathbb{N}_{\leq N+4} \times U_{\beta_3}$ be trivializations of chambers such that the sheets $\bigcup_{i=N+1}^{N+2} i \times U_{\beta_2}$ and $\bigcup_{j=N+1}^{N+4} j \times U_{\beta_3}$ belong to a swallowtail singularity as shown in Figure 23. Using the sheet data for cusp singularities, restrictions of injections to these sheets give

$$\begin{aligned} \sigma_2(N + 1) &= N + 3, & \sigma_2(N + 2) &= N + 4, \\ \sigma_3(N + 1) &= N + 1, & \sigma_3(N + 2) &= N + 2, \\ \sigma_5(N + 1) &= N + 1, & \sigma_5(N + 2) &= N + 4. \end{aligned}$$

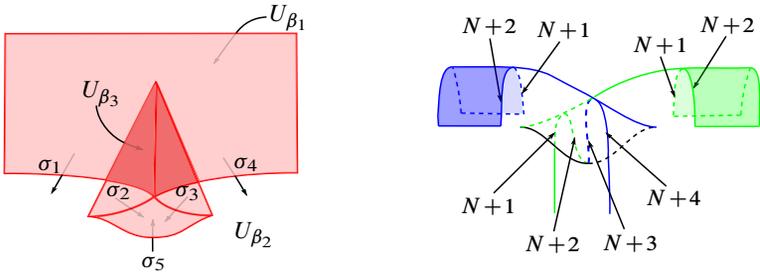


Figure 23: The swallowtail-1 sheet data.

Definition 3.18 [22, Definition 1.49] A spatial diagram is a triple $(\Delta, \Gamma, \mathcal{S})$ consisting of a 3-dimensional graphic Δ , a chambering foam Γ for Δ subordinate to a Δ -compatible cover $\mathcal{O} = \{O_\alpha\}_{\alpha \in J}$ of I^3 , and a sheet data \mathcal{S} associated to the pair (Δ, Γ) .

Any spatial diagram $(\Delta, \Gamma, \mathcal{S})$ produces a compact 3-dimensional manifold with corners M with $\partial M = \Sigma_1 \sqcup \bar{\Sigma}_2$, where Σ_1 and Σ_2 are $\langle 2 \rangle$ -surfaces. Similar to the previous two sections, in the case of a tuple (Δ^G, Γ) , the associated sheet data can be improved to yield a relative homotopy class from M to X and $\langle 2 \rangle$ - X -surfaces (Σ_1, R_1, P_1) and (Σ_2, R_2, P_2) . We call such sheet data carrying X -manifold data to sheets G -sheet data and denote it with \mathcal{S}^G . Then, generalizing a spatial diagram, we define a G -spatial diagram as a triple $(\Delta^G, \Gamma, \mathcal{S}^G)$.

Proposition 3.19 Let $(\Phi_1^G, \Gamma_1, \mathcal{S}_1^G)$ and $(\Phi_2^G, \Gamma_2, \mathcal{S}_2^G)$ be G -planar diagrams and let (Σ_1, R_1, P_1) and (Σ_2, R_2, P_2) be the constructed cobordism type $\langle 2 \rangle$ - X -surfaces, respectively. Then (Σ_1, R_1, P_1) is X -homeomorphic to (Σ_2, R_2, P_2) relative to boundary if and only if there exists a G -spatial diagram $(\Delta^G, \Gamma, \mathcal{S}^G)$ which restricts to $(\Phi_1^G, \Gamma_1, \mathcal{S}_1^G)$ and $(\Phi_2^G, \Gamma_2, \mathcal{S}_2^G)$ on the components of $I^2 \times \partial I$.

Proof (\implies) For a given such X -homeomorphism \mathcal{F} , we obtain a G -spatial diagram by first taking a generic map on the mapping cylinder of \mathcal{F} and then choosing a compatible chambering foam.

(\impliedby) The properties of the stratification imply that the boundary components of the manifold constructed from the G -spatial diagram are diffeomorphic relative to boundary and the compatibility of G -labels of the arcs on the constructed manifold implies that they are X -homeomorphic. □

We define a relation among G -planar diagrams by $(\Phi_1^G, \Gamma_1, \mathcal{S}_1^G) \sim (\Phi_2^G, \Gamma_2, \mathcal{S}_2^G)$ if there exists a G -spatial diagram $(\Delta^G, \Gamma, \mathcal{S}^G)$ restricting to the given G -planar diagrams on the components of $I^2 \times \partial I$. It is not hard to see that \sim is an equivalence relation. Since generalized movie moves provide (nontrivial) local relations on G -planar diagrams, the equivalence relation \sim can be described using these moves as follows:

Proposition 3.20 *Two G -planar diagrams are equivalent if and only if they can be related by a finite sequence of isotopies or movie moves in Figures 17, 18, 19, 21 and 22.³*

Proposition 3.19 implies the following theorem, which is the first main step towards the classification of 2-dimensional extended X -HFTs:

Theorem 3.21 (G -planar decomposition theorem) *The relative X -homeomorphism classes of cobordism type (2) - X -surfaces are in bijection with the equivalence classes of G -planar diagrams.*

4 The classification of 2-dimensional extended X -HFTs

4.1 New bicategories arising from diagrams

In this section, we use the G -planar decomposition theorem to introduce symmetric monoidal X -cobordism bicategories with diagrams.

Definition 4.1 An object of an X -cobordism bicategory with diagrams $X\text{Bord}_2^{\text{PD}}$ is a triple $((M, \widehat{M}_1, \widehat{M}_2, \widehat{g}_2), \overline{M}, \omega)$, where $(M, \widehat{M}_1, \widehat{M}_2, \widehat{g}_2)$ is an object of $X\text{Bord}_2$, \overline{M} is a finite set of ordered oriented points, and $\omega: M \rightarrow \overline{M}$ is an orientation-preserving bijection.

A 1-morphism is a triple $((A, \widehat{A}_0, \widehat{A}_1, T, \widehat{p}_1), L, \nu)$, where $(A, \widehat{A}_0, \widehat{A}_1, T, \widehat{p}_1)$ is an X -haloed 1-cobordism, $L = (\Psi^G, \Gamma, \mathcal{S}^G)$ is a G -linear diagram, and $\nu: (A, T, p) \rightarrow (\overline{A}, \overline{T}, \overline{p})$ is an X -homeomorphism over I with $\nu(T) = \overline{T}$, where $(\overline{A}, \overline{T}, \overline{p})$ is the pointed 1-cobordism constructed from the G -linear diagram. The composition of two composable triples is componentwise: the composition of 1-morphisms in $X\text{Bord}_2$, the composition of diagrams described below, and the extension of two X -homeomorphisms over I , respectively.

A 2-morphism is a triple $([(S, \widehat{S}, R, \widehat{F})], P, \kappa)$, where $[(S, \widehat{S}, R, \widehat{F})]$ is an isomorphism class of an X -haloed 2-cobordism, $P = [(\Phi^G, \Gamma, \mathcal{S}^G)]$ is an equivalence class of a

³By referencing these figures we mean the list of all generalized movie moves described in this section.

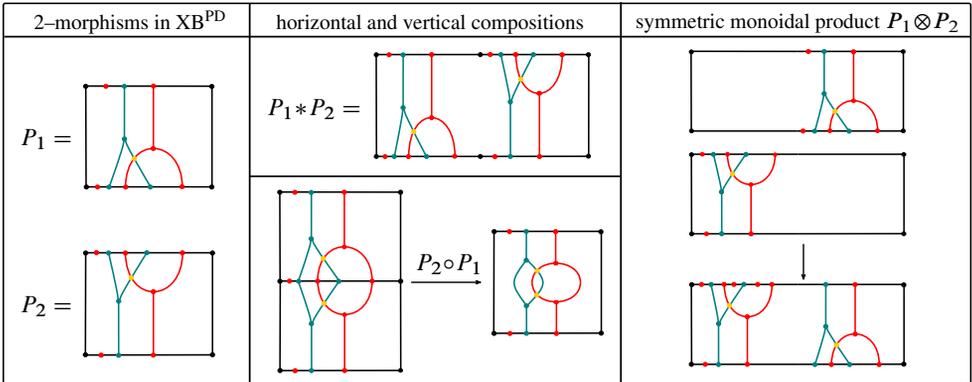


Figure 24: Compositions and symmetric monoidal product of 2-morphisms in \mathbf{XB}^{PD} .

G -planar diagram, and $\kappa : (S, R, F) \rightarrow (\bar{S}, \bar{R}, \bar{F})$ is an X -homeomorphism over I^2 , where $(\bar{S}, \bar{R}, \bar{F})$ is a cobordism type $\langle 2 \rangle$ - X -manifold constructed from a representative $(\Phi^G, \Gamma, \mathcal{S}^G)$. The composition of two composable triples is componentwise, similar to the composition of 1-morphisms.

The second bicategory \mathbf{XB}^{PD} is defined by forgetting X -haloed manifolds and cobordisms in $X\text{Bord}_2^{\text{PD}}$ and taking isotopy classes of G -linear diagrams. In order to define isotopic G -linear diagrams, we first need to explain compositions and monoidal products of diagrams.

Horizontal compositions of G -linear and G -planar diagrams are given by the horizontal concatenation of diagrams, where both G -sheet data agree and form new G -sheet data. Vertical composition of equivalence classes of G -planar diagrams is vertical concatenation of diagrams followed by an isomorphism $I \cup_{\text{pt}} I \cong I$ and forgetting the G -linear diagram on the face along which two G -planar diagrams are concatenated. Figure 24 shows an example of horizontal and vertical compositions of 2-morphisms in \mathbf{XB}^{PD} whose labels are omitted.

A symmetric monoidal structure on \mathbf{XB}^{PD} is defined as follows. Let $P_1 = (\Phi_1^G, \Gamma_1, \mathcal{S}_1^G)$ and $P_2 = (\Phi_2^G, \Gamma_2, \mathcal{S}_2^G)$ be two G -planar diagrams on $[m, n] \times I$ and on $[a, b] \times I$ for $m, n, a, b \in \mathbb{Z}$, respectively. Let V_{left} be the leftmost chamber of P_1 and V_{right} be the rightmost chamber of P_2 . Then $P_1 \otimes P_2$ is defined by stretching V_{left} to the left by $b - a$ units, stretching V_{right} to the right by $n - m$ units, and joining the stretched diagrams (see Figure 24). The G -sheet data and the labels of the resulting diagram are modified

accordingly. The symmetric monoidal structure on G -linear diagrams can be deduced from this description. It is not hard to see that the described symmetric monoidal product of diagrams is compatible with the disjoint union of X -haloed manifolds.

Recall that objects of \mathbf{XB}^{PD} are finite set of ordered oriented points, 1-morphisms are isotopy classes of G -linear diagrams, and 2-morphisms are equivalence classes of G -planar diagrams. The notion of isotopy between G -linear diagrams is generated by the following identifications. Let $L = (\Psi^G, \Gamma, \mathcal{S}^G)$ be any G -linear diagram, \emptyset be the empty G -linear diagram for the empty 1-manifold, and id_a be the identity G -linear diagram of the ordered set a . Then $L = L \otimes \emptyset = \emptyset \otimes L$ and $L = L \circ \text{id}_a = \text{id}_b \circ L$, where $L: a \rightarrow b$. In this case, it is not hard to see that \mathbf{XB}^{PD} is a strict 2-category.

Lemma 4.2 *Both $\mathbf{XBord}_2^{\text{PD}}$ and \mathbf{XB}^{PD} are symmetric monoidal bicategories under disjoint union of X -haloed manifolds with operation \otimes on diagrams.*

The proof for \mathbf{XB}^{PD} is very similar to the proof of Lemma 2.6. The case of $\mathbf{XBord}_2^{\text{PD}}$ follows from the compatibility of symmetric monoidal structures. Considering the G -planar decomposition theorem, a natural question is whether the symmetric monoidal bicategory \mathbf{XB}^{PD} defined by using diagrams is symmetric monoidally equivalent to X -cobordism bicategory \mathbf{XBord}_2 . We give a positive answer using the following theorem:

Theorem 4.3 (Whitehead theorem for symmetric monoidal bicategories [22, Theorem 2.25]) *Let \mathcal{B} and \mathcal{C} be symmetric monoidal bicategories. A symmetric monoidal 2-functor $F: \mathcal{B} \rightarrow \mathcal{C}$ is a symmetric monoidal equivalence if and only if it is an equivalence of underlying bicategories. That is, F is essentially surjective on objects, essentially full on 1-morphisms, and fully faithful on 2-morphisms.*

Proposition 4.4 *The forgetful 2-functors F and G given by forgetting X -haloed cobordisms and diagrams, respectively,*

$$\mathbf{XB}^{\text{PD}} \xleftarrow[\simeq]{F} \mathbf{XBord}_2^{\text{PD}} \xrightarrow[\simeq]{G} \mathbf{XBord}_2,$$

are symmetric monoidal equivalences.

Proof For any given finite set W of ordered oriented points or a compact oriented 0-manifold with cooriented codimension two X -halation $(Y, \hat{Y}_0, \hat{Y}_1, \hat{g})$, there exist objects in $\mathbf{XBord}_2^{\text{PD}}$ whose images under F and G are isomorphic to W and $(Y, \hat{Y}_0, \hat{Y}_1, \hat{g})$, respectively. For any given X -haloed 1-cobordism, there exists a Morse function with distinct critical values leading to a G -linear diagram and any G -linear diagram

produces an X -haloed⁴ 1-cobordism. Thus, by Proposition 3.5, each 2-functor is (essentially) full on 1-morphisms. Lastly, by the G -planar decomposition theorem, the 2-functors F and G are fully faithful on 2-morphisms. \square

Proposition 4.4 implies that the X -cobordism bicategory $X\text{Bord}_2$ is symmetric monoidally equivalent to XB^{PD} . The advantage of XB^{PD} is being a computadic unbiased semistrict symmetric monoidal 2-category. In the appendix, we provide the definition of computadic unbiased semistrict symmetric monoidal 2-category and prove this claim, whose precise statement is given below (see Theorem 4.5). This result is an important step of the classification, which we want to describe here.

The fact that XB^{PD} is a computadic symmetric monoidal bicategory roughly means that there exist four sets — namely generating objects \mathcal{XG}_0 , generating 1-morphisms \mathcal{XG}_1 , generating 2-morphisms \mathcal{XG}_2 , and generating relations \mathcal{XR} among 2-morphisms forming the presentation \mathbb{XP} — such that there exists a (canonical) isomorphism of symmetric monoidal bicategories $F_{\text{uss}}(\mathbb{XP}) \rightarrow \text{XB}^{\text{PD}}$, where $F_{\text{uss}}(\mathbb{XP})$ is constructed from \mathbb{XP} . Therefore, we have

$$(1) \quad F_{\text{uss}}(\mathbb{XP}) \xrightarrow[\simeq]{\exists} \text{XB}^{\text{PD}} \xleftarrow[\simeq]{F} X\text{Bord}_2^{\text{PD}} \xrightarrow[\simeq]{G} X\text{Bord}_2$$

and the cofibrancy theorem states that symmetric monoidal 2-functors out of $F_{\text{uss}}(\mathbb{XP})$ are determined by the images of generating sets subject to the relations. Thus, the classification of 2-dimensional extended X -HFTs up to symmetric monoidal equivalence reduces to understanding images of generators in \mathbb{XP} satisfying relations. The following theorem lists the presentation $\mathbb{XP} = (\mathcal{XG}_0, \mathcal{XG}_1, \mathcal{XG}_2, \mathcal{XR})$ of XB^{PD} and its proof is given in Section A.3.

Theorem 4.5 *The symmetric monoidal bicategory XB^{PD} is a computadic unbiased semistrict symmetric monoidal 2-category with presentation*

$$\mathbb{XP} = (\mathcal{XG}_0, \mathcal{XG}_1, \mathcal{XG}_2, \mathcal{XR})$$

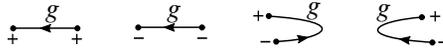
given by the diagram versions⁵ of elements in Figures 25 and 26, and pairs of G -planar diagrams corresponding to equalities in Figure 22, where the labels $g_1, g_2, g_3, g_4, g, g'$ and g'' are indexed over G so that $g_1g_2g_3g_4 = e$.

⁴Halation can be encoded into G -sheet data by equipping trivializations of chambers with halations.
⁵For generating 2-morphisms, we mean equivalence classes of G -planar diagrams. We consider G -linear and G -planar diagrams whose chambering sets and graphs are trivial, corresponding to covers with single elements.

generating objects:



generating 1-morphisms:



generating 2-morphisms:

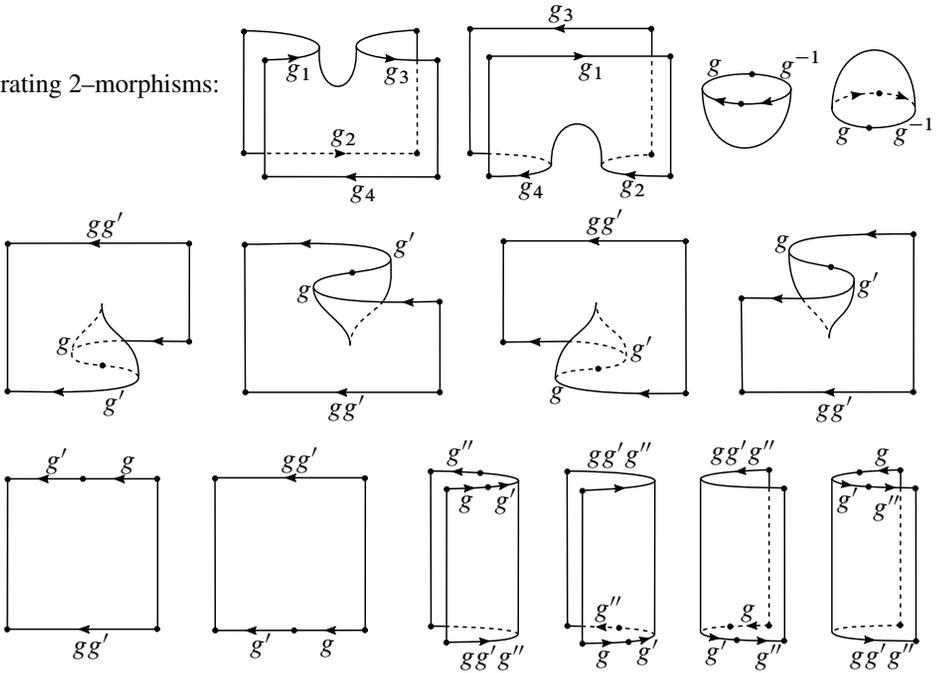


Figure 25: Generating objects (\mathcal{XG}_0), 1-morphisms (\mathcal{XG}_1), and 2-morphisms (\mathcal{XG}_2).

4.2 The cofibrancy theorem

Knowing the equivalence $\mathbf{XB}^{\text{PD}} \simeq \mathbf{XBord}_2$, the classification of 2-dimensional extended X -HFTs mainly concerns the understanding of symmetric monoidal 2-functors defined on \mathbf{XB}^{PD} . The cofibrancy theorem [22] is a coherence theorem for such 2-functors. More precisely, this theorem allows replacing (weak) symmetric monoidal 2-functors defined on computadic symmetric monoidal bicategories with their strict versions naturally. Furthermore, such strict 2-functors are determined by the images of generating sets of a presentation subject to the relations.

The cofibrancy theorem holds for any computadic monoidal bicategory (see [22; 21]) and specifically for stricter versions of symmetric monoidal bicategories. In this section, we only focus on its version for computadic unbiased semistrict symmetric monoidal 2-categories, which is the key step of the classification. In the following, we denote

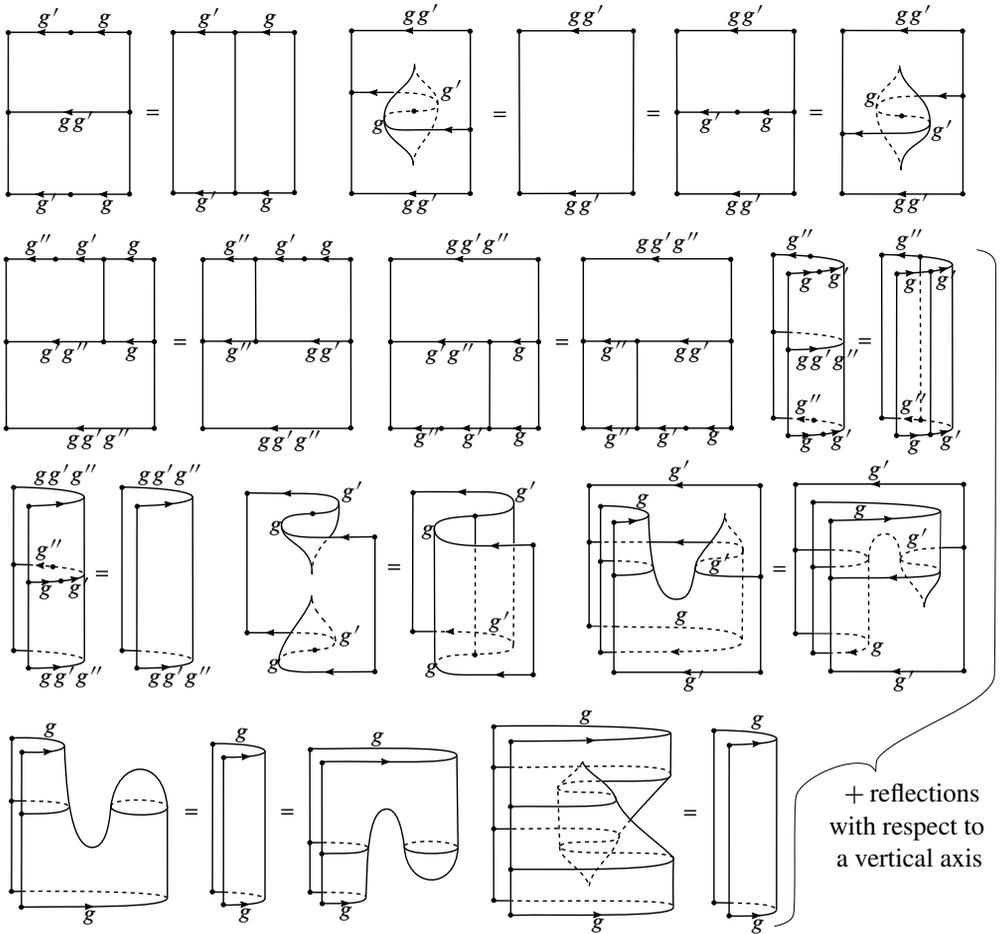


Figure 26: Generating relations $(\mathcal{X}\mathcal{R})$ among 2-morphisms.

the collection of objects of a symmetric monoidal bicategory \mathcal{C} by \mathcal{C}_0 , 1-morphisms by \mathcal{C}_1 , and 2-morphisms by \mathcal{C}_2 .

Definition 4.6 Let $(\mathbb{P} = (\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{R}), s, t)$ be an unbiased semistrict symmetric presentation and \mathcal{C} be a symmetric monoidal bicategory. The bicategory $\mathbb{P}(\mathcal{C})$ of \mathbb{P} -data in \mathcal{C} is defined as follows:

- The objects of $\mathbb{P}(\mathcal{C})$ are triples of assignments (A_0, A_1, A_2) such that $A_i : \mathcal{G}_i \rightarrow \mathcal{C}_i$ for $i = 0, 1, 2$ satisfying the following conditions. These assignments extend canonically to $\text{BW}^{\text{USS}}(\mathcal{G}_0)$, $\text{BS}^{\text{USS}}(\mathcal{G}_1)/\sim$, and $\text{PG}^{\text{USS}}(\mathcal{G}_2)/\sim$ using the monoidal product and braiding of \mathcal{C} as follows. For any $x = a_1 \dots a_n \in \text{BW}^{\text{USS}}(\mathcal{G}_0)$, we have $A_0(x) = A_0(a_1) \otimes \dots \otimes A_0(a_n)$, $\text{id}_x = \text{id}_{A_0(x)}$ and $\beta_{a, \sigma(a)}^\sigma$ for $\sigma \in \mathcal{S}_n$ is given by writing σ as

a product of adjacent transpositions first and then applying the braiding of \mathcal{C} to each of them. Composition and monoidal product of 1–morphisms follow similarly by extending A_1 to $BS^{uss}(\mathcal{G}_1)/\sim$. In the extension to $PG^{uss}(\mathcal{G}_2)$ interchanger 2–morphisms ϕ are taken as the identity and the extension to the remaining 2–morphisms uses the naturality of the braiding of \mathcal{C} and the corresponding modifications. These extensions are required to be globular, ie $A_0 \circ p = p \circ A_1$ and $A_1 \circ p = p \circ A_2$, where p is used for the source s and target t maps. Among all extensions, only those with $A_2(x) = A_2(y)$ for all $(x, y) \in \mathcal{R}$ are considered.

- The 1–morphisms from $A = (A_0, A_1, A_2)$ to $B = (B_0, B_1, B_2)$ are pairs of assignments (α_0, α_1) such that $\alpha_i: \mathcal{G}_i \rightarrow \mathcal{C}_{i+1}$ for $i = 0, 1$, $s(\alpha_0(a)) = A_0(a)$, $t(\alpha_0(b)) = B_0(b)$ and $\alpha_1(f): B_1(f) \circ \alpha_0(a) \xrightarrow{\cong} \alpha_0(b) \circ A_1(f)$ for all $f: a \rightarrow b \in \mathcal{G}_1$. These assignments extend canonically to $BW^{uss}(\mathcal{G}_0)$ using the monoidal product of \mathcal{C} and to $BS^{uss}(\mathcal{G}_1)/\sim$ using induction on the number of monoidal products and compositions as follows. For $x = a_1 \dots a_n \in BW^{uss}(\mathcal{G}_0)$, $\alpha_1(\text{id}_x)$ is the composition

$$B_1(\text{id}_x) \circ \alpha_0(x) = \text{id}_{B_0(x)} \circ \alpha_0(x) \xrightarrow{\ell^{\mathcal{C}}} \alpha_0(x) \xrightarrow{(r^{\mathcal{C}})^{-1}} \alpha_0(x) \circ \text{id}_{A_0(x)} = \alpha_0(x) \circ A_1(\text{id}_x),$$

where ℓ and r are the left and right unitors of the underlying bicategory of \mathcal{C} . For $x = a_1 \dots a_n \in BW^{uss}(\mathcal{G}_0)$ and $\sigma \in S_n$, the 2–morphism $\alpha_1(\beta_{x, \sigma(x)}^{\sigma})$ is given by the components of the braiding (equivalence transformation) of \mathcal{C} on the 1–morphism $\alpha_0(x)$. Next, for elements $f: a \rightarrow b$ and $f': a' \rightarrow b'$ in $BW^{uss}(\mathcal{G}_1)$, $\alpha_1(f \otimes f')$ is defined as the composition

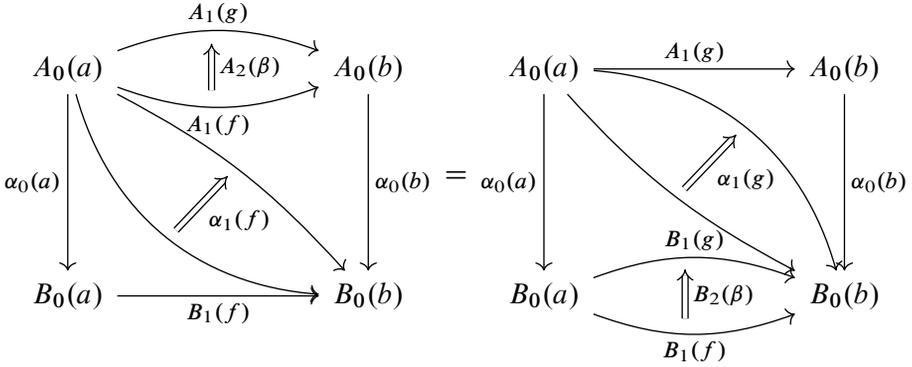
$$\begin{aligned} (B_1(f) \otimes B_1(f')) \circ (\alpha_0(a) \otimes \alpha_0(a')) &\rightarrow (B_1(f) \circ \alpha_0(a)) \otimes (B_1(f') \circ \alpha_0(a')) \\ &\xrightarrow{\alpha_1(f) \otimes \alpha_1(f')} (\alpha_0(b) \circ A_1(f)) \otimes (\alpha_0(b') \circ A_1(f')) \\ &\rightarrow (\alpha_0(b) \otimes \alpha_0(b')) \circ (A_1(f) \otimes A_1(f')). \end{aligned}$$

For elements $f: a \rightarrow b$ and $g: b \rightarrow c$ in $BW^{uss}(\mathcal{G}_1)$, $\alpha_1(g \circ f)$ is defined as the composition

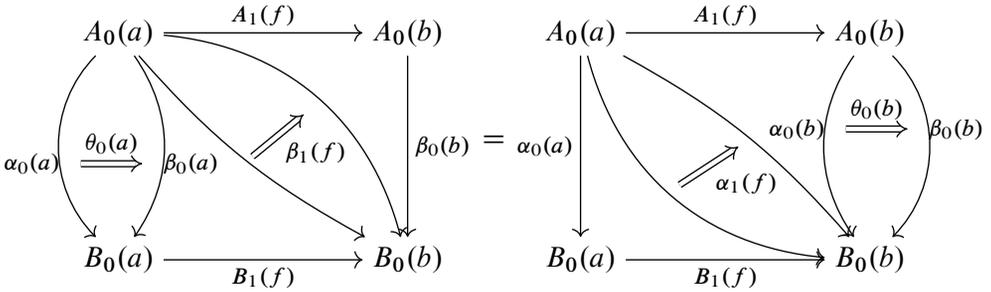
$$\begin{aligned} (B_1(g) \circ B_1(f)) \circ \alpha_0(a) &\xrightarrow{\hat{a}^{\mathcal{C}}} B_1(g) \circ (B_1(f) \circ \alpha_0(a)) \xrightarrow{\text{id} * \alpha_1(f)} B_1(g) \circ (\alpha_0(b) \circ A_1(f)) \\ &\xrightarrow{(\hat{a}^{\mathcal{C}})^{-1}} (B_1(g) \circ \alpha_0(b)) \circ A_1(f) \xrightarrow{\alpha_1(g) * \text{id}} (\alpha_0(c) \circ A_1(g)) \circ A_1(f) \\ &\xrightarrow{\hat{a}^{\mathcal{C}}} \alpha_0(c) \circ (A_1(g) \circ A_1(f)), \end{aligned}$$

where $\hat{a}^{\mathcal{C}}$ is the associator of the underlying bicategory of \mathcal{C} . These assignments are also required to be natural with respect to equivalence classes of paragraphs. That

is, for any $[\beta] \in \text{PG}^{\text{USS}}(\mathcal{G}_2)/\sim$ with $\beta: f \rightarrow g$, we have $(\text{id}_{\alpha_0(b)} * A_2(\beta)) \circ \alpha_1(f) = \alpha_1(g) \circ (B_2(\beta) * \text{id}_{\alpha_0(a)})$; equivalently,



- The 2-morphisms from $\alpha = (\alpha_0, \alpha_1)$ to $\beta = (\beta_0, \beta_1)$ are assignments $\theta_0: \mathcal{G}_0 \rightarrow \mathcal{C}_2$ such that $\theta_0(a): \alpha_0(a) \rightarrow \beta_0(a)$ for all $a \in \mathcal{G}_0$, where $\alpha, \beta: A \rightarrow B$ for $A = (A_0, A_1, A_2)$ and $B = (B_0, B_1, B_2)$. These assignments extend canonically to $\text{BW}^{\text{USS}}(\mathcal{G}_0)$ using the monoidal product of \mathcal{C} and they are required to be natural with respect to $\text{BW}^{\text{USS}}(\mathcal{G}_1)$. That is, for any $f: a \rightarrow b \in \text{BW}^{\text{USS}}(\mathcal{G}_1)$, we have $\beta_1(f) \circ (\theta_0(a) * \text{id}_{B_1(f)}) = (\text{id}_{A_1(f)} * \theta_0(b)) \circ \alpha_1(f)$; equivalently,



With the trivial coherence data $(\chi, \iota, \mathcal{W}, \mathcal{G}, \mathcal{R}, \mathcal{U})$, any object of $\mathbb{P}(\mathcal{C})$ considered with the extensions gives rise to a strict symmetric monoidal 2-functor $F_{\text{USS}}(\mathbb{P}) \rightarrow \mathcal{C}$ (see [22; 25]). Similarly, any 1-morphism and 2-morphism of $\mathbb{P}(\mathcal{C})$ considered with their extensions yield a strict⁶ symmetric monoidal transformation and a symmetric monoidal modification, respectively (see [22, Section 2.3]).

Definition 4.7 Let \mathcal{C} be a symmetric monoidal bicategory and $F_{\text{USS}}(\mathbb{P})$ be a computadic unbiased semistrict symmetric monoidal 2-category for a given presentation

⁶Those monoidal transformations whose monoidal structure 1- and 2-morphisms are identities.

$\mathbb{P} = (\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{R})$. The bicategory $\text{SymMon}(\mathbb{F}_{\text{uss}}(\mathbb{P}), \mathcal{C})$ has symmetric monoidal 2-functors as objects, symmetric monoidal transformations as 1-morphisms, and symmetric monoidal modifications as 2-morphisms.

By construction of $\mathbb{P}(\mathcal{C})$, we have a strict inclusion functor

$$\iota : \mathbb{P}(\mathcal{C}) \hookrightarrow \text{SymMon}(\mathbb{F}_{\text{uss}}(\mathbb{P}), \mathcal{C})$$

given by the associated 2-functors, transformations, and modifications. The cofibrancy theorem below states that ι is an equivalence of bicategories:

Theorem 4.8 (cofibrancy theorem [22, Theorem 2.78]) *Let \mathcal{C} be a symmetric monoidal bicategory and let $\mathbb{F}_{\text{uss}}(\mathbb{P})$ be a computadic unbiased semistrict symmetric monoidal 2-category constructed from an unbiased semistrict presentation $\mathbb{P} = (\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{R})$. Then the inclusion $\iota : \mathbb{P}(\mathcal{C}) \hookrightarrow \text{SymMon}(\mathbb{F}_{\text{uss}}(\mathbb{P}), \mathcal{C})$ is an equivalence of bicategories.*

The following lemma, taken from [22, Lemma 2.15], implies that, for any symmetric monoidal bicategory \mathcal{C} , the bicategories $\text{SymMon}(X\text{Bord}_2, \mathcal{C})$ and $\text{SymMon}(\text{XB}^{\text{PD}}, \mathcal{C})$ of symmetric monoidal 2-functors, symmetric monoidal transformations, and modifications are equivalent:

Lemma 4.9 [22] *Let \mathcal{M} and \mathcal{M}' be symmetric monoidal bicategories, and $H : \mathcal{M} \rightarrow \mathcal{M}'$ be a symmetric monoidal 2-functor which is an equivalence. Then there is a canonical equivalence $H^* : \text{SymMon}(\mathcal{M}', \mathcal{B}) \rightarrow \text{SymMon}(\mathcal{M}, \mathcal{B})$ of bicategories given by composition on the level of objects, and by symmetric monoidal whiskering on the level of 1- and 2-morphisms.*

We denote the bicategory $\text{SymMon}(X\text{Bord}_2, \mathcal{C})$ by $\mathcal{E}\text{-}\mathcal{H}\mathcal{F}\mathcal{J}(X, \mathcal{C})$ and state the classification of 2-dimensional extended HFTs with target $X \simeq K(G, 1)$ as follows:

Theorem 4.10 *Let $\mathbb{X}\mathbb{P}$ be the presentation of XB^{PD} given in Theorem 4.5. Then, for any symmetric monoidal bicategory \mathcal{C} , there is an equivalence of bicategories $\mathcal{E}\text{-}\mathcal{H}\mathcal{F}\mathcal{J}(X, \mathcal{C}) \simeq \mathbb{X}\mathbb{P}(\mathcal{C})$.*

Proof Theorem 4.5 gives a presentation $\mathbb{X}\mathbb{P}$ of XB^{PD} as a computadic unbiased semistrict symmetric monoidal 2-category. By the cofibrancy theorem, we have $\text{SymMon}(\text{XB}^{\text{PD}}, \mathcal{C}) \simeq \mathbb{X}\mathbb{P}(\mathcal{C})$. Using the symmetric monoidal equivalence between XB^{PD} and $X\text{Bord}_2$ in Proposition 4.4 and Lemma 4.9, we obtain the result. \square

4.3 $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFTs

Every 2-dimensional extended X -HFT gives a nonextended one by restricting to oriented X -circles and X -cobordisms between them. A natural question is how the classification of 2-dimensional extended HFTs is related to Turaev’s classification of 2-dimensional HFTs by crossed Frobenius G -algebras [27]. To understand this relation, we study extended HFTs taking values in $\text{Alg}_{\mathbb{k}}^2$, which has \mathbb{k} -algebras as objects, bimodules as 1-morphisms, and bimodule maps as 2-morphisms for a commutative ring \mathbb{k} with unity.

The symmetric monoidal structure of $\text{Alg}_{\mathbb{k}}^2$ is given by tensoring over \mathbb{k} . We denote (E, C) -bimodule D by ${}_E D_C$ and omit the symbol \mathbb{k} when either C or E is \mathbb{k} . We regard ${}_E D_C$ as a 1-morphism from C to E , which is in line with the composition in $X\text{Bord}_2$ (see Figure 3). Composition of 1-morphisms ${}_E D_C$ and ${}_C B_A$ is the bimodule ${}_E (D \otimes_C B)_A$.

Before studying $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFTs, we recall necessary algebraic notions and introduce quasibiangular G -algebras. Recall that a G -algebra over a commutative ring \mathbb{k} is an associative \mathbb{k} -algebra K equipped with a decomposition $K = \bigoplus_{g \in G} K_g$ such that $K_g K_h \subseteq K_{gh}$ for any $g, h \in G$. In this case, K_e is the principal component and K is called strongly graded if $K_g K_h = K_{gh}$ or, equivalently, the natural map $K_g \otimes_{K_e} K_h \rightarrow K_{gh}$ is an isomorphism for all $g, h \in G$. The opposite G -algebra of K is defined as $K^{\text{op}} = \bigoplus_{g \in G} K_{g^{-1}}$, where the order of multiplication is reversed.

Definition 4.11 [28] Let $K = \bigoplus_{g \in G} K_g$ be a G -algebra over a commutative ring \mathbb{k} . Recall that an inner product on K is a symmetric bilinear form $\eta: K \otimes K \rightarrow \mathbb{k}$ satisfying $\eta(ab, c) = \eta(a, bc)$ for any $a, b, c \in K$ such that $\eta|_{K_g \otimes K_h}$ is nondegenerate when $gh = e$ and zero otherwise. A Frobenius G -algebra is a G -algebra K with an inner product η and components of K are finitely generated projective \mathbb{k} -modules.

Let $(K = \bigoplus_{g \in G} K_g, \eta)$ be a Frobenius G -algebra over \mathbb{k} . Each nondegenerate form $\eta|_{K_g \otimes K_{g^{-1}}}$ yields an element $\eta_g^- = \sum_{i \in I_g} p_i^g \otimes q_i^g \in K_g \otimes K_{g^{-1}}$, called an inner product element, where I_g is finite and η_g^- is characterized by $a = \sum_{i \in I_g} \eta(a, q_i^g) p_i^g$ for any $a \in K_g$. Since η is symmetric, we have $\sum_i p_i^{g^{-1}} \otimes q_i^{g^{-1}} = \sum_i q_i^g \otimes p_i^g$ for all $g \in G$.

Recall that an associative \mathbb{k} -algebra A is separable if there exists an element $a = \sum_{i=1}^n p_i \otimes q_i \in A \otimes_{\mathbb{k}} A^{\text{op}}$, called separability idempotent, such that $\sum_{i=1}^n p_i q_i = 1$

and $ab = ba$ for all $b \in A$. A separable algebra A is called *strongly separable* if the separability idempotent is symmetric, ie $a = \sum_{i=1}^n p_i \otimes q_i = \sum_{i=1}^n q_i \otimes p_i$.

Lemma 4.12 *Let $(K = \bigoplus_{g \in G} K_g, \eta)$ be a Frobenius G -algebra with inner product elements $\{\eta_g^- = \sum_i p_i^g \otimes q_i^g\}_{g \in G}$ and $z \in K_e$. Then, for any $g, h \in G$ and $b \in K_{g^{-1}}$, we have*

$$(2) \quad \sum_i p_i^h \otimes zq_i^h b = \sum_j bp_j^{gh} \otimes zq_j^{gh}.$$

In particular, for any $b \in K$ and $c \in K_{h^{-1}}$, we have $\sum_j p_j^g bzq_j^g c = \sum_k cp_k^{hg} bzq_k^{hg}$.

Proof Since both sides belong to $K_h \otimes K_{h^{-1}g^{-1}}$, it is enough to check that they give the same functionals on the dual \mathbb{k} -module $K_{h^{-1}} \otimes K_{gh}$. For any $x \in K_{h^{-1}}$ and $y \in K_{gh}$, applying $x \otimes y$ to the left-hand side of (2) and using the cyclic symmetry property of η , we obtain

$$\begin{aligned} \sum_i \eta(p_i^h, x)\eta(zq_i^h b, y) &= \sum_i \eta(x, p_i^h)\eta(q_i^h, byz) = \eta\left(x, \sum_i \eta(byz, q_i^h)p_i^h\right) \\ &= \eta(x, byz). \end{aligned}$$

Similarly, applying $x \otimes y$ to the right-hand side of (2), we have

$$\begin{aligned} \sum_j \eta(bp_j^{gh}, x)\eta(zq_j^{gh}, y) &= \sum_j \eta(xb, p_j^{gh})\eta(q_j^{gh}, yz) = \eta\left(xb, \sum_j \eta(yz, q_j^{gh})p_j^{gh}\right) \\ &= \eta(xb, yz). \quad \square \end{aligned}$$

We generalize biangular G -algebras, which were introduced by Turaev [27], as follows:

Definition 4.13 A strongly graded Frobenius G -algebra (K, η) is called *quasibiangular* if there exists a central element $z \in K_e$, ie $za = az$ for all $a \in K_e$, such that, for the collection of inner product elements $\{\sum_i p_i^g \otimes q_i^g\}_{g \in G}$, the equations $\sum_i p_i^g zq_i^g = 1$ hold for all $g \in G$.

Remark By Lemma 4.12, the principal component of a quasibiangular G -algebra is a separable algebra with separability idempotent $\sum_i p_i^e \otimes zq_i^e$. A biangular G -algebra is a quasibiangular G -algebra with $z = 1$. Similarly, the principal component of a biangular G -algebra is strongly separable.

One way of studying an algebra is to study the category of modules over it. Recall that Morita equivalence of algebras is the equivalence of categories of modules. In the case of a graded algebra, one studies the category of graded modules. An equivalence of

such categories is called a graded Morita equivalence (see [3, Theorem 3.2]), which was introduced by Boisen [3] as follows:

Definition 4.14 [3] A G -graded Morita equivalence ζ between G -algebras $K = \bigoplus_{g \in G} K_g$ and $L = \bigoplus_{g \in G} L_g$ is a quadruple $({}_L U_K, {}_K V_L, \tau, \mu)$, where ${}_L U_K = \bigoplus_{g \in G} U_g$ is a graded (L, K) -bimodule — that is, $L_g U_h K_{g'} \subset U_{ghg'}$ — ${}_K V_L = \bigoplus_{g \in G} V_g$ is a graded (K, L) -bimodule, and $\tau: {}_K K K \rightarrow {}_K V \otimes_L U_K$ and $\mu: {}_L U \otimes_K V_L \rightarrow {}_L L L$ are graded (K, K) - and (L, L) -bimodule maps, respectively, such that the compositions

$$\begin{aligned} {}_L U_K &\rightarrow {}_L U \otimes_K {}_K K \xrightarrow{\text{id} \otimes \tau} {}_L U \otimes_K (V \otimes_L U_K) \rightarrow ({}_L U \otimes_K V) \otimes_L U_K \\ &\xrightarrow{\mu \otimes \text{id}} {}_L L \otimes_L U_K \rightarrow {}_L U_K, \\ {}_K V_L &\rightarrow {}_K K \otimes_K V_L \xrightarrow{\tau \otimes \text{id}} ({}_K V \otimes_L U) \otimes_K V_L \rightarrow {}_K V \otimes_L (U \otimes_K V_L) \\ &\xrightarrow{\text{id} \otimes \mu} {}_K V \otimes_L L L \rightarrow {}_K V_L \end{aligned}$$

are id_U and id_V , respectively. When τ and μ are invertible as G -graded bimodule maps, ζ is called a G -graded Morita context.

Definition 4.15 Let $\zeta = ({}_L U_K, {}_K V_L, \tau, \mu)$ and $\zeta' = ({}_L U'_K, {}_K V'_L, \tau', \mu')$ be two G -graded Morita equivalences. An equivalence of G -graded Morita equivalences ζ and ζ' is a pair of G -graded bimodule maps $\xi: {}_L U_K \rightarrow {}_L U'_K$ and $\rho: {}_K V_L \rightarrow {}_K V'_L$ such that $\mu = \mu' \circ (\xi \otimes \rho)$ and $\tau' = (\rho \otimes \xi) \circ \tau$.

Lemma 4.16 [9] Assume that G -algebras $K = \bigoplus_{g \in G} K_g$ and $L = \bigoplus_{g \in G} L_g$ are G -graded Morita equivalent. Then, if K is strongly graded, then L is also strongly graded.

Next, we transfer the inner product of one Frobenius G -algebra to another using a graded Morita context between them. As the first step we recall the following lemma:

Lemma 4.17 [22; 10] Any Morita context $\zeta = ({}_L U_K, {}_K V_L, \tau, \eta)$ between \mathbb{k} -algebras K and L induces a canonical isomorphism of \mathbb{k} -modules $\zeta_*: K/[K, K] \rightarrow L/[L, L]$.

An explicit formula for the isomorphism ζ_* in the lemma is provided in [10]. The inner product η of a Frobenius G -algebra (K, η) is determined at its principal component by $\eta(a, b \cdot 1) = \eta(ab, 1)$. This allows us to denote (K, η) by (K, Λ) , where $\Lambda: K_e \rightarrow \mathbb{k}$ is a nondegenerate trace. Since η is symmetric, Λ factors through $K_e/[K_e, K_e]$.

Lemma 4.12 implies that, for a symmetric Frobenius algebra (K_e, Λ_e) , an inner product element $\sum_i p_i^e \otimes q_i^e$ can be considered as the image of $1 \otimes 1$ under a bimodule map $\xi: K_e^1(K_e)_{K_e^2} \otimes_{K_e^3} (K_e)_{K_e^4} \rightarrow K_e^1(K_e)_{K_e^4} \otimes_{K_e^3} (K_e)_{K_e^2}$, where numbers indicate module actions, ie $K_e^i = K_e$ for $i = 1, 2, 3, 4$. In the case of a quasibiangular G -algebra $(K = \bigoplus_{g \in G} K_g, \Lambda)$, inner product elements $\{\sum_i p_i^g \otimes q_i^g\}_{g \in G \setminus \{e\}}$ are the images of $1 \otimes 1$ under the composition

$$(3) \quad K_e^1(K_e)_{K_e^2} \otimes_{K_e^3} (K_e)_{K_e^4} \rightarrow K_e^1(K_e)_{K_e^2} \otimes_{K_e^3} (K_{g^{-1}} \otimes_{K_e} K_e \otimes_{K_e} K_g)_{K_e^4} \\ \rightarrow K_e^1(K_e \otimes_{K_e} K_g)_{K_e^4} \otimes_{K_e^3} (K_{g^{-1}} \otimes_{K_e} K_e)_{K_e^2} \rightarrow K_e^1(K_g)_{K_e^4} \otimes_{K_e^3} (K_{g^{-1}})_{K_e^2},$$

where the second homomorphism is the identity on $K_{g^{-1}}$ and K_g , and ξ on $K_e \otimes K_e$. In the following, we consider inner product elements as the images of $1 \otimes 1$ under the above bimodule maps.

Definition 4.18 Let (K, Λ_K) and (L, Λ_L) be quasibiangular G -algebras over \mathbb{k} with collections of inner product elements $\{\eta_g^K\}_{g \in G}$ and $\{\eta_g^L\}_{g \in G}$, respectively. A G -graded Morita context $\zeta = ({}_L U_K, {}_K V_L, \tau, \mu)$ between K and L is said to be *compatible* if $\Lambda_L = (\zeta_{\{e\}})_* \Lambda_K$ and $\eta_g^L = (\zeta_{\{e\}})_*(\eta_g^K)$ for all $g \in G$, where $(\zeta_{\{e\}})_*(\eta_g^K)$ consists of inner product elements for $(L, (\zeta_e)_* \Lambda_K)$ given by $\xi'(1 \otimes 1)$ under the commutative diagram

$$\begin{array}{ccc} L_e(U_e \otimes K_e \otimes V_e)_{L_e} \otimes_{\mathbb{k}} L_e(U_e \otimes K_e \otimes V_e)_{L_e} & \xrightarrow{\mu_{\{e\}}} & L_e(L_e)_{L_e} \otimes_{\mathbb{k}} L_e(L_e)_{L_e} \\ \text{id} \otimes \xi \otimes \text{id} \downarrow & & \downarrow \xi' \\ L_e(U_e \otimes K_e \otimes V_e)_{L_e} \otimes_{\mathbb{k}} L_e(U_e \otimes K_e \otimes V_e)_{L_e} & \xrightarrow{\mu_{\{e\}}} & L_e(L_e)_{L_e} \otimes_{\mathbb{k}} L_e(L_e)_{L_e} \end{array}$$

The remaining inner product elements are obtained from ξ' as described above.

Theorem 4.19 Any $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFT $Z: X\text{Bord}_2 \rightarrow \text{Alg}_{\mathbb{k}}^2$ whose precomposition $\text{XB}^{\text{PD}} \xrightarrow{\cong} X\text{Bord}_2 \xrightarrow{Z} \text{Alg}_{\mathbb{k}}^2$ gives a strict symmetric monoidal 2-functor determines a triple (A, B, ζ) , where A and B are quasibiangular G -algebras, and ζ is a compatible G -graded Morita context between A and B^{op} . Conversely, for any such triple (A, B, ζ) , there exists an $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended X -HFT.

Proof Let $Z: X\text{Bord}_2 \rightarrow \text{Alg}_{\mathbb{k}}^2$ be such a 2-dimensional extended HFT. The cofibrancy theorem implies that there exists an object Z' in $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ such that $\iota(Z')$ is the composition $\text{XB}^{\text{PD}} \xrightarrow{\cong} X\text{Bord}_2 \xrightarrow{Z} \text{Alg}_{\mathbb{k}}^2$, where $\iota: \mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2) \rightarrow \text{SymMon}(\text{XB}^{\text{PD}}, \text{Alg}_{\mathbb{k}}^2)$ is the equivalence of bicategories.

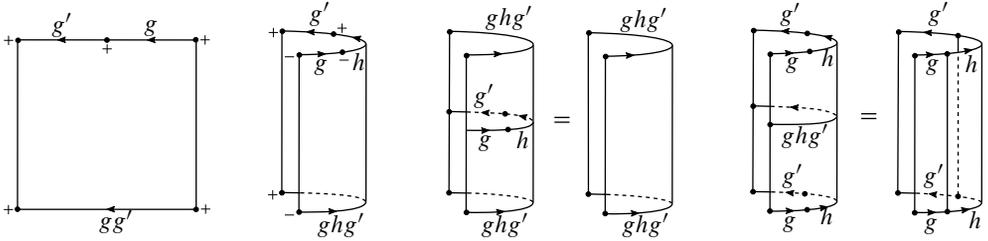


Figure 27: Part of generators and relations giving G -algebra and G -graded module.

We have \mathbb{k} -algebras $Z'(\bullet^+) = A_e$ and $Z'(\bullet^-) = B_e$ in $Z'_0(\mathcal{XG}_0)$, corresponding to two generating objects of $\mathbb{X}\mathbb{P}$. There are four types of generating 1-morphisms and each is indexed by the elements of G . For every $g \in G$, they give the bimodules, in $Z'_1(\mathcal{XG}_1)$,

$$\begin{aligned} Z'(+ \xrightarrow{g} +) &= A_g \quad (A_e, A_e)\text{-bimodule,} \\ Z'(- \xrightarrow{g} -) &= B_g \quad (B_e, B_e)\text{-bimodule,} \\ Z'(\overset{+}{\curvearrowright} g) &= M_g \quad (B_e \otimes_{\mathbb{k}} A_e, \mathbb{k})\text{-bimodule,} \\ Z'(g \overset{-}{\curvearrowleft}) &= N_g \quad (\mathbb{k}, A_e \otimes_{\mathbb{k}} B_e)\text{-bimodule.} \end{aligned}$$

The first 2-morphism in Figure 27 defines a G -graded product on the bimodule $\bigoplus_{g \in G} A_g$. Associativity of this product is the obvious relation in Figure 26. Denote the corresponding G -algebra by $A = \bigoplus_{g \in G} A_g$. The first relation in Figure 26 shows that the bimodule map

$$(4) \quad A_e(A_{g'}) \otimes_{A_e} (A_g)_{A_e} \xrightarrow{\cong} A_e(A_{gg'})_{A_e}$$

is invertible for all $g, g' \in G$. Since multiplication in the G -algebra A is defined using (4), we have $A_g A_{g'} = A_{gg'}$ for all $g, g' \in G$, ie A is strongly graded. Similar arguments for (B_e, B_e) -bimodules $\{B_g\}_{g \in G}$ yield another strongly graded G -algebra $B = \bigoplus_{g \in G} B_g$.

Using the opposite algebra, we can turn algebra actions on bimodules around. More precisely, a left B_e -action on $A_e \otimes_{B_e} M_g$ can be turned into a right B_e^{op} -action and the right B_e -action on $(N_g)_{A_e \otimes B_e}$ can be turned into a left B_e^{op} -action. The second 2-morphism in Figure 27 gives

$$(5) \quad B_e \otimes_{A_e} (B_{g^{-1}} \otimes_{\mathbb{k}} A_{g'}) \otimes_{B_e \otimes A_e} M_h \xrightarrow{\cong} B_e \otimes_{A_e} M_{g^{-1}hg'}.$$

Turning B_e -actions on B_g around gives $A_e(M_{ghg'})_{B_e^{\text{op}}}$ and the collection of all such bimodule maps turns $\{A_e(M_g)_{B_e^{\text{op}}}\}_{g \in G}$ into a G -graded (A, B^{op}) -bimodule $M = \bigoplus_{g \in G} M_g$. Similarly, reflections of this 2-morphism and the corresponding relations

with respect to a vertical axis show that $N = \bigoplus_{g \in G} N_g$ is a G -graded (B^{op}, A) -bimodule.

There are four types of cusp generators and each is indexed by two elements of G . For every $g, g' \in G$ they give the bimodule maps, in $Z'_2(\mathcal{XG}_2)$,

$$\begin{aligned} f_1^{gg'} &: A_e(A_{gg'})_{A_e} \rightarrow A_e M_g \otimes_{B_e^{\text{op}}} (N_{g'})_{A_e}, \\ f_2^{gg'} &: B_e^{\text{op}} N_g \otimes_{A_e} (M_{g'})_{B_e^{\text{op}}} \rightarrow B_e^{\text{op}} (B_{gg'}^{\text{op}})_{B_e^{\text{op}}}, \\ f_3^{gg'} &: B_e^{\text{op}} (B_{gg'}^{\text{op}})_{B_e^{\text{op}}} \rightarrow B_e^{\text{op}} N_g \otimes_{A_e} (M_{g'})_{B_e^{\text{op}}}, \\ f_4^{gg'} &: A_e M_g \otimes_{B_e^{\text{op}}} (N_{g'})_{A_e} \rightarrow A_e (A_{gg'})_{A_e}, \end{aligned}$$

given in the order of cusp generators in Figure 25. These bimodule maps are required to satisfy the relations in \mathcal{XR} . Relations containing cusp generators indicate that these bimodule maps are two-sided inverses, ie $f_1^{gg'} = (f_4^{gg'})^{-1}$ and $f_2^{gg'} = (f_3^{gg'})^{-1}$. It is not hard to see that for each i the collection $\{f_i^{gg'}\}_{g, g' \in G}$ of bimodule maps forms a G -graded bimodule map f_i . The collection of swallowtail morphisms corresponds to the compositions of graded bimodule maps

$$\begin{aligned} B^{\text{op}} N_A &\rightarrow B^{\text{op}} N \otimes_A A A \xrightarrow{\text{id} \otimes f_1} B^{\text{op}} N \otimes_A M \otimes_{B^{\text{op}}} N_A \xrightarrow{f_2 \otimes \text{id}} B^{\text{op}} B^{\text{op}} \otimes_{B^{\text{op}}} N_A \rightarrow B^{\text{op}} N_A, \\ A M_{B^{\text{op}}} &\rightarrow A A \otimes_A M_{B^{\text{op}}} \xrightarrow{f_1 \otimes \text{id}} A M \otimes_{B^{\text{op}}} N \otimes_A M_{B^{\text{op}}} \xrightarrow{\text{id} \otimes f_2} A M \otimes_{B^{\text{op}}} B_{B^{\text{op}}}^{\text{op}} \rightarrow A M_{B^{\text{op}}}. \end{aligned}$$

Swallowtail relations imply that both compositions equal to identity bimodule maps of N and M , respectively. In other words, $\zeta = (B^{\text{op}} N_A, A M_{B^{\text{op}}}, f_1, f_2)$ is a G -graded Morita context. Using ζ , we can replace B^{op} -module actions with A -module actions as follows. A right (left) B^{op} -module can be turned into a right (left) A -module by tensoring with ${}_{B^{\text{op}}}(N)_A$ (${}_A M_{B^{\text{op}}}$), such as tensoring ${}_A M_{B^{\text{op}}}$ with ${}_{B^{\text{op}}}(N)_A$ yields ${}_A M \otimes_{B^{\text{op}}} N_A$, which is isomorphic to ${}_A A A$ via f_4 .

The remaining generators are Morse generators consisting of saddles, cup, and cap 2-morphisms. The collection of bimodule maps in $Z'_2(\mathcal{XG}_2)$ for the first saddle morphism in Figure 25 yields a graded bimodule map of the form

$$A \otimes_B M \otimes_{\mathbb{k}} N_{A \otimes B} \rightarrow A \otimes_B (A \otimes_{\mathbb{k}} B)_{A \otimes B}.$$

By turning the B -module actions around, we obtain

$$A \otimes_{B^{\text{op}}} (M \otimes_{\mathbb{k}} N)_{A \otimes B^{\text{op}}} \rightarrow A \otimes_{B^{\text{op}}} (A \otimes_{\mathbb{k}} B)_{A \otimes B^{\text{op}}},$$

where the left B^{op} -action is on N and the right B^{op} -action is on M . As pointed out above, B^{op} -module actions can be replaced by A -module actions and we get a graded

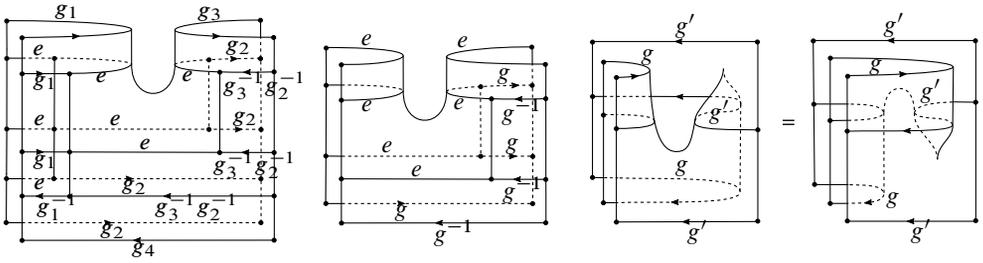


Figure 28: Saddle morphisms and cusp flip relation.

$(A_1 \otimes A_3, A_2 \otimes A_4)$ -bimodule map of the form

$$\xi : A_1 A_{A_2} \otimes_{\mathbb{k}} A_3 A_{A_4} \rightarrow A_1 A_{A_4} \otimes_{\mathbb{k}} A_3 A_{A_2},$$

where numbers indicate module actions, ie $A_i = A$ for $i = 1, 2, 3, 4$. The first morphism in Figure 28 shows that an arbitrary saddle morphism can be obtained from an e -labeled saddle morphism and other generating 2-morphisms. This implies that the graded bimodule map ξ is determined at $1 \otimes 1 \in A_e \otimes A_e$, which we denote by a finite sum $\sum_i p_i^e \otimes q_i^e$, and it satisfies $\sum_i a p_i^e \otimes q_i^e = \sum_i p_i^e \otimes q_i^e a$ for all $a \in A$. Similarly, we denote the image of $1 \otimes 1$ under the second 2-morphism in Figure 28 by $\eta_g^A = \sum_i p_i^g \otimes q_i^g$ for all $g \in G$ (compare with (3)).

In the same way, the collection of bimodule maps in $Z_2'(\mathcal{X}\mathcal{G}_2)$ for the second saddle morphism gives a graded $(A_1 \otimes A_3, A_2 \otimes A_4)$ -bimodule map of the form

$$\eta : A_1 A_{A_2} \otimes_{\mathbb{k}} A_3 A_{A_4} \rightarrow A_1 A_{A_4} \otimes_{\mathbb{k}} A_3 A_{A_2}.$$

The cusp flip relation shown in Figure 28 implies that $\xi = \eta$. Before considering cup and cap generators, note that, using ζ , we can assign the collection of all g - and g^{-1} -labeled circles to $\bigoplus_{g \in G} A_g \otimes_{(A_e \otimes A_e^{\text{op}})} A_{g^{-1}}$. The collections of 2-morphisms

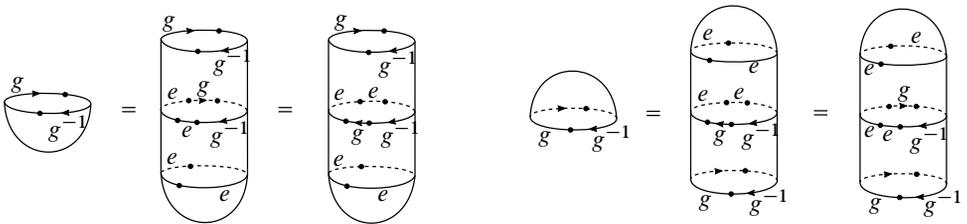


Figure 29: Cup and cap morphisms on nonprincipal components.

in Figure 29 give the bimodule maps

$$\Lambda: \bigoplus_{g \in G} A_g \otimes_{A_e \otimes A_e^{\text{op}}} A_{g^{-1}} \rightarrow \mathbb{k}, \quad u: \mathbb{k} \rightarrow \bigoplus_{g \in G} A_g \otimes_{A_e \otimes A_e^{\text{op}}} A_{g^{-1}},$$

respectively. Figure 29 implies that cup and cap morphisms are determined on the principal component. Since $A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_e = A_e/[A_e, A_e]$, cup morphism on the principal component can be considered as a symmetric linear map $\Lambda: A_e \rightarrow \mathbb{k}$. Additionally, Figure 29 shows that, on nonprincipal components, cup morphism is given by multiplication followed by Λ , leading to a symmetric \mathbb{k} -bilinear map $\eta_g: A_g \otimes A_{g^{-1}} \rightarrow \mathbb{k}$. Morse relations involving cup morphism indicate the nondegeneracy of η_g as follows. Assuming $\sum_j \beta_g^j \otimes 1 \otimes \beta_{g^{-1}}^j$ is the image of 1 under $A_e \xrightarrow{\sim} A_g \otimes_{A_e} A_e \otimes_{A_e} A_{g^{-1}}$, the first (left) 2-morphism in Figure 30 corresponds to the composition

$$\begin{aligned} a \mapsto 1 \otimes 1 \otimes a \mapsto \sum_j \beta_g^j \otimes \left(\sum_i p_i^e \otimes q_i^e \right) \otimes \beta_{g^{-1}}^j \otimes a \mapsto \sum_i p_i^g \otimes q_i^g \otimes a \\ \mapsto \sum_i p_i^g \eta(q_i^g, a) \end{aligned}$$

and the Morse relation implies that it is equivalent to id_{A_g} . Similarly, reflection of this morphism with label g^{-1} gives $b = \sum_i \eta_g(b, p_i^g) q_i^g$ for any $b \in A_{g^{-1}}$, which shows that η_g is nondegenerate. Thus, (A, η_A) is a Frobenius G -algebra, where $(\eta_A)|_{A_g \otimes A_h}$ is η_g when $h = g^{-1}$ and zero otherwise.

The remaining Morse relations contain cap morphisms which are determined on the principal component. For any $c \in A_e$, assuming $u(1)|_{A_e \otimes A_e} = \sum_j a_j \otimes b_j$, the second 2-morphism in Figure 30 corresponds to the compositions

$$c \otimes \sum_j a_j \otimes b_j \mapsto \sum_{i,j} c p_i^e \otimes a_j q_i^e \otimes b_j \mapsto \sum_{i,j} c p_i^e b_j a_j q_i^e \mapsto c \sum_i p_i^e z q_i^e,$$

where $z = \sum_j b_j a_j \in A_e$. The Morse relation implies that $\sum_i p_i^e z q_i^e = 1$ and consequently $\sum_i p_i^e \otimes z q_i^e$ is a separability idempotent of the algebra A_e . Thus, (A_e, η_e) is a separable symmetric Frobenius algebra, as shown in [22]. Similarly, we have $\sum_i p_i^g z q_i^g = 1$ using the saddle whose image gives η_g^A . Until now we used ζ to replace B^{op} -actions by A -actions. By changing the roles of A and B , we obtain a quasibiangular G -algebra B and ζ is a compatible graded Morita context between B and A^{op} .

Thus, any object in $\mathbb{X}\mathbb{P}^{\text{PD}}$ determines a triple (A, B, ζ) . Conversely, for any such triple, one constructs an object of $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ by assigning values to generating objects, 1-morphisms, and 2-morphisms of $\mathbb{X}\mathbb{P}$ satisfying generating relations using

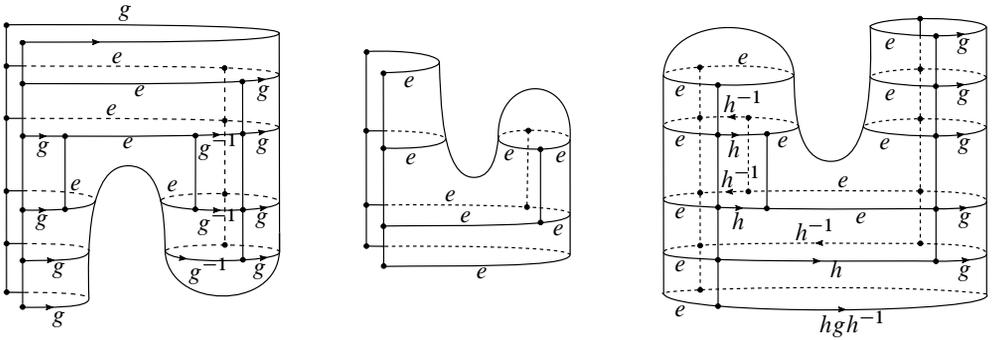


Figure 30: Compositions of generating 2-morphisms forming Morse relations.

the above arguments. Then, by the cofibrancy theorem, this object gives a strict symmetric monoidal 2-functor $\mathbf{XB}^{\text{PD}} \rightarrow \text{Alg}_{\mathbb{k}}^2$ whose composition with the equivalence $X\text{Bord}_2 \xrightarrow{\sim} \mathbf{XB}^{\text{PD}}$ produces the desired extended X -HFT. \square

Remark The cofibrancy theorem implies that any symmetric monoidal 2-functor $Z: \mathbf{XB}^{\text{PD}} \rightarrow \text{Alg}_{\mathbb{k}}^2$ can be strictified. That is, Z is equivalent to a strict symmetric monoidal 2-functor. From now on, by an $\text{Alg}_{\mathbb{k}}^2$ -valued extended X -HFT giving triple (A, B, ζ) we mean the triple coming from the corresponding strict symmetric monoidal 2-functor.

Turaev [28] defined the G -center of a biangular G -algebra. We extend this notion to a G -center of a quasibiangular G -algebra (K, η) as $Z_G(K) = \bigoplus_{g \in G} \Psi(K_g)$, where $\Psi(a) = \sum_i p_i^e a q_i^e$ for inner product elements $\{\sum_i p_i^g \otimes q_i^g\}_{g \in G}$. In general, the G -center is not commutative and it differs from the usual center of the algebra. However, it has a crossed Frobenius G -algebra structure, which is defined as follows:

Definition 4.20 [28] A Frobenius G -algebra $(L = \bigoplus_{g \in G} L_g, \eta)$ is *crossed* if L is endowed with a group homomorphism $\varphi: G \rightarrow \text{Aut}(L)$ satisfying the following conditions:

- (i) φ is conjugation type, ie $\varphi_h(L_g) = L_{hgh^{-1}}$ and $\varphi_h|_{L_h} = \text{id}_{L_h}$ for every $g, h \in G$.
- (ii) $ba = \varphi_h(a)b$ for any $a \in L$ and $b \in L_h$.
- (iii) $\text{Tr}(\mu \varphi_h: L_g \rightarrow L_g) = \text{Tr}(\varphi_{g^{-1}} \mu_c: L_h \rightarrow L_h)$ for all $g, h \in G$ and $c \in L_{hgh^{-1}h^{-1}}$, where $\mu_c: L \rightarrow L$ is left multiplication by c and Tr is the trace of a map.
- (iv) η is invariant under φ .

Lemma 4.21 Let (K, η) be a quasibiangular G -algebra with a central element $z \in K_e$ and a collection of inner product elements $\{\sum_i p_i^g \otimes q_i^g\}_{g \in G}$. Then $Z_G(K)$ is a unital G -algebra with multiplication $\sum_i p_i^e a q_i^e \cdot \sum_i p_i^e b q_i^e = \sum_{i,j} p_i^e a q_i^e p_j^e b q_j^e z^{-1}$ for all $a, b \in K$ and the triple $(Z_G(K), \eta|_{Z_G(K)}, \{\varphi_g|_{Z_G(K)}\}_{g \in G})$ is a crossed Frobenius G -algebra, where $\varphi_g(a) = \sum_i p_i^g a z q_i^g$ for all $a \in K$ and all $g \in G$.

Proof The unit of $Z_G(K)$ is $\Psi(z^2) = z$. By Lemma 4.12, we have the equality

$$(6) \quad \sum_i p_i^h b z' q_i^h c = \sum_i c p_i^{g^h} b z' q_i^{g^h}$$

for all $c \in K_{g^{-1}}$, $z' \in K_e$, $b \in K$ and $g, h \in G$. Taking $z' = 1$ and $g = h = e$ gives

$$(7) \quad \begin{aligned} \Psi(a \Psi(b) z^{-1}) &= \sum_{i,j} p_i^e a p_j^e b q_j^e z^{-1} q_i^e = \sum_{i,j} p_i^e a p_j^e b q_j^e q_i^e z^{-1} \\ &= \sum_{i,j} p_i^e a q_i^e p_j^e b q_j^e z^{-1} = \Psi(a) \cdot \Psi(b), \end{aligned}$$

which implies that $Z_G(K)$ is closed under multiplication. Restriction of η to $Z_G(K)$ is an inner product and hence $(Z_G(K), \eta|_{Z_G(K)})$ is a Frobenius G -algebra. For any $b \in K$ and for all $h \in G$, we have

$$\Psi(z \varphi_h(b)) = \sum_j p_j^e z \left(\sum_i p_i^h b z q_i^h \right) q_j^e = \sum_{i,j} p_j^e z q_j^e p_i^h b z q_i^h = \sum_i p_i^h b z q_i^h = \varphi_h(b),$$

which shows that $\varphi_h(K) \subset Z_G(K)$. Similarly, for any $\sum_i p_i^e a q_i^e \in Z_G(K)$, we have

$$\varphi_e \left(\sum_i p_i^e a q_i^e \right) = \sum_j p_j^e \left(\sum_i p_i^e a q_i^e \right) z q_j^e = \sum_{i,j} p_j^e z q_j^e p_i^e a q_i^e = \sum_i p_i^e a q_i^e,$$

showing $\varphi_e|_{Z_G(K)} = \text{id}_{Z_G(K)}$. Note that, for any $g \in G$ and $a \in K$, we have

$$\varphi_g(\Psi(a)) = \sum_j p_j^g \left(\sum_i p_i^e a q_i^e \right) z q_j^g = \sum_j p_j^g z q_j^g \sum_i p_i^e a q_i^e = \sum_i p_i^g a q_i^g$$

and using this we have the equality, for all $\bar{a} = \Psi(a), \bar{b} = \Psi(b) \in Z_G(K)$ and $g \in G$,

$$\begin{aligned} \varphi_g(\bar{a} \cdot \bar{b}) &= \sum_k p_k^g \left(\sum_{i,j} p_i^e a q_i^e p_j^e b q_j^e z^{-1} \right) z q_k^g \\ &= \left(\sum_i p_i^g a q_i^g \right) \left(\sum_j p_j^g b q_j^g \right) \left(\sum_k p_k^g q_k^g \right) = \varphi_g(\bar{a}) \cdot \varphi_g(\bar{b}), \end{aligned}$$

showing φ_g is an algebra homomorphism. For the last equality, we have $\sum_k p_k^g q_k^g = z^{-1}$ since $\sum_k p_k^g q_k^g = \sum_{i,j} p_i^e \beta_g^j \beta_{g^{-1}}^j q_i^e = \sum_i p_i^e q_i^e = z^{-1}$, where $\sum_j \beta_g^j \otimes \beta_{g^{-1}}^j \in K_g \otimes_{K_e} K_{g^{-1}}$ is the inverse of $1 \in K_e$ under the product map $K_h \otimes_{K_e} K_{h^{-1}} \rightarrow K_e$ (compare the first equality with the bimodule map corresponding to the second 2-morphism in Figure 28). We also have

$$\varphi_g(\varphi_h(\Psi(b))) = \sum_j p_j^g \left(\sum_i p_i^h b q_i^h \right) z q_j^g = \sum_j p_j^g z q_j^g \sum_i p_i^{g h} b q_i^{g h} = \varphi_{gh}(\Psi(b))$$

for all $g, h \in G$ and $b \in K$, which also implies that $\varphi_{g^{-1}}$ is the inverse of φ_g for all $g \in G$. For all $\bar{a} = \Psi(a), \bar{b} = \Psi(b) \in Z_G(K)$ and $g \in G$, using the cyclic symmetry of η , we have

$$\begin{aligned} \eta(\varphi_g(\bar{a}), \bar{b}) &= \eta\left(\sum_i p_i^g \bar{a} z q_i^g, \sum_j p_j^e b q_j^e\right) = \eta\left(\bar{a}, \sum_{i,j} z q_i^g p_j^e b q_j^e p_i^g\right) \\ &= \eta\left(\bar{a}, \sum_{i,k} p_k^{g^{-1}} b q_k^{g^{-1}} z q_i^g p_i^g\right) = \eta(\bar{a}, \varphi_{g^{-1}}(\bar{b})), \end{aligned}$$

showing the inner product η is invariant under $\varphi: G \rightarrow \text{Aut}(Z_G(K))$. For any $\bar{c} = \Psi(c) \in Z_G(K)_h$, we have

$$\varphi_h(\bar{c}) = \sum_i p_i^h c q_i^h = \sum_{i,j} p_i^e \beta_h^j c \beta_{h^{-1}}^j q_i^e = \sum_i p_i^e c q_i^e = \bar{c},$$

where $\sum_j \beta_h^j \otimes \beta_{h^{-1}}^j \in K_h \otimes K_{h^{-1}}$ is the inverse of $1 \in K_e$ under the product bimodule map. This shows that φ_h acts by the identity on $Z_G(K)_h$ for all $h \in G$. Equation (6) gives $\varphi_g(a)b = b\varphi_{h^{-1}g}(a)$ for $a \in K, b \in K_h$ and $g, h \in G$. In this case, by taking $g = h$, we have $\varphi_h(a)b = ba$. Let $\mu_c: K \rightarrow K$ be multiplication by $c \in K$; then, for any $g, h \in G$ and $c \in K_{ghg^{-1}h^{-1}}$, we have⁷

$$\begin{aligned} \text{Tr}(\mu_c \varphi_h: K_g \rightarrow K_g) &= \sum_i \eta(c \varphi_h(p_i^g), q_i^g) = \sum_{i,j} \eta(c p_j^h p_i^g z q_j^h, q_i^g) \\ &= \sum_{i,j} \eta(q_i^g c p_j^h z p_i^g, q_j^h) = \sum_j \eta(\varphi_{g^{-1}}(c p_j^h), q_j^h) \\ &= \text{Tr}(\varphi_{g^{-1}} \mu_c: K_h \rightarrow K_h). \quad \square \end{aligned}$$

Any 2-dimensional extended HFT produces a nonextended one by restricting it to a symmetric monoidal full subcategory $X\mathbb{C}\text{ob}_2$ of $X\text{Bord}_2$, defined as follows. The objects of $X\mathbb{C}\text{ob}_2$ are $\{\overset{g}{\underset{g}{\rightrightarrows}}\}_{g \in G}$, the empty 1-morphism in $X\text{Bord}_2$, and disjoint

⁷See [28] for the first equality.

unions of these 1–morphisms. The morphisms of $X\mathbb{C}\text{ob}_2$ are the 2–morphisms of $X\text{Bord}_2$ among these 1–morphisms. We define a symmetric monoidal functor $D: X\mathbb{C}\text{ob}_2 \rightarrow X\text{Cob}_2$ by $\begin{array}{c} e \rightarrow g \\ \leftarrow \end{array} \mapsto \begin{array}{c} \leftarrow g \\ \rightarrow \end{array}$ for any $g \in G$. On morphisms, D forgets a point on each boundary component and takes the corresponding relative homotopy class. Using the definitions, it is not hard to see that D is an equivalence of categories. By the restriction of $Z: X\text{Bord}_2 \rightarrow \text{Alg}_{\mathbb{k}}^2$ to $X\mathbb{C}\text{ob}_2$ above, we mean precomposing Z with $D^{-1}: X\text{Cob}_2 \rightarrow X\mathbb{C}\text{ob}_2$.

Corollary 4.22 *Let $Z: X\text{Bord}_2 \rightarrow \text{Alg}_{\mathbb{k}}^2$ be an extended HFT giving (A, B, ζ) . Then, the nonextended HFT obtained from Z by restricting to $X\mathbb{C}\text{ob}_2$ is the nonextended HFT associated to the G –center of the quasibiangular G –algebra (A, η_A) .*

Proof Proceeding with the notation used in the proof of [Theorem 4.19](#), the image of a g –labeled circle under Z is given by

$$A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_g = \{b \in A_g \mid a \cdot b = b \cdot a \text{ for all } a \in A_e\}.$$

The G –center of (A, η_A) is $Z_G(A) = \bigoplus_{g \in G} \Psi(A_g)$. For any $a \in A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_g$, we have

$$a = 1 \cdot a = \left(\sum_i p_i^e z q_i^e \right) a = \sum_i p_i^e a z q_i^e = \Psi(a z) \in \Psi(A_g)$$

and, for any $\sum_i p_i^e a q_i^e \in \Psi(A_g)$ and $b \in A_e$, we have

$$\left(\sum_i p_i^e a q_i^e \right) b = \sum_i p_i^e a q_i^e b = \sum_i b p_i^e a q_i^e = b \left(\sum_i p_i^e a q_i^e \right),$$

where the middle equality is the result of [Lemma 4.12](#). Thus, we have $A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_g = \Psi(A_g)$ for all $g \in G$. The third 2–morphism in [Figure 30](#) gives the crossed structure on the restricted HFT and it corresponds to the sequence of compositions

$$\begin{aligned} 1 \otimes a \mapsto 1 \otimes \sum_j a_j b_j \otimes a &\mapsto \sum_j 1 \otimes \beta_h^j z \beta_{h-1}^j \otimes a \mapsto 1 \otimes \sum_j \beta_h^j \left(\sum_i p_i^e \otimes z q_i^e \right) \beta_{h-1}^j \otimes a \\ &\mapsto 1 \otimes \sum_i p_i^h \otimes z q_i^h \otimes a \mapsto 1 \otimes \sum_i p_i^h a z q_i^h, \end{aligned}$$

which coincides with the crossed structure of $Z_G(A)$. □

Example 4.23 Let \mathbb{k} be an algebraically closed field. Then separable \mathbb{k} –algebras are the same as semisimple \mathbb{k} –algebras. By the Artin–Wedderburn structure theorem, any separable algebra is isomorphic to a product of finitely many matrix algebras

over \mathbb{k} . Consider the G -algebra $A = \bigoplus_{g \in G} A_g$ whose principal component is a product $A_e = \prod_{i=1}^n M_{k_i}(\mathbb{k})$ of $k_i \times k_i$ matrix algebras over \mathbb{k} such that each k_i is invertible in \mathbb{k} and each component is given by $A_g = \ell_g A_e$, where ℓ_g is a basis, ie for any $a \in A_g$, there exists $b \in A_e$ such that $a = \ell_g b$. Define an inner product η on A as

$$\eta(a, b) = \begin{cases} r \operatorname{Tr}(L_{ab}: A_e \rightarrow A_e) & \text{when } ab \in A_e, \\ 0 & \text{otherwise,} \end{cases}$$

where $r \in \mathbb{k}$ is invertible and L_{ab} is the left multiplication by ab map. We can express the inner product concretely as $\eta(\ell_g \prod_{i=1}^n A_i, \ell_{g^{-1}} \prod_{i=1}^n B_i) = r \sum_{i=1}^n k_i \operatorname{Tr}(A_i B_i)$. For each $g \in G$, an inner product element can be chosen as

$$\eta_g^- = r^{-1} \prod_{i=1}^n k_i^{-1} \sum_{\alpha, \beta=1}^{k_i} \ell_g E_{\alpha, \beta} \otimes \ell_{g^{-1}} E_{\beta, \alpha} \in A_g \otimes A_{g^{-1}},$$

where $E_{\alpha, \beta}$ is the (α, β) -elementary matrix. In this case, the central element $z \in A_e$ is given by $(rI_{k_1}, \dots, rI_{k_n})$, where I_{k_i} denotes the $k_i \times k_i$ identity matrix. Note that $\prod_{i=1}^n k_i^{-1} \sum_{\alpha, \beta=1}^{k_i} E_{\alpha, \beta} \otimes E_{\beta, \alpha}$ is a separability idempotent of A_e . Thus, the map $\Psi: A_g \rightarrow A_g$ is given by

$$\Psi\left(\ell_g \prod_{i=1}^n A_i\right) = r^{-1} \prod_{i=1}^n k_i^{-1} \sum_{\alpha, \beta=1}^{k_i} E_{\alpha, \beta} (\ell_g A_i) E_{\beta, \alpha} = r^{-1} \prod_{i=1}^n k_i^{-1} \ell_g \operatorname{Tr}(A_i) I_{k_i},$$

which is a projection onto its center $\ell_g \mathbb{k}^n$.

4.4 The bicategory of 2-dimensional extended X -HFTs

Until now we have studied the objects of $\mathbb{XP}(\operatorname{Alg}_{\mathbb{k}}^2)$. **Theorem 4.10** implies that studying 1- and 2-morphisms of $\mathbb{XP}(\operatorname{Alg}_{\mathbb{k}}^2)$ leads us to a bicategory equivalent to $\mathcal{E}\text{-}\mathcal{HFJ}(X, \operatorname{Alg}_{\mathbb{k}}^2)$. Let Z_0 and Z_1 be extended HFTs with target X giving triples (A, B, ζ) and (A', B', ζ') , respectively. A 1-morphism $\alpha: Z_0 \rightarrow Z_1$ in $\mathbb{XP}(\operatorname{Alg}_{\mathbb{k}}^2)$ gives 1-morphisms $\alpha_0(\bullet^+) = {}_{A'_e} R_{A_e}$ and $\alpha_0(\bullet^-) = {}_{B'_e} S_{B_e}$, and 2-morphisms

$$\begin{aligned} \alpha_1(+ \xrightarrow{g} +): {}_{A'_e} A'_g \otimes_{A'_e} R_{A_e} &\rightarrow {}_{A'_e} R \otimes_{A_e} (A_g)_{A_e}, \\ \alpha_1(- \xrightarrow{g} -): {}_{B'_e} B'_g \otimes_{B'_e} S_{B_e} &\rightarrow {}_{B'_e} S \otimes_{B_e} (B_g)_{B_e}, \\ \alpha_1(\overset{+}{\curvearrowright} \xrightarrow{g}): {}_{A'_e \otimes B'_e} (M'_g)_{\mathbb{k}} &\rightarrow {}_{A'_e \otimes B'_e} (R \otimes S) \otimes_{A_e \otimes B_e} (M_g)_{\mathbb{k}}, \\ \alpha_1(g \xleftarrow{\curvearrowright} +): {}_{\mathbb{k}} N'_g \otimes_{B'_e \otimes A'_e} &(S \otimes R)_{B_e \otimes A_e} \rightarrow {}_{\mathbb{k}} (N_g)_{B_e \otimes A_e}, \end{aligned}$$

which are isomorphisms for all $g \in G$ and G -graded bimodules M, M', N and N' are the components of ζ and ζ' . These morphisms are natural with respect to generating 2-morphisms. Naturality with respect to graded multiplication

$$\begin{array}{ccc} & g' & g \\ & \downarrow & \downarrow \\ + & \square & + \\ & g & g' \end{array}$$

leads to the commutativity of the diagram

$$\begin{array}{ccc} A'_e(A'_{g'}) \otimes_{A'_e} A'_g \otimes_{A'_e} R_{A_e} & \xrightarrow{\alpha_1 \left(\begin{array}{ccc} & g' & g \\ & \downarrow & \downarrow \\ + & \square & + \\ & g & g' \end{array} \right)} & A'_e R \otimes_{A_e} A_{g'} \otimes_{A_e} (A_g)_{A_e} \\ Z_1 \left(\begin{array}{ccc} & g' & g \\ & \downarrow & \downarrow \\ + & \square & + \\ & g & g' \end{array} \right) \downarrow & & \downarrow Z_0 \left(\begin{array}{ccc} & g' & g \\ & \downarrow & \downarrow \\ + & \square & + \\ & g & g' \end{array} \right) \\ A'_e A'_{gg'} \otimes_{A'_e} R_{A_e} & \xrightarrow{\alpha_1 \left(\begin{array}{ccc} & g & g' \\ & \downarrow & \downarrow \\ + & \square & + \end{array} \right)} & A'_e R \otimes_{A_e} (A_{gg'})_{A_e} \end{array}$$

for all $g, g' \in G$. We denote bimodules $A'_e A'_g \otimes_{A'_e} R_{A_e}$ and $A'_e R \otimes_{A_e} (A_g)_{A_e}$ by $A'_e (R'_g)_{A_e}$ and $A'_e (R''_g)_{A_e}$, respectively. Commutativity of the above diagram implies that they are naturally isomorphic. Thus, we can use one of them and denote it by R_g . Similarly, S_g denotes a (B'_e, B_e) -bimodule. These assignments and naturality with respect to

$$\left\{ \begin{array}{ccc} & g' & g \\ & \downarrow & \downarrow \\ + & \square & + \\ & g & g' \end{array} \right\}_{g, g' \in G}$$

turn these bimodules into G -graded (A', A) - and (B', B) -bimodules $R = \bigoplus_{g \in G} R_g$ and $S = \bigoplus_{g \in G} S_g$, respectively. Similarly, naturality with respect to G -module generators turns collections $\{\alpha_1(\begin{smallmatrix} + & \xrightarrow{\quad} & g \end{smallmatrix})\}_{g \in G}$ and $\{\alpha_1(g \xleftarrow{\quad} +)\}_{g \in G}$ into G -graded $(A' \otimes B', \mathbb{k})$ - and $(\mathbb{k}, B \otimes A)$ -bimodule maps, respectively.

Using $\alpha_0(\bullet^-)$, we define a 1-morphism $\alpha'_0(\bullet^+) = A_e R'_{A'_e}$ by

$$\begin{aligned} \alpha'_0(\bullet^+) &= [Z_1(\begin{smallmatrix} \leftarrow & + \end{smallmatrix}) \otimes \text{id}_{Z_0(\bullet^+)}] \circ [\alpha(\bullet^-) \otimes \sigma_{Z_0(\bullet^+), Z_1(\bullet^+)}] \circ [Z_0(\begin{smallmatrix} + & \xrightarrow{\quad} \end{smallmatrix}) \otimes \text{id}_{Z_1(\bullet^+)}], \\ \alpha'_0(\bullet^+) &= [(N'_e)_{B'_e \otimes A'_e} \otimes_{\mathbb{k} A_e} (A_e)_{A_e}] \otimes_{B'_e \otimes A'_e \otimes A_e} [B'_e S_{B_e} \otimes_{\mathbb{k}} \sigma_{A_e, A'_e}] \\ &\quad \otimes_{B_e \otimes A_e \otimes A'_e} [B_e \otimes_{A_e} M_e \otimes_{\mathbb{k}} A'_e (A'_e)_{A'_e}], \end{aligned}$$

where σ is the symmetric braiding of $\text{Alg}_{\mathbb{k}}^2$. Using $\alpha'_0(\bullet^+)$, we define a 2-morphism

$$\begin{aligned} \alpha'_1(\begin{smallmatrix} + & \xrightarrow{g} & + \end{smallmatrix}) &= Z_0(\begin{smallmatrix} + & \xrightarrow{g} & + \end{smallmatrix}) \circ \alpha'_0(\bullet^+) \rightarrow \alpha'_0(\bullet^+) \circ Z_1(\begin{smallmatrix} + & \xrightarrow{g} & + \end{smallmatrix}), \\ \alpha'_1(\begin{smallmatrix} + & \xrightarrow{g} & + \end{smallmatrix}) &= A_e A_e \otimes_{A_e} R'_{A'_e} \rightarrow A_e R' \otimes_{A'_e} (A'_g)_{A'_e}. \end{aligned}$$

Using naturality, R' is turned into a G -graded (A, A') -bimodule $R' = \bigoplus_{g \in G} R'_g$. The 1-morphism $\alpha_0(\bullet^-)$ can be obtained from $\alpha'_0(\bullet^+)$ by applying the composition

$Z_1(\square) \circ \text{id}_{\alpha_0(\bullet^-)} \circ Z_0(\square)$ to the 1-morphism

$$[Z_1(\overset{\leftarrow}{\curvearrowright}) \otimes Z_0(\overset{\leftarrow}{\curvearrowright}) \otimes \text{id}_{Z_1(\bullet^-)}] \circ [\alpha_0(\bullet^-) \otimes \sigma_{Z_0(\bullet^+), Z_1(\bullet^+)} \otimes \sigma_{Z_1(\bullet^-), Z_0(\bullet^-)}] \\ \circ [Z_0(\overset{\rightarrow}{\curvearrowright}) \otimes Z_1(\overset{\rightarrow}{\curvearrowright}) \otimes \text{id}_{Z_0(\bullet^-)}]$$

and, similarly, $\alpha_1(\overset{\leftarrow}{\curvearrowright})$ can be obtained from $\alpha'_1(\overset{\rightarrow}{\curvearrowright})$. Likewise, using $\alpha'_0(\bullet^+)$ in the images of cusp generators under Z_0 , the 2-morphisms $\alpha'_1(\overset{\rightarrow}{\curvearrowright})$ and $\alpha'_1(g \overset{\leftarrow}{\curvearrowright})$ are defined, and $\alpha_1(\overset{\rightarrow}{\curvearrowright})$ and $\alpha_1(g \overset{\leftarrow}{\curvearrowright})$ can be obtained from these 2-morphisms.

As in the proof of Theorem 4.19, using G -graded Morita contexts ζ and ζ' , graded bimodules M and M' can be replaced by ${}_A \otimes_{A^{\text{op}}} A$ and ${}_{A'} \otimes_{(A')^{\text{op}}} A'$. We can also replace the graded bimodule S by R' using $\alpha'_0(\bullet^+)$. Thus, naturality with respect to G -module generators turns the collection $\{\alpha'_1(\overset{\rightarrow}{\curvearrowright})\}_{g \in G}$ into a bimodule map ${}_{A'} R' \otimes_A R'_{A'} \rightarrow {}_{A'} R' \otimes_A R'_{A'}$. Similarly, the collection $\{\alpha'_1(g \overset{\leftarrow}{\curvearrowright})\}_{g \in G}$ is turned into a bimodule map ${}_{A'} R' \otimes_{A'} R_A \rightarrow {}_A A_A$. Naturality with respect to cusp generators indicates that the compositions

$${}_{A'} R_A \rightarrow {}_{A'} R' \otimes_{A'} R'_{A'} \xrightarrow{\alpha'_1(\overset{\rightarrow}{\curvearrowright}) \otimes \text{id}} {}_{A'} R' \otimes_A R' \otimes_{A'} R_A \xrightarrow{\text{id} \otimes \alpha'_1(\overset{\leftarrow}{\curvearrowright})} {}_{A'} R' \otimes_A A_A \rightarrow {}_{A'} R_A, \\ {}_A R'_{A'} \rightarrow {}_A R' \otimes_{A'} R'_{A'} \xrightarrow{\text{id} \otimes \alpha'_1(\overset{\rightarrow}{\curvearrowright})} {}_A R' \otimes_{A'} R' \otimes_A R'_{A'} \xrightarrow{\alpha'_1(\overset{\leftarrow}{\curvearrowright}) \otimes \text{id}} {}_A A \otimes_A R'_{A'} \rightarrow {}_A R'_{A'}$$

are id_R and $\text{id}_{R'}$, respectively. In other words, α gives a G -graded Morita context between A and A' . Similarly, one can define $\alpha'_0(\overset{\leftarrow}{\curvearrowright})$ and obtain a G -graded Morita context between B and B' . Naturality with respect to Morse generators indicates that G -graded Morita contexts are compatible. Hence, α leads to two compatible G -graded Morita contexts. In the theory of bicategories, this means that both $\alpha_0(\bullet^+)$ and $\alpha_0(\bullet^-)$ are parts of two adjoint equivalences. Since an adjoint equivalence is the same as an equivalence (see [22, Proposition A.27]), Z_0 and Z_1 are equivalent extended HFTs.

Let $\alpha^1, \alpha^2: Z_0 \rightarrow Z_1$ be 1-morphisms in $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ and $\theta: \alpha^1 \rightarrow \alpha^2$ be a 2-morphism in $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$. Assume that Z_0 and Z_1 give triples (A, B, ζ) and (A', B', ζ') as before and 1-morphisms give $\alpha_0^1(\bullet^+) = {}_{A'_e} R_{A_e}$ and $\alpha_0^2(\bullet^+) = {}_{A'_e} P_{A_e}$. Then $\theta_0(\bullet^+): {}_{A'_e} R_{A_e} \rightarrow {}_{A'_e} P_{A_e}$ and the naturality of $\theta_0(\bullet^+)$ with respect to $\overset{\rightarrow}{\curvearrowright}$ is the commutativity of the diagram

$$\begin{array}{ccc} {}_{A'_e} A'_g \otimes_{A'_e} R_{A_e} & \xrightarrow{\alpha_1^1(\overset{\rightarrow}{\curvearrowright})} & {}_{A'_e} R \otimes_{A_e} (A_g)_{A_e} \\ \theta_0(\bullet^+) \downarrow & & \downarrow \theta_0(\bullet^+) \\ {}_{A'_e} A'_g \otimes_{A'_e} P_{A_e} & \xrightarrow{\alpha_1^2(\overset{\rightarrow}{\curvearrowright})} & {}_{A'_e} P \otimes_{A_e} (A_g)_{A_e} \end{array}$$

which shows that $\theta_0(\bullet^+)$ is a G -graded bimodule map. Assuming $(\alpha'_0)^1(\bullet^+) = {}_{A_e}R'_{A'_e}$ and $(\alpha'_0)^2(\bullet^+) = {}_{A'_e}P'_{A_e}$, we similarly have a graded bimodule map $\theta'_0(\bullet^+): {}_{A_e}R'_{A'_e} \rightarrow {}_{A'_e}P'_{A_e}$ using $\theta_0(\bullet^-)$ and $(\alpha'_0)^i(\bullet^+)$ for $i = 1, 2$. Naturality with respect to $\overset{+}{\dashrightarrow}g$ and $g\overset{-}{\dashleftarrow}+$ corresponds to the commutativity of these bimodule maps with the unit and counit of the adjunctions. In other words, θ leads to an equivalence of graded Morita contexts. In the same way, using B and B' , one gets another equivalence of graded Morita contexts.

Motivated by these observations, we define a bicategory Frob^G and a forgetful 2-functor $\mathcal{F}': \mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2) \rightarrow \text{Frob}^G$ as follows. The bicategory Frob^G has quasibiangular G -algebras as objects, compatible G -graded Morita contexts as 1-morphisms, and equivalences of G -graded Morita contexts as 2-morphisms. The forgetful 2-functor \mathcal{F}' maps an object of $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ giving (A, B, ζ) to A . On 1-morphisms, \mathcal{F}' maps $\alpha: Z_0 \rightarrow Z_1$ to a compatible G -graded Morita context between quasibiangular G -algebras whose principal components are $Z_0(\bullet^+)$ and $Z_1(\bullet^+)$. On 2-morphisms, \mathcal{F}' maps $\theta: \alpha^1 \rightarrow \alpha^2$ to an equivalence of the compatible G -graded Morita contexts. Composing \mathcal{F}' with the equivalence $\mathcal{E}\text{-}\mathcal{H}\mathcal{F}\mathcal{J}(X, \text{Alg}_{\mathbb{k}}^2) \simeq \mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$, we define \mathcal{F} .

Theorem 4.24 *The 2-functor \mathcal{F} is an equivalence of bicategories $\mathcal{E}\text{-}\mathcal{H}\mathcal{F}\mathcal{J}(X, \text{Alg}_{\mathbb{k}}^2) \simeq \text{Frob}^G$.*

Proof It is enough to show that \mathcal{F}' is an equivalence and we use the Whitehead theorem (Theorem 4.3). For a given quasibiangular G -algebra A , the triple $(A, A^{\text{op}}, \text{id})$ gives an object Z of $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ such that $\mathcal{F}'(Z) = A$. Let α be a compatible G -graded Morita context between quasibiangular G -algebras A and A' . Then triples $(A, (A')^{\text{op}}, \alpha)$ and $(A', A^{\text{op}}, \alpha)$ give objects Z_0 and Z_1 in $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ such that $\mathcal{F}'(\alpha') = \alpha$, where $\alpha': Z_0 \rightarrow Z_1$.

For any two 1-morphisms $\alpha^1, \alpha^2: Z_0 \rightarrow Z_1$, we claim that

$$\mathcal{F}'(\alpha^1, \alpha^2): \text{Hom}(\alpha^1, \alpha^2) \rightarrow \text{Hom}(\mathcal{F}'(\alpha^1), \mathcal{F}'(\alpha^2))$$

is an injection. Assume that different 2-morphisms $\theta^1, \theta^2: \alpha^1 \rightarrow \alpha^2$ in $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ give the same equivalence of G -graded Morita contexts. This means that pairs $(\theta_0^1(\bullet^-), (\theta_0^1)^1(\bullet^-))$ and $(\theta_0^2(\bullet^-), (\theta_0^2)^2(\bullet^-))$ give different graded bimodule maps while the images of θ^1 and θ^2 under \mathcal{F}' , $(\theta_0^1(\bullet^+), (\theta_0^1)^1(\bullet^+))$ and $(\theta_0^2(\bullet^+), (\theta_0^2)^2(\bullet^+))$, give the same graded bimodule maps. This is a contradiction because each $(\theta_0^i)^i(\bullet^-)$ is obtained from $(\theta_0)^i(\bullet^+)$ and each $(\theta_0^i)^i(\bullet^+)$ is obtained from $(\theta_0)^i(\bullet^-)$ for $i = 1, 2$.

For the surjectivity, let $\theta: \mathcal{F}'(\alpha^1) \rightarrow \mathcal{F}'(\alpha^2)$ be an equivalence of graded Morita contexts. Then the equivalence of graded Morita contexts $(\theta_0(\bullet^-), \theta'_0(\bullet^-))$ can be obtained from $\theta_0(\bullet^+), \theta'_0(\bullet^+), (\alpha'_0)^1(\bullet^-)$, and $(\alpha'_0)^2(\bullet^-)$. \square

Corollary 4.25 *Two triples (A_1, B_1, ζ_1) and (A_2, B_2, ζ_2) produce equivalent extended X -HFTs if and only if there exists a compatible G -graded Morita context between A_1 and A_2 .*

Lastly, we comment on the relation between extended HFTs whose targets are related by covering maps. Let $Y \simeq K(H, 1)$ be a pointed CW-complex for a nontrivial subgroup $H \leq G$ and $p: (Y, y) \rightarrow (X, x)$ be a covering. Then, any Y -manifold can be turned into an X -manifold by postcomposing a representative of the characteristic map with p . This gives a symmetric monoidal 2-functor $\iota_H: Y\text{Bord}_2 \rightarrow X\text{Bord}_2$ and precomposing any extended X -HFT with ι_H yields an extended Y -HFT. Moreover, for any symmetric monoidal bicategory \mathcal{C} , precomposition of a \mathcal{C} -valued extended X -HFT with ι lifts to a 2-functor $\text{SymMon}(X\text{Bord}_2, \mathcal{C}) \rightarrow \text{SymMon}(Y\text{Bord}_2, \mathcal{C})$ by forgetting the naturality of transformations with respect to $G \setminus H$ -labeled 1-morphisms. Correspondingly, there is a 2-functor $\mathbb{X}\mathbb{P}(\mathcal{C}) \rightarrow \mathbb{Y}\mathbb{P}(\mathcal{C})$, where $\mathbb{X}\mathbb{P}$ and $\mathbb{Y}\mathbb{P}$ are the presentations of $X\text{Bord}_2$ and $Y\text{Bord}_2$, respectively. When \mathcal{C} is $\text{Alg}_{\mathbb{k}}^2$, the functor $\text{Frob}^G \rightarrow \text{Frob}^H$ is given by forgetting the $G \setminus H$ components of quasibiangular G -algebras, compatible G -graded Morita contexts, and equivalences of G -graded Morita contexts. In other words, a G -graded Morita context can be considered as a collection of Morita contexts indexed by the subgroups of G (see [3]).

4.5 The $(G \times \text{SO}(2))$ -structured cobordism hypothesis

A different approach to categorical classification of (fully) extended oriented HFTs is given by the structured cobordism hypothesis due to Lurie [15]. The cobordism hypothesis [1; 15; 2] was conjectured by Baez and Dolan in their seminal paper [2]. Lurie [15] reformulated the cobordism hypothesis using (∞, n) -categories and generalized it to a structured cobordism hypothesis using homotopy fixed points.

The bordism category involved in the structured cobordism hypothesis consists of manifolds with corresponding structures. For a topological group Γ and a fixed continuous homomorphism $\chi: \Gamma \rightarrow O(n)$, let $\zeta_\chi: \mathbb{R}^n \times_\Gamma E\Gamma$ denote the associated rank n vector bundle over the classified space $B\Gamma$. Then, a Γ -structure on a manifold M of dimension $k \leq n$ consists of a continuous map $f: M \rightarrow B\Gamma$ and an isomorphism

$TM \oplus \mathbb{R}^{n-k} \rightarrow f^* \zeta_\chi$ of vector bundles, where \mathbb{R}^{n-k} is a trivial rank $n - k$ vector bundle. In the following theorem the category Bord_n^Γ consists of manifolds equipped with Γ -structures for a fixed homomorphism χ .

(Γ, χ) -structured cobordism hypothesis (Lurie [15]) Let \mathcal{C} be a symmetric monoidal (∞, n) -category (see [5]) and Bord_n^Γ be the symmetric monoidal Γ -structured cobordism (∞, n) -category for a group Γ . Then there is a canonical equivalence of (∞, n) -categories

$$\text{Fun}^\otimes(\text{Bord}_n^\Gamma, \mathcal{C}) \xrightarrow{\sim} ((\mathcal{C}^{\text{fd}})^\sim)^{h\Gamma},$$

where Fun^\otimes is the (∞, n) -category of symmetric monoidal functors between symmetric monoidal (∞, n) -categories, \mathcal{C}^{fd} is the (∞, n) -subcategory of fully dualizable objects with duality data, $(\mathcal{C}^{\text{fd}})^\sim$ is the underlying ∞ -groupoid and $((\mathcal{C}^{\text{fd}})^\sim)^{h\Gamma}$ is the ∞ -groupoid of homotopy Γ -fixed points given by

$$((\mathcal{C}^{\text{fd}})^\sim)^{h\Gamma} = \text{Hom}_\Gamma(E\Gamma, (\mathcal{C}^{\text{fd}})^\sim),$$

where $E\Gamma$ is a weakly contractible ∞ -groupoid equipped with a free Γ -action.

Remark A 2-dimensional EHFT with target $X \simeq K(G, 1)$, ie a classifying space BG , is a $(G \times \text{SO}(2))$ -structured 2-dimensional ETFT, where $\chi: G \times \text{SO}(2) \rightarrow O(2)$ is given by $(g, A) \mapsto A$.

When \mathbb{k} is an algebraically closed field of characteristic zero, Davidovich [6] showed that, for a finite group G , homotopy $(G \times \text{SO}(2))$ -fixed points in $(\text{Alg}_{\mathbb{k}}^{\text{fd}})^\sim$ are given by G -equivariant algebras. A G -equivariant algebra is a strongly graded Frobenius G -algebra with semisimple principal component. Her methods do not particularly require G to be finite and can be extended to discrete groups directly. Since the notions of separability and semisimplicity for a \mathbb{k} -algebra are equivalent when \mathbb{k} is an algebraically closed field of characteristic zero, the objects of Frob^G and the objects of $((\text{Alg}_{\mathbb{k}}^{\text{fd}})^\sim)^{h(G \times \text{SO}(2))}$ coincide.

Assume that \mathbb{k} is an algebraically closed field of characteristic zero. The Artin-Wedderburn theorem implies that any separable \mathbb{k} -algebra is isomorphic to a product of matrix algebras over \mathbb{k} . Let $A_e = \text{End}(V_1) \times \text{End}(V_2) \times \dots \times \text{End}(V_n)$ be such an algebra, where V_1, V_2, \dots, V_n are finite-dimensional \mathbb{k} -vector spaces. Recall that $A = \bigoplus_{g \in G} A_g$ is strongly graded by the generators, leading to bimodule isomorphisms $\{\tau_{g, g'}: A_{g'} \otimes_{A_e} A_g \xrightarrow{\cong} A_{gg'}\}_{g, g' \in G}$; that is, each A_g is an invertible (A_e, A_e) -bimodule. Under the above assumption on A_e , these isomorphisms form a function

$\tau: G \times G \rightarrow (\mathbb{k}^*)^n$. Moreover, the relations involving these generators give the commutative diagram, for all $g, g', g'' \in G$,

$$\begin{array}{ccccc}
 (A_{g''} \otimes_{A_e} A_{g'}) \otimes_{A_e} A_g & \xrightarrow{\tau(g',g'') \otimes \text{id}} & A_{g'g''} \otimes_{A_e} A_g & \xrightarrow{\tau(g,g'g'')} & A_{gg'g''} \\
 \cong \downarrow & & & & \downarrow \text{id} \\
 A_{g''} \otimes_{A_e} (A_{g'} \otimes_{A_e} A_g) & \xrightarrow{\text{id} \otimes \tau(g,g')} & A_{g''} \otimes_{A_e} A_{gg'} & \xrightarrow{\tau(gg',g'')} & A_{gg'g''}
 \end{array}$$

and isotopy classes of G -linear diagrams generate the relations, which can be expressed as the commutative diagram, for all $g \in G$,

$$\begin{array}{ccccc}
 A_g \otimes_{A_e} A_e & \xrightarrow{\tau(e,g)} & A_g & \xleftarrow{\tau(g,e)} & A_e \otimes_{A_e} A_g \\
 & \searrow \cong & \downarrow \text{id} & \swarrow \cong & \\
 & & A_g & &
 \end{array}$$

which imply that τ is a normalized 2-cocycle. Davidovich [6] showed that any invertible (A_e, A_e) -bimodule is isomorphic to one of the form

$$\text{Hom}_{\mathbb{k}}(V_{\sigma(1)}, V_1) \times \text{Hom}_{\mathbb{k}}(V_{\sigma(2)}, V_2) \times \cdots \times \text{Hom}_{\mathbb{k}}(V_{\sigma(n)}, V_n)$$

for some permutation $\sigma \in S_n$; denote this bimodule by A^σ . Since the direct sum $A = \bigoplus_{g \in G} A_g$ forms a G -algebra, permutations indeed form a homomorphism $\sigma: G \rightarrow S_n$.

It is known that all traces on a matrix algebra are given as some (nonzero) constant multiple of the matrix trace. Thus, in the case of $A_e = \text{End}(V_1) \times \text{End}(V_2) \times \cdots \times \text{End}(V_n)$ there are constants $r_i \in \mathbb{k}^*$ for $i = 1, \dots, n$ and the inner product of the quasibimodular G -algebra $A = \bigoplus_{g \in G} A^\sigma(g)$ is given by $\eta(f, g) = \text{Tr}(r \circ (g \circ^\sigma f))$ for any $f \in A_g$ and $g \in A_{g^{-1}}$, where $r = (r_1 \text{id}_{V_1}, \dots, r_n \text{id}_{V_n})$ and \circ^σ is the composition of morphisms under σ such as $f_i \circ g_{\sigma(i)}$ for $f_i \in \text{Hom}_{\mathbb{k}}(V_{\sigma(i)}, V_i)$ and $g_{\sigma(i)} \in \text{Hom}_{\mathbb{k}}(V_{\sigma(\sigma(i))}, V_{\sigma(i)})$. Since the inner product is invariant under cyclic order, ie $\eta(f, g \cdot h) = \eta(h \cdot f, g) = \eta(h, f \cdot g)$, the vector $r \in (\mathbb{k}^*)^n$ must satisfy $\text{Im}(\sigma) \subseteq \text{Stab}_{S_n}(r)$, where S_n acts on r by permuting the entries. More explicitly, as an example, consider the products $(h \circ^\sigma g) \circ^\sigma f$ and $g \circ^\sigma (f \circ^\sigma h)$ for $f \in A_g, g \in A_{g'}$ and $h \in A_{(gg')^{-1}}$. Then the corresponding traces of these morphisms in $A_e = \text{End}(V_1) \times \text{End}(V_2) \times \cdots \times \text{End}(V_n)$ are related by the permutation $\sigma(gg') \in S_n$.

Using the above arguments, when \mathbb{k} is an algebraically closed field of characteristic zero, we can conclude that, up to an isomorphism, a quasibimodular G -algebra

$(A = \bigoplus_{g \in G} A_g, \eta)$ is determined by a Morita class of the principal component ($n \geq 1$), a normalized 2-cocycle $\tau: G \times G \rightarrow (\mathbb{k}^*)^n$, a homomorphism $\sigma: G \rightarrow S_n$, and an element $r \in (\mathbb{k}^*)^n$ with $\text{Im}(\sigma) \subseteq \text{Stab}_{S_n}(r)$.

Let $(A_2 M_{A_1, A_1} N_{A_2}, \kappa, \mu)$ be a graded Morita context between two quasibiangular G -algebras (A_1, η_1) and (A_2, η_2) which are determined by the normalized 2-cocycles τ_i , homomorphisms $\sigma_i: G \rightarrow S_n$, and elements $r_i \in (\mathbb{k}^*)^n$ with $\text{Im}(\sigma_i) \subseteq \text{Stab}_{S_n}(r_i)$ for $i = 1, 2$. Then M and N are invertible (A_2, A_1) - and (A_1, A_2) -bimodules, respectively, which means there exists $\sigma \in S_n$ such that M_e is isomorphic to

$$M^\sigma = \text{Hom}(V_{\sigma(1)}, W_1) \times \text{Hom}(V_{\sigma(2)}, W_2) \times \cdots \times \text{Hom}(V_{\sigma(n)}, W_n),$$

where $(A_i)_g = \text{Hom}(V_{\sigma_i^g(1)}, V_1) \times \cdots \times \text{Hom}(V_{\sigma_i^g(n)}, V_n)$ for all $g \in G$ and $\sigma_i^g = \sigma_i(g) \in S_n$ for $i = 1, 2$. Being a graded (A_2, A_1) -bimodule forces σ to satisfy $\sigma_2^g = \sigma \sigma_1^g \sigma^{-1}$ for all $g \in G$. In this case, nonprincipal components are given as $M_g = \text{Hom}(V_{\sigma'_g(1)}, W_1) \times \cdots \times \text{Hom}(V_{\sigma'_g(n)}, W_n)$, where $\sigma'_g = \sigma \sigma_1^g = \sigma_2^g \sigma$ for all $g \in G$. Using the similar arguments for the invertible (A_1, A_2) -bimodule N , we obtain $N_g = \text{Hom}(W_{\sigma''_g(1)}, V_1) \times \cdots \times \text{Hom}(W_{\sigma''_g(n)}, V_n)$ for $\sigma''_g = \sigma_1^{g^{-1}} \sigma^{-1} = \sigma^{-1} \sigma_2^{g^{-1}}$ for all $g \in G$.

Transferring the Frobenius form via the graded Morita context amounts to finding a central element in $(A_2)_e$ corresponding to $r_1 \in (\mathbb{k}^*)^n$. Using the identity component M_e , this element is given by $\sigma(r_1)(\text{id}_{W_1}, \text{id}_{W_2}, \dots, \text{id}_{W_n})$. Thus, we have the equality $\sigma(r_1) = r_2 \in (\mathbb{k}^*)^n$. The bimodule isomorphisms $\kappa: {}_{A_1}(A_1)_{A_1} \rightarrow {}_{A_1}N \otimes_{A_2} M_{A_1}$ and $\mu: {}_{A_2}M \otimes_{A_1} N_{A_2} \rightarrow {}_{A_2}(A_2)_{A_2}$ lead to a map $\phi: G \rightarrow (\mathbb{k}^*)^n$ and the graded Morita context equations produce a map $\phi: (A_1)_g \rightarrow (A_2)_g$ for all $g \in G$ such that the diagram

$$\begin{array}{ccc} (A_1)_g \otimes (A_1)_h & \xrightarrow{\phi_1(g,h)} & (A_1)_{gh} \\ \phi(g)\phi(h) \downarrow & & \downarrow \phi(gh) \\ (A_2)_g \otimes (A_2)_h & \xrightarrow{\phi_2(g,h)} & (A_2)_{gh} \end{array}$$

commutes for all $g, h \in G$. This means that normalized 2-cocycles $\phi_1, \phi_2: G \times G \rightarrow (\mathbb{k}^*)^n$ differ by a coboundary $\partial\phi$. Thus, we conclude that quasibiangular G -algebras, up to compatible G -graded Morita contexts, are in bijection with

$$\coprod_{r=1}^{\infty} \coprod_{[r] \in (\mathbb{k}^*)^n / S_n} H^2(G; (\mathbb{k}^*)^n) \times \text{Hom}(G, \text{Stab}_{S_n}(r)) / \sim,$$

where the equivalence \sim is given by conjugation. Using [Theorem 4.24](#), we derive the following proposition, which was previously proven by Davidovich [\[6\]](#):

Proposition 4.26 [6] *The set of equivalence classes of fully extended oriented 2–dimensional G –equivariant TFTs, ie EHFTs with $K(G, 1)$ target, with values in $\text{Alg}_{\mathbb{k}}^2$ is in bijection with*

$$\pi_0 \text{Fun}^{\otimes}(\text{Bord}_2^{G \times \text{SO}(2)}, \text{Alg}_{\mathbb{k}}^2) \cong \coprod_{r=1}^{\infty} \coprod_{[r] \in (\mathbb{k}^*)^n / S_n} H^2(G; (\mathbb{k}^*)^n) \times \text{Hom}(G, \text{Stab}_{S_n}(r)) / \sim,$$

where the equivalence \sim is given by conjugation.

The fact that $\text{Alg}_{\mathbb{k}}^2$ –valued fully extended oriented 2–dimensional G –equivariant TFTs are classified in two different ways — namely using the structured cobordism hypothesis and without using it — is an important step towards the verification of the $(G \times \text{SO}(2))$ –structured cobordism hypothesis for $\text{Alg}_{\mathbb{k}}^2$ –valued such TFTs. In this case, the $(G \times \text{SO}(2))$ –structured cobordism hypothesis gives an equivalence of bigroupoids $\mathcal{E}\text{–}\mathcal{HFT}(X, \text{Alg}_{\mathbb{k}}^2) \simeq ((\text{Alg}_{\mathbb{k}}^{\text{fd}}) \sim)^{h(G \times \text{SO}(2))}$. We have $\mathcal{E}\text{–}\mathcal{HFT}(X, \text{Alg}_{\mathbb{k}}^2) \simeq \text{Frob}^G$ by [Theorem 4.24](#) and [Davidovich \[6\]](#) showed that the fundamental bigroupoid of $((\text{Alg}_{\mathbb{k}}^{\text{fd}}) \sim)^{h(G \times \text{SO}(2))}$ is equivalent to the bigroupoid $\text{Grp}_2(BG, \mathcal{G}_{\text{ori}})$ of 2–functors, transformations, and modifications. Here the bigroupoid BG has one object, $|G|$ 1–morphisms, and only identity 2–morphisms, and the bigroupoid \mathcal{G}_{ori} has semisimple Frobenius algebras as objects, invertible bimodules compatible with Frobenius forms as 1–morphisms, and invertible bimodule maps as 2–morphisms (see [Proposition 3.3.2](#) in [\[6\]](#)).

[Davidovich](#) showed that every 2–functor $F : BG \rightarrow \mathcal{G}_{\text{ori}}$ gives rise to a quasibiangular G –algebra, and vice versa [\[6, Proposition 3.4.5\]](#). We define a 2–functor $\mathcal{F} : \text{Frob}^G \rightarrow \text{Grp}_2(BG, \mathcal{G}_{\text{ori}})$ as follows. The image of a quasibiangular G –algebra is the corresponding functor described above. Next, let $(A_2 M_{A_1, A_1} N_{A_2}, \kappa, \mu)$ be a G –graded compatible Morita context between two quasibiangular G –algebras (A_1, η_1) and (A_2, η_2) which are determined by two triples (τ_i, σ_i, r_i) for $i = 1, 2$ such that $M_e \cong M^\sigma$, as in the proof of [Proposition 4.26](#) above. Then we define $\mathcal{F}((A_2 M_{A_1, A_1} N_{A_2}, \kappa, \mu))$ to be the natural transformation between the corresponding 2–functors producing the bimodule M^σ (see the proof of [Proposition 4.26](#) given in [\[6\]](#)). Lastly, the image of an equivalence of G –graded Morita contexts under \mathcal{F} is the modification producing the invertible bimodule map $M^\sigma \rightarrow M^{\sigma'}$ between the corresponding bimodules described above.

[Proposition 4.26](#) implies that \mathcal{F} is essentially surjective on objects. The fact that any natural transformation $\eta : F_1 \rightarrow F_2$ between functors $F_1, F_2 : BG \rightarrow \mathcal{G}_{\text{ori}}$ is isomorphic to one producing a bimodule of the form M^σ (see the proof of [Proposition 4.26](#) given

in [6]) implies that \mathcal{F} is essentially full on 1–morphisms. Using similar arguments as in the proof of [Theorem 4.24](#), one can show that \mathcal{F} is fully faithful on 2–morphisms. Consequently, the 2–functor $\mathcal{F}: \text{Frob}^G \rightarrow \text{Grp}_2(BG, \mathcal{G}_{\text{ori}})$ is an equivalence by the Whitehead theorem for bicategories ([Theorem 4.3](#)).

Corollary 4.27 *For any discrete group G and any algebraically closed field \mathbb{k} of characteristic zero, the $(G \times \text{SO}(2))$ –structured cobordism hypothesis for $\text{Alg}_{\mathbb{k}}^2$ –valued oriented EHFTs with target $X \simeq K(G, 1)$ holds true.*

5 Extended unoriented X –HFTs and their classifications

In this section, by allowing X –manifolds and X –cobordisms to be nonorientable, we define 2–dimensional extended unoriented X –HFTs and classify them. The definition and classification of these theories are parallel with the oriented case and we only describe the changes. All of the manifolds and cobordisms in this section do not have any orientation data and they are not necessarily orientable.

5.1 The unoriented X –cobordism bicategory and its presentation

In this section, we define 2–dimensional extended unoriented HFTs with target $X \simeq K(G, 1)$, where every element of G has order two. To avoid repetition, from now on, we assume that G is such a group and X is a pointed $K(G, 1)$ –space. The restriction to such groups is not essential but for convenience.⁸ As in the oriented case, the 2–dimensional extended unoriented X –cobordism bicategory plays the essential role and it is defined using X –halations (see [Section 2.2](#)) as follows:

Definition 5.1 *The 2–dimensional extended unoriented X –cobordism bicategory $X\text{Bord}_2^{\text{un}}$ has*

- quadruples $(M, \widehat{M}_1, \widehat{M}_2, \widehat{g}_2)$ consisting of compact 0–manifolds equipped with two cooriented X –halations as objects,
- quintuples $(A, \widehat{A}_0, \widehat{A}_1, T, \widehat{p}_1)$ consisting of 1–dimensional marked X –manifolds equipped with two cooriented X –halations as 1–morphisms,
- isomorphism classes $[(S, \widehat{S}, R, \widehat{F})]$ of quadruples consisting of cobordism type (2) – X –surfaces equipped with a codimension zero X –halation as 2–morphisms.

⁸Order two elements appear as a result of new generators and we avoid keeping track of which elements are required to have order two and which ones are not. See [12] for the case of nonextended HFTs with arbitrary aspherical targets.

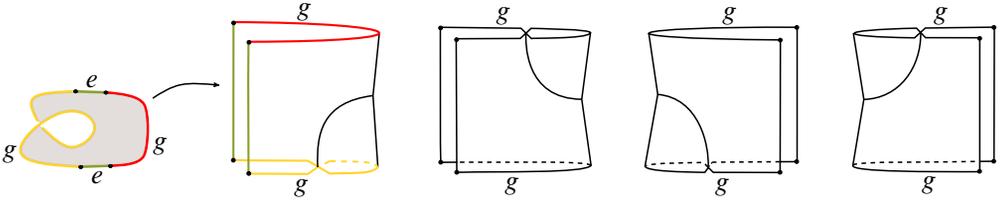


Figure 31: Additional generating 2-morphisms of $\mathcal{X}G_2^{\text{un}}$.

Similar to $X\text{Bord}_2$, the disjoint union operation is the symmetric monoidal product for $X\text{Bord}_2^{\text{un}}$.

Definition 5.2 For a symmetric monoidal bicategory \mathcal{C} , a \mathcal{C} -valued 2-dimensional extended unoriented homotopy field theory with target X is a symmetric monoidal 2-functor from $X\text{Bord}_2^{\text{un}}$ to \mathcal{C} .

It is not hard to modify G -linear, G -planar, and G -spatial diagrams for the unoriented setting. We only need to consider sheets with no additional orientation data. In this case, each diagram produces an unoriented version of the corresponding X -manifold. There are new sheet data for the fold singularity coming from $\langle 2 \rangle$ - X -surface representatives of a Möbius band. These $\langle 2 \rangle$ - X -surfaces are shown in Figure 31. This extra sheet data produces new relations coming from different possible gluings of these generators with themselves and with the earlier (orientable) generators. Figure 32 shows these relations of 2-morphisms of $X\text{Bord}_2^{\text{un}}$ instead of the corresponding unoriented G -planar diagrams. By generalizing the equivalence relations on the set of unoriented G -planar

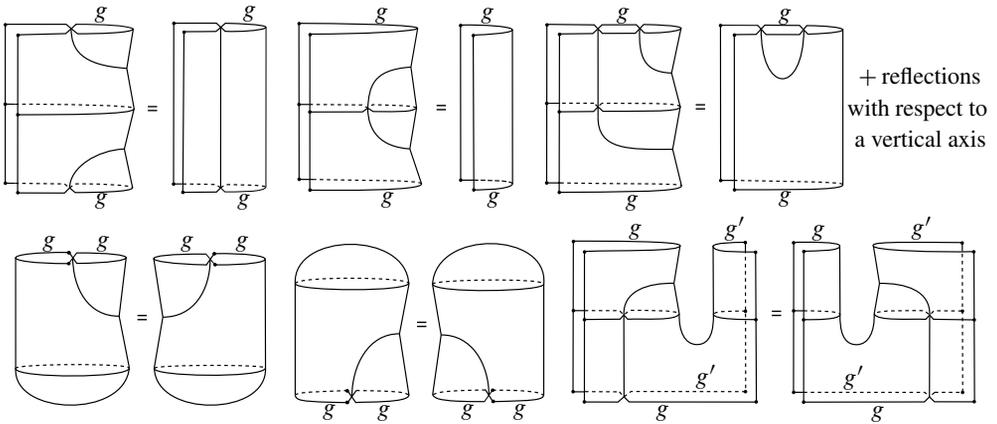


Figure 32: Additional generating relations of $\mathcal{X}\mathcal{R}^{\text{un}}$.

diagrams given by unoriented G -spatial diagrams, we obtain the following unoriented G -planar decomposition theorem:

Theorem 5.3 *The relative X -homeomorphism classes of unoriented cobordism type $\langle 2 \rangle$ - X -surfaces are in bijection with the equivalence classes of unoriented G -planar diagrams.*

Parallel to the oriented case, we define a new symmetric monoidal bicategory $\mathbb{X}B^{PD,un}$, which has unoriented points as objects, unoriented G -linear diagrams as 1-morphisms, and equivalence classes of unoriented G -planar diagrams as 2-morphisms. The unoriented G -planar decomposition theorem implies that $\mathbb{X}B^{PD,un}$ is symmetric monoidally equivalent to $X\text{Bord}_2^{un}$ via the Whitehead theorem for symmetric monoidal bicategories. The correspondence between G -planar diagrams and the string diagrams for unbiased semistrict symmetric monoidal bicategories gives the following result:

Theorem 5.4 *The symmetric monoidal bicategory $\mathbb{X}B^{PD,un}$ is a computadic unbiased semistrict symmetric monoidal 2-category with the presentation*

$$\mathbb{X}P^{un} = (\mathcal{X}G_0^{un}, \mathcal{X}G_1^{un}, \mathcal{X}G_2^{un}, \mathcal{X}R^{un})$$

which has one generating object, $\{\bullet\}$, G -linear diagrams of $\{\cdot \xrightarrow{g}, \xrightarrow{g} \cdot, \cdot \xleftarrow{g}, \xleftarrow{g} \cdot\}_{g \in G}$ as the generating 1-morphisms, G -planar diagram versions of elements in Figures 25 and 31 as the generating 2-morphisms, and G -planar diagram versions of pairs in Figures 26 and 32 as the generating relations.

The bicategory $\mathcal{E}\text{-}\mathcal{HFT}^{un}(X, \mathcal{C})$ of 2-dimensional extended unoriented X -HFTs has \mathcal{C} -valued unoriented extended X -HFTs as objects, symmetric monoidal transformations as 1-morphisms, and symmetric monoidal modifications as 2-morphisms. For any given symmetric monoidal bicategory \mathcal{C} , using the cofibrancy theorem, we state the classification of \mathcal{C} -valued 2-dimensional extended unoriented X -HFTs as the equivalence of bicategories $\mathcal{E}\text{-}\mathcal{HFT}^{un}(X, \mathcal{C}) \simeq \mathbb{X}P^{un}(\mathcal{C})$.

Remark There is a symmetric monoidal 2-functor $\text{Forget}^{or}: X\text{Bord}_2 \rightarrow X\text{Bord}_2^{un}$ given by forgetting the orientation. In the same way, any oriented or unoriented 2-dimensional extended TFT leads to an oriented or unoriented extended HFT, respectively,

by forgetting the X -manifold data. The diagram

$$\begin{array}{ccc}
 \mathcal{E}\text{-}\mathcal{TFT}^{\text{un}}(\mathcal{C}) & \xrightarrow{\text{Forget}^{\text{or}}} & \mathcal{E}\text{-}\mathcal{TFT}(\mathcal{C}) \\
 \text{Forget}^X \downarrow & & \downarrow \text{Forget}^X \\
 \mathcal{E}\text{-}\mathcal{HFT}^{\text{un}}(X, \mathcal{C}) & \xrightarrow{\text{Forget}^{\text{or}}} & \mathcal{E}\text{-}\mathcal{HFT}(X, \mathcal{C})
 \end{array}$$

indicates the universality of unoriented 2-dimensional extended TFTs in this context, where $\mathcal{E}\text{-}\mathcal{TFT}^{\text{un}}(\mathcal{C})$ and $\mathcal{E}\text{-}\mathcal{TFT}(\mathcal{C})$ are defined similarly using $\text{Bord}_2^{\text{un}}$ and Bord_2 , respectively.

5.2 $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended unoriented X -HFTs

Tagami [26] classified 2-dimensional nonextended unoriented HFTs by extended crossed Frobenius G -algebras (see also [12]). Similar to the oriented case, our goal is to understand the relation between his classification and the restriction of $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended unoriented HFTs to circles and cobordisms between them.

Firstly, we introduce necessary algebraic notions. Let K be a G -algebra and V be a (K, K^{op}) -bimodule. The *conjugate* of V is the (K, K^{op}) -bimodule \underline{V} obtained by turning actions around. Similarly, the *conjugate* of a graded Morita context $\zeta = ({}_{K^{\text{op}}}U_K, {}_K V_{K^{\text{op}}}, \tau, \mu)$ is given by $\underline{\zeta} = ({}_{K^{\text{op}}}\underline{U}_K, {}_K \underline{V}_{K^{\text{op}}}, \underline{\tau}, \underline{\mu})$. We generalize stellar algebras introduced in [22] to stellar G -algebras as follows:

Definition 5.5 A *stellar G -algebra* is a G -algebra $K = \bigoplus_{g \in G} K_g$ equipped with a G -graded Morita context $\zeta = ({}_{K^{\text{op}}}U_K, {}_K V_{K^{\text{op}}}, \tau, \mu)$ together with an isomorphism of G -graded Morita contexts $\sigma : \zeta \cong \underline{\zeta}$ such that $\sigma \circ \underline{\sigma}$ is the identity isomorphism, where $\underline{\sigma}$ is the induced isomorphism between $\underline{\zeta}$ and $\underline{\underline{\zeta}}$.

The stellar structure on a G -algebra can be transferred along a graded Morita context as follows. Let $\rho = ({}_K U'_L, {}_L V'_K, \kappa, \nu)$ be a G -graded Morita context between G -algebras K and L and let (K, ζ, σ) be a stellar structure on K with $\zeta = ({}_{K^{\text{op}}}U_K, {}_K V_{K^{\text{op}}}, \tau, \mu)$. Then $(L, \rho_* \zeta, \rho_* \sigma)$ is a stellar algebra, where

$$\rho_* \zeta = ({}_{L^{\text{op}}}\underline{U}' \otimes_{K^{\text{op}}} U \otimes_K U'_L, {}_L V' \otimes_K V \otimes_{K^{\text{op}}} \underline{V}'_{L^{\text{op}}}, \underline{\kappa} \otimes \tau \otimes \kappa, \underline{\nu} \otimes \mu \otimes \nu)$$

and $\rho_* \sigma : \rho_* \zeta \cong \underline{\rho_* \zeta}$ is given by σ .

Definition 5.6 Let (K, ζ, σ) be a stellar G -algebra with $\zeta = ({}_K\text{op}U_K, {}_K V_{K\text{op}}, \tau, \mu)$ and let (K, η) be a quasibiangular G -algebra. The stellar structure is said to be *compatible* with the quasibiangular G -algebra if there exists an element $\sum_j a_j \otimes b_j \in K_e \otimes K_e$ giving the central element $z = \sum_j b_j a_j$ such that the diagrams⁹

$$\begin{array}{ccccccc}
 K \otimes K & \xrightarrow{\tau \otimes \text{id}} & (V \otimes U) \otimes K & \xrightarrow{\sigma \otimes \text{id}} & (\underline{V} \otimes \underline{U}) \otimes K & \xrightarrow{\underline{\tau}^{-1} \otimes \text{id}} & K \otimes K \\
 \text{id} \otimes \tau \downarrow & & & & & & \downarrow \eta \\
 K \otimes (V \otimes U) & \xrightarrow{\text{id} \otimes \sigma} & K \otimes (\underline{V} \otimes \underline{U}) & \xrightarrow{\text{id} \otimes \underline{\tau}^{-1}} & K \otimes K & \xrightarrow{\eta} & \mathbb{k} \\
 & & & & & & \\
 \mathbb{k} & \xrightarrow{\iota} & K \otimes K & \xrightarrow{\tau \otimes \text{id}} & (V \otimes U) \otimes K & \xrightarrow{\sigma \otimes \text{id}} & (\underline{V} \otimes \underline{U}) \otimes K \\
 \iota \downarrow & & & & & & \downarrow \underline{\tau}^{-1} \otimes \text{id} \\
 K \otimes K & \xrightarrow{\text{id} \otimes \tau} & K \otimes (V \otimes U) & \xrightarrow{\text{id} \otimes \sigma} & K \otimes (\underline{V} \otimes \underline{U}) & \xrightarrow{\text{id} \otimes \underline{\tau}^{-1}} & K \otimes K \\
 & & & & & & \\
 K \otimes K & \xrightarrow{\tau \otimes \text{id}} & (V \otimes U) \otimes K & \xrightarrow{\sigma \otimes \text{id}} & (\underline{V} \otimes \underline{U}) \otimes K & \xrightarrow{\underline{\tau}^{-1} \otimes \text{id}} & K \otimes K \\
 \text{id} \otimes \tau \downarrow & & & & & & \downarrow \xi \\
 K \otimes (V \otimes U) & \xrightarrow{\text{id} \otimes \sigma} & K \otimes (\underline{V} \otimes \underline{U}) & \xrightarrow{\text{id} \otimes \underline{\tau}^{-1}} & K \otimes K & \xrightarrow{\xi} & K \otimes K
 \end{array}$$

commute, where $\iota(1)|_{K_e \otimes K_e} = \sum_j a_j \otimes b_j$ and $\xi: {}_{K_1}K_{K_2} \otimes {}_{K_3}K_{K_4} \rightarrow {}_{K_1}K_{K_4} \otimes {}_{K_3}K_{K_2}$ is a graded bimodule map with $K_i = K$ for $i = 1, 2, 3, 4$ and $\xi(1) = \sum_i p_i^e \otimes q_i^e$ is an inner product element of the principal component. We call such a compatible quadruple (K, η, ζ, σ) a *quasibiangular stellar G -algebra*.

Definition 5.7 A morphism of quasibiangular stellar G -algebras $(K, \eta^k, \zeta^K, \sigma^K)$ and $(L, \eta^L, \zeta^L, \sigma^L)$ is a compatible G -graded Morita context $\rho = ({}_K U_L, {}_L V_K, \tau, \mu)$ together with an equivalence of G -graded Morita contexts $\phi: \zeta^L \rightarrow \rho_* \zeta^K$ such that $\rho_* \sigma^K \circ \phi = \underline{\phi} \circ \sigma^L$, where $\underline{\phi}: \zeta^L \rightarrow \rho_* \zeta^K$. Two such morphisms (ρ, ϕ) and (ρ', ϕ') are isomorphic if there exists an equivalence of G -graded Morita contexts $\alpha: \rho \rightarrow \rho'$ such that $\phi' = \alpha \circ \phi$ and $\underline{\phi}' = \underline{\alpha} \circ \underline{\phi}$ for $\underline{\alpha}: \rho \rightarrow \rho'$.

Theorem 5.8 Let G be a group with each nonidentity element having order 2 and X be a $K(G, 1)$ -space. Any $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended unoriented X -HFT $Z: X\text{Bord}_2^{\text{un}} \rightarrow \text{Alg}_{\mathbb{k}}^2$ whose precomposition $X\text{B}^{\text{PD,un}} \xrightarrow{\cong} X\text{Bord}_2^{\text{un}} \xrightarrow{Z} \text{Alg}_{\mathbb{k}}^2$ gives a

⁹Tensors in diagrams are taken over K, K^{op} or $K \otimes_{\mathbb{k}} K^{\text{op}}$.

strict symmetric monoidal 2–functor determines a quasibiangular stellar G –algebra (A, η, ζ, σ) . Conversely, for any quasibiangular stellar G –algebra (A, η, ζ, σ) , there exists an $\text{Alg}_{\mathbb{k}}^2$ –valued 2–dimensional extended unoriented X –HFT.

Proof Let $Z : X\text{Bord}_2^{\text{un}} \rightarrow \text{Alg}_{\mathbb{k}}^2$ be such a 2–dimensional extended unoriented HFT. The cofibrancy theorem implies that there exists an object Z' in $\mathbb{X}\mathbb{P}^{\text{un}}(\text{Alg}_{\mathbb{k}}^2)$ such that $\iota(Z')$ is the composition $X\text{B}^{\text{PD,un}} \xrightarrow{\simeq} X\text{Bord}_2^{\text{un}} \xrightarrow{Z} \text{Alg}_{\mathbb{k}}^2$, where $\iota : \mathbb{X}\mathbb{P}^{\text{un}}(\text{Alg}_{\mathbb{k}}^2) \rightarrow \text{SymMon}(X\text{B}^{\text{PD,un}}, \text{Alg}_{\mathbb{k}}^2)$ is the equivalence of bicategories.

Following the proof of [Theorem 4.19](#), we have a strongly graded G –algebra $A = \bigoplus_{g \in G} A_g$, where $Z'(\bullet) = A_e$. We also have G –graded $(A \otimes A, \mathbb{k})$ – and $(\mathbb{k}, A \otimes A)$ –bimodules $M = \bigoplus_{g \in G} M_g$ and $N = \bigoplus_{g \in G} N_g$, respectively. By turning actions around, we obtain the (A, A^{op}) –bimodule M and (A^{op}, A) –bimodule N .

Bimodule maps in $Z'_2(\mathcal{X}^{\text{un}}\mathcal{G}_2)$ corresponding to cusp generators (subject to relations) yield a G –graded Morita context $\zeta = ({}_{A^{\text{op}}}N_A, {}_A M_{A^{\text{op}}}, f_1, f_2)$ between A and A^{op} , where $f_1 : {}_A A A \rightarrow {}_A M \otimes_{A^{\text{op}}} N_A$ and $f_2 : {}_{A^{\text{op}}} N \otimes_A M_{A^{\text{op}}} \rightarrow {}_{A^{\text{op}}} A_{A^{\text{op}}}^{\text{op}}$ are invertible G –graded bimodule maps. Bimodule maps in $Z'_2(\mathcal{X}^{\text{un}}\mathcal{G}_2)$ for the Morse generators satisfying the relations imply that (A, η) is a quasibiangular G –algebra. The generators in [Figure 31](#) give the graded bimodule maps, in $Z'_2(\mathcal{X}^{\text{un}}\mathcal{G}_2)$,

$$\begin{aligned} \sigma_1 : {}_A M_{A^{\text{op}}} &\rightarrow {}_A \underline{M}_{A^{\text{op}}}, & \sigma_2 : {}_{A^{\text{op}}} N_A &\rightarrow {}_{A^{\text{op}}} \underline{N}_A, \\ \sigma'_1 : {}_A \underline{M}_{A^{\text{op}}} &\rightarrow {}_A M_{A^{\text{op}}}, & \sigma'_2 : {}_{A^{\text{op}}} \underline{N}_A &\rightarrow {}_{A^{\text{op}}} N_A. \end{aligned}$$

These graded bimodule maps are subject to the relations in [Figure 32](#). Thereby, we have $\sigma'_1 \circ \sigma_1 = \text{id}_M$, $\sigma_1 \circ \sigma'_1 = \text{id}_{\underline{M}}$, $\sigma'_2 \circ \sigma_2 = \text{id}_N$, and $\sigma_2 \circ \sigma'_2 = \text{id}_{\underline{N}}$. These isomorphisms of bimodules lead to an isomorphism $\sigma : \underline{\zeta} \cong \zeta$. Applying σ to $\underline{\zeta}$ gives another isomorphism $\underline{\sigma} : \underline{\zeta} \rightarrow \zeta$, whose composition with σ gives $\sigma \circ \underline{\sigma} : \underline{\zeta} \cong \zeta$. The third relation in the first row of [Figure 32](#) and its reflection indicate that compositions of bimodule maps $\underline{M} \rightarrow \underline{M} \rightarrow M$ and $\underline{N} \rightarrow \underline{N} \rightarrow N$ are identity maps.

Thus, additional generators and relations among them lead to a stellar structure (ζ, σ) on the quasibiangular G –algebra A . The remaining relations imply compatibility, giving the quasibiangular stellar G –algebra (A, η, ζ, σ) . For any quasibiangular stellar G –algebra, one constructs an object of $\mathbb{X}\mathbb{P}^{\text{un}}(\text{Alg}_{\mathbb{k}}^2)$ by assigning values to generating objects, 1–morphisms, and 2–morphisms of $\mathbb{X}\mathbb{P}$ satisfying generating relations using the above arguments. Then, this object gives a strict symmetric monoidal 2–functor $X\text{B}^{\text{PD,un}} \rightarrow \text{Alg}_{\mathbb{k}}^2$ whose composition with the equivalence $X\text{Bord}_2^{\text{un}} \xrightarrow{\simeq} X\text{B}^{\text{PD,un}}$ produces the desired unoriented extended X –HFT. □

Similar to the oriented case, every 2–dimensional extended unoriented HFT with target X produces a nonextended one by precomposition, $X\text{Cob}_2^{\text{un}} \rightarrow X\text{Cob}_2^{\text{un}} \rightarrow \text{Alg}_{\mathbb{k}}^2$, where $X\text{Cob}_2^{\text{un}}$ and $X\text{Cob}_2^{\text{un}}$ are defined just as $X\text{Cob}_2$ and $X\text{Cob}_2$ using unoriented X –manifolds. In the unoriented case, extended crossed Frobenius G –algebras play an important role in the study of 2–dimensional nonextended unoriented X –HFTs and they are defined as follows:

Definition 5.9 [26] Let (K, η, φ) be a crossed Frobenius G –algebra over \mathbb{k} . An *extended* structure on K consists of a \mathbb{k} –module homomorphism $\Phi: K \rightarrow K$ and a family of elements $\{\theta_g \in K_e\}_{g \in G}$ satisfying the following conditions:

- (1) $\Phi(K_g) \subset K_g$ and $\Phi(\theta_g) = \theta_g$ for all $g \in G$.
- (2) $\Phi \circ \varphi_g = \varphi_g \circ \Phi$ for all $g \in G$.
- (3) $\Phi(vw) = \Phi(w)\Phi(v)$ for any $v, w \in K$ and $\Phi(1_K) = 1_K$.
- (4) $\Phi^2 = \text{id}$.
- (5) $\eta \circ (\Phi \otimes \Phi) = \eta$.
- (6) For any $g, h, l \in G$ and $v \in K_{gh}$, we have

$$m \circ (\Phi \circ \varphi_l) \circ \Delta_{g,h}(v) = \varphi_l(\theta_{gl}\theta_l v),$$

$$m \circ (\varphi_l \otimes \Phi) \circ \Delta_{g,h}(v) = \varphi_l(\theta_{hl}\theta_l v),$$

where $\Delta_{g,h}: K_{gh} \rightarrow K_g \otimes K_h$ is defined by $(\text{id}_g \otimes \eta) \circ (\Delta_{g,h} \otimes \text{id}_h) = m$. Such a map $\Delta_{g,h}$ is uniquely determined since η is nondegenerate and each K_g is finitely generated.

- (7) $\Phi(\theta_h v) = \varphi_{hg}(\theta_{hg} v)$ for any $g, h \in G$ and $v \in K_g$.
- (8) $\varphi_h(\theta_g) = \theta_g$ for any $g, h \in G$.
- (9) For any $g, h, l \in G$, we have $\theta_g \theta_h \theta_l = q(1)\theta_{ghl}$, where $q: \mathbb{k} \rightarrow K_e$ is defined as follows: let $\{a_i \in K_{gh}\}_{i=1}^n$ and $\{b_i \in K_{gh}\}_{i=1}^n$ be families of elements of K_{gh} satisfying $\sum_i \eta(b_i \otimes v)a_i = \varphi_{hl}(v)$ for any $v \in K_{gh}$. As in (3), such a_i and b_i are uniquely determined and $q(1) = \sum_i a_i b_i$.

Theorem 5.10 (Tagami [26]) *Let G be a group with each nonidentity element having order 2. There is a bijection between the isomorphism classes of 2–dimensional unoriented HFTs with target $X \simeq K(G, 1)$ and the isomorphism classes of extended crossed Frobenius G –algebras.*

Corollary 5.11 Assume that $Z : X \text{Bord}_2^{\text{un}} \rightarrow \text{Alg}_{\mathbb{K}}^2$ determines a quasibiangular stellar G -algebra (A, η, ζ, σ) . The stellar structure (ζ, σ) gives an extended structure on the crossed Frobenius G -algebra $Z_G(A)$. Moreover, the corresponding 2-dimensional X -HFT is the unoriented X -HFT obtained by restricting Z to $X \text{Cob}_2^{\text{un}}$.

Proof We have a crossed Frobenius G -algebra $(Z_G(A), \eta|_{Z_G(A)}, \{\varphi|_{Z_G(A)}\}_{g \in G})$. By Tagami’s classification, the 2-dimensional unoriented HFT given by the restriction of Z to circles and cobordisms between them induces an extended structure on $Z_G(A)$. We claim that the homomorphism Φ and elements $\{\theta_g \in Z_G(A)_e\}_{g \in G}$ come from the stellar structure (ζ, σ) on A .

In [26], for each $g \in G$, the restriction $\Phi|_{Z_G(A)_g} : Z_G(A)_g \rightarrow Z_G(A)_g$ is the involution induced by an orientation-reversing homeomorphism of a g -labeled circle. In the extended case, this morphism is given by additional 2-morphisms (Figure 31). More precisely, $\Phi|_{Z_G(A)_g} : A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_g \rightarrow A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_g$ is defined by $\Phi(a \otimes b) = a \otimes \Phi_g(b)$, where Φ_g is defined so that the diagram

$$\begin{array}{ccc}
 A_e M_g \otimes_{A_e^{\text{op}}} (N_e)_{A_e} & \xrightarrow{\cong} & A_e (A_g)_{A_e} \\
 \sigma_1 \otimes \text{id} \downarrow & & \downarrow \Phi_g = Z \left(\begin{array}{c} g \\ \square \\ g \end{array} \right) \\
 A_e \underline{M}_g \otimes_{A_e^{\text{op}}} (N_e)_{A_e} & \longrightarrow & A_e (A_g)_{A_e}
 \end{array}$$

commutes. It is not hard to see that Φ reverses the orientation of the oriented (input) circle. In [26], for every $g \in G$, the element θ_g is the image of HFT under the Möbius strip whose boundary is labeled by $g^2 = e$, where the Möbius strip is considered as the cobordism from the empty 1-manifold to the boundary circle. In the extended case, $\theta_g \in A_e \otimes_{A_e \otimes A_e^{\text{op}}} A_e$ is the image of $1 \in \mathbb{K}$ under the composition of a $\{g, g\}$ -labeled cap morphism followed by new generators (see Figure 32), which is composed with module actions turning boundary labels into $\{e, e\}$ (see Figure 29).

The involution Φ and elements $\{\theta_g\}_{g \in G}$ are defined according to their topological description given in [26]. Hence, $(Z_G(A), \eta|_{Z_G(A)}, \{\varphi_g|_{Z_G(A)}\}_{g \in G}, \Phi, \{\theta_g\}_{g \in G})$ is an extended crossed Frobenius G -algebra, which, by definition, corresponds to the restriction of $Z : X \text{Bord}_2^{\text{un}} \rightarrow \text{Alg}_{\mathbb{K}}^2$ to X -circles and unoriented X -cobordisms between them. □

5.3 The bicategory of 2-dimensional extended unoriented X -HFTs

In order to upgrade Theorem 5.8 to an equivalence of bicategories, we study morphisms in the bicategory $\mathbb{X}P^{\text{un}}(\text{Alg}_{\mathbb{K}}^2)$. Let α be a 1-morphism from Z_0 to Z_1 giving quasi-

biangular stellar G -algebras (A, η, ζ, σ) and $(A', \eta', \zeta', \sigma')$, respectively. We know from the oriented case that α gives a compatible G -graded Morita context ξ between G -algebras A and A' . Assuming $\alpha_0(\bullet) = {}_{A_e}R_{A'_e}$ and $\xi = ({}_A R_{A'}, {}_{A'} R_A, \tau, \mu)$, naturality with respect to the first generator in Figure 31 is the commutativity of the diagram

$$\begin{array}{ccc}
 A'_e \underline{(M'_g)}_{(A'_e)^{op}} & \xrightarrow{\alpha_1(\rhd g)} & A'_e R' \otimes_{A_e} M_g \otimes_{A_e^{op}} \underline{R'}_{(A'_e)^{op}} \\
 \sigma' = Z_1 \left(\begin{array}{c} \boxed{g} \\ \downarrow \\ \boxed{g} \end{array} \right) \downarrow & & \downarrow \xi_* \sigma = Z_0 \left(\begin{array}{c} \boxed{g} \\ \downarrow \\ \boxed{g} \end{array} \right) \\
 A'_e \underline{(M'_g)}_{(A'_e)^{op}} & \xrightarrow{\alpha_1(\triangleright \triangleright g)} & A'_e R' \otimes_{A_e} \underline{M}_g \otimes_{A_e^{op}} \underline{R'}_{(A'_e)^{op}}
 \end{array}$$

where M and M' are components of the graded Morita contexts ζ and ζ' , respectively. There are similar commutative diagrams for the remaining three generators. These diagrams indicate that the G -graded Morita context ξ gives an equivalence of G -graded Morita contexts ζ' and $\xi_* \zeta$ with $\alpha \circ \sigma' = \xi_* \sigma \circ \alpha$. In other words, α leads to a morphism of stellar G -algebras (see Definition 5.7).

Let $\theta: \alpha^1 \rightarrow \alpha^2$ be a 2-morphism in $\mathbb{X}\mathbb{P}(\text{Alg}_{\mathbb{k}}^2)$ with $\theta_0(\bullet) = {}_{A_e}R_{A'_e} \rightarrow {}_{A_e}P_{A'_e}$. In the oriented case we observed that θ induces an equivalence of G -graded Morita contexts $\xi = ({}_A R_{A'}, {}_{A'} R_A, \tau, \mu)$ and $\rho = ({}_A P_{A'}, {}_{A'} P_A, \kappa, \nu)$. Naturality of $\theta_0(\bullet)$ with respect to $\triangleright g$ is the commutativity of the diagram

$$\begin{array}{ccc}
 A'_e \underline{(M'_g)}_{(A'_e)^{op}} & \xrightarrow{\alpha_1^1(\triangleright g)} & A'_e R' \otimes_{A_e} \underline{M}_g \otimes_{A_e^{op}} \underline{R'}_{(A'_e)^{op}} \\
 \text{id} \downarrow & & \downarrow \theta_0(\bullet \bullet) \\
 A'_e \underline{(M'_g)}_{(A'_e)^{op}} & \xrightarrow{\alpha_1^2(\triangleright g)} & A'_e P' \otimes_{A_e} \underline{M}_g \otimes_{A_e^{op}} \underline{P'}_{(A'_e)^{op}}
 \end{array}$$

and there is a similar diagram for the naturality with respect to $g \triangleleft$. Naturality for $\{\overset{g}{\longleftarrow}\}_{g \in G}$ and $\{\overset{g}{\longrightarrow}\}_{g \in G}$ gives $\alpha^2 = \theta \circ \alpha^1$ and naturality for $\{\triangleright g\}_{g \in G}$ and $\{g \triangleleft\}_{g \in G}$ gives $\underline{\alpha}_2 = \underline{\theta} \circ \underline{\alpha}^1$. In other words, θ gives an isomorphism of stellar G -algebra morphisms (see Definition 5.7).

These observations lead us to define a bicategory Frob_*^G , which has quasibiangular stellar G -algebras as objects, their morphisms as 1-morphisms, and isomorphisms of quasibiangular stellar G -algebra morphisms as 2-morphisms. The above arguments imply that there exists a 2-functor $\mathcal{F}': \mathbb{X}\mathbb{P}^{\text{un}}(\text{Alg}_{\mathbb{k}}^2) \rightarrow \text{Frob}_*^G$. Composing \mathcal{F}' with the equivalence $\mathcal{E}\text{-}\mathcal{H}\mathcal{F}\mathcal{T}^{\text{un}}(X, \text{Alg}_{\mathbb{k}}^2) \simeq \mathbb{X}\mathbb{P}^{\text{un}}(\text{Alg}_{\mathbb{k}}^2)$, we define the 2-functor \mathcal{F} .

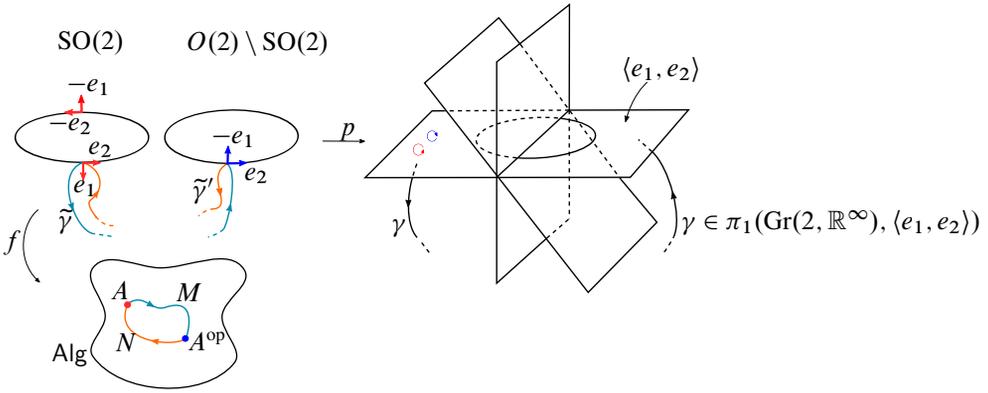


Figure 33: A reflection-invariant map.

Theorem 5.12 *The 2-functor $\mathcal{F}: \mathcal{E}\text{-}\mathcal{HFT}^{\text{un}}(X, \text{Alg}_{\mathbb{k}}^2) \rightarrow \text{Frob}_*^G$ is an equivalence of bicategories.*

Proof The proof follows from the above arguments and the Whitehead theorem for bicategories. □

5.4 The $(G \times O(2))$ -structured cobordism hypothesis

Parallel to oriented case, we want to compare [Theorem 5.12](#) with the classification given by the $(G \times O(2))$ -structured cobordism hypothesis. To do this, we need to understand homotopy $(G \times O(2))$ -fixed points in $(\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim}$, which are given by

$$((\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim})^{h(G \times O(2))} = \text{Map}_G(EG, \text{Map}_{O(2)}(EO(2), \text{Alg})),$$

where G acts on invariant maps trivially and Alg is the 2-type corresponding to the ∞ -groupoid $(\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim}$. Recall that the unoriented Grassmannian $\text{Gr}(2, \mathbb{R}^\infty)$ is a model for $BO(2)$ and the Stiefel manifold $V(2, \mathbb{R}^\infty)$ is one for $EO(2)$. The universal principal $O(2)$ -bundle $p: V(2, \mathbb{R}^\infty) \rightarrow \text{Gr}(2, \mathbb{R}^\infty)$ is given by $p((e_1, e_2)) = \langle e_1, e_2 \rangle$, ie the plane generated by the orthonormal 2-frame (e_1, e_2) .

Lemma 5.13 *The reflection-invariant maps in $\text{Map}(V(2, \mathbb{R}^\infty), \text{Alg})$ determine stellar structures on \mathbb{k} -algebras.*

Proof A reflection ω in $O(2)$ acts on Alg by sending a \mathbb{k} -algebra A to its opposite algebra A^{op} . Let f be a reflection-invariant map with $f((e_1, e_2)) = A$. Let γ be a representative of the nontrivial element of $\pi_1(\text{Gr}(2, \mathbb{R}^\infty), \langle e_1, e_2 \rangle) \cong \mathbb{Z}/2\mathbb{Z}$. Lift γ to $\tilde{\gamma}$ starting at (e_1, e_2) and ending at $\omega((e_1, e_2))$ (see [Figure 33](#)). Then $f(\tilde{\gamma})$ is a

(A^{op}, A) -bimodule M and invariance under ω means $f(\omega(\tilde{\gamma})) = {}_{A^{\text{op}}}M_A = \omega(M)$. Lifting γ to $\tilde{\gamma}'$ starting at $\omega((e_1, e_2))$ gives a path ending at (e_1, e_2) . Similarly, $f(\tilde{\gamma}')$ is an (A, A^{op}) -bimodule N and we have $f(\omega(\tilde{\gamma}')) = {}_AN_{A^{\text{op}}} = \omega(N)$. Loops $\tilde{\gamma}' * \tilde{\gamma}$ and $\tilde{\gamma} * \tilde{\gamma}'$ bound disks since $V(2, \mathbb{R}^\infty)$ is contractible. In other words, there exist basepoint-fixing homotopies which take these loops to constant loops at (e_1, e_2) and $\omega((e_1, e_2))$, respectively. Images of the second homotopy and the time reversed version of the first homotopy under f yield invertible bimodule maps $\mu: {}_{A^{\text{op}}}M \otimes_A N_{A^{\text{op}}} \rightarrow {}_{A^{\text{op}}}A_{A^{\text{op}}}^{\text{op}}$ and $\tau: {}_AA_A \rightarrow {}_AN \otimes_{A^{\text{op}}} M_A$, respectively. The compositions of homotopies corresponding to the conditions on μ and τ to form a Morita context are the constant homotopies of paths $\tilde{\gamma}$ and $\tilde{\gamma}'$. Thus, $\zeta = ({}_{A^{\text{op}}}M_A, {}_AN_{A^{\text{op}}}, \tau, \mu)$ is a Morita context. Similarly, loops $\tilde{\gamma} * \omega(\tilde{\gamma})$ and $\tilde{\gamma}' * \omega(\tilde{\gamma}')$ bound, which implies that there is an equivalence of Morita contexts $\sigma: \zeta \cong \underline{\zeta}$. Since the order of reflection is two, we have $\sigma \circ \underline{\sigma} = \text{id}$. Thus, any reflection-invariant map leads to stellar algebra structures on algebras. \square

Lemma 5.14 *For an algebraically closed field \mathbb{k} of characteristic zero, the homotopy $(G \times O(2))$ -fixed points of $(\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim}$ are quasibiangular stellar G -algebras.*

Proof Serre automorphism trivializes the homotopy $SO(2)$ -action (see [6]), which turns the space of homotopy $(G \times SO(2))$ -fixed points into

$$\text{Map}_G(EG, \text{Map}(\tilde{\text{Gr}}(2, \mathbb{R}^\infty), \text{Alg})).$$

Davidovich [6] showed that homotopy $SO(2)$ -fixed points are semisimple symmetric Frobenius \mathbb{k} -algebras. Then, understanding homotopy $O(2)$ -fixed points means understanding invariance under reflections. Using Lemma 5.13, we conclude that homotopy $O(2)$ -fixed points are finite-dimensional semisimple symmetric Frobenius \mathbb{k} -algebras with a stellar structure.

The stellar structure is compatible with the Frobenius form as follows. A Frobenius form on a \mathbb{k} -algebra A is determined by a central element, which is the image of 1 under a bimodule map $z: {}_AA_A \rightarrow {}_AA_A$. Geometrically, $z(1)$ is an element of

$$\pi_2(\text{Map}(BSO(2), \text{Alg}_r), f) \cong (\mathbb{k}^\times)^r,$$

where the algebra $A \in \text{Alg}_r \subset \text{Alg} = \coprod_{r=1}^\infty \text{Alg}_r$ is isomorphic to

$$\text{End}(V_1) \times \text{End}(V_2) \times \cdots \times \text{End}(V_r)$$

under the Artin–Wedderburn isomorphism for finite-dimensional \mathbb{k} -vector spaces V_1, \dots, V_r .

Compatibility means that (horizontal) composition of z with ζ yields z again. Geometrically, this corresponds to conjugating the representing sphere based at f with loops in Alg_r given by bimodules of ζ . Since this loop is contractible, conjugation does not change $z(1)$ in the second homotopy group. Thus, we have a compatible stellar structure, and following Davidovich's methods [6] we obtain that, for a discrete group G , homotopy $(G \times O(2))$ -fixed points are quasibiangular stellar G -algebras. \square

The above lemma is an important step toward the verification of the $(G \times O(2))$ -structured cobordism hypothesis for $\text{Alg}_{\mathbb{k}}^2$ -valued 2-dimensional extended unoriented HFTs with $K(G, 1)$ target. This version of the cobordism hypothesis states the equivalence of bigroupoids Frob_*^G and $((\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim})_{\leq 2}^{h(G \times O(2))}$, where $((\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim})_{\leq 2}^{h(G \times O(2))}$ is the fundamental bigroupoid of $((\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim})^{h(G \times O(2))}$. Lemma 5.14 implies that the objects of these bigroupoids coincide. The next step is to give an explicit (algebraic) description of the bigroupoid $((\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim})_{\leq 2}^{h(G \times O(2))}$ and to write down the 2-functor $\mathcal{F}: \text{Frob}_*^G \xrightarrow{\cong} ((\text{Alg}_{\mathbb{k}}^{\text{fd}})^{\sim})_{\leq 2}^{h(G \times O(2))}$, similar to the oriented case. Here we skip these steps, which may later appear elsewhere.

Appendix Unbiased semistrict symmetric monoidal 2-categories

In this section we recall unbiased semistrict monoidal 2-categories and their computadic versions, and prove Theorem 4.5. Our main reference is [22, Section 2.10]. A similar exposition is given in [21, Appendix].

A.1 String diagrams for bicategories

The symmetric monoidal bicategory XB^{PD} is not a fully weak symmetric monoidal bicategory but a certain stricter version. The strict bicategories we are interested in are unbiased semistrict symmetric monoidal 2-categories introduced by Schommer-Pries [22]. To recall their definition, we first review string diagrams for bicategories.

Alternative to pasting diagrams, string diagrams are tools describing morphisms in a bicategory. Instead of arrows between objects and 1-morphisms, a string diagram consists of regions, arcs, and vertices. Each region represents an object and each arc represents a 1-morphism between objects whose corresponding regions share this arc as a common boundary. Each vertex represents a 2-morphism between 1-morphisms

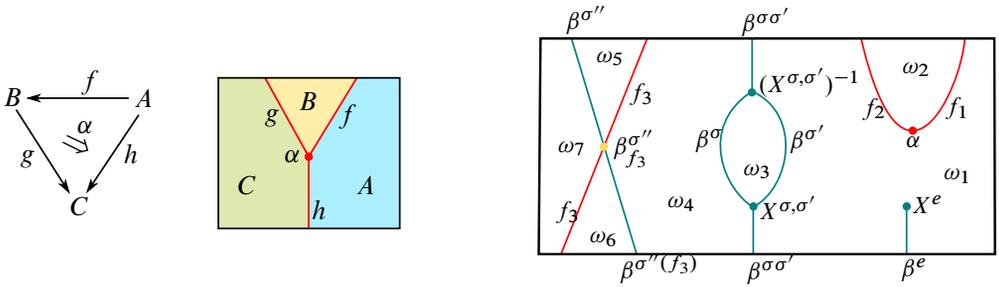


Figure 34: Pasting diagram with the corresponding string diagram and a string diagram for an unbiased semistrict symmetric monoidal 2-category.

whose corresponding arcs are connected with each other via this vertex. In Figure 34, left, a pasting diagram and the corresponding string diagram is shown. Note that we read string diagrams from right to left and from top to bottom.

Unbiased semistrict symmetric monoidal 2-categories are strict enough to admit a version of a string diagram. An example of such a string diagram is shown in Figure 34, in which regions are labeled with objects $\{\omega_i\}_{i=1}^7$, red arcs are labeled with 1-morphisms $\{f_j\}_{j=1}^3$, and a red vertex is labeled with a 2-morphism α . However, there are additional strings and vertices of different colors coming from the coherence morphisms of an unbiased semistrict symmetric monoidal 2-category.

Definition A.1 [22, Definition 2.32] An unbiased semistrict symmetric monoidal 2-category is a triple (\mathcal{C}, β, X) , where $\mathcal{C} = (\mathcal{C}, \otimes, \iota, \alpha, \lambda, \rho, \mathcal{P}, \mathcal{M}, \mathcal{L}, \mathcal{R})$ is a monoidal bicategory (see [25, Appendix]) such that:

- (i) The underlying bicategory is a strict 2-category.
- (ii) The transformations α, λ and ρ and modifications $\mathcal{P}, \mathcal{M}, \mathcal{L}$ and \mathcal{R} are identities.
- (iii) The monoidal product $\otimes = (\otimes, \phi_{(f, f'), (g, g')}, \phi_{(a, a')}) : \mathcal{C} \times \mathcal{C} \rightarrow \mathcal{C}$ is cubical. That is, the interchanger

$$\phi_{(f, f'), (g, g')}^\otimes : (f \otimes f') \circ (g \otimes g') \rightarrow (f \circ g) \otimes (f' \circ g')$$

is the identity if either f or g' is the identity 1-morphism and $\phi_{(a, a')}^\otimes : \text{id}_a \otimes a' \rightarrow \text{id}_a \otimes \text{id}_{a'}$ is the identity for all objects a and a' .

Secondly, β denotes a collection of transformations (turquoise edges and yellow point in Figure 34)

$$\{\beta^\sigma : (\mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \dots \otimes \mathcal{C}_n \rightarrow \mathcal{C}) \rightarrow (\mathcal{C}_{\sigma(1)} \otimes \mathcal{C}_{\sigma(2)} \otimes \dots \otimes \mathcal{C}_{\sigma(n)} \rightarrow \mathcal{C})\}_{\sigma \in S_n, n \geq 0},$$

where $\mathcal{C}_i = \mathcal{C}$ for all $i = 1, \dots, n$ and $S_0 := \{1\}$ with $\beta^1 : (1 \hookrightarrow \mathcal{C}) \rightarrow (1 \hookrightarrow \mathcal{C})$ being the identity transformation between inclusions of functors. Lastly, X denotes a collection of invertible modifications (turquoise points in Figure 34)

$$X^{\sigma, \sigma'} : (\beta^\sigma * 1) \circ \beta^{\sigma'} \rightarrow \beta^{\sigma\sigma'} \quad \text{and} \quad X^e : \text{id} \rightarrow \beta^e$$

for every $\sigma, \sigma' \in S_n$ and identity element $e \in S_n$ such that:

- (i) The transformations $\{\beta^\sigma\}_{\sigma \in S_n, n \geq 0}$ and modifications $\{X^{\sigma, \sigma'}, X^e\}_{\sigma, \sigma', e \in S_n, n \geq 0}$ satisfy the conditions

$$\begin{aligned} \beta^{\text{id} \sqcup \sigma} &= \text{id} \otimes \beta^\sigma, & \beta^{\sigma \sqcup \text{id}} &= \beta^\sigma \otimes \text{id}, \\ X^{(\text{id} \sqcup \sigma), (\text{id} \sqcup \sigma')} &= \text{id} * X^{\sigma, \sigma'}, & X^{(\sigma \sqcup \text{id}), (\sigma' \sqcup \text{id})} &= X^{\sigma, \sigma'} * \text{id} \end{aligned}$$

and the first three conditions on Figure 35 for all $\sigma, \sigma', \sigma'', e \in S_n$ and $n > 0$,

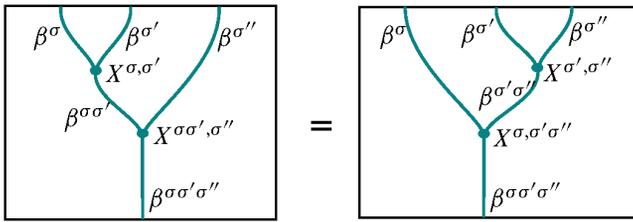
- (ii) For a fixed $n > 0$ and a collection of n natural numbers $\{k_i\}_{i=1}^n$, let $\tilde{\sigma} \in S_N$ be given by the operadic product¹⁰ $\sigma \circ (\tau_i)$, where $N = \sum_i k_i$, $\sigma \in S_n$, and $\tau_i \in S_{k_i}$ for all $i = 1, 2, \dots, n$. Then the 2-morphism $\beta^\sigma_{(\beta^{\tau_i})} = \beta^{\sqcup \tau_i} \circ \beta^\sigma \rightarrow \beta^\sigma \circ \beta^{\sqcup \tau_i}$ satisfies the equality given by the last condition in Figure 35 for all $n > 0$, $\sigma \in S_n$, and $\tau_i \in S_{k_i}$. In particular, when $\tau_i = e$ for all $i = 1, \dots, n$, we have $\beta^{\tilde{\sigma}} = \beta^\sigma$, $X^{\tilde{e}} = X^e$, and $X^{\tilde{\sigma}, \tilde{\sigma}'} = X^{\sigma, \sigma'}$ for all $\sigma, \sigma', e \in S_n$,
- (iii) The transformations $\{\beta^\sigma\}_{\sigma \in S_n, n \geq 0}$ and modifications $\{X^{\sigma, \sigma'}, X^e\}_{\sigma, \sigma', e \in S_n, n \geq 0}$ satisfy the conditions given by the reflections of diagrams in Figure 35 with respect to a horizontal axis.

In order to prove Theorem 4.5, we first need to show that the symmetric monoidal bicategory XB^{PD} is an unbiased semistrict symmetric monoidal 2-category. Recall that objects of XB^{PD} are finite sets of ordered oriented points, 1-morphisms are isotopy classes of G -linear diagrams, and 2-morphisms are equivalence classes of G -planar diagrams.

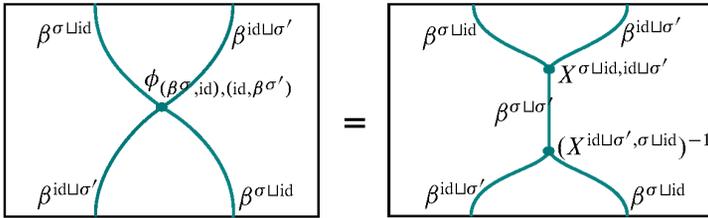
Lemma A.2 *Chambering sets, graphs, and foams equip XB^{PD} with the structure of an unbiased semistrict symmetric monoidal 2-category.*

Proof Recall that compositions of morphisms in XB^{PD} are given by the concatenation of diagrams. Since 1-morphisms are isotopy classes of G -linear diagrams and 2-morphisms are equivalence classes of G -planar diagrams, the underlying bicategory is

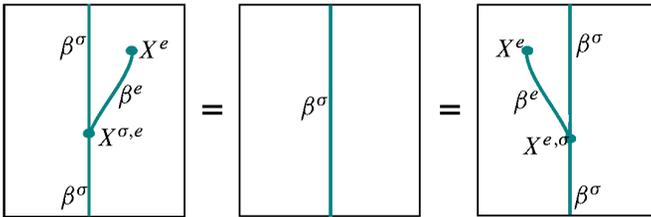
¹⁰By an operadic product we mean the composition $\sigma(\tau_1, \dots, \tau_n)$.



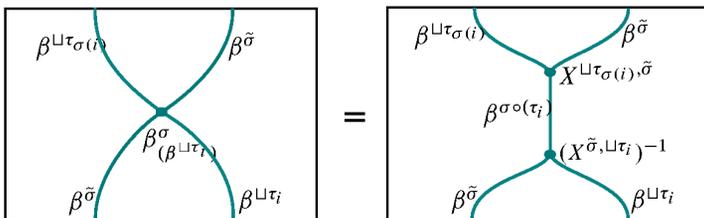
$$X^{\sigma \sigma', \sigma''} \circ (X^{\sigma, \sigma'} * 1) = X^{\sigma, \sigma' \sigma''} \circ (1 * X^{\sigma', \sigma''})$$



$$\phi_{(\beta^\sigma, \text{id}), (\text{id}, \beta^{\sigma'})} = (X^{\text{id} \sqcup \sigma', \sigma \sqcup \text{id}})^{-1} \circ X^{\sigma \sqcup \text{id}, \text{id} \sqcup \sigma'}$$



$$X^{\sigma, e} \circ (1 * X^e) = \text{id}_{\beta^\sigma} = X^{e, \sigma} \circ (X^e * 1)$$



$$\beta^{\sigma}_{(\beta^{\sqcup \tau_i})} = (X^{\tilde{\sigma}, \sqcup \tau_i})^{-1} \circ X^{\sqcup \tau_\sigma(i), \tilde{\sigma}}$$

Figure 35: Some of the axioms of an unbiased semistrict symmetric monoidal 2–category.

a strict 2–category. The symmetric monoidal structure of XB^{PD} is cubical by definition. The transformations α , λ and ρ and modifications \mathcal{P} , \mathcal{L} , \mathcal{M} and \mathcal{R} are identities since 2–morphisms are equivalence classes of G –planar diagrams. The local models CP

relations among string diagrams	corresponding chambering foams

Figure 36: Relations between string diagrams for unbiased semistrict symmetric monoidal 2-categories and the corresponding spatial foams.

and CK₄ of chambering foam shown in Figure 20 give two conditions in Figure 35, left. For the remaining two conditions, recall that a chambering graph can only have univalent and trivalent vertices. □

Let (\mathcal{C}, β, X) be an unbiased semistrict symmetric monoidal 2-category. The invertibility of modifications $\{X^{\sigma, \sigma'}, X^e\}_{\sigma, \sigma', e \in S_n, n \geq 0}$ and axioms of unbiased semistrict symmetric monoidal 2-category generate relations between structure morphisms. These relations are given¹¹ in Figure 36, left, in terms of string diagrams. Since chambering foams are responsible for the relations between boundary chambering graphs, in Figure 36, right, chambering foams corresponding to these relations are shown.

A.2 Computadic unbiased semistrict symmetric monoidal 2-categories

To finish the proof of Theorem 4.5, we need to show that the unbiased semistrict symmetric monoidal 2-category \mathbf{XB}^{PD} is computadic. As the next step, we review computadic unbiased semistrict symmetric monoidal 2-categories and show that \mathbf{XB}^{PD} is an example. Such a 2-category is constructed from a certain presentation (an unbiased semistrict symmetric monoidal 3-computad), which we call an *unbiased semistrict presentation*. This type of presentation \mathbb{P} consists of four sets $(\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{R})$ together with source and target maps $s, t : \mathcal{G}_1 \rightarrow \mathbf{BW}^{\text{USS}}(\mathcal{G}_0)$ and $s, t : \mathcal{G}_2 \rightarrow \mathbf{BS}^{\text{USS}}(\mathcal{G}_1)$, which we

¹¹Different labelings of string diagrams are possible and each possible labeling is a relation.

describe below. For a given such $\mathbb{P} = (\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{R}, s, t)$, the four sets are respectively called *generating objects*, *generating 1–morphisms*, *generating 2–morphisms*, and *generating relations among 2–morphisms*. The following series of definitions start with the ingredients of $F_{\text{uss}}(\mathbb{P})$ and continue with the definition of each ingredient in the given order.

Definition A.3 For a given unbiased semistrict presentation $\mathbb{P} = (\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, \mathcal{R}, s, t)$, the objects of $F_{\text{uss}}(\mathbb{P})$ are *binary words* in \mathcal{G}_0 , the 1–morphisms are *binary sentences* in \mathcal{G}_1 , and the 2–morphisms are *equivalence classes of paragraphs* in \mathcal{G}_2 .

Definition A.4 Let \mathcal{G}_0 be a set. The set $\text{BW}^{\text{uss}}(\mathcal{G}_0)$ of *binary words in \mathcal{G}_0* contains the symbol ι , the elements of \mathcal{G}_0 , and \otimes products, ie $a \otimes b \in \text{BW}^{\text{uss}}(\mathcal{G}_0)$ for all $a, b \in \text{BW}^{\text{uss}}(\mathcal{G}_0)$ such that, for any $a \in \mathcal{G}_0$, the elements $\iota \otimes a$, a , and $a \otimes \iota$ are identified.

Since binary words in \mathcal{G}_0 form the objects of $F_{\text{uss}}(\mathbb{P})$, the set \mathcal{G}_1 of generating 1–morphisms is equipped with source and target maps $s, t: \mathcal{G}_1 \rightarrow \text{BW}^{\text{uss}}(\mathcal{G}_0)$.

Definition A.5 Let \mathcal{G}_1 be a set equipped with maps $s, t: \mathcal{G}_1 \rightarrow \text{BW}^{\text{uss}}(\mathcal{G}_0)$. The set $\text{BW}^{\text{uss}}(\mathcal{G}_1)$ of *binary words in \mathcal{G}_1* contains elements of \mathcal{G}_1 , id_a and $\beta_{a, \sigma(a)}^\sigma$ for any $a \in \text{BW}^{\text{uss}}(\mathcal{G}_0)$ and $\sigma \in \mathcal{S}_n$, where a is a word of length n . The extension of the source and target maps to these elements are $s(\text{id}_a) = a$, $t(\text{id}_a) = a$, $s(\beta_{a, \sigma(a)}^\sigma) = a$ and $t(\beta_{a, \sigma(a)}^\sigma) = \sigma(a)$.

Definition A.6 Let $\text{BW}^{\text{uss}}(\mathcal{G}_1)$ be a set of binary words in \mathcal{G}_1 with $s, t: \text{BW}^{\text{uss}}(\mathcal{G}_1) \rightarrow \text{BW}^{\text{uss}}(\mathcal{G}_0)$. The set $\underline{\text{BS}}^{\text{uss}}(\mathcal{G}_1)$ contains binary words in \mathcal{G}_1 , compositions $g \circ f$ for any $f, g \in \text{BW}^{\text{uss}}(\mathcal{G}_1)$ with $s(g) = t(f)$, and monoidal products $f \otimes g$ for any $f, g \in \text{BW}^{\text{uss}}(\mathcal{G}_1)$. The source and target maps extend naturally to $\underline{\text{BS}}^{\text{uss}}(\mathcal{G}_1)$ by

- $s(g \circ f) = s(f)$ and $t(g \circ f) = t(g)$ for any $g \circ f \in \underline{\text{BS}}^{\text{uss}}(\mathcal{G}_1)$,
- $s(f \otimes g) = s(f) \otimes s(g)$ and $t(f \otimes g) = t(f) \otimes t(g)$ for any $f, g \in \text{BW}^{\text{uss}}(\mathcal{G}_1)$.

The set $\text{BS}^{\text{uss}}(\mathcal{G}_1)$ of *binary sentences in \mathcal{G}_1* is the quotient $\underline{\text{BS}}^{\text{uss}}(\mathcal{G}_1)/\sim$, where \sim is the smallest equivalence relation generated by the identifications $f \otimes \text{id}_\iota \sim f$, $f \sim \text{id}_\iota \otimes f$, $f \circ \text{id}_a \sim f \sim \text{id}_b \circ f$, and $f \otimes f' \sim (\text{id}_b \otimes f') \circ (f \otimes \text{id}_{a'})$ for any $f, f' \in \underline{\text{BS}}^{\text{uss}}(\mathcal{G}_1)$ with $s(f) = a$, $t(f) = b$, $s(f') = a'$, and $t(f') = b'$ (see [22, Lemma 2.81]).

symbol	source	target
id_f	f	f
$\phi_{(f_1, f_2), (f_3, f_4)}^\otimes$	$(f_1 \otimes f_2) \circ (f_3 \otimes f_4)$	$(f_1 \circ f_3) \otimes (f_2 \circ f_4)$
$\phi_{a, a'}^\otimes$	$\text{id}_{a \otimes a'}$	$\text{id}_a \otimes \text{id}_{a'}$
$r\beta_f^\sigma$	$f \circ \beta^\sigma$	$\beta^{\bar{\sigma}} \circ f$
$l\beta_f^\sigma$	$\beta^\sigma \circ f$	$f \circ \beta^{\bar{\sigma}}$
$X^{\sigma, \sigma'}$	$(\beta^\sigma * 1) \circ \beta^{\sigma'}$	$\beta^{\sigma \sigma'}$
X^e	id	β^e

Table 1: Binary words in \mathcal{G}_2 .

Since binary sentences in \mathcal{G}_1 form the 1-morphisms of $\text{F}_{\text{uss}}(\mathbb{P})$, the set \mathcal{G}_2 of generating 2-morphisms is equipped with source and target maps $s, t : \mathcal{G}_2 \rightarrow \text{BS}^{\text{uss}}(\mathcal{G}_1)$ satisfying $s \circ s = s \circ t$ and $t \circ s = t \circ t$. Then, an unbiased semistrict symmetric monoidal 2-computad $\mathbb{P}^{\mathcal{G}}$ consists of generating sets $\mathcal{G}_0, \mathcal{G}_1$ and \mathcal{G}_2 together with maps $s, t : \mathcal{G}_1 \rightarrow \text{BW}^{\text{uss}}(\mathcal{G}_0)$, and $s, t : \mathcal{G}_2 \rightarrow \text{BS}^{\text{uss}}(\mathcal{G}_1)$ satisfying $s \circ s = s \circ t$ and $t \circ s = t \circ t$.

Definition A.7 Let \mathcal{G}_2 be a set equipped with $s, t : \mathcal{G}_2 \rightarrow \text{BS}^{\text{uss}}(\mathcal{G}_1)$ satisfying $s \circ s = s \circ t$ and $t \circ s = t \circ t$. The set $\text{BW}^{\text{uss}}(\mathcal{G}_2)$ of binary words in \mathcal{G}_2 contains every element of \mathcal{G}_2 and the symbols in Table 1 for every $f \in \text{BW}^{\text{uss}}(\mathcal{G}_1)$ and every $f_1, f_2, f_3, f_4 \in \text{BS}^{\text{uss}}(\mathcal{G}_1)$ with $s(f_1) = t(f_3)$ and $s(f_2) = t(f_4)$, and for all $\sigma, e \in S_n, \bar{\sigma} \in S_m$ for $n, m \geq 0$. Moreover, $\text{BW}^{\text{uss}}(\mathcal{G}_2)$ contains the inverses of symbols in Table 1 except for symbols containing the β^σ . The set of preparagraphs $\text{PG}^{\text{uss}}(\mathcal{G}_2)$ is constructed from $\text{BW}^{\text{uss}}(\mathcal{G}_2)$ by adding compositions and monoidal products as above. Similar to $\text{BS}^{\text{uss}}(\mathcal{G}_1)$, there are certain identifications on $\text{PG}^{\text{uss}}(\mathcal{G}_2)$ generated by $\phi_{(\text{id}, f_2), (f_3, f_4)}^\otimes = \text{id}$, $\phi_{(f_1, f_2), (f_3, \text{id})}^\otimes = \text{id}$ and $\phi_{(a, a')}^\otimes = \text{id}$ for all $f_1, f_2, f_3, f_4 \in \text{B}^{\text{uss}}(\mathcal{G}_1)$ and $a, a' \in \text{BW}^{\text{uss}}(\mathcal{G}_0)$. We consider the smallest equivalence relation \sim on $\text{PG}^{\text{uss}}(\mathcal{G}_2)$ generated by these identifications along with identifications

$$\{p \circ p^{-1} = \text{id}_{t(p)}, p^{-1} \circ p = \text{id}_{s(p)}\}_{p \in \text{PG}^{\text{uss}}(\mathcal{G}_2)}$$

and those for β and X coming from the definition of unbiased semistrict symmetric monoidal 2-category. The quotient set is denoted by $\text{PG}^{\text{uss}}(\mathcal{G}_2)$ and called the set of paragraphs in \mathcal{G}_2 .

Definition A.8 The set \mathcal{R} of generating relations among 2-morphisms for an unbiased semistrict symmetric monoidal 2-computad $\mathbb{P}^{\mathcal{G}} = (\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, s, t)$ consists of pairs

(F, G) of paragraphs in \mathcal{G}_2 in $F(\mathbb{P}^{\mathcal{G}})$ with $s(F) = s(G)$ and $t(F) = t(G)$. An unbiased semistrict presentation or unbiased semistrict symmetric monoidal 3-computad \mathbb{P} consists of an unbiased semistrict symmetric monoidal 2-computad $\mathbb{P}^{\mathcal{G}}$ and a set \mathcal{R} of generating relations among 2-morphisms for $\mathbb{P}^{\mathcal{G}}$.

The 2-morphisms of the computadic unbiased semistrict symmetric monoidal 2-category $F_{\text{uss}}(\mathbb{P})$ are the \simeq -equivalence classes of paragraphs in \mathcal{G}_2 , where \simeq is the smallest equivalence relation on $\text{PG}^{\text{uss}}(\mathcal{G}_2)$ such that \simeq is generated by \mathcal{R} and closed under compositions and monoidal products. An unbiased semistrict symmetric monoidal 2-category (\mathcal{C}, β, X) is called *computadic* if there exists a strict symmetric monoidal equivalence $\mathcal{F}: F_{\text{uss}}(\mathbb{P}) \rightarrow \mathcal{C}$ for an unbiased semistrict presentation \mathbb{P} .

In simpler terms, $F_{\text{uss}}(\mathbb{P})$ can be described as follows. The objects of $F_{\text{uss}}(\mathbb{P})$ are words in \mathcal{G}_0 . There are two kinds of elementary 1-morphisms, which can be described as

- (i) $\beta_{a, \sigma(a)}^\sigma: a \rightarrow \sigma(a)$, where $\sigma \in S_n$ for $n \geq 0$ and a is a word of length n ,
- (ii) $\text{id}_a \otimes f \otimes \text{id}_b$, where $f \in \mathcal{G}_1$ and $a, b \in \text{BW}^{\text{uss}}(\mathcal{G}_0)$,

so that nonidentity 1-morphisms of $F_{\text{uss}}(\mathbb{P})$ are given by compositions of elementary 1-morphisms. The 2-morphisms of $F_{\text{uss}}(\mathbb{P})$ are the equivalence classes of string diagrams, where two string diagrams are equivalent if they are related by finitely many (local) moves which come from the generating relations \mathcal{R} , Figures 35 and 36, and the naturality of β and X with generating morphisms. Compositions of morphisms are given by horizontal and vertical concatenations of string diagrams while the (cubical) monoidal product is given by stretching out diagrams from different horizontal directions and merging them (see Figure 24).

Example A.9 Consider an unbiased semistrict presentation

$$\mathbb{X}\mathbb{P} = (\mathcal{X}\mathcal{G}_0, \mathcal{X}\mathcal{G}_1, \mathcal{X}\mathcal{G}_2, \mathcal{X}\mathcal{R})$$

whose generating sets are given as $\mathcal{X}\mathcal{G}_0 = \{\bullet^+, \bullet^-\}$, $\mathcal{X}\mathcal{G}_1 = \{ \overset{\bullet}{\longleftarrow} F_2^g, \overset{\bullet}{\longrightarrow} F_1^g, \overset{\bullet}{\longleftarrow} P_g, \overset{\bullet}{\longrightarrow} N_g \}$, ie G -linear diagrams of $\{ \overset{\bullet}{\longleftarrow} g, g \overset{\bullet}{\longleftarrow}, + \overset{\bullet}{\longleftarrow} g, +, - \overset{\bullet}{\longleftarrow} g, - \}$ without chambering sets, and $\mathcal{X}\mathcal{G}_2$ consists of G -planar diagrams without chambering graphs of generating 2-morphisms in Figure 25. The set of relations $\mathcal{X}\mathcal{R}$ consists of pairs of G -planar diagrams corresponding to pairs of $\langle 2 \rangle$ - X -surfaces in Figure 26.

An object of $F_{\text{uss}}(\mathbb{X}\mathbb{P})$ is either ι or words in \bullet^+ and \bullet^- . Each 1-morphism is a composition of the following two types of elementary 1-morphisms: $\beta_{a, \sigma(a)}^\sigma$, where

$\sigma \in S_n$ and a is a word of length n ; and a G -linear diagram whose 1-morphism is labeled with $\text{id}_{a_1} \otimes \cdots \otimes f \otimes \cdots \otimes \text{id}_{a_n}$ for some $0 < k \leq n$, where a_i is either \bullet^+ or \bullet^- and $f \in \mathcal{G}_1$. Here we use $\text{id}_{\bullet^+} = \overrightarrow{P_e}$, $\text{id}_{\bullet^-} = \overrightarrow{N_e}$, and $f \otimes f' = (\text{id}_b \otimes f') \circ (f \otimes \text{id}_{a'})$. The 2-morphisms of $F_{\text{uss}}(\mathbb{X}\mathbb{P})$ are equivalence classes of paragraphs $\text{PG}^{\text{uss}}(\mathcal{X}\mathcal{G}_2)$, where the equivalence relation is generated by the set of generating relations $\mathcal{X}\mathcal{R}$ (see Figures 17, 18, 19, and 26), and the string diagrams are given in Figures 22, 35, and 36.

A.3 Proof of Theorem 4.5

Note that the string diagram interpretation of elements of $\text{PG}^{\text{uss}}(\mathcal{X}\mathcal{G}_2)$ coincides with the string diagram interpretation of 2-morphisms of XB^{PD} except for (possible) black points on G -linear diagrams (see Figure 14). More precisely, the sets of labels for regions coincide, in both string diagrams there are two types of 1-morphisms whose sets of labels and possible intersecting patterns coincide, and in both string diagrams there are three types of vertices whose sets of labels coincide for each type of vertex. Lastly, equivalence relations on both string diagrams are generated by the same local moves or, equivalently, by the same movie moves. This observation suggests an isomorphism between $F_{\text{uss}}(\mathbb{X}\mathbb{P})$ and XB^{PD} , namely a symmetric monoidal equivalence preserving the unbiased semistrict symmetric monoidal structures. The following lemma shows that this is indeed the case and finishes the proof of Theorem 4.5:

Lemma A.10 *There exists a canonical isomorphism $\Theta : F_{\text{uss}}(\mathbb{X}\mathbb{P}) \rightarrow \text{XB}^{\text{PD}}$ of unbiased semistrict symmetric monoidal 2-categories.*

Proof Comparing the descriptions of unbiased semistrict symmetric monoidal 2-categories $F_{\text{uss}}(\mathbb{X}\mathbb{P})$ and XB^{PD} given above, it is not hard to define the 2-functor Θ . On the level of objects Θ maps ι to the empty set and words in set $\{\bullet^+, \bullet^-\}$ to the finite ordered oriented points given by the words.

On 1-morphisms, it is enough to specify the images of elementary 1-morphisms $\{\beta^\sigma, \text{id}_{a_1} \otimes \cdots \otimes f \otimes \cdots \otimes \text{id}_{a_n}\}$, where $\sigma \in S_n$, $n \geq 0$, $f \in \mathcal{X}\mathcal{G}_1$, and $a_i \in \{\bullet^+, \bullet^-\}$ for $1 \leq i \leq n$. For $\sigma \in S_n$ and a word a , the 1-morphism $\Theta(\beta_{a,\sigma(a)}^\sigma)$ is a G -linear diagram whose chambering set has only one element labeled by β^σ . The latter 1-morphism is mapped to a G -linear diagram described in Example A.9. Recall that 1-morphisms of $F_{\text{uss}}(\mathbb{X}\mathbb{P})$ are equivalence classes determined by certain identifications and 1-morphisms of XB^{PD} are isotopy classes of G -linear diagrams. It is not hard to see that the above assignments are well defined on 1-morphisms.

The 2–functor Θ maps the equivalence class $[P]$ of a paragraph $P \in \text{PG}^{\text{USS}}(\mathcal{X}\mathcal{G}_2)/\sim$ to the equivalence class of the string diagram corresponding to P . This assignment makes sense because, as mentioned in [Example A.9](#), any representative string diagram can be interpreted in both 2–categories. Since in both the 2–categories $\text{F}_{\text{USS}}(\mathbb{X}\mathbb{P})$ and XB^{PD} string diagrams are considered up to the same lists of local moves (movie moves), this assignment is well defined.

We use the Whitehead theorem for symmetric monoidal bicategories ([Theorem 4.3](#)) to show that Θ is a symmetric monoidal equivalence. It is clear that Θ is essentially surjective on objects. We claim that Θ is essentially full on 1–morphisms. To prove this, it is enough to show that every G –linear diagram is isomorphic to a composition of 1–morphisms $\{\Theta(\beta_{a,\sigma(a)}^\sigma), \Theta(\text{id}_{a_1} \otimes \cdots \otimes \text{id}_{a_n}), \Theta(\text{id}_{a_1} \otimes \cdots \otimes f \otimes \cdots \otimes \text{id}_{a_n})\}$ for some $n \geq 0$, $\sigma \in S_n$, and $f \in \mathcal{X}\mathcal{G}_1$. For a given G –linear diagram it is obvious how to write it as a composition of these diagrams except for extra black points. Recall that when we compose G –linear diagrams we do not remove black points along which two diagrams are concatenated. However, there are invertible G –planar diagrams which remove these points (see [Figure 10](#)). Therefore, up to invertible 2–morphisms, every G –linear diagram can be written as a composition of the above 1–morphisms.

Recall that G –planar diagrams are formed using generic maps and the X –manifold data of cobordism type $\langle 2 \rangle$ – X –surfaces. Thus, any G –planar diagram can be obtained from generating 2–morphisms in [Figure 25](#) under horizontal and vertical compositions, and symmetric monoidal product operation. This implies that, for any G –planar diagram, there exists a paragraph such that their equivalence classes are matched by Θ . Consequently, θ is fully faithful on 2–morphisms.

Hence, the Whitehead theorem implies that Θ is an equivalence. By definition, Θ preserves the unbiased semistrict symmetric monoidal structures. That is, XB^{PD} is a computadic unbiased semistrict symmetric monoidal 2–category. \square

References

- [1] **D Ayala, J Francis**, *The cobordism hypothesis*, preprint (2017)
- [2] **J C Baez, J Dolan**, *Higher-dimensional algebra and topological quantum field theory*, *J. Math. Phys.* 36 (1995) 6073–6105 [MR](#) [Zbl](#)
- [3] **P Boisen**, *Graded Morita theory*, *J. Algebra* 164 (1994) 1–25 [MR](#) [Zbl](#)
- [4] **U Bunke, P Turner, S Willerton**, *Gerbes and homotopy quantum field theories*, *Algebr. Geom. Topol.* 4 (2004) 407–437 [MR](#) [Zbl](#)

- [5] **D Calaque, C Scheimbauer**, *A note on the (∞, n) -category of cobordisms*, *Algebr. Geom. Topol.* 19 (2019) 533–655 [MR](#) [Zbl](#)
- [6] **O Davidovich**, *State sums in two dimensional fully extended topological field theories*, PhD thesis, University of Texas at Austin (2011)
- [7] **DS Freed**, *Higher algebraic structures and quantization*, *Comm. Math. Phys.* 159 (1994) 343–398 [MR](#) [Zbl](#)
- [8] **M Golubitsky, V Guillemin**, *Stable mappings and their singularities*, *Graduate Texts in Math.* 14, Springer (1973) [MR](#) [Zbl](#)
- [9] **J Haefner**, *Graded Morita theory for infinite groups*, *J. Algebra* 169 (1994) 552–586 [MR](#) [Zbl](#)
- [10] **J Hesse, C Schweigert, A Valentino**, *Frobenius algebras and homotopy fixed points of group actions on bicategories*, *Theory Appl. Categ.* 32 (2017) 652–681 [MR](#) [Zbl](#)
- [11] **DC Isaksen**, *Calculating limits and colimits in pro-categories*, *Fund. Math.* 175 (2002) 175–194 [MR](#) [Zbl](#)
- [12] **A Kapustin, A Turzillo**, *Equivariant topological quantum field theory and symmetry protected topological phases*, *J. High Energy Phys.* (2017) art. id. 006 [MR](#) [Zbl](#)
- [13] **G Laures**, *On cobordism of manifolds with corners*, *Trans. Amer. Math. Soc.* 352 (2000) 5667–5688 [MR](#) [Zbl](#)
- [14] **RJ Lawrence**, *Triangulations, categories and extended topological field theories*, from “Quantum topology” (L H Kauffman, R A Baadhio, editors), *Ser. Knots Everything* 3, World Sci., River Edge, NJ (1993) 191–208 [MR](#) [Zbl](#)
- [15] **J Lurie**, *On the classification of topological field theories*, from “Current developments in mathematics, 2008” (D Jerison, B Mazur, T Mrowka, W Schmid, R Stanley, S-T Yau, editors), *International*, Somerville, MA (2009) 129–280 [MR](#) [Zbl](#)
- [16] **J Milnor**, *Lectures on the h -cobordism theorem*, Princeton Univ. Press (1965) [MR](#) [Zbl](#)
- [17] **L Müller**, *Extended functorial field theories and anomalies in quantum field theories*, PhD thesis, Heriot-Watt University (2020)
- [18] **L Müller, R J Szabo**, *Extended quantum field theory, index theory, and the parity anomaly*, *Comm. Math. Phys.* 362 (2018) 1049–1109 [MR](#) [Zbl](#)
- [19] **L Müller, R J Szabo**, *'t Hooft anomalies of discrete gauge theories and non-abelian group cohomology*, *Comm. Math. Phys.* 375 (2020) 1581–1627 [MR](#) [Zbl](#)
- [20] **L Müller, L Woike**, *Parallel transport of higher flat gerbes as an extended homotopy quantum field theory*, *J. Homotopy Relat. Struct.* 15 (2020) 113–142 [MR](#) [Zbl](#)
- [21] **P Pstragowski**, *On dualizable objects in monoidal bicategories, framed surfaces and the cobordism hypothesis*, master’s thesis, University of Bonn (2014)

- [22] **C J Schommer-Pries**, *The classification of two-dimensional extended topological field theories*, PhD thesis, University of California, Berkeley (2009) [MR](#) Available at <https://www.proquest.com/docview/304843359>
- [23] **C Schweigert, L Woike**, *Extended homotopy quantum field theories and their orbifoldization*, *J. Pure Appl. Algebra* 224 (2020) art. id. 106213 [MR](#) [Zbl](#)
- [24] **MA Shulman**, *Constructing symmetric monoidal bicategories*, preprint (2010)
- [25] **K Sözer**, *Two-dimensional extended homotopy quantum field theories*, PhD thesis, Indiana University (2020) [MR](#) Available at <https://www.proquest.com/docview/2404618779>
- [26] **K Tagami**, *Unoriented HQFT and its underlying algebra*, *Topology Appl.* 159 (2012) 833–849 [MR](#) [Zbl](#)
- [27] **V Turaev**, *Homotopy field theory in dimension 2 and group algebras*, preprint (1999)
- [28] **V Turaev**, *Homotopy quantum field theory*, EMS Tracts in Mathematics 10, Eur. Math. Soc., Zürich (2010) [MR](#) [Zbl](#)
- [29] **MB Young**, *Orientation twisted homotopy field theories and twisted unoriented Dijkgraaf–Witten theory*, *Comm. Math. Phys.* 374 (2020) 1645–1691 [MR](#) [Zbl](#)

*Le laboratoire de mathématiques Paul Painlevé, Université de Lille
Lille, France*

kursat.sozer@univ-lille.fr

Received: 11 September 2019 Revised: 5 January 2021

Efficient multisections of odd-dimensional tori

THOMAS KINDRED

Rubinstein and Tillmann generalized the notions of Heegaard splittings of 3–manifolds and trisections of 4–manifolds by defining *multisections* of PL n –manifolds, which are decompositions into $k = \lfloor \frac{1}{2}n \rfloor + 1$ n –dimensional 1–handlebodies with nice intersection properties. For each odd-dimensional torus T^n , we construct a multisection which is *efficient* in the sense that each 1–handlebody has genus n , which we prove is optimal; each multisection is *symmetric* with respect to both the permutation action of S_n on the indices and the \mathbb{Z}_k translation action along the main diagonal. We also construct such a trisection of T^4 , lift all symmetric multisections of tori to certain cubulated manifolds, and obtain combinatorial identities as corollaries.

57K50, 57M99, 57N99, 57R10, 57R15; 05A10

1 Introduction

Every closed 3–manifold¹ X admits a decomposition into two 3–dimensional 1–handlebodies² glued along their boundaries. Gay and Kirby extended this classical notion of *Heegaard splittings* by proving that every closed 4–manifold admits a *trisection*, ie a decomposition $X = \bigcup_{i \in \mathbb{Z}_3} X_i$ where each X_i is a 4–dimensional 1–handlebody, each $X_i \cap X_{i+1}$ is a 3–dimensional 1–handlebody, and $X_0 \cap X_1 \cap X_2$ is a closed surface. Rubinstein and Tillmann [9] then extended these decompositions to arbitrary dimension by proving that every closed (PL) manifold of arbitrary dimension admits a *PL multisection*:

Definition 1.1 A *PL multisection* of a closed manifold X of dimension $n = 2k - 1$ (resp. $2k - 2$) is a decomposition $X = \bigcup_{i \in \mathbb{Z}_k} X_i$, where:

¹Unless stated otherwise, all *manifolds* are piecewise-linear (PL), compact, connected, and orientable. A manifold X is *closed* if $\partial X = \emptyset$. A general reference by Rourke and Sanderson is [8].

²A d –dimensional h –*handlebody* is a d –manifold obtained by gluing d –dimensional r –handles for various $r = 0, \dots, h$. Since we work in the PL category, the gluing maps must be PL and the attaching regions must be PL submanifolds.

- Each X_i is an n -dimensional 1-handlebody.
- $\bigcap_{i \in \mathbb{Z}_k} X_i$ is a closed $(n+1-k)$ -dimensional submanifold.
- $\bigcap_{i \in I} X_i$ is an $(n+1-|I|)$ -dimensional $|I|$ - (resp. $(|I|-1)$ -) handlebody for each $I \subset \mathbb{Z}_k$ with $2 \leq |I| \leq k-1$.³

One may define *smooth* multisections of *smooth* manifolds analogously: the only extra condition is that, for each nonempty $I \subset \mathbb{Z}_k$, the inclusion of $X_I = \bigcap_{i \in I} X_i$ into X is a smooth embedding, with corners.⁴ Lambert-Cole and Miller proved that every smooth 5-manifold admits a smooth trisection [6]. In dimensions $n \geq 6$, the topic is wide open. In particular:

Question 1 Does every closed smooth manifold of arbitrary dimension admit a smooth multisection?

The distinction between PL multisections and smooth ones comes down to that of PL and smooth handle decompositions.⁵ This is because any PL multisection $X = \bigcup_{i \in \mathbb{Z}_k} X_i$ gives rise to a nice PL handle decomposition (see Proposition 2.5) coming from handle decompositions of the various X_I ; requiring each inclusion $X_I \hookrightarrow X$ to be smooth (with corners) ensures that the gluings in this handle decomposition are smooth. Henceforth, unless stated otherwise, all multisections are PL.

The topology of a closed manifold X of dimension $n \neq 2$ bounds the *efficiency* of its multisection as follows. Let $g(X_i)$ denote the *genus* of X_i .⁶

Definition 1.2 The *efficiency* of a multisection $X = \bigcup_{i \in \mathbb{Z}_k} X_i$ is

$$\frac{1 + \text{rank } \pi_1(X)}{1 + \max_i g(X_i)}$$

A multisection is *efficient* if its efficiency is 1.

³Rubinstein and Tillmann state this condition differently, requiring that each $\bigcap_{i \in I} X_i$ is an $(n+1-|I|)$ -dimensional submanifold with an $|I|$ - (resp. $(|I|-1)$ -) dimensional spine, where a spine of a manifold N is a subpolyhedron $P \subset \text{int}(N)$ onto which N collapses. Certainly any h -handlebody has an h -dimensional spine. Conversely, given a spine P of N , we may assume that N is triangulated and P is a simplicial subcomplex which admits no elementary collapses; then N is PL homeomorphic to a regular neighborhood R of P in N , and R has handle decomposition consisting of one r -handle for each r -simplex in P .

⁴More precisely, for nonempty $I \subset \mathbb{Z}_k$, the set of corner points of X_I must be $\text{corners}(X_I) = \bigcup_{i, j \notin I; i \neq j} X_I \cap X_i \cap X_j$.

⁵While any smooth structure determines a (smooth) handle decomposition, and conversely, a PL handle decomposition does not necessarily determine a smooth structure.

⁶ X_i is an n -dimensional 1-handlebody, so we have $X_i \cong \natural^g(S^1 \times D^{n-1})$ for some $g = g(X_i)$. (Throughout, we denote PL homeomorphism by \cong .)

We will show:

Corollary 2.7 *In any dimension $n \neq 2$, no multisection of any manifold has efficiency greater than 1, and in any efficient multisection $X = \bigcup_{i \in \mathbb{Z}_k} X_i$, all X_i have the same genus, $g(X_i) = \text{rank } \pi_1(X)$.⁷*

This notion of an *efficient* multisection generalizes a notion introduced by Lambert-Cole and Meier in [5]. They call a trisection of a simply connected 4-manifold X *efficient* if the genus of the central surface Σ equals $b_2(X)$. Indeed, one always has $g(\Sigma) \leq b_2(X)$, and equality holds if and only if each piece of the trisection is a 4-ball.

We close the introduction with an outline of the paper.

- **Section 2** establishes several general properties of multisections.
- **Section 3** begins a detailed investigation of multisections of *odd-dimensional tori*, starting with detailed descriptions of the multisections of T^n for $n = 3, 4, 5$. Roughly stated, the main result is:

Theorem 7.10 *Each $n = (2k-1)$ -torus admits an efficient multisection which is symmetric with respect to the S_n permutation action on the indices and the \mathbb{Z}_k translation action along the main diagonal.*

The full version of **Theorem 7.10** gives a simple expression (1) for each piece X_i of this multisection. The hard part is describing a handle decomposition of $X_I = \bigcap_{i \in I} X_i$ for arbitrary n and $I \subsetneq \mathbb{Z}_k$.

- **Section 4** introduces three types of building blocks; under our main construction, each handle of each X_I will be a product of such blocks.
- **Section 5** describes further examples of X_I under our construction, each featuring a new complication in its handle decomposition.
- **Section 6** proves several combinatorial facts about our main construction. In particular, **Section 6.2** proves that $T^n = \bigcup_{i \in \mathbb{Z}_k} X_i$, and **Section 6.4** establishes a closed expression (2) for arbitrary X_I . Also, **Section 6.3** establishes two combinatorial corollaries, which may be of independent interest.

⁷In dimension two, efficiency is strictly bounded above by 2; this bound is sharp, since any surface of even genus g admits a multisection with efficiency $(1 + 2g)/(1 + g)$.

- [Section 7](#) describes a handle decomposition of arbitrary X_I from our main construction, confirms the details of this decomposition, shows that the central intersection $\bigcap_{i \in \mathbb{Z}_k} X_i$ is a closed k -manifold, and puts everything together to prove [Theorem 7.10](#).
- [Section 8](#) extends [Theorem 7.10](#) to certain cubulated manifolds.
- [Appendix A](#) features tables, several detailing follow-up examples for the complications introduced in [Sections 3](#) and [5](#), others detailing aspects of the handle decomposition described in [Section 7.1](#).
- [Appendix B](#) describes four other ways one might try to multisection T^n .

Acknowledgments Thank you to Mark Brittenham, Charlie Frohman, Hugh Howards, Peter Lambert-Cole, and Maggie Miller for helpful discussions. Thank you to the referee for numerous suggestions to improve the clarity and exposition of the paper. Special thanks to Alex Zupan for helpful discussions throughout the project, especially during its early stages, when we collaborated to find efficient trisections of T^4 and T^5 .

2 Multisections and their efficiency

In this section, we describe a way of obtaining a (PL) handle decomposition of a manifold given a multisection (see [Proposition 2.5](#)), and we deduce, with the exception of 2-manifolds, that no multisection has efficiency greater than 1 (see [Corollary 2.7](#)). We begin, however, by describing examples of multisections in arbitrary dimension.

2.1 Simple examples of multisections

Example 2.1 For $n = 2k - 1$, the n -sphere

$$S^n = \partial \prod_{i=0}^{k-1} D^2 = \bigcup_{i=0}^{k-1} \left(\prod_{j=0}^{i-1} D^2 \times S^1 \times \prod_{j=i+1}^{k-1} D^2 \right)$$

admits a multisection in which each

$$X_i = \prod_{j=0}^{i-1} D^2 \times S^1 \times \prod_{j=i+1}^{k-1} D^2$$

is an n -dimensional 1-handlebody of genus 1. In dimension 3, this is the genus-1 Heegaard splitting of $S^3 = D^2 \times D^2$ with central surface $S^1 \times S^1$. In arbitrary

dimension n , the central intersection is the k -torus $\prod_{j=0}^{k-1} S^1$, and more generally, for each $I \subset \mathbb{Z}_k$ with $1 \leq |I| = \ell \leq k - 1$, the intersection

$$X_I = \bigcap_{j \in I} X_j = \prod_{j=0}^{k-1} \left\{ \begin{array}{l} S^1 \text{ if } j \in I, \\ D^2 \text{ if } j \notin I \end{array} \right\} \cong \prod_{j=0}^{\ell-1} S^1 \times \prod_{j=\ell}^{k-1} D^2 \cong T^\ell \times D^{2(k-\ell)}$$

is a thickened ℓ -torus. In dimension 5, Lambert-Cole and Miller use this construction and a second trisection of S^5 , whose central intersection is a 3-sphere rather than a 3-torus, to show that, unlike Heegaard splittings of 3-manifolds and trisections of 4-manifolds, trisections of a given 5-manifold need not be stably equivalent [6].

Example 2.2 [9] Using homogeneous coordinates $[z_0 : \dots : z_{k-1}]$ on $\mathbb{C}\mathbb{P}^{k-1}$, one can define a multisection by

$$X_i = \{[z_0 : \dots : z_{k-1}] : |z_i| \geq |z_j| \text{ for } j = 0, \dots, k - 1\}.$$

Then each X_I with $|I| = \ell$ is related by permutation to a thickened torus

$$\bigcap_{i=0}^{\ell-1} X_i = \left\{ [1 : z_1 : \dots : z_{k-1}] \mid \begin{array}{l} |z_j| = 1 \text{ for } j = 1, \dots, \ell - 1, \\ |z_j| \leq 1 \text{ for } j = \ell, \dots, k - 1 \end{array} \right\} \cong T^{\ell-1} \times D^{2(k-\ell)}.$$

In particular, the central intersection is the k -torus

$$\{[1 : z_1 : \dots : z_{k-1}] : |z_1| = \dots = |z_{k-1}| = 1\}.$$

These symmetric multisections are also efficient, since each X_i has genus 0.

2.2 General properties of multisections

Proposition 2.3 Let Z_i be an n -dimensional h_i -handlebody, $i = 1, 2$, and let $\phi: Y_1 \rightarrow Y_2$ glue compact $Y_i \subset \partial Z_i$, such that $Y_1 \cong Y_2$ is an h -handlebody. Then $Z = Z_1 \cup_\phi Z_2$ is an h' -handlebody for $h' = \max\{h_1, h_2, h + 1\}$.

Proof By taking a regular neighborhood N of $Y = \phi(Y_1) = \phi(Y_2)$ in Z , where $N \equiv Y \times I$, we may identify $Z \setminus \text{int}(N)$ with $Z_1 \sqcup Z_2$, which is a 2-component h'' -handlebody, where $h'' = \max\{h_1, h_2\}$. Then, for each i -handle $H \equiv D^i \times D^{n-1-i}$ in Y for $0 \leq i \leq h$, we can glue on $H \times I$ along $\partial(D^i \times I) \times D^{n-1-i} \cong S^i \times D^{n-1-i}$, and so attaching $H \times I$ is the same as attaching an $(i + 1)$ -handle, where $i + 1 \leq h + 1$. \square

Proposition 2.4 Let $X = \bigcup_{i \in \mathbb{Z}_k} X_i$ be a multisection of a closed manifold of dimension $n = 2k - 1$ (resp. $n = 2k - 2$). Then for each $1 \leq j \leq i \leq k - 1$

$$\bigcup_{t=0}^{j-1} X_t \cap \bigcap_{t=j}^i X_t$$

is a $(2k + j - i - 2)$ -dimensional $(i + j)$ -handlebody (resp. $(2k + j - i - 3)$ -dimensional $(i + j - 1)$ -handlebody).

Proof We address the odd-dimensional case, arguing by lexicographical induction on (i, j) . The even-dimensional case follows analogously. When $(i, j) = (1, 1)$, the proposition is true by definition, since $X_0 \cap X_1$ is a 2-handlebody.

Let $(i, j) > (1, 1)$. Assume, for each $(r, s) < (i, j)$, that $(X_0 \cup \dots \cup X_{s-1}) \cap X_s \cap \dots \cap X_r$ is a $(2k + s - r - 2)$ -dimensional $(r + s)$ -handlebody. Let

$$Z_1 = \bigcup_{t=0}^{j-2} X_t \cap \bigcap_{t=j}^i X_t \quad \text{and} \quad Z_2 = \bigcap_{t=j-1}^i X_t,$$

so that

$$\bigcup_{t=0}^{j-1} X_t \cap \bigcap_{t=j}^i X_t = Z_1 \cup Z_2.$$

So, by induction, Z_1 is a $(2k + j - i - 2)$ -dimensional $(i + j - 2)$ -handlebody, and, by the definition of multisection, Z_2 is a $(2k + j - i - 2)$ -dimensional $(i + 1 - j)$ -handlebody. Further,

$$Z_1 \cap Z_2 = \bigcup_{t=0}^{j-2} X_t \cap \bigcap_{t=j-1}^i X_t,$$

which, by induction, is a $(2k + j - i - 3)$ -dimensional $(i + j - 1)$ -handlebody. Therefore, by Proposition 2.3, $Z_1 \cup Z_2$ is a $(2k + j - i - 2)$ -dimensional h -handlebody, where

$$h = \max\{i + j - 2, i + 1 - j, i + j\} = i + j. \quad \square$$

Proposition 2.5 Let $X = \bigcup_{i \in \mathbb{Z}_k} X_i$ be a multisection of a closed manifold of dimension $n = 2k - 1$ (resp. $n = 2k - 2$). Then X admits a handle decomposition in which each X_j contributes only r -handles for $r \leq 2j + 1$ (resp. $r \leq 2j$).

Proof We address the odd-dimensional case; the even-dimensional case follows analogously. Arguing by induction on i , we will show that $X_0 \cup \dots \cup X_i$ admits a

handle decomposition in which each X_j contributes only r -handles for $r \leq 2j + 1$. The base case is trivial. For the induction step, consider

$$(X_0 \cup \cdots \cup X_{i-1}) \cup_{(X_0 \cup \cdots \cup X_{i-1}) \cap X_i} X_i.$$

By induction, $X_0 \cup \cdots \cup X_{i-1}$ admits a handle decomposition in which each X_j contributes only r -handles for $r \leq 2j + 1$. Extend this to the required handle decomposition of $X_0 \cup \cdots \cup X_i$ as follows. Let N be a collared neighborhood of $(X_0 \cup \cdots \cup X_{i-1}) \cap X_i$ in X_i . As in the proof of [Proposition 2.3](#), first construct the disjoint union

$$(X_0 \cup \cdots \cup X_{i-1}) \sqcup (X_i \setminus \text{int}(N)),$$

thereby contributing 0- and 1-handles to X_i , as $X_i \setminus \text{int}(N)$ is PL homeomorphic to X_i . Second, glue in N , thereby contributing r -handles for $r = 1, \dots, 2i + 1$, since $(X_0 \cup \cdots \cup X_{i-1}) \cap X_i$ is a $2i$ -handlebody by [Proposition 2.4](#). \square

2.3 Efficiency of multisections

Next, we consider the efficiency of multisections in light of [Proposition 2.5](#). Recall [Definition 1.2](#).

Proposition 2.6 *In dimension $n \neq 2$, any multisection $X = \bigcup_{i \in \mathbb{Z}_k} X_i$ obeys*

$$\min_{i \in \mathbb{Z}_k} g(X_i) \geq \text{rank } \pi_1(X).$$

Proof Given a multisection of X , label the pieces so that $g(X_{k-1}) \leq g(X_i)$ for all i . Construct a handle structure on X as guaranteed by [Proposition 2.5](#). All the n - and $(n-1)$ -handles are in X_{k-1} , since $n \neq 2$. Flip X upside down. Now all the 0- and 1-handles are in X_{k-1} , so

$$\text{rank } \pi_1(X) \leq \text{rank } \pi_1(X_{k-1}) = g(X_{k-1}) = \min_{i \in \mathbb{Z}_k} g(X_i). \quad \square$$

Corollary 2.7 *In any dimension $n \neq 2$, no multisection of any manifold has efficiency greater than 1, and in any efficient multisection $X = \bigcup_{i \in \mathbb{Z}_k} X_i$, all X_i have the same genus, $g(X_i) = \text{rank } \pi_1(X)$.*

3 Motivating examples

In this section, we describe our multisections of T^3 , T^4 , and T^5 in detail. We also establish notation that will be used throughout the rest of the paper.

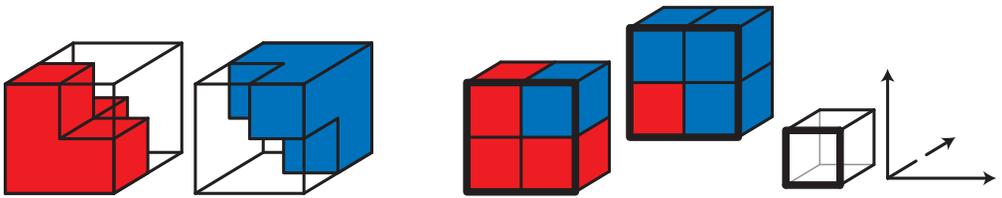


Figure 1: A Heegaard splitting of T^3 .

3.1 Intuitive approach to T^3 , T^4 , and T^5

Figure 1 illustrates an efficient Heegaard splitting of the 3–torus, which suggests viewing T^3 as $(\mathbb{R}/2\mathbb{Z})^3$; then the splitting is determined by a partition of the eight unit cubes with vertices in the lattice $(\mathbb{Z}/2\mathbb{Z})^3$. Moreover, *this* partition satisfies two symmetry properties. First, the permutation action of S_3 on the indices in T^3 fixes each piece of the splitting. Second, the \mathbb{Z}_2 translation action along the main diagonal of T^3 switches the two pieces: $X_i + (1, 1, 1) = X_{i+1}$.

How might one construct efficient trisections of T^n for $n = 4, 5$ with symmetry properties analogous to Figure 1’s splitting of T^3 ? To begin, one might view these T^n as $(\mathbb{R}/3\mathbb{Z})^n$ — rather than, say, $(\mathbb{R}/2\mathbb{Z})^n$, because we seek a trisection rather than a splitting — and seek an appropriate partition of the 3^n unit cubes with vertices in the lattice $(\mathbb{Z}/3\mathbb{Z})^n$. From now on, for brevity, we will refer to these unit cubes as *subcubes* of T^n .

To start forming this partition, one might assign each subcube $[i, i + 1]^n$ to X_i (because of the translation action). Next, one might assign subcubes of the forms $[i, i + 1]^{n-1}[i + 1, i + 2]$ and $[i, i + 1]^{n-1}[i - 1, i]$ to X_i as well, and extend these assignments using the permutation action on the indices. At this point, each X_i is indeed an n –dimensional 1–handlebody, and so the rest of the partition should be constructed in a way that preserves this fact, while also giving rise to the needed intersection properties. Figure 2 illustrates this intermediate stage in the case of T^4 .⁸

⁸All combinatorial data conveyed in Figures 2 and 3 comes from the arrangements of the nine 3×3 squares outlined in bold; beyond this, the style of the illustration reflects the fact that each pictured subcube is a 4–cube. A model 4–cube is also drawn, next to coordinate axes. The solid axes represent directions in which abutting subcubes are shown in contact with each other (understanding that the interval that appears as $[0, 3]$ actually represents the circle $\mathbb{R}/3\mathbb{Z}$); the dashed axes represent directions in which abutting subcubes align at a distance in the figure. Similarly, Figure 1 shows a model 3–cube and coordinate axes, and Figure 5 a model 5–cube and coordinate axes.

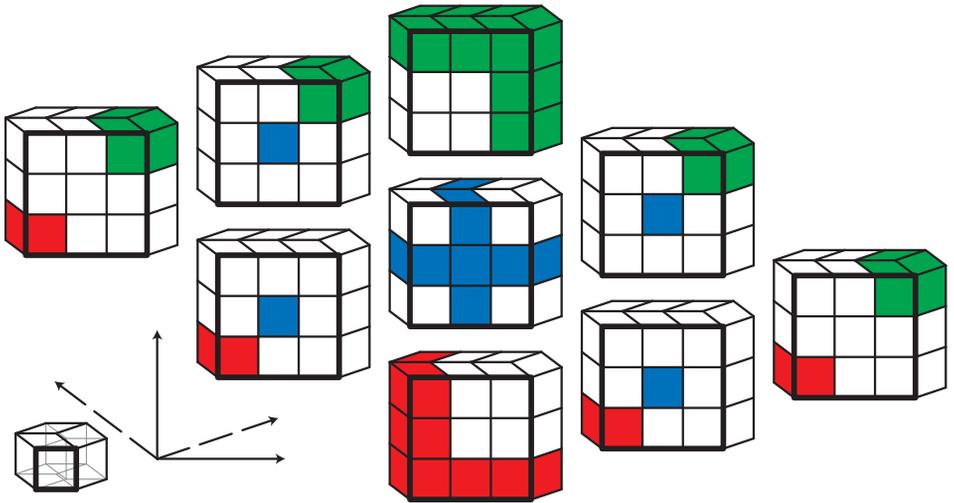


Figure 2: Start partitioning the subcubes of $T^4 = (\mathbb{R}/3\mathbb{Z})^4$ like this, giving three 4-dimensional 1-handlebodies.

For T^4 , the symmetry properties imply that the remaining partition is determined by the assignments of the subcubes $[0, 1]^2[1, 2][2, 3]$ and $[0, 1]^2[1, 2]^2$. Assigning both subcubes to X_0 and extending symmetrically gives the decomposition of T^4 illustrated in Figures 3 and 4. Section 3.3 will confirm that this decomposition is indeed a trisection.

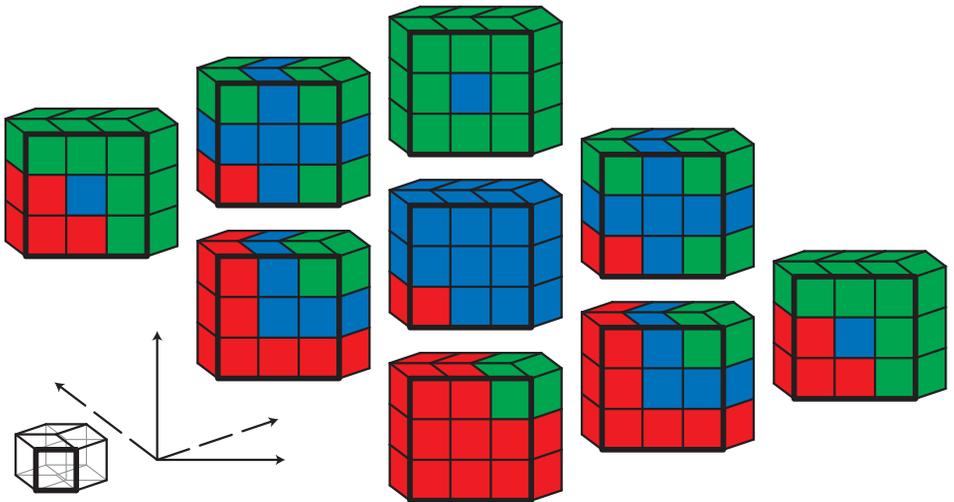


Figure 3: Partitioning the 3^4 subcubes of $T^4 = (\mathbb{R}/3\mathbb{Z})^4$ like this gives a symmetric efficient trisection.

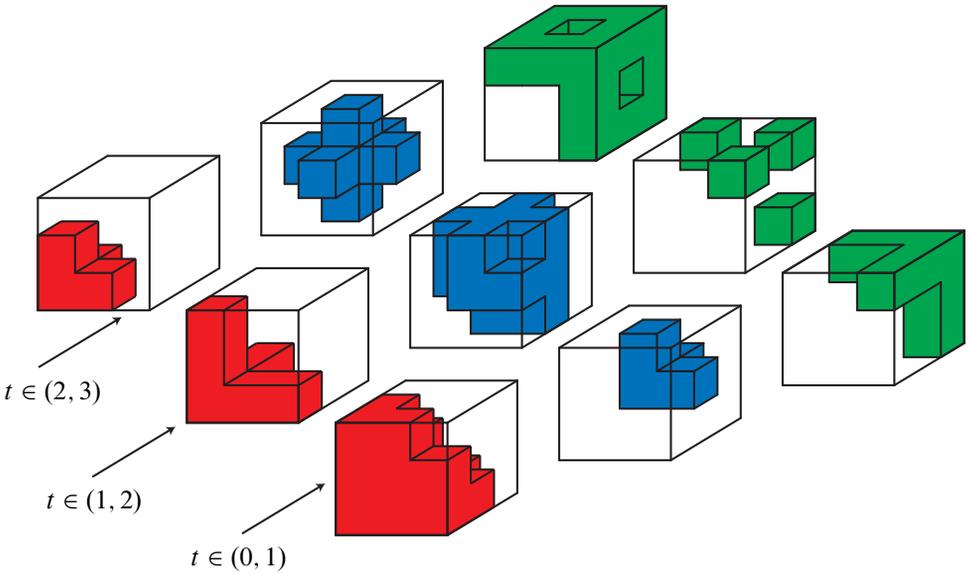


Figure 4: In the multisection of T^4 from Figure 3, each slice $T^3 \times \{t\}$, $t \in (\mathbb{R}/3\mathbb{Z}) \setminus \mathbb{Z}_3$, intersects X_0, X_1, X_2 like this.

A similar approach leads to the decomposition of T^5 shown in Figure 5. Section 3.4 will confirm that this, too, is a trisection.

3.2 Notation

Notation 3.1 Let $X, Y \subset Z$ be compact subspaces of a topological space. Denote “ X cut along Y ” by $X \parallel Y$. In every example where we use this notation, $X \parallel Y$ equals the closure of $X \setminus Y$ in Z . (The general construction is somewhat more complicated.)

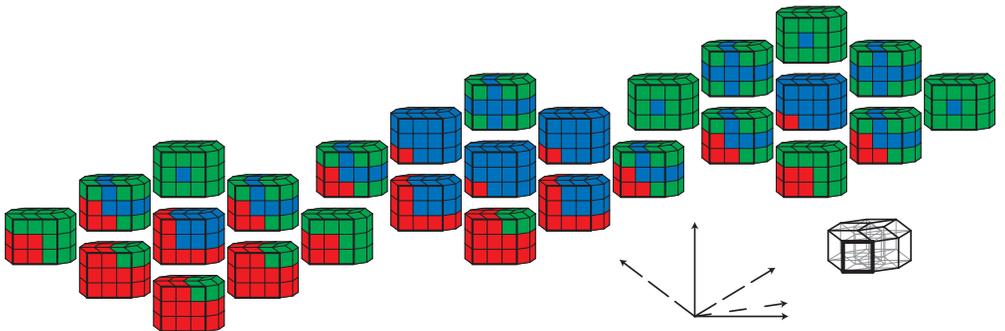


Figure 5: Partitioning the 3^5 subcubes of $T^5 = (\mathbb{R}/3\mathbb{Z})^5$ like this gives a symmetric efficient trisection.

Given $n = 2k - 1, 2k - 2$, view the n -torus T^n as $(\mathbb{R}/k\mathbb{Z})^n$. Let S_n denote the permutation group on n elements.

Notation 3.2 Given $\vec{x} = (x_1, \dots, x_n) \in T^n$ and $\sigma \in S_n$, write

$$\vec{x}_\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

Also, given $U \subset T^n$ and $\vec{v} \in T^n$, write

$$U + \vec{v} = \{\vec{u} + \vec{v} : \vec{u} \in U\}.$$

The symmetric group S_n acts on T^n by permuting the indices via $\sigma : \vec{x} \mapsto \vec{x}_\sigma$. Because we are interested in subsets of T^n which are fixed by this action:

Notation 3.3 For any subset $U \subset T^n$, write

$$\langle U \rangle = \{\vec{x}_\sigma : \vec{x} \in U, \sigma \in S_n\} \subset T^n.$$

Note, for any $U \subset T^n$, that $\langle U \rangle$ is fixed by the action of S_n on T^n . We can state our main result explicitly:

Theorem 7.10 For $n = 2k - 1$, the n -torus $T^n = (\mathbb{R}/k\mathbb{Z})^n = [0, k]^n / \sim$ admits an efficient multisection $T^n = \bigcup_{i \in \mathbb{Z}_k} X_i$ defined by

$$(1) \quad X_0 = \langle [0, 1]^2 \cdots [0, k - 1]^2 [0, k] \rangle, \quad X_i = X_0 + (i, \dots, i), \quad i \in \mathbb{Z}_k.$$

By construction, the decomposition is symmetric with respect to the permutation action on the indices and the translation action on the main diagonal.

Anticipating the concrete and (somewhat) low-dimensional nature of the examples in Sections 3 and 5 and Appendix A, we give the first few intervals $[i, i + 1]$ for $i \in \mathbb{Z}_k$ special notations:

Notation 3.4 Write

$$[0, 1] = \alpha, \quad [1, 2] = \beta, \quad [2, 3] = \gamma, \quad [3, 4] = \delta, \quad [4, 5] = \varepsilon, \quad [5, 6] = \zeta, \quad [6, 7] = \eta.$$

To further abbreviate our notation, we often omit \times symbols and use exponents to denote repeated factors. For example, we can describe the two pieces of the Heegaard splitting of T^3 from Figure 1 like this:

$$X_0 = \alpha^3 \cup \alpha^2 \beta \cup \alpha \beta \alpha \cup \beta \alpha^2, \quad X_1 = \beta^3 \cup \beta^2 \alpha \cup \beta \alpha \beta \cup \alpha \beta^2.$$

Using Notation 3.3, we can further abbreviate this notation:

$$X_0 = \alpha^3 \cup \langle \alpha^2 \beta \rangle = \langle \alpha^2 [0, 2] \rangle, \quad X_1 = \beta^3 \cup \langle \alpha \beta^2 \rangle = \langle [0, 2] \beta^2 \rangle.$$

We often omit the braces around singleton factors. For example, in T^3

$$X_0 \cap X_1 = \langle [0, 1] \times [1, 2] \times \{0\} \rangle \cup \langle [0, 1] \times [1, 2] \times \{1\} \rangle = \langle \alpha\beta 0 \rangle \cup \langle \alpha\beta 1 \rangle.$$

We also extend [Notation 3.3](#) in the way suggested by the following example:

$$\langle 0\alpha \rangle \beta^2 = (\{0\} \times \alpha \times \beta \times \beta) \cup (\alpha \times \{0\} \times \beta \times \beta).$$

More precisely, if we decompose T^n as a product $T^n = T^{n_1} \times \dots \times T^{n_p}$ and $U_i \subset T^{n_i}$ for $i = 1, \dots, p$, then

$$\langle U_1 \rangle \cdots \langle U_p \rangle = \{(\vec{x}_{\sigma_1}^1, \vec{x}_{\sigma_2}^2, \dots, \vec{x}_{\sigma_p}^p) : \vec{x}^i \in T^{n_i}, \sigma_i \in S_{n_i}, i = 1, \dots, p\},$$

where, extending [Notation 3.2](#) and writing $\vec{x}^i = (x_1^i, \dots, x_{n_i}^i)$,

$$\vec{x}_{\sigma_i}^i = (x_{\sigma_i(1)}^i, \dots, x_{\sigma_i(n_i)}^i).$$

Starting in dimension 7, some handle decompositions will require subdividing unit subintervals $\alpha, \beta, \gamma, \delta, \dots$ into halves or thirds. Anticipating this:

Notation 3.5 Write

$$\alpha^- = [0, \frac{1}{2}], \quad \alpha^+ = [\frac{1}{2}, 1], \quad \dots, \quad [\eta^-] = [6, \frac{13}{2}], \quad \eta^+ = [\frac{13}{2}, 7]$$

and

$$\alpha_3^- = [0, \frac{1}{3}], \quad \alpha_3^\circ = [\frac{1}{3}, \frac{2}{3}], \quad \alpha_3^+ = [\frac{2}{3}, 1], \quad \dots, \quad \eta_3^\circ = [\frac{19}{3}, \frac{20}{3}], \quad \eta_3^+ = [\frac{20}{3}, 7].$$

Because of the symmetry of our main construction under the \mathbb{Z}_k translation action on T^n , it will suffice, when considering X_I from that construction, to allow I to be arbitrary *only up to cyclic permutation*. In order to utilize this convenience:

Notation 3.6 Given $I \subset \mathbb{Z}_k$ with $|I| = \ell > 0$, write $X_I = \bigcap_{i \in I} X_i$, and write $I = \{i_s\}_{s \in \mathbb{Z}_\ell}$ such that

$$0 \leq i_0 < i_1 < \dots < i_{\ell-1} \leq k - 1.$$

Definition 3.7 Let $I = \{i_s\}_{s \in \mathbb{Z}_\ell}$ as in [Notation 3.6](#). For each $r \in \mathbb{Z}_\ell$, define $I^r = \{i + r : i \in I\} \subset \mathbb{Z}_k$. Write each $I^r = \{i_s^r\}_{s \in \mathbb{Z}_\ell}$ with $0 \leq i_0^r < i_1^r < \dots < i_{\ell-1}^r \leq k - 1$. Say that I is *simple* if, for each $r \in \mathbb{Z}_\ell$, we have $I \leq I^r$ under the lexicographical ordering of their elements, ie if each $I_r \neq I$ has some $s \in \mathbb{Z}_\ell$ with $i_t = i_t^r$ for each $t = 0, \dots, s - 1$ and $i_s < i_s^r$.

Notation 3.8 Given simple $I = \{i_s\}_{s \in \mathbb{Z}_\ell} \subsetneq \mathbb{Z}_k$ as in [Notation 3.6](#), define

$$T = \{s \in \mathbb{Z}_\ell : i_s - 1 \notin I\}.$$

Write $T = \{t_r\}_{r \in \mathbb{Z}_m}$ with $0 = t_0 < \dots < t_m < \ell$ (see [Observation 3.10](#)). For each $r \in \mathbb{Z}_m$, write $I_r = \{i_{t_r}, \dots, i_{t_{r+1}-1}\}$. Then

$$I = I_1 \sqcup \dots \sqcup I_m,$$

and, for each $r = 0, \dots, m - 1$, we have $|I_r| = \max I_r + 1 - \min I_r$ (each block I_r comprises consecutive indices) and $\min I_{r+1} \geq \max I_r + 2$ (the blocks are nonconsecutive).

Given $i_* \in I$ (denoted specifically by i_*), denote the block I_r containing i_* by I_* .

Convention 3.9 Throughout, reserve the notations $n, k, \alpha, \dots, \eta, \alpha^-, \dots, \eta^+, \alpha_3^-, \dots, \eta_3^+, I, X_I, \ell, T$, and m for the way they are used in [Notations 3.4–3.8](#), assume, unless otherwise stated, that $I \subset \mathbb{Z}_k$ is simple, and reserve, for any $s \in \mathbb{Z}_\ell$ or $r \in \mathbb{Z}_m$, the notations i_s, t_r, I_r, i_* , and I_* for the way they are used in [Notations 3.6](#) and [3.8](#).

Observation 3.10 Given $I \subsetneq \mathbb{Z}_k$, we have $i_0 = 0, i_{\ell-1} \leq k - 2$, and $|I_0| \geq |I_r|$ for each $r \in \mathbb{Z}_m$; if $|I_0| = |I_r|$, then $|I_1| \geq |I_{r+1}|$.

Given $I \subset \mathbb{Z}_k$ and $s \in \mathbb{Z}_\ell$, write

$$(i_1, \dots, \hat{i}_s, \dots, i_\ell) = (i_1, \dots, i_{s-1}, i_{s+1}, \dots, i_\ell) \subset T^{\ell-1}.$$

We now have enough notation to describe a closed formula for the X_I coming from our main construction [\(1\)](#):

Lemma 6.13 Given nonempty $I \subseteq \mathbb{Z}_k$, X_I is given by

$$(2) \quad \bigcup_{i_* \in I} \left\langle (i_1, \dots, \hat{i}_*, \dots, i_\ell) \prod_{r \in \mathbb{Z}_\ell} [i_r, i_r + 1]^2 \cdots [i_r, i_{r+1} - 1]^2 [i_r, i_{r+1}] \right\rangle.$$

In particular,

$$(3) \quad \bigcap_{i \in \mathbb{Z}_k} X_i = \bigcup_{i_* \in \mathbb{Z}_k} \left\langle (0, \dots, \hat{i}_*, \dots, k - 1) \prod_{i \in \mathbb{Z}_k} [i, i + 1] \right\rangle.$$

We will prove [Lemma 6.13](#) in [Section 6.4](#).

3.3 Trisection of T^4

The decomposition of T^4 from [Figure 3](#) is given by

$$(4) \quad \begin{aligned} X_0 &= \langle \alpha^2[0, 2][0, 3] \rangle = \langle \alpha^4 \rangle \cup \langle \alpha^3 \beta \rangle \cup \langle \alpha^3 \gamma \rangle \cup \langle \alpha^2 \beta^2 \rangle \cup \langle \alpha^2 \beta \gamma \rangle, \\ X_i &= X_0 + (i, i, i, i). \end{aligned}$$

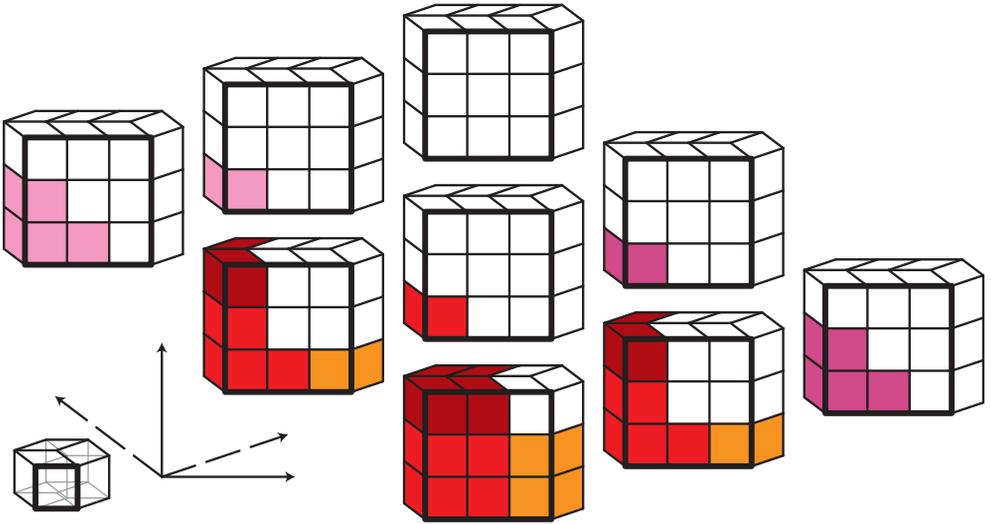


Figure 6: A handle decomposition of X_0 in Figure 3's trisection of T^4 . Each of the five handles has a different color.

It is evident from Figure 3 that $X_0 \cup X_1 \cup X_2 = T^4$. Also, $I = \{0\}$ and $I = \{0, 1\}$ are the only proper subsets of $\{0, 1, 2\}$ which are simple. Therefore, to check that (4) determines a trisection of T^4 it suffices to prove that X_0 is a 4-dimensional 1-handlebody and $X_0 \cap X_1$ is a 3-dimensional 1-handlebody with $\partial(X_0 \cap X_1) = X_0 \cap X_1 \cap X_2$.

Indeed, Figure 6 shows a handle decomposition of X_0 in which $\langle \alpha^2[0, 2]^2 \rangle$ is a 0-handle and $\langle \alpha^2[0, 2]\gamma \rangle$ supplies four 1-handles, each a permutation of $\langle \alpha^2[0, 2]\gamma \rangle$. More precisely, each 1-handle is given, in terms of some permutation $\sigma \in S_4$ (using Notation 3.2), by

$$(5) \quad \{ \vec{x}_\sigma : \vec{x} \in \langle \alpha^2[0, 2]\gamma \rangle \}.$$

Now consider

$$(6) \quad X_0 \cap X_1 = \langle \alpha 1 \beta [1, 3] \rangle \cup \langle 0 \alpha \beta^2 \rangle.$$

We claim that this is a 3-dimensional 1-handlebody in which

- $Y_1 = \langle \alpha 1 \beta^2 \rangle$ is the 0-handle;
- $Y_2 = \langle 0 \alpha \beta^2 \rangle$ gives six 1-handles, all permutations of $Y_2^* = \langle 0 \alpha \rangle \beta^2$;
- $Y_3 = \langle \alpha 1 \beta \gamma \rangle$ gives four 1-handles, all permutations of $Y_3^* = \langle \alpha 1 \beta \rangle \gamma$.

Figure 7 shows this decomposition of $X_0 \cap X_1$:

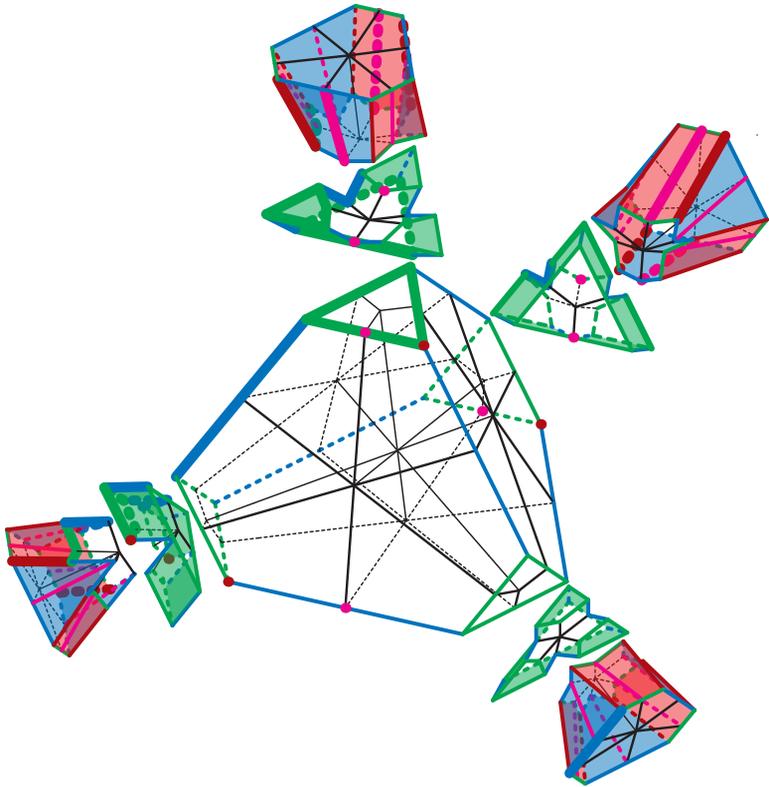


Figure 7: A handle decomposition of $X_0 \cap X_1$ in our trisection of T^4 . The trisection diagram on $\partial(X_0 \cap X_1) = X_0 \cap X_1 \cap X_2 = \langle \alpha\beta 02 \rangle \cup \langle \alpha\gamma 12 \rangle \cup \langle \beta\gamma 01 \rangle$ has two types of red curves; one of each is bold. The same holds for blue and green.

- The shape in the center (which looks like a truncated tetrahedron) is the 0–handle $\langle \alpha 1 \beta^2 \rangle$, comprising 12 cubes, each a permutation of $\alpha 1 \beta^2$ (see (5) and the paragraph before it). The interior lattice point is $(1, 1, 1)$, and each triangular-looking face is a permutation of $0 \langle 1 \beta^2 \rangle$. Each blue segment on $\partial \langle \alpha 1 \beta^2 \rangle$ is a permutation of $\langle \alpha 1 \rangle 2^2$.
- Each of the four three-pronged pieces is a permutation of $0 \langle \alpha \beta^2 \rangle$, glued to the 0–handle along $0 \langle 1 \beta^2 \rangle$. The twelve cubes that form these pieces are then glued in pairs: eg $0 \alpha \beta^2$ and $\alpha 0 \beta^2$ meet along the face $00 \beta^2$, and the other pairs are permutations of this. The union of each pair of cubes (a permutation of) $Y_2^* = \langle 0 \alpha \rangle \beta^2$, is a 1–handle which is glued to the 0–handle along (the corresponding permutation of) $\langle 01 \rangle \beta^2$. Note that Y_2^* intersects other permutations of Y_2^* , but only within $Y_2^* \cap Y_1$. Therefore, attaching Y_2^* to Y_1 amounts to attaching six 1–handles.

- Each of the four remaining pieces is a permutation of $Y_3^* = \langle \alpha 1 \beta \rangle \gamma$ and attaches to Y_1 and Y_2 , along (the corresponding permutations of) $\langle \alpha 1 \beta \rangle 2 \subset \langle \alpha 1 \beta^2 \rangle$ and $\langle \alpha 1 \beta \rangle 0 \subset \langle \alpha \beta^2 \rangle 0$, respectively.

For emphasis, some key details of this decomposition which will be instructive toward the odd-dimensional case (we will justify some of these details in Section 4) are

$$Y_1 = Y_1^* = \langle \alpha 1 \beta^2 \rangle \cong D^3,$$

so Y_1 is a 0-handle,

$$Y_2^* = \langle 0 \alpha \rangle \beta^2 \cong D^1 \times D^2,$$

$$Y_2^* \cap (Y_2 \setminus Y_2^*) \subset Y_2^* \cap Y_1 = (\partial \langle 0 \alpha \rangle) \times \beta^2 = \langle 0 1 \rangle \beta^2 \cong S^0 \times D^2,$$

so attaching Y_2 to Y_1 amounts to attaching a collection of 1-handles, and

$$Y_3^* = \langle \alpha 1 \beta \rangle \gamma \cong D^2 \times D^1,$$

$$Y_3^* \cap (Y_3 \setminus Y_3^*) \subset Y_3^* \cap (Y_1 \cup Y_2) = \langle \alpha 1 \beta \rangle \times \partial \gamma \cong D^2 \times S^0,$$

so attaching Y_3 to $Y_1 \cup Y_2$ amounts to attaching a collection of 1-handles. Thus, $X_0 \cap X_1$ is a 4-dimensional 1-handlebody. Note in Figure 7 that $\partial(X_0 \cap X_1)$ is the central surface

$$(7) \quad X_0 \cap X_1 \cap X_2 = \langle \alpha \beta 0 2 \rangle \cup \langle \alpha \gamma 1 2 \rangle \cup \langle \beta \gamma 0 1 \rangle,$$

which is colored in Figure 7 according to the color scheme from (7). Moreover, the red (resp. blue, green) line segments in Figure 7 form the “red (resp. blue, green) curves” in a trisection diagram for this trisection, and so Figure 7 is, in fact, a trisection diagram; see [2; 7].

Note that what we have actually shown is that Figures 3, 4, and 7 give a combinatorial description of an efficient trisection of T^4 . Thus, since the PL and smooth categories coincide in dimension 4, T^4 has a smooth structure for which we have described a trisection. Most likely, this is the standard smooth structure on T^4 , but we have not yet proven this, nor will we in this paper.

One way to prove this would be to describe a (smooth=PL) isotopy (ie a sequence of handleslides on the central surface) between our trisection and another trisection of the standard T^4 , such as either of those due to Koenig or Williams, the former obtained by viewing T^4 as $T^3 \times S^1$ [4], the latter by viewing T^4 as $T^2 \times T^2$ [11]. There may well be isotopies between our constructions and theirs, but attempting to construct such isotopies explicitly is messy, in part because the central surface has genus 10, and so it remains an open question as to whether or not *all* efficient trisections of T^4 are mutually isotopic. In other words: does the following theorem, proven using minimal surface theory, extend to dimension four?

Y_z	Y_z^*	h	z	glue to
$\langle \alpha^2[0, 2]^2 \rangle$	$\langle \alpha^2[0, 2]^2 \rangle$	0	1	
$\langle \alpha^2[0, 2]^\gamma \rangle$	$\langle \alpha^2[0, 2]^\gamma \rangle$	1	2	1

Table 1: X_0 from the trisection of T^4 .

Theorem 3.11 (Frohman [1]) *Up to isotopy, T^3 has a unique minimal-genus Heegaard splitting.*

Question 2 Up to isotopy, does T^4 have a unique efficient trisection?

Question 3 Does T^4 admit exotic smooth structures? If it does, then which of these exotic structures are compatible with efficient trisections?

3.4 Trisection of T^5

The decomposition of T^5 from Figure 5 is given by

$$(8) \quad X_0 = \langle \alpha^2[0, 2]^2[0, 3] \rangle, \quad X_i = X_0 + (i, i, i, i, i).$$

The handle decompositions of X_I for $I = \{0\}, \{0, 1\}$ are quite similar to those from T^4 . Focus first on $I = \{0\}$, ie on the handle decomposition of X_0 . Note the single factor of $[0, 3]$ in (8). As we will explain shortly, the handle decomposition of X_0 here comes from the decomposition of the interval

$$[0, 3] = [0, 2] \cup \gamma,$$

and likewise for X_0 from the trisection of T^4 . These handle decompositions appear in Tables 1 and 2. These and subsequent tables are organized as follows. In each z^{th} row, Y_z is a union of handles of index h , Y_z^* is an *example* of such an h -handle, and the entry in the column *glue to* lists those indices z' for which Y_z^* glues to $Y_{z'}$ along at least one face of codimension 1. The other handles from Y_z are related to Y_z^* by permutation; for details, see Section 7.1.5.

Note in both Tables 1 and 2 that $Y_1 = Y_1^*$ is star-shaped in a particularly nice way (more detail to come in Section 4), and hence is a ball which we may view as a 0-handle.

J	Y_z	Y_z^*	h	z	glue to
\emptyset	$\langle \alpha^2[0, 2]^3 \rangle$	$\langle \alpha^2[0, 2]^3 \rangle$	0	1	
$\{0\}$	$\langle \alpha^2[0, 2]^2 \gamma \rangle$	$\langle \alpha^2[0, 2]^2 \gamma \rangle$	1	2	1

Table 2: X_0 from the trisection of T^5 .

Y_z	Y_z^*	h	z	glue to
$\langle \alpha 1 \beta^2 \rangle$	$\langle \alpha 1 \beta^2 \rangle$	0	1	
$\langle 0 \alpha \beta^2 \rangle$	$\langle 0 \alpha \rangle \beta^2$	1	2	1
$\langle \alpha 1 \beta \gamma \rangle$	$\langle \alpha 1 \beta \rangle \gamma$	1	3	1,2

Table 3: From the trisection of T^4 : $X_0 \cap X_1 = \langle \alpha 1 \beta [1, 3] \rangle \cup \langle 0 \alpha \beta^2 \rangle$.

J	i_*	Y_z	Y_z^*	h	z	glue to
\emptyset	0	$\langle \alpha 1 \beta^3 \rangle$	$\langle \alpha 1 \beta^3 \rangle$	0	1	
	1	$\langle 0 \alpha \beta^3 \rangle$	$\langle 0 \alpha \rangle \beta^3$	1	2	1
$\{0\}$	0	$\langle \alpha 1 \beta^2 \gamma \rangle$	$\langle \alpha 1 \beta^2 \rangle \gamma$	1	3	1,2
	1	$\langle \gamma 0 \alpha \beta^2 \rangle$	$\langle \gamma 0 \alpha \rangle \beta^2$	2	4	2,3

Table 4: From the trisection of T^5 : $X_0 \cap X_1 = \langle \alpha 1 \beta^2 [1, 3] \rangle \cup \langle 0 \alpha \beta^2 [1, 3] \rangle$.

Then Y_2^* is the product of the same sort of star-shaped ball with the interval γ and glues to Y_1 along the product of that ball with $\partial \gamma$. The red γ here, and all red henceforth, indicates a positive contribution to the handle index h .

Next, consider X_I for $I = \{0, 1\}$ from T^4 and T^5 . Similarly to the former (recall (6)), the latter is given by

$$(9) \quad X_0 \cap X_1 = \langle \alpha 1 \beta^2 [1, 3] \rangle \cup \langle 0 \alpha \beta^2 [1, 3] \rangle.$$

Handle decompositions are summarized in Tables 3 and 4, which are organized in largely the same way as Tables 1 and 2.

Regarding the first columns of Table 4, each Y_z there corresponds to a pair (J, i_*) , where $J \subset \{\min I_r\} = \{0\}$ ⁹ and $i_* \in I = \{0, 1\}$. For details see Section 7.1.2.

3.5 The difficulty with T^6

Suppose we try to quadrisect T^6 in the same way, viewing T^6 as $(\mathbb{R}/4\mathbb{Z})^6 = [0, 4]^6 / \sim$ and partitioning the 4^6 resulting subcubes into four classes. The first problem is that no such partition is symmetric with respect to both the permutation action of \mathbb{Z}_6 on the indices and the translation action of \mathbb{Z}_4 along the main diagonal. To see this, consider the subcube $\alpha^3 \gamma^3$. The problem is that

$$\alpha^3 \gamma^3 + (2, 2, 2, 2, 2, 2) = \gamma^3 \alpha^3 \subset \langle \alpha^3 \gamma^3 \rangle.$$

⁹Recall from Notation 3.8 that $\{\min I_r\} = \{i_t : t \in T\} = \{i_t \in I : i_t - 1 \notin I\}$, so eg $\{\min I_r\} = \{0\}$ if $I = \{0\}$, $I = \{0, 1\}$ or $I = \{0, 1, 2\}$, and $\{\min I_r\} = \{0, 2\}$ if $I = \{0, 2\}$.

Fundamentally, the problem is that $k = 4$ and $n = 2k - 2 = 6$ are not relatively prime. (In odd dimensions this trouble does not arise, since k and $2k - 1$ are relatively prime.) Perhaps there is a less symmetric way to partition the subcubes of $[0, 4]^6/\sim$ which gives a quadrisection of T^6 , but trial and error suggests to the author that this is unlikely.

Conjecture 4 No partition of the subcubes of $[0, 4]^6/\sim$ gives a quadrisection of T^6 .

Question 5 Does the 6-dimensional torus admit an *efficient* quadrisection?

4 Star-shaped building blocks

This section introduces three types of building blocks, each of which is PL homeomorphic to a ball.¹⁰ In Section 7, when we describe and then justify the handle decomposition of arbitrary X_I in arbitrary odd dimension, this will be particularly helpful. The main idea is that we will decompose arbitrary X_I into many pieces. Each piece will be a product of such building blocks, and hence PL homeomorphic to a ball (see Lemma 7.6). Of course, we will still need to describe how all these balls are glued together and explain why this gives a handle decomposition.

In fact, we saw all three types of building blocks in Section 3. For example, denoting PL homeomorphism by \cong , the factors $\alpha^2[0, 2] \cong D^3$, $\alpha^2[0, 2]^2 \cong D^4$, and $\alpha^2[0, 2]^3 \cong D^5$ from Tables 1 and 2 are examples of the first type of building block; see (10). The factor $\langle 0\alpha \rangle \cong D^1$ of Y_2^* in Tables 3 and 4 is an example of the second type of building block; see (11). The factor $\langle \gamma 0\alpha \rangle \cong D^2$ from Y_4^* in Table 4 is an example of the third type, as are those factors $\langle \alpha 1\beta^r \rangle \cong D^{r+1}$, which appear in four places in Tables 3 and 4.

Given $\vec{p}, \vec{q} \in \mathbb{R}^n$, denote the convex hull of $\{\vec{p}, \vec{q}\}$ by

$$[\vec{p}, \vec{q}] = \{t\vec{p} + (1-t)\vec{q} : 0 \leq t \leq 1\}.$$

Let $\vec{p} \in Y \subset \mathbb{R}^n$. Define the *scope* of \vec{p} in Y to be the largest star of \vec{p} in Y :

$$\text{scope}(Y; \vec{p}) = \{\vec{q} \in Y : [\vec{p}, \vec{q}] \subset Y\}.$$

Say that Y is *star-shaped* about \vec{p} if $Y = \text{scope}(Y; \vec{p})$. The *link* of \vec{p} in Y is

$$\text{lk}_Y(\vec{p}) = \{\vec{v} \in \mathbb{R}^n : |\vec{v}| = 1, [\vec{p}, \vec{p} + \varepsilon\vec{v}] \subset Y \text{ for some } \varepsilon > 0\}.$$

Thus, Y is a d -dimensional PL submanifold of \mathbb{R}^n near \vec{p} if and only if either

- $\text{lk}_Y(\vec{p}) \cong S^{d-1}$, in which case \vec{p} is in the *interior* of Y ; or
- $\text{lk}_Y(\vec{p}) \cong D^{d-1}$, in which case $\vec{p} \in \partial Y$.

¹⁰In the PL category, an n -ball D^n is any manifold PL homeomorphic to the standard n -simplex, and an n -sphere S^n is any manifold PL homeomorphic to ∂D^n .

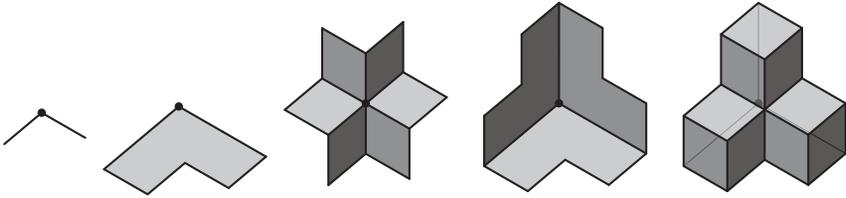


Figure 8: Left to right: $\langle 0\alpha \rangle$, $\langle \alpha[0, 2] \rangle$, $\langle \alpha 1\beta \rangle$, $\langle 0\alpha[0, 2] \rangle$, and $\langle \alpha^2[0, 2] \rangle$.

Suppose $Y = \text{scope}(Y; \vec{p})$ and $\text{lk}_Y(\vec{p}) \cong S^{d-1}$, so Y is star-shaped about \vec{p} and is a PL d -submanifold of \mathbb{R}^n near \vec{p} . In this situation, we say Y is *strongly star-shaped* about \vec{p} if moreover, for every point $\vec{q} \in Y$, every point $\vec{x} \in [\vec{p}, \vec{q}] \setminus \{\vec{q}\}$ satisfies $\text{lk}_Y(\vec{x}) \cong S^{d-1}$. This extra requirement implies that, for each $\vec{q} \in \text{link}_Y(\vec{p})$, the ray from \vec{p} through \vec{q} contains at most one point of ∂Y . Moreover:

Proposition 4.1 *If $Y \subset \mathbb{R}^n$ is compact and strongly star-shaped about $\vec{p} \in Y$, then Y is PL homeomorphic to a compact ball.*

Proof By definition, there is a PL homeomorphism $\phi: S^{d-1} \rightarrow \text{lk}_Y(\vec{p})$. There is also a map $\psi: Y \setminus \{\vec{p}\} \rightarrow \text{lk}_Y(\vec{p})$ given by $\psi: \vec{q} \mapsto (\vec{q} - \vec{p})/|\vec{q} - \vec{p}|$.¹¹ Denote the restriction $\psi|_{\partial Y}$ by Ψ . The assumptions that Y is compact and *strongly* star-shaped about \vec{p} imply that Ψ has a well-defined continuous inverse map, and hence is a PL homeomorphism. Define a polar coordinate system $\Phi: Y \rightarrow D^d$ by $\Phi: \vec{p} \mapsto \vec{0}$ and, for $\vec{q} \neq \vec{p}$,

$$\Phi: \vec{q} \mapsto \frac{|\vec{q} - \vec{p}|}{|\Psi^{-1} \circ \psi(\vec{q}) - \vec{p}|} \cdot \phi^{-1} \circ \psi(\vec{q}).$$

This map Φ is a PL homeomorphism, because the inverse map $D^d \rightarrow Y$ is

$$\Phi^{-1}: r\vec{\theta} \mapsto \vec{p} + r|\Psi^{-1} \circ \phi(\vec{\theta}) - \vec{p}| \cdot \phi(\vec{\theta}). \quad \square$$

In $T^n = (\mathbb{R}/k\mathbb{Z})^n$, for $d \leq n - 1$, identify $T^d = (\mathbb{R}/k\mathbb{Z})^d$ with $(\mathbb{R}/k\mathbb{Z})^d \times \{\vec{0}\} \subset T^n$, and likewise for T^{d+1} . For any $0 < a_1 \leq \dots \leq a_d < k$ (not necessarily integers), define

$$(10) \quad C_1 = \left\langle \prod_{r=1}^d [0, a_r] \right\rangle \subset T^d,$$

$$(11) \quad C_2 = \left\langle \{0\} \times \prod_{r=1}^d [0, a_r] \right\rangle \subset T^{d+1},$$

¹¹We use the product metric on \mathbb{R}^n : if $\vec{p} = (p_1, \dots, p_n)$ and $\vec{q} = (q_1, \dots, q_n)$, then $|\vec{q} - \vec{p}| = \max_i |q_i - p_i|$.

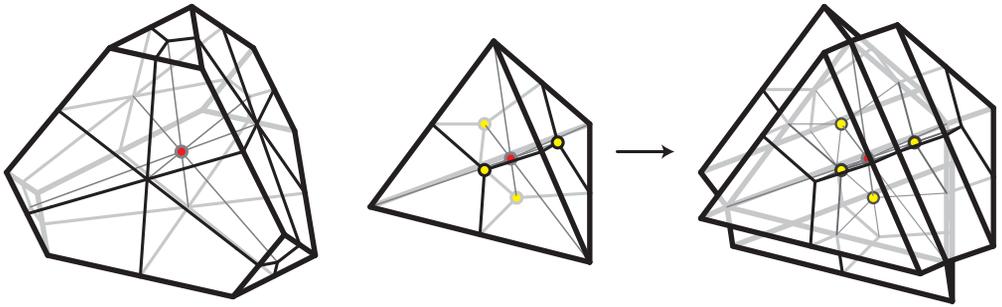


Figure 9: Left to right: $\langle \alpha\beta^2 \rangle$ and $\langle 0\alpha^3 \rangle \rightarrow \langle 0\alpha^2[0, 2] \rangle$.

and

$$(12) \quad C_3 = \left\langle [0, a_1] \times \{a_1\} \times \prod_{r=2}^d [a_1, a_r] \right\rangle \subset T^{d+1}.$$

Figures 8 and 9 show low-dimensional examples of these *building blocks*. In Figure 8, $\langle \alpha[0, 2] \rangle$ and $\langle \alpha^2[0, 2] \rangle$ are examples of C_1 , $\langle 0\alpha \rangle$ and $\langle 0\alpha[0, 2] \rangle$ are examples of C_2 , and $\langle \alpha\beta \rangle$ is an example of C_3 . In Figure 9, $\langle 0\alpha^3 \rangle$ and $\langle 0\alpha^2[0, 2] \rangle$ are examples of C_2 , and $\langle \alpha\beta^2 \rangle$ is an example of C_3 .

Lemma 4.2 $C_1, C_2,$ and C_3 from (10)–(12) are PL homeomorphic to D^d .

Proof Let $b = \frac{1}{2}(k + a_d)$. Then $C_1 \subset [0, b]^d$ and $C_2, C_3 \subset [0, b]^{d+1}$, where $b < k$, so we may view C_1 as a subset of \mathbb{R}^d , and C_2 and C_3 as subsets of \mathbb{R}^{d+1} . Let $a = \frac{1}{2}a_1$, $\vec{p}_1 = (a, \dots, a) \in \mathbb{R}^d$, $\vec{p}_2 = \vec{0} \in \mathbb{R}^{d+1}$, and $\vec{p}_3 = (a_1, \dots, a_1) \in \mathbb{R}^{d+1}$. Then, for $i = 1, 2, 3$, C_i is compact and strongly star-shaped about \vec{p}_i , with $\text{link}_{C_i}(\vec{p}_i) \cong S^{d-1}$, and hence PL homeomorphic to D^d by Proposition 4.1. \square

5 Further examples

As noted in the introduction, the hardest part of verifying our multisection of T^n , in arbitrary odd dimension n , is describing the handle decomposition of X_I for arbitrary $I \subset \mathbb{Z}_k$. That task will follow three main steps. First, Lemma 6.13 will establish a closed formula (2) for arbitrary X_I . Second, Section 7.1 will describe how (in several steps) to decompose X_I into pieces, each of which is a product of the building blocks from Section 4, and will describe an order on these pieces. Third, Section 7.2 will establish several properties of the resulting decomposition, eventually proving that it is an appropriate handle decomposition of X_I and thus verifying Theorem 7.10.

J	Y_z	Y_z^*	h	z	glue to
\emptyset	$\langle \alpha^2[0, 2]^2[0, 3]^2[0, 4]^3 \rangle$	$\langle \alpha^2[0, 2]^2[0, 3]^2[0, 4]^3 \rangle$	0	1	
$\{0\}$	$\langle \alpha^2[0, 2]^2[0, 3]^2[0, 4]^2\varepsilon \rangle$	$\langle \alpha^2[0, 2]^2[0, 3]^2[0, 4]^2\varepsilon \rangle$	1	2	1

Table 5: X_0 from the quadrisection of T^7 .

To prepare, this section describes a few more examples, each of which confronts and resolves an additional complication in the handle decomposition of some X_I in some dimension. This section contains no proofs and little narration. Instead, the reader is encouraged to peruse the tables in order to build intuition for the denser sections that follow. Indeed, assuming only the correctness of the formula (2), the reader should now be able to use their understanding of the building blocks from Section 4 to check the correctness of the handle decompositions, as detailed in the last five columns of the tables (starting with Y_z).

The harder part will be understanding how each handle decomposition has been constructed. This is the purpose of the columns in each table which precede Y_z , which we do not attempt to describe in detail until Section 7.

5.1 Quadrisection of T^7

The next several examples come from the decomposition of T^7 given by $X_0 = \langle \alpha^2[0, 2]^2[0, 3]^2[0, 4] \rangle$ and $X_i = X_0 + (i, i, i, i, i, i, i)$. The handle decompositions of X_I for $I = \{0\}, \{0, 1\}$, summarized in Tables 5 and 6, respectively, follow the same pattern in dimension seven (and all higher odd dimensions) as in dimension five (recall Tables 2 and 4 and the attending discussions). More instructive examples follow.

5.1.1 X_I when $I = \{0, 2\}$

From the quadrisection of T^7 , consider

$$X_0 \cap X_2 = \langle \alpha^2[0, 2]\gamma^2[2, 4] \rangle \cup \langle \alpha^2[0, 2]2\gamma^2[2, 4] \rangle.$$

J	i_*	Y_z	Y_z^*	h	z	glue to
\emptyset	0	$\langle \alpha 1\beta^2[1, 3]^3 \rangle$	$\langle \alpha 1\beta^2[1, 3]^3 \rangle$	0	1	
	1	$\langle 0\alpha\beta^2[1, 3]^3 \rangle$	$\langle 0\alpha \rangle \langle \beta^2[1, 3]^3 \rangle$	1	2	1
$\{0\}$	0	$\langle \alpha 1\beta^2[1, 3]^2\delta \rangle$	$\langle \alpha 1\beta^2[1, 3]^2\delta \rangle$	1	3	1, 2
	1	$\langle \delta 0\alpha\beta^2[1, 3]^2 \rangle$	$\langle \delta 0\alpha \rangle \langle \beta^2[1, 3]^2 \rangle$	2	4	2, 3

Table 6: X_I for $I = \{0, 1\}$ from the quadrisection of T^7

J	i_*	V	V^-	Y_z	Y_z^*	h	z	glue to
\emptyset	0	\emptyset	\emptyset	$\langle \alpha^3 2\gamma^3 \rangle$	$\alpha^3 \langle 2\gamma^3 \rangle$	0	1	
	2	\emptyset		$\langle 0\alpha^3 \gamma^3 \rangle$	$\langle 0\alpha^3 \rangle \gamma^3$	0	2	
$\{0\}$	0	\emptyset	\emptyset	$\langle \alpha^3 2\gamma^2 \delta \rangle$	$\alpha^3 \langle 2\gamma^2 \rangle \delta$	1	3	1,2
	2	$\{0\}$	\emptyset	$\langle \delta^+ 0\alpha^3 \gamma^2 \rangle$	$\langle \delta^+ 0\alpha^3 \rangle \gamma^2$	0	4	
$\{2\}$	0	$\{2\}$	\emptyset	$\langle \delta^- 0\alpha^3 \gamma^2 \rangle$	$\delta^- \langle 0\alpha^3 \rangle \gamma^2$	1	5	2,4
			$\{2\}$	$\langle \alpha^2 \beta^+ 2\gamma^3 \rangle$	$\alpha^2 \langle \beta^+ 2\gamma^3 \rangle$	0	6	
	2	\emptyset	\emptyset	$\langle \alpha^2 \beta^- 2\gamma^3 \rangle$	$\alpha^2 \beta^- \langle 2\gamma^3 \rangle$	1	7	1,6
			$\{2\}$	$\langle 0\alpha^2 \beta \gamma^3 \rangle$	$\langle 0\alpha^2 \rangle \beta \gamma^3$	1	8	1,2
$\{0, 2\}$	0	$\{2\}$	\emptyset	$\langle \alpha^2 \beta^+ 2\gamma^2 \delta \rangle$	$\alpha^2 \langle \beta^+ 2\gamma^2 \rangle \delta$	1	9	6,8
			$\{2\}$	$\langle \alpha^2 \beta^- 2\gamma^2 \delta \rangle$	$\alpha^2 \beta^- \langle 2\gamma^2 \rangle \delta$	2	10	3,7,8,9
	2	$\{0\}$	\emptyset	$\langle \delta^+ 0\alpha^2 \beta \gamma^2 \rangle$	$\langle \delta^+ 0\alpha^2 \rangle \beta \gamma^2$	1	11	3,4
			$\{0\}$	$\langle 0\alpha^2 \beta \gamma^2 \delta^- \rangle$	$\langle 0\alpha^2 \rangle \beta \gamma^2 \delta^-$	2	12	3,5,8,11

Table 7: X_I for $I = \{0, 2\}$ from the quadrisection of T^7 .

Table 7 summarizes a handle decomposition $X_I = Y_1 \cup \dots \cup Y_{12}$. As with X_I for $I = \{0, 1\}$, the decomposition of X_I for $I = \{0, 2\}$ is organized largely according to $\{(J, i_*) : J \subset \{\min I_r\}, i_* \in I\}$. With $Y_4, Y_5, Y_7, Y_8, Y_{10}$, and Y_{11} here, we have $J \setminus \{\min I_*\} \neq \emptyset$, requiring us to split a unit interval into subintervals, in this case halves. Details on how this is done, including the definitions and purposes of the sets $V^- \subset V \subset I$, appear in Section 7.1, especially Table 10, and in Tables 11 and 12 in Appendix A.

5.1.2 X_I when $I = \{0, 1, 2\}$ Still in dimension seven, consider

$$X_0 \cap X_1 \cap X_2 = \langle \alpha^2 [0, 2] \gamma^2 [2, 4]^2 [2, 5] \rangle \cup \langle \alpha^2 [0, 2] 2\gamma^2 [2, 4]^2 [2, 5] \rangle.$$

Table 8 summarizes a handle decomposition $X_I = Y_1 \cup \dots \cup Y_{12}$. Again, the decomposition of X_I for $I = \{0, 1, 2\}$ is organized largely according to

$$\{(J, i_*) : J \subset \{\min I_r\}, i_* \in I\}.$$

Here, we have a block I_r (in this case $I_r = I$) with $|I_r| \geq 3$, requiring us at times to split a unit interval into thirds, as seen here in Y_6 – Y_8 and Y_{14} – Y_{16} . Details on this and the set U appear in Section 7.1, especially Table 10, and in Tables 11 and 12 in Appendix A.

Another new complication arises here in Y_1 – Y_4 and Y_9 – Y_{12} , where $i_* + 2 \in I_*$, requiring us to split certain unit intervals into halves according to a different rule than in Section 5.1.1. Again, all the rules for splitting unit intervals into halves and thirds are detailed in Section 7.1, especially Table 10, and in Tables 11 and 12 in Appendix A.

J	i_*	U	V	V^-	Y_z^*	h	z	glue to	
\emptyset	0	\emptyset	$\{1,2\}$	$\{1\}$	$\alpha^-1\langle\beta^+2\gamma^3\rangle$	0	1		
				\emptyset	$\langle\alpha^+1\rangle\langle\beta^+2\gamma^3\rangle$	1	2	1	
				$\{1,2\}$	$\alpha^-\langle1\beta^-\rangle\langle2\gamma^3\rangle$	1	3	1	
				$\{2\}$	$\langle\alpha^+1\beta^-\rangle\langle2\gamma^3\rangle$	2	4	2,3	
	1	\emptyset	\emptyset	\emptyset	$\langle0\alpha\rangle\langle\beta2\gamma^3\rangle$	1	5	1,3	
					$0\alpha_3^0\langle1\beta\rangle\gamma^3$	1	6	5	
	2	$\{1\}$	\emptyset	\emptyset	$\langle0\alpha_3^-\rangle\langle1\beta\rangle\gamma^3$	2	7	5,6	
					$0\langle\alpha_3^+1\beta\rangle\gamma^3$	2	8	5,6	
	$\{0\}$	0	\emptyset	$\{1,2\}$	$\{1\}$	$\delta\alpha^-1\langle\beta^+2\gamma^2\rangle$	1	9	1,6,7
					\emptyset	$\delta\langle\alpha^+1\rangle\langle\beta^+2\gamma^2\rangle$	2	10	2,6,8
$\{1,2\}$					$\delta\alpha^-\langle1\beta^-\rangle\langle2\gamma^2\rangle$	2	11	3,6,7	
$\{2\}$					$\delta\langle\alpha^+1\beta^-\rangle\langle2\gamma^2\rangle$	3	12	4,6,8	
1		\emptyset	\emptyset	\emptyset	$\langle\delta0\alpha\rangle\langle\beta2\gamma^2\rangle$	2	13	5,9,11	
					$\langle\delta0\rangle\alpha_3^0\langle1\beta\rangle\gamma^2$	2	14	6,13	
2		$\{1\}$	\emptyset	\emptyset	$\langle\delta0\alpha_3^-\rangle\langle1\beta\rangle\gamma^2$	3	15	7,13,14	
					$\langle\delta0\rangle\langle\alpha_3^+1\beta\rangle\gamma^2$	3	16	8,13,14	

Table 8: X_I for $I = \{0, 1, 2\}$ from the quadrisection of T^7 .

5.2 X_I when $I = \{0, 1, 2, 3, 5\}$ from T^{13}

There is one more complication, which arises, first in dimension 11, whenever X_I for $I = I_1 \sqcup \dots \sqcup I_m$, has some $I_r \not\equiv i_*$ with $|I_r| \geq 3$. In fact, though, the *difficulty* of this complication only becomes apparent in dimension 13. From the septisection of T^{13} , consider X_I for $I = \{0, 1, 2, 3, 5\}$, which is given by

$$\langle\alpha1\beta2\gamma3\delta^2[3, 5]5\epsilon^2[5, 7]\rangle \cup \langle0\alpha\beta2\gamma3\delta^2[3, 5]5\epsilon^2[5, 7]\rangle \cup \langle0\alpha1\beta\gamma3\delta^2[3, 5]5\epsilon^2[5, 7]\rangle \\ \cup \langle0\alpha1\beta2\gamma\delta^2[3, 5]5\epsilon^2[5, 7]\rangle \cup \langle0\alpha1\beta2\gamma3\delta^2[3, 5]\epsilon^2[5, 7]\rangle.$$

In this example, the new complication arises when $i_* = 5$ and $0 \notin J$, ie in the part of X_I given by

$$\langle0\alpha1\beta2\gamma3\delta^2[3, 5]\epsilon^3\rangle,$$

part of which appears in the first several Y_z in the handle decomposition of this X_I ; see Table 9. The tricky part here is how to order the pieces Y_z ; see Section 7.1.4, especially (23).

Also see Table 18 in Appendix A, which summarizes the start of the handle decomposition of X_I for $I = \{0, 1, 2, 3, 4, 6\}$ from T^{15}

V^-	Y_z^*	h	z	glue to
\emptyset	$0\langle\alpha^+1\rangle\langle\beta^+2\rangle\langle\gamma^+3\delta^3\rangle\zeta^3$	0	1	
$\{1\}$	$\langle 0\alpha^- \rangle 1\langle\beta^+2\rangle\langle\gamma^+3\delta^3\rangle\zeta^3$	1	2	1
$\{1, 2\}$	$\langle 0\alpha^- \rangle \langle 1\beta^- \rangle 2\langle\gamma^+3\delta^3\rangle\zeta^3$	1	3	2
$\{2\}$	$0\langle\alpha^+1\beta^- \rangle 2\langle\gamma^+3\delta^3\rangle\zeta^3$	2	4	1,3
$\{2, 3\}$	$0\langle\alpha^+1\beta^- \rangle \langle 2\gamma^- \rangle \langle 3\delta^3 \rangle \zeta^3$	1	5	4
$\{1, 2, 3\}$	$\langle 0\alpha^- \rangle \langle 1\beta^- \rangle \langle 2\gamma^- \rangle \langle 3\delta^3 \rangle \zeta^3$	2	6	3,5
$\{1, 3\}$	$\langle 0\alpha^- \rangle 1\langle\beta^+2\gamma^- \rangle \langle 3\delta^3 \rangle \zeta^3$	2	7	2,6
$\{3\}$	$0\langle\alpha^+1\rangle\langle\beta^+2\gamma^- \rangle \langle 3\delta^3 \rangle \zeta^3$	3	8	1,5,7

Table 9: From the septisection of T^{13} : the start of the handle decomposition of X_I when $I = \{0, 1, 2, 3, 5\}$. Here, $J = \emptyset, i_* = 5, U = \emptyset,$ and $V = \{1, 2, 3\}$.

6 Combinatorics

This section proves several combinatorial facts about the decompositions of T^n . In particular, Section 6.2 proves that $T^n = \bigcup_{i \in \mathbb{Z}_k} X_i$, and Section 6.4 establishes a closed expression (2) for arbitrary X_I . Also, Section 6.3 establishes two combinatorial corollaries, which may be of independent interest but otherwise are not needed in this paper.

6.1 Notation

Because each X_i from our construction (1) is symmetric under the permutation action of S_n on the indices in T^n , it will often suffice, when considering an arbitrary point $\vec{x} = (x_1, \dots, x_n) \in (\mathbb{R}/k\mathbb{Z})^n = T^n$, to assume that \vec{x} is *monotonic* in the sense that $x_1 \leq x_2 \leq \dots \leq x_n \leq k + x_1$.

Denoting the main diagonal of T^n by Δ , each monotonic point $\vec{x} = (x_1, \dots, x_n) \in T^n \setminus \Delta$ corresponds to a unique point $(x_1, \dots, x_n) \in \mathbb{R}^n$ with $0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq x_1 + k < 2k$. For such \vec{x} , extend the point $(x_1, \dots, x_n) \in \mathbb{R}^n$ to a point $\vec{x}_\infty = (x_r)_{r \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ by defining, for each $r \in \mathbb{Z}_k$ and $m \in \mathbb{Z}$,

$$x_{r+mn} = x_r + mk.$$

We will mainly be interested in $0 \leq x_1 \leq \dots \leq x_{2n}$, where

$$x_{2n} = x_n + k \leq x_1 + 2k < 3k.$$

With this setup, for any monotonic $\vec{x} \in T^n \setminus \Delta$, define the following *cutoff indices* $a_r(\vec{x}), b_r(\vec{x}) \in \mathbb{Z}$ for each $r \in \mathbb{Z}$:

$$a_r(\vec{x}) = \min\{a : x_{a+1} \geq r\} \quad \text{and} \quad b_r(\vec{x}) = \min\{b : x_{b+1} > r\}.$$

Note that, in all cases, we have $a_0(\vec{x}) \leq 0$, with equality if and only if $x_n \neq k \equiv 0 \in \mathbb{R}/k\mathbb{Z}$. The main point is:

Observation 6.1 *Let $\vec{x} \in T^n \setminus \Delta$ be monotonic. Then $\vec{x} \in [0, 1]^2 \cdots [0, k - 1]^2 [0, k]$ if and only if $b_s(\vec{x}) \geq 2s$ for every $s = 0, \dots, k - 1$.*

Note that $b_0(\vec{x}) \geq 0$ in all cases. In order to apply the principle of **Observation 6.1** more broadly, for each $r \in \mathbb{Z}$ write

$$\vec{x}_r = (x_{1+a_r(\vec{x})}, x_{2+a_r(\vec{x})}, \dots, x_{n+a_r(\vec{x})}).$$

The point regarding monotonic points off the main diagonal is:

Observation 6.2 *If $\vec{x} \in T^n \setminus \Delta$ is monotonic and $r \in \mathbb{Z}$, then*

$$r \leq x_{1+a_r(\vec{x})} \leq \dots \leq x_{n+a_r(\vec{x})} < r + k,$$

and the following conditions are equivalent:

- $\vec{x}_r \in [r, r + 1]^2 \cdots [r, r + k - 1]^2 [r, r + k]$.
- $b_{r+s}(\vec{x}_r) \geq 2s$ for every $s = r + 1, \dots, r + k$.
- $b_{r+s}(\vec{x}) \geq a_r(\vec{x}) + 2s$ for every $s = r + 1, \dots, r + k$.

Observations **6.1** and **6.2** apply more generally:

Observation 6.3 *If $\vec{x} \in X_r \subset T^n \setminus \Delta$, then there is a permutation $\sigma \in S_n$ such that $\vec{x}_\sigma \in [r, r + 1]^2 \cdots [r, r + k - 1]^2 [r, r + k]$ is monotonic.*

Note also that either class of cutoff indices provides two-sided bounds for the other class:

Observation 6.4 *If $\vec{x} \in T^n$ is nonzero and monotonic and $r \in \mathbb{Z}$, then*

$$\dots \leq a_r(\vec{x}) \leq b_r(\vec{x}) \leq a_{r+1}(\vec{x}) \leq b_{r+1}(\vec{x}) \leq \dots$$

with $a_r(\vec{x}) = b_r(\vec{x})$ if and only if $x_{a_r(\vec{x})+1} \notin \mathbb{Z}_k$, and $b_r(\vec{x}) = a_{r+1}(\vec{x})$ if and only if $x_{b_r(\vec{x})+1} \geq r + 1$.

Note that $x_{b_r(\vec{x})+1}$ is the first coordinate in \vec{x} that exceeds r . Here is another convenient property:

Observation 6.5 Any nonzero monotonic $\vec{x} \in T^n$ and $r \in \mathbb{Z}_{\geq 0}$ satisfy

$$(13) \quad a_{r+k}(\vec{x}) = n + a_r(\vec{x}) \quad \text{and} \quad b_{r+k}(\vec{x}) = n + b_r(\vec{x}).$$

Noting that $X_r \cap \Delta = \{(x, \dots, x) : x \in [r, r + 1]\}$, we can express each X_r in terms of cutoff indices as follows:

Proposition 6.6 Let $\vec{x} \in T^n \setminus \Delta$ be monotonic, and let $r \in \mathbb{Z}_k$. Then $\vec{x} \in X_r$ if and only if $\vec{x}_r \in [r, r + 1]^2 \cdots [r, r + k - 1]^2 [r, r + k]$. In particular,

$$(14) \quad X_r \setminus \Delta = \langle \text{monotonic } \vec{x} : b_{r+s}(\vec{x}) \geq a_r(\vec{x}) + 2s \text{ for } s = 0, \dots, k - 1 \rangle.$$

Proof Write $\vec{x}_r = (x_1, \dots, x_n)$. Note that $r \leq x_1 \leq \dots \leq x_n < r + k$. To show that $\vec{x}_r \in [r, r + 1]^2 \cdots [r, r + k - 1]^2 [r, r + k]$ if and only if $\vec{x}_r \in X_r$, we prove both containments. One is trivial. For the other, suppose $\vec{x}_r \notin [r, r + 1]^2 \cdots [r, r + k - 1]^2 [r, r + k]$. Then **Observation 6.2** implies that $b_{r+s}(\vec{x}_r) < 2s$ for some $s = 0, \dots, k - 1$, so

$$r + s < x_{2s}, \dots, x_n < r + k.$$

Thus, at least $n + 1 - 2s$ of the coordinates of \vec{x} lie in the open interval $(r + s, r + k)$. Yet, $2s$ of the n factors of $[r, r + 1]^2 \cdots [r, r + k - 1]^2 [r, r + k]$ are disjoint from that open interval, which is a contradiction. **Observation 6.3** now implies that $\vec{x} \in X_r \setminus \Delta$ if and only if \vec{x} is an element of the right side of (14). □

6.2 The X_r have disjoint interiors and cover T^n

Proposition 6.7 With the setup from **Theorem 7.10**, X_r and X_s have disjoint interiors whenever $0 \leq r < s \leq k - 1$.

This is implied by **Lemma 6.13**, but the following proof is much easier than that of the lemma; we include it for expository reasons.

Proof By the symmetry of the construction, we may assume that $r = 0$. Assume for contradiction that the interiors of X_r and X_s intersect. Then $X_r \cap X_s$ has positive measure, so there is a monotonic point $\vec{x} = (x_1, \dots, x_n) \in X_0 \cap X_j$ such that for every $i = 1, \dots, n$ we have $x_i \notin \mathbb{Z}_k$.

So $a_i(\vec{x}) = b_i(\vec{x})$ for each $i = 1, \dots, n$, by **Observation 6.4**. In particular, since $\vec{x} \in X_0$, we have $a_0 = b_0 = 0$, and $a_s = b_s \geq 2s$ by **Proposition 6.6**. But then, since $\vec{x} \in X_s$ and $a_s \geq 2s$, **Observation 6.5** and **Proposition 6.6** give the following contradiction:

$$n = n + b_0 = b_k = b_{s+(k-s)}, \quad n \geq a_s + 2(k - s), \quad n \geq 2k. \quad \square$$

Lemma 6.8 We have $X_0 \cup \dots \cup X_{k-1} = T^n$.

Proof Let $\vec{x} \in T^n$. We will prove that $\vec{x} \in X_s$ for some s . If $\vec{x} = (x, \dots, x) \in \Delta$, then $\vec{x} \in X_{[x]}$. Assume instead that $\vec{x} \in T^n \setminus \Delta$. Also assume without loss of generality that \vec{x} is monotonic with $a_0(\vec{x}) = 0$. Throughout this proof, denote each $a_s(\vec{x})$ by a_s and each $b_s(\vec{x})$ by b_s .

Let $s_0 = 0$, so that $a_{s_0} = a_0 = 0$. If $b_s \geq 2s = 2s - a_{s_0}$ for all $s = 1, \dots, k - 1$, then $\vec{x} \in X_0 = X_{s_0}$. Otherwise, choose the smallest s_1 such that $b_{s_1} < 2s_1$. Thus, $b_s \geq 2s$ whenever $s < s_1$, so, by **Observation 6.4**,

$$2s_1 - 2 \leq b_{s_1-1} \leq a_{s_1} \leq b_{s_1} \leq 2s_1 - 1.$$

Continue in this way: for each s_t , choose the minimum $s_{t+1} = s_t + 1, \dots, k - 1$ such that $b_{s_{t+1}} < a_{s_t} + 2(s_{t+1} - s_t)$, if such s_{t+1} exists. Eventually this process terminates with some s_u , so that

- $b_s \geq a_{s_t} + 2(s - s_t)$ whenever $s_t \leq s \leq s_{t+1}$ for $t = 0, \dots, u - 1$,
- $b_s \geq a_{s_t} + 2(s - s_u)$ whenever $s_u \leq s \leq k - 1$, and
- $b_{s_{t+1}} < a_{s_t} + 2(s_{t+1} - s_t)$ for each $t = 0, \dots, u - 1$.

Hence, for each $t = 0, \dots, u - 1$, **Observation 6.4** gives

$$a_{s_t} + 2(s_{t+1} - 1 - s_t) \leq b_{s_{t+1}-1} \leq a_{s_{t+1}} \leq b_{s_{t+1}} \leq a_{s_t} + 2(s_{t+1} - s_t) - 1.$$

Subtracting a_{s_t} from the first, middle, and last expressions gives

$$2(s_{t+1} - s_t) - 2 \leq a_{s_{t+1}} - a_{s_t} \leq 2(s_{t+1} - s_t) - 1.$$

Therefore, for any $t = 0, \dots, u - 1$,

$$a_{s_u} - a_{s_t} = \sum_{r=t}^{u-1} (a_{s_{r+1}} - a_{s_r}) \leq \sum_{r=t}^{u-1} (2(s_{r+1} - s_r) - 1) = 2(s_u - s_t) - (u - t),$$

$$a_{s_u} - a_{s_t} \leq 2(s_u - s_t) - 1.$$

Rearranging gives

$$(15) \quad a_{s_u} - 2s_u \leq a_{s_t} - 2s_t - 1.$$

We claim that $\vec{x} \in X_{s_u}$. This is true if (and only if) $b_s \geq a_{s_u} + 2(s - s_u)$ for each $s = s_u, \dots, s_u + k - 1$. Fix some $s = k, \dots, s_u + k - 1$. Then $s_t \leq s - k \leq s_{t+1} - 1$ for

some $t = 0, \dots, u - 1$. By construction, we have $b_{s-k} \geq a_{s_t} + 2(s - k - s_t)$. Together with (13) and (15), this gives

$$b_s = b_{s-k} + n \geq a_{s_t} + 2(s - k - s_t) + 2k - 1 = (a_{s_t} - 2s_t - 1) + 2s \geq (a_{s_u} - 2s_u) + 2s = a_{s_u} + 2(s - s_u). \quad \square$$

6.3 Combinatorial corollaries

This subsection establishes two combinatorial corollaries, which may be of independent interest but otherwise are not needed in this paper.

We have proven that the pieces X_r of the multisection of T^n have disjoint interiors and cover T^n . Also, $X_r = X_0 + (r, \dots, r)$, so all X_r have the same number of unit cubes. Since there are k^n unit cubes in $T^n = (\mathbb{R}/k\mathbb{Z})^n$, each X_r contains k^{n-1} unit cubes. By counting these unit cubes a different way, we obtain the following:¹²

Corollary 6.9 For any $n = 2k - 1$,

$$(16) \quad k^{n-1} = \sum_{i_0=2}^n \binom{n}{i_0} \sum_{i_2=4-i_0}^{n-i_0} \binom{n-i_0}{i_1} \sum_{i_3=6-i_0-i_1}^{n-i_0-i_1} \binom{n-i_0-i_1}{i_2} \dots \sum_{i_{k-2}=2k-2-\sum_{j=0}^{k-3} i_j}^{n-\sum_{j=0}^{k-3} i_j} \binom{n-\sum_{j=0}^{k-3} i_j}{i_{k-1}}.$$

Note that (16) is also the number of spanning trees of the complete bipartite graph $K_{j,j}$ where $j = k$ [10].

Proof X_0 consists of k^{n-1} subcubes, each of the form $\prod_{r=1}^n [w_r, w_r + 1]$ for some $w_1, \dots, w_n \in \mathbb{Z}_k$. For each $s = 0, \dots, k - 2$, there are at least $2s + 2$ indices among $r = 1, \dots, n$ with $w_r \in \{0, \dots, s\}$, and conversely any subcube of that form with this property will be in X_0 . (This characterization follows from the expression (1) for X_0 .) In (16), $i_s = \#\{r : w_r = s\}$, so $i_0 \geq 2$, $i_0 + i_1 \geq 4$, and so on. \square

As noted above, each subcube of T^n has the form $\prod_{r=1}^n [w_r, w_r + 1]$ for some $w_1, \dots, w_n \in \mathbb{Z}_k$. Say that two subcubes $\prod_{r=1}^n [w_r, w_r + 1]$ and $\prod_{r=1}^n [w'_r, w'_r + 1]$ have the same combinatorial type if (w'_1, \dots, w'_n) is a permutation of (w_1, \dots, w_n) . Counting combinatorial cube types in three different ways yields:

¹²By definition, if $a, b \in \mathbb{Z}$ with $b < 0$, then $\binom{a}{b} = 0$.

Corollary 6.10 For any $n = 2k - 1$, we have

$$\begin{aligned}
 (17) \quad k \sum_{i_0=2}^n \sum_{i_1=\max\{0,4-i_0\}}^{n-i_0} \sum_{i_2=\max\{0,6-i_0-i_1\}}^{n-i_0-i_1} \cdots \sum_{i_{k-2}=\max\{0,2k-2-\sum_{j=0}^{k-3} i_j\}}^{n-\sum_{j=0}^{k-3} i_j} 1 \\
 = \sum_{i_0=0}^n \sum_{i_1=0}^{n-i_0} \sum_{i_2=0}^{n-i_0-i_1} \cdots \sum_{i_{k-2}=0}^{n-\sum_{j=0}^{k-3} i_j} 1 = \binom{3k-2}{k-1}.
 \end{aligned}$$

Proof The first expression is k times the number of *cube types* in X_0 , counted using the same principle and notation as in [Corollary 6.9](#). The second counts the number of cube types in T^n , each of which we may write in the form $\prod_{r=0}^{k-1} [r, r + 1]^{i_r}$, and is thus characterized by a tuple (i_0, \dots, i_{k-1}) with $\sum_{r=0}^{k-1} i_r = n$. The third counts the number of cube types in T^n by writing $a_0 = 0$ and $a_k = 3k - 1$, and associating each $A = \{a_1, \dots, a_{k-1}\} \subset \{1, \dots, 3k - 2\}$ satisfying $a_1 < \dots < a_{k-1}$ with the cube type

$$\prod_{i=1}^k \prod_{j=a_{i-1}+1}^{a_i-1} [i - 1, i]. \quad \square$$

See [\[10\]](#) for [other interpretations](#) of (17).

6.4 Verification of the formula $X_I = (2)$

Next, we will use the cutoff indices $a_r(\vec{x})$ and $b_r(\vec{x})$ to verify (2). To prepare this, we define subsets $C_{I,s} \subset T^n$ as follows. Let $I \subset \mathbb{Z}_k$ following [Convention 3.9](#), with $s \in \mathbb{Z}_\ell$, and write $i_s = i_*$. Then define¹³

$$\begin{aligned}
 (18) \quad C_{I,s} = & \left(\prod_{t=0}^{s-1} \{i_t\} \times [i_t, i_t + 1]^2 \times \cdots \times [i_t, i_{t+1} - 1]^2 \times [i_t, i_{t+1}] \right) \\
 & \times [i_*, i_* + 1]^2 \times \cdots \times [i_*, i_{s+1} - 1]^2 \times [i_*, i_{s+1}] \\
 & \times \left(\prod_{t=s+1}^{\ell-1} \{i_t\} \times [i_t, i_t + 1]^2 \times \cdots \times [i_t, i_{t+1} - 1]^2 \times [i_t, i_{t+1}] \right).
 \end{aligned}$$

Note the “missing” $\{i_*\}$ at the start of the second line; this corresponds to the \hat{i}_* in (2). Observe that the expression on the right side of (2) equals

$$\bigcup_{s \in \mathbb{Z}_\ell} \langle C_{I,s} \rangle.$$

¹³The first line in (18) contributes no factors to $C_{I,s}$ if $s = 0$, and likewise for the third line if $s = \ell - 1$. In particular, if $I = \{0\}$, then $s = 0$ and $C_{I,s} = [0, 1]^2 \times \cdots \times [0, k - 1]^2 [0, k]$, so $\langle C_{I,s} \rangle = X_0$.

Proposition 6.11 *Let $I \subset \mathbb{Z}_k$ follow [Convention 3.9](#), $s \in \mathbb{Z}_\ell$, and $C_{I,s}$ be as in (18). Suppose $\vec{x} \in T^n \setminus \Delta$ is monotonic. Then $\vec{x} \in C_{I,s}$ if and only if all of the following conditions hold:*

- $b_t(\vec{x}) \geq 2t + 1$ for $0 \leq t < i_*$,
- $b_t(\vec{x}) \geq 2t$ for $i_* \leq t \leq k - 1$,
- $a_t(\vec{x}) \leq 2t$ for $t = i_0, \dots, i_*$, and
- $a_t(\vec{x}) \leq 2t - 1$ for $t = i_{s+1}, \dots, i_{\ell-1}$.

Proof This follows from the definitions upon consideration of each entry in \vec{x} . □

Also note the following generalization of [Observation 6.3](#):

Observation 6.12 *Let $I \subset \mathbb{Z}_k$ follow [Convention 3.9](#), $s \in \mathbb{Z}_\ell$, and $C_{I,s}$ be as in (18). Suppose $\vec{x} \in \langle C_{I,s} \rangle$. Then there is a permutation $\sigma \in S_n$ such that $\vec{x}_\sigma \in C_{I,s}$ is monotonic.*

Lemma 6.13 *Given nonempty $I \subset \mathbb{Z}_k$ (following [Convention 3.9](#)),*

$$(19) \quad X_I = \bigcup_{s \in \mathbb{Z}_\ell} \langle C_{I,s} \rangle.$$

In particular,

$$(3) \quad \bigcap_{i_* \in \mathbb{Z}_k} X_i = \bigcup_{i_* \in I} \left\langle (i_1, \dots, \hat{i}_*, \dots, i_\ell) \prod_{i \in \mathbb{Z}_k} [i, i + 1] \right\rangle.$$

Note that the formula (19) is equivalent to (2).

Proof We argue by induction on ℓ . When $\ell = 1$, $X_I = X_0 = \langle C_{I,0} \rangle = (2)$.

Assume now that $\ell > 1$. First, we will show that

$$(20) \quad X_I \subset \bigcup_{s \in \mathbb{Z}_\ell} \langle C_{I,s} \rangle.$$

Let $\vec{x} \in X_I$, and define $I' = I \setminus \{i_{\ell-1}\}$. Note that I' is simple and $X_I = X_{I'} \cap X_{i_{\ell-1}}$. Since $\vec{x} \in X_{I'}$, the induction hypothesis implies that $\vec{x} \in \langle C_{I',s_0} \rangle$ for some $s_0 \in \mathbb{Z}_{\ell-1}$. By [Observation 6.12](#), there exists $\sigma \in S_n$ such that \vec{x}_σ is monotonic and $\vec{x}_\sigma \in C_{I',s_0}$.

[Proposition 6.11](#) implies that

- $b_t(\vec{x}_\sigma) \geq 2t + 1$ for $0 \leq t \leq i_{s_0} - 1$,
- $b_t(\vec{x}_\sigma) \geq 2t$ for $i_{s_0} \leq t \leq k - 1$,
- $a_t(\vec{x}_\sigma) \leq 2t$ for $t = i_0, \dots, i_{s_0}$, and
- $a_t(\vec{x}_\sigma) \leq 2t - 1$ for $t = i_{s_0+1}, \dots, i_{\ell-2}$.

If also $a_{i_{\ell-1}}(\vec{x}_\sigma) \leq 2i_{\ell-1} - 1$, then Proposition 6.11 implies that $\vec{x}_\sigma \in C_{I,s_0}$. In that case, we are done proving the forward containment. Assume instead that $a_{i_{\ell-1}}(\vec{x}_\sigma) \geq 2i_{\ell-1}$. We now split into two cases:

Case 1 Assume $a_{i_{\ell-1}}(\vec{x}_\sigma) = 2i_{\ell-1}$. We claim that $\vec{x}_\sigma \in C_{I,\ell-1}$. By Proposition 6.11, since \vec{x}_σ is monotonic, it will suffice to show

- (a) $b_t(\vec{x}_\sigma) \geq 2t + 1$ for $0 \leq t \leq i_{\ell-1} - 1$,
- (b) $b_t(\vec{x}_\sigma) \geq 2t$ for $i_{\ell-1} \leq t \leq k - 1$, and
- (c) $a_t(\vec{x}_\sigma) \leq 2t$ for $t = i_0, \dots, i_{\ell-1}$.

Observation 6.5, Proposition 6.6, and the facts that $\vec{x}_\sigma \in X_{i_{\ell-1}}$ and $a_{i_{\ell-1}}(\vec{x}_\sigma) = 2i_{\ell-1}$ imply, for each $t = 0, \dots, i_{\ell-1} - 1$, that

$$b_t(\vec{x}_\sigma) = b_{t+k}(\vec{x}_\sigma) - n \geq 2(t + k) + a_{i_{\ell-1}}(\vec{x}_\sigma) - 2i_{\ell-1} - n \geq 2t + 1.$$

This verifies (a). Taking $t = i_{\ell-1}, \dots, k - 1$, similar reasoning confirms (b):

$$b_t(\vec{x}_\sigma) \geq a_{i_{\ell-1}}(\vec{x}_\sigma) + 2(t - i_{\ell-1}) \geq 2t.$$

Finally, we have $a_t(\vec{x}_\sigma) \leq 2t$ for each $t = i_0, \dots, i_{\ell-1}$. For $t = i_0, \dots, i_{\ell-2}$, this is because $\vec{x}_\sigma \in X_I$; for $t = i_{\ell-1}$, it is our assumption in Case 1. Thus, in Case 1, (a), (b), and (c) hold, and so $\vec{x}_\sigma \in C_{I,\ell-1}$.

Case 2 Assume instead that $a_{i_{\ell-1}}(\vec{x}_\sigma) \geq 2i_{\ell-1} + 1$. Writing $\vec{x}_\sigma = (x_1, \dots, x_n)$, we claim in this case that $x_1 = x_2 = 0 \equiv k$ and that $\vec{y} = (x_2, \dots, x_n, x_1) \in C_{I,\ell-1}$. By similar reasoning to Case 1,

$$b_0(\vec{x}_\sigma) = b_k(\vec{x}_\sigma) - n \geq a_{i_{\ell-1}}(\vec{x}_\sigma) + 2(k - i_{\ell-1}) - n \geq 2.$$

Thus, $x_1 = x_2 = 0 \equiv k$. Define \vec{y} as above. Note that, since \vec{x}_σ is monotonic, \vec{y} is also monotonic. It remains to show that $\vec{y} \in C_{I,\ell-1}$. The arguments are almost identical to those in Case 1, except that we need to check that $a_{i_{\ell-1}}(\vec{y}) \leq 2i_{\ell-1}$. Using Observations 6.4 and 6.5 and the fact that $a_{i_{\ell-1}}(\vec{y}) = a_{i_{\ell-1}}(\vec{x}_\sigma) - 1$, we compute

$$\begin{aligned} a_{i_{\ell-1}}(\vec{y}) &= a_{i_{\ell-1}}(\vec{x}_\sigma) - 1 \leq b_{k-1}(\vec{x}_\sigma) - 2(k-1-i_{\ell-1}) - 1 \leq a_k(\vec{x}_\sigma) - 2k + 1 + 2i_{\ell-1} \\ &= a_0(\vec{x}_\sigma) + (n+1-2k) + 2i_{\ell-1} = a_0(\vec{x}_\sigma) + 2i_{\ell-1} \leq 2i_{\ell-1}. \end{aligned}$$

This completes the proof of the forward containment (20). For the reverse containment, keep the same subset $I \subset \mathbb{Z}_k$ from the start of the induction step of the proof, fix some $s \in \mathbb{Z}_\ell$, let $\vec{x} \in C_{I,s}$ be monotonic, and let $t \in I = \{i_0, \dots, i_{\ell-1}\}$. We will show for each $r = 0, \dots, k - 1$ that $b_{t+r}(\vec{x}) \geq a_t(\vec{x}) + 2r$. Proposition 6.6 will then imply that $\vec{x} \in X_t$. Since t is arbitrary, this will imply that $\vec{x} \in X_I$, completing the proof. We will

split into cases, but first note, since \vec{x} is monotonic, that Proposition 6.11 implies

- $b_t(\vec{x}) \geq 2t + 1$ for $0 \leq t \leq i_s - 1$,
- $b_t(\vec{x}) \geq 2t$ for $i_s \leq t \leq k - 1$,
- $a_t(\vec{x}) \leq 2t$ for $t = i_0, \dots, i_s$, and
- $a_t(\vec{x}) \leq 2t - 1$ for $t = i_{s+1}, \dots, i_{\ell-2}$.

Case 1 If $t + r \leq k - 1$, then $b_{t+r}(\vec{x}) \geq 2(t + r) \geq a_t(\vec{x}) + 2r$.

Case 2 If instead $t + r \geq k$ and $t + r \leq k + i_s - 1$, then

$$b_{t+r}(\vec{x}) = n + b_{t+r-k}(\vec{x}) \geq n + 2(t + r - k) + 1 = 2t + 2r + (n + 1 - 2k),$$

$$b_{t+r}(\vec{x}) \geq a_t(\vec{x}) + 2r.$$

Case 3 Similarly, if $t + r \geq k$ and $t \geq i_{s+1}$, then

$$b_{t+r}(\vec{x}) = n + b_{t+r-k}(\vec{x}) \geq n + 2(t + r - k) = (2t - 1) + 2r + (n + 1 - 2k),$$

$$b_{t+r}(\vec{x}) \geq a_t(\vec{x}) + 2r.$$

Are there other cases? If there were, they would satisfy $t + r \geq k + i_s$ and $t \leq i_s$, giving

$$k + i_s \leq t + r \leq i_s + r \quad \text{so} \quad k \leq r.$$

Yet $r \leq k - 1$ by assumption. Therefore, in every case, $b_{t+r}(\vec{x}) \geq a_t(\vec{x}) + 2r$, and so $\vec{x} \in X_{i_t}$ for arbitrary $t \in I$. Thus, $\vec{x} \in X_I$. This completes the proof of the reverse containment, and thus of the equality in (2)=(19). □

7 General construction

This section confirms the remaining details of our main construction and completes the proof of our main result, Theorem 7.10. Namely, Section 7.1 describes how to decompose arbitrary X_I , and Section 7.2 shows that this decomposition does in fact give an appropriate handle structure for X_I .

Section 7 uses Notations 3.3, 3.6 and 3.8 and Convention 3.9.

7.1 Handle decompositions: the general case

Throughout Section 7.1, fix arbitrary $I = \{i_s\}_{s \in \mathbb{Z}_\ell} = \bigsqcup_{r \in \mathbb{Z}_m} I_r \subsetneq \mathbb{Z}_k$, following Convention 3.9. Recall in particular that $T = \{t \in \mathbb{Z}_\ell : i_t - 1 \notin I\} = \{t_r\}_{r \in \mathbb{Z}_m}$, so that $\{\min I_r\}_{r \in \mathbb{Z}_m} = \{i_t\}_{t \in T}$.

7.1.1 Overview In Section 7.1, we will decompose X_I into handles in several steps as follows. First, we will decompose X_I into pieces X_{I,J,i_*} determined by all pairs

(J, i_*) where $J \subset \{\min I_r\}$ and $i_* \in I$. Second, for fixed (J, i_*) , we will define disjoint subsets $U, V \subset I$ for the purpose of dividing each interval $[i - 1, i]$ for $i \in I$ into thirds if $i \in U$, into halves if $i \in V$, or neither if $i \notin U, V$. Third, still fixing (J, i_*) , after dividing certain intervals into halves and thirds as just described, we will decompose each piece X_{I,J,i_*} into pieces $X_{I,J,i_*,V^-,U^\circ,U^-}$; these pieces are determined by all triples (V^-, U°, U^-) where $V^- \subset V, U^\circ \subset U$, and $U^- \subset U \setminus U^\circ$. For each of the first three steps, we will describe what to do within each block I_r ; then we will take a product across all blocks and extend by permutations of the indices.

Fourth, we will order the possibilities of the tuple $(J, i_*, V^-, U^\circ, U^-)$, thus determining an order on the pieces $X_{I,J,i_*,V^-,U^\circ,U^-}$. The order will be lexicographical, and will thus require defining orders on $\{J \subset \{\min I_r\}\}, \{i_* \in I\}, \{V^- \subset V\}, \{U^\circ \subset U\}$, and $\{U^- \subset U \setminus U^\circ\}$. Of these five orders, only the third will be somewhat complicated. Once we define this order, we will use it to relabel the various pieces $X_{I,J,i_*,V^-,U^\circ,U^-}$ as Y_z , with $z = 1, 2, 3, \dots$. Fifth and finally, we will decompose each Y_z into handles, one of which we denote by Y_z^* (each handle H from Y_z is related to Y_z^* by $H = \{\vec{x}_\tau : \vec{x} \in Y_z^*\}$ for some fixed permutation $\tau \in S_n$).

7.1.2 Decomposing X_I according to (J, i_*) Fix arbitrary $J \subset \{\min I_r\}$ and $i_* \in I$ for all of Section 7.1.2. Momentarily fixing arbitrary $r \in \mathbb{Z}_m$, write

$$(21) \quad a = \min I_r, \quad b = \max I_r, \quad c = \min I_{r+1}, \quad \text{and} \quad \widehat{C}_r = \prod_{j=b+1}^{c-1} [b, j]^2,$$

and define

$$(22) \quad C_r = \begin{cases} [a - 1, a] & \text{if } i_* = a \in J, \\ [a - 1, a] \times \{a\} & \text{if } i_* \neq a \in J, \\ \{a\} & \text{if } i_* \neq a \notin J, \\ \text{(no factor)} & \text{if } i_* = a \notin J \end{cases} \\ \times \prod_{i=a+1}^b \begin{cases} [i - 1, i] \times \{i\} & \text{if } i \neq i_*, \\ [i - 1, i] & \text{if } i = i_* \end{cases} \times \begin{cases} \widehat{C}_r \times [b, c - 1] & \text{if } c \notin J, \\ \widehat{C}_r & \text{if } c \in J \end{cases}.$$

Now the piece of X_I corresponding to the pair (J, i_*) is given by

$$X_{I,J,i_*} = \left\langle \prod_{r \in \mathbb{Z}_m} C_r \right\rangle.$$

7.1.3 The index subsets $U, V \subset I$ Fix arbitrary $J \subset \{\min I_r\}$ and $i_* \in I$ for all of Section 7.1.3. For each $r \in \mathbb{Z}_m$, define subsets $U_r, V_r \subset I_r$ following Table 10 (or

	$i_* \notin I_r,$ $a \notin J$	$i_* \notin I_r,$ $a \in J$	$i_* \in I_r,$ $i_* \leq b-2$	$i_* \in I_r,$ $i_* \geq b-1$
U_r	\emptyset	$I_r \setminus \{a, b\}$	$I_r \setminus \{a, i_*, i_* + 1, b\}$	$I_r \setminus \{a, i_*, b\}$
V_r	$I_r \setminus \{a\}$	$\{b\}$	$\{i_* + 1, b\}$	\emptyset
$I_r \setminus (U_r \cup V_r)$	$\{a\}$	$\{a\} \setminus \{b\}$	$\{a, i_*\}$	$\{a, i_*, b\}$

Table 10: The index subsets $U_r, V_r \subset I_r$ when $I_r = \{a, \dots, b\}$.

equivalently according to Tables 11 and 12 in Appendix A, which present U_r and V_r more explicitly). Note that $\min I_r \notin (U_r \cup V_r)$ unless $I_r \neq I_*$ and $\min I_r = \max I_r \in J$. See Table 7 for an example of this exceptional case: X_I for $I = \{0, 2\}$, from T^7 .

Define

$$U = \bigcup_{r \in \mathbb{Z}_m} U_r \quad \text{and} \quad V = \bigcup_{r \in \mathbb{Z}_m} V_r.$$

Next, decompose each X_{I,J,i_*} into pieces $X_{I,J,i_*,V^-,U^\circ,U^-}$ as follows. Write

$$2^V = \{V^- \subset V\} \quad \text{and} \quad 2^U = \{U^\circ \subset U\},$$

and, given $U^\circ \subset U$,

$$2^{U \setminus U^\circ} = \{U^- \subset U \setminus U^\circ\}.$$

Given $V^- \subset V$, write $V^+ = V \setminus V^-$, and given $U^\circ \subset U$ and $U^- \subset U \setminus U^\circ$, write $U^+ = U \setminus (U^\circ \cup U^-)$. Then $V = V^- \sqcup V^+$ and $U = U^- \sqcup U^\circ \sqcup U^+$. Momentarily fixing $r \in \mathbb{Z}_m$, write a, b, c , and \hat{C}_r as in (21), and for each $i \in I_r$ define

$$\rho_i = \begin{cases} [i-1, i-\frac{2}{3}] & \text{if } i \in U^-, \\ [i-\frac{2}{3}, i-\frac{1}{3}] & \text{if } i \in U^\circ, \\ [i-\frac{1}{3}, i] & \text{if } i \in U^+, \\ [i-1, i-\frac{1}{2}] & \text{if } i \in V^-, \\ [i-\frac{1}{2}, i] & \text{if } i \in V^+, \\ [\max I_{r-1}, i-1] & \text{if } i = a \notin J \cup V, \\ [i-1, i] & \text{otherwise.} \end{cases}$$

Note that $\rho_i \subset [i-1, i]$ for each $i = a+1, \dots, b$, that $\rho_a \subset [a-1, a]$ if $a \in J$, and that $\rho_c = [b, c-1]$ if $c \notin J$. Still fixing $r \in \mathbb{Z}_m$, define

$$X_{I,J,i_*,V^-,U^\circ,U^-,r} = \left\{ \begin{array}{ll} \rho_a & \text{if } i_* = a \in J, \\ \rho_a \times \{a\} & \text{if } i_* \neq a \in J, \\ \{a\} & \text{if } i_* \neq a \notin J, \\ \text{(no factor)} & \text{if } i_* = a \notin J \end{array} \right\} \times \prod_{i=a+1}^b \left\{ \begin{array}{ll} \rho_i \times \{i\} & \text{if } i \neq i_*, \\ \rho_i & \text{if } i = i_* \end{array} \right\} \times \left\{ \begin{array}{ll} \hat{C}_r \times \rho_c & \text{if } c \notin J, \\ \hat{C}_r & \text{if } c \in J \end{array} \right\}.$$

The piece of X_I corresponding to the tuple $(J, i_*, V^-, U^\circ, U^-)$ is

$$X_{I,J,i_*,V^-,U^\circ,U^-} = \left\langle \prod_{r \in \mathbb{Z}_m} X_{I,J,i_*,V^-,U^\circ,U^-,r} \right\rangle.$$

Note that $X_{I,J,i_*} = \bigcup_{V^-,U^\circ,U^-} X_{I,J,i_*,V^-,U^\circ,U^-}$.

7.1.4 Ordering the pieces $X_{I,J,i_*,V^-,U^\circ,U^-}$ Next, we define orders $<$ on $I, 2^V, 2^U, 2^{U \setminus U^\circ}$, and $\{J \subset \{\min I_r\}\}$ and use these to order the pieces $X_{I,J,i_*,V^-,U^\circ,U^-}$ lexicographically. We then relabel them as Y_1, Y_2, Y_3, \dots .

Order $\{J \subset \{\min I_r\}\}$ and 2^U partially by inclusion so that $J' < J$ if $J' \subsetneq J$ and $U'^\circ < U^\circ$ if $U'^\circ \subsetneq U^\circ$; extend these partial orders arbitrarily to total orders. Define an arbitrary total order $<$ on $2^{U \setminus U^\circ}$. Partially order I so that $i < i'$ if $i \in I_r, i' \in I_s$, and $i - \min I_r < i_s - \min I_s$; extend arbitrarily to a total order on I .

It remains to order 2^V . This will be slightly more complicated. To do this, we first define a total order $<_r$ on 2^{V_r} for each $r \in \mathbb{Z}_m$. First consider the case $I_r \ni i_*$, ie $I_r = I_*$. If $i_* \geq \max I_* - 1$, we have $V_r = \emptyset$, so there is nothing to do. Otherwise, we have $i_* \leq \max I_* - 2$ and $V_r = \{i_* + 1, \max I_*\}$; in this case, order 2^{V_r} as follows:

$$\{i_* + 1\} <_r \emptyset <_r \{i_* + 1, \max I_*\} <_r \{\max I_*\}.$$

Now consider the case $I_r \not\ni i_*$. Define $<_r$ on 2^{V_r} recursively by $V_r^- <_r V_r'^-$ if

- $\max V_r^- < \max V_r'^-$, or
- $\max V_r^- = \max V_r'^-$ and $V_r'^- \setminus \{\max V_r'^-\} <_r V_r^- \setminus \{\max V_r^-\}$.

Note the reversal of order on the line above. If we assume without loss of generality that $V_r = \{0, \dots, b\}$, we can write the order explicitly:

$$(23) \quad \emptyset <_r \{0\} <_r \{0, 1\} <_r \{1\} <_r \{1, 2\} <_r \{0, 1, 2\} <_r \{0, 2\} <_r \{2\} <_r \{2, 3\} <_r \{0, 2, 3\} <_r \{0, 1, 2, 3\} <_r \{1, 2, 3\} <_r \{1, 3\} <_r \dots <_r \{0, 1, 2, b\} <_r \{1, 2, b\} <_r \{1, b\} <_r \{0, 1, b\} <_r \{0, b\} <_r \{b\}.$$

See Tables 9 and 18, and the part of Table 16 where $i_* = 4$.

Use the orderings $<_r$ on 2^{V_r} to define a partial order on 2^V by declaring $V^- < V'^-$ if

- $V^- \cap I_r <_r V'^- \cap I_r$ for some r , and
- there is no r for which $V'^- \cap I_r <_r V^- \cap I_r$.

Extend \prec arbitrarily to a total order on 2^V . This determines a total order on

$$(24) \quad \{(J, i_*, V^-, U^\circ, U^-)\}_{J \subset T, i_* \in I, V^- \subset V, U^\circ \subset U, U^- \subset U \setminus U^\circ},$$

and thus on the pieces $X_{I, J, i_*, V^-, U^\circ, U^-}$. Relabel these pieces as Y_z for $z = 1, \dots, \#(24)$, according to this order.

7.1.5 Decomposing each Y_z into handles Each Y_z is now given by an expression of the form

$$(25) \quad \left\langle \prod_{r=1}^n \chi_r \right\rangle,$$

where each χ_r is either a closed interval or a singleton. Fixing arbitrary z , use the expression (25) to define the coarsest equivalence relation \sim on $\{1, \dots, n\}$ that obeys the following property: whenever $\chi_r \subset \chi_s$, we have $r \sim s$. Denote the set of equivalence classes under \sim by $P = \{R_1, \dots, R_p\}$, and for each $r = 1, \dots, p$ write $\langle \prod_{s \in R_r} \chi_s \rangle = \xi_r$. Define

$$(26) \quad Y_z^* = \prod_{r=1}^p \xi_r.$$

In Section 7.2, we will see that each Y_z^* is a handle, and that attaching Y_z to $\bigcup_{s=1}^{z-1} Y_s$ amounts to attaching a collection of handles, each related to Y_z^* as follows. Let

$$G = \{\sigma \in S_n : \vec{x}_\sigma \in Y_z^* \text{ whenever } \vec{x} \in Y_z^*\} = S_{n|R_1|} \times \dots \times S_{n|R_p|}$$

consist of the permutations on the indices of T^n which fix Y_z^* setwise. Then there is a one-to-one correspondence between the left cosets of G and the handles in Y_z :

$$\tau G \leftrightarrow \{\vec{x}_\tau : \vec{x} \in Y_z^*\}.$$

Example 7.1 Consider $X_I \subset T^9$ where $I = \{0, 1, 2, 3\}$, which is detailed in Tables 14 and 15. Note that $T = \{0\}$. In particular, consider the first and twelfth rows of Table 14 (after the headings), where $J = \emptyset$, $i_* = 0$, $U = \{2\}$, and $V = \{1, 3\}$. The first row of Table 14 corresponds to

$$(27) \quad Y_1 = X_{I, J, s, V^-, U^\circ, U^-} = \langle \alpha^- 1 \beta_3^\circ 2 \gamma^+ 3 \delta^3 \rangle,$$

where $V^- = \{1\}$, $U^\circ = \{2\}$, and $U^- = \emptyset$ with

$$\begin{aligned} \chi_1 &= \alpha^- = [0, \frac{1}{2}], & \chi_2 &= \{1\}, & \chi_3 &= \beta_3^\circ = [\frac{4}{3}, \frac{5}{3}], & \chi_4 &= \{2\}, \\ \chi_5 &= \gamma^+ = [\frac{5}{2}, 3], & \chi_6 &= \{3\}, & \chi_7 &= \chi_8 = \chi_9 = \delta = [3, 4]. \end{aligned}$$

The ensuing partition of $\{1, \dots, 9\}$ gives

$$P = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5, 6, 7, 8, 9\}\},$$

and so

$$Y_1^* = \chi_1 \times \chi_2 \times \chi_3 \times \chi_4 \times \langle \chi_5 \times \chi_6 \times \chi_7 \times \chi_8 \times \chi_9 \rangle = \alpha^{-1} \beta_3^{\circ} 2 \langle \gamma^+ 3\delta^3 \rangle,$$

where

$$\xi_1 = \alpha^-, \quad \xi_2 = \{1\}, \quad \xi_3 = \beta_3^{\circ}, \quad \xi_4 = \{2\}, \quad \text{and} \quad \xi_5 = \langle \gamma^+ 3\delta^3 \rangle.$$

The twelfth row of Table 14 corresponds to

$$Y_{12} = X_{I,J,s,V^-,U^{\circ},U^-} = \langle \alpha^+ 1 \beta_3^+ 2 \gamma^- 3 \delta^3 \rangle,$$

where $V^- = \{3\}$, $U^{\circ} = \emptyset = U^-$. The ensuing partition of $\{1, \dots, 9\}$ gives

$$P = \{\{1, 2\}, \{3, 4, 5\}, \{6, 7, 8, 9\}\},$$

and so

$$Y_{12}^* = \langle \chi_1 \times \chi_2 \rangle \times \langle \chi_3 \times \chi_4 \times \chi_5 \rangle \times \langle \chi_6 \times \chi_7 \times \chi_8 \times \chi_9 \rangle = \underbrace{\langle \alpha^+ 1 \rangle}_{\xi_1} \underbrace{\langle \beta_3^+ 2 \gamma^- \rangle}_{\xi_2} \underbrace{\langle 3 \delta^3 \rangle}_{\xi_3}.$$

7.2 Properties of handle decompositions

7.2.1 Combinatorics

Proposition 7.2 *Let $i \in I_s \cap V^-$ for some $s \in \mathbb{Z}_m$, where $i_* \notin I_s$. Define $b = \max I_s$ and $c = \max(I_s \cap V^-)$. Let $V'^- = V^- \setminus \{i\}$. Then $V'^- \prec V^-$ if and only if $|V^- \cap \{i + 1, \dots, b\}|$ is even.*

Proof We use induction on $c - i$. When $c - i = 0$, we have $c = i > \max(I_s \cap V^- \setminus \{i\})$ and $I_r \cap V^- = I_r \cap V^- \setminus \{i\}$ for all $r \neq s$, so $V'^- \prec V^-$.

Now assume $c - i = t > 0$, and that the claim holds if $\max(I_s \cap V^-) - i < t$. Let $W^- = V^- \setminus \{c\}$ and $W'^- = V'^- \setminus \{c\}$. Then $|V^- \cap \{i + 1, \dots, b\}|$ and $|W^- \cap \{i + 1, \dots, b\}|$ have opposite parities. Also, by construction, $V^- \prec V'^-$ if and only if $W'^- \prec W^-$. The result now follows by induction. □

Notation 7.3 Denote the symmetric difference of sets R and S by

$$R \ominus S = (R \setminus S) \cup (S \setminus R).$$

Proposition 7.4 *Let $A \subset V$ be such that $V^- \prec V^- \ominus \{a\}$ for each $a \in A$. Then $V^- \prec V^- \setminus A$.*

Proof Suppose first that $A \subset I_s$ for some $s \in \mathbb{Z}_m$. Write $A = \{a_1, \dots, a_q\}$ with $\min I_s \leq a_1 \leq \dots \leq a_q \leq \max I_s = b$. Assume that $i_* \notin I_s$ and $|I_s| \geq 3$ (the other cases are trivial). Proposition 7.2 implies, for each $a \in A$, that $|V^- \cap \{a + 1, \dots, b\}|$ is odd if and only if $a \in V^-$. For each $r = 1, \dots, q$, denote the symmetric difference $V^- \ominus \{a_1, \dots, a_r\}$ by V_r^- . Then, $|V_a^- \cap \{a + 1, \dots, b\}| = |V^- \cap \{a + 1, \dots, b\}|$ for each $a = 0, \dots, q - 1$. Since this quantity is odd if and only if $a \in V^-$, Proposition 7.2 implies

$$V^- \prec V_1^- \prec \dots \prec V_q^- = V^- \setminus A.$$

For the general case, apply this argument repeatedly for each $s \in \mathbb{Z}_m$. □

7.2.2 Topology

Observation 7.5 *In X_I , if Y_z comes from $(J, i_*, V^-, U^\circ, U^-)$ and Y_w comes from $(J, i_*, V^-, U^\circ, U'^-)$, then $Y_z \cap Y_w = \emptyset$ unless $U^- = U'^-$. That is, if $U^- \neq U'^-$, then*

$$X_{I, J, i_*, V^-, U^\circ, U^-} \cap X_{I, J, i_*, V^-, U^\circ, U'^-} = \emptyset.$$

Lemma 7.6 *Each factor ξ_r in the expression (26) for Y_z^* has one of the forms described in Lemma 4.2, and thus is PL homeomorphic to $D^{d(r)}$ for some $d(r) \geq 0$.*

Moreover, $\sum_{r=1}^p d(r) = n + 1 - |I|$, so $Y_z^* \cong D^{n+1-|I|}$.

Proof Regarding the first claim, we examine the equivalence relation \sim that led to (26). Suppose $\chi_r \subset \chi_{r'}$. Then, by construction, either χ_r is a singleton (in $I \setminus \{i_*\}$) and $\chi_{r'}$ is an interval with this singleton as an endpoint, or else $\chi_r \supset [\max I_s, \max I_s + 1]$ for some $s \in \mathbb{Z}_\ell$. Moreover, by construction, if $\chi_{r'}$ contains a point of $I \setminus \{i_*\}$, then it contains only one such point and it contains no interval of the form $[\max I_s, \max I_s + 1]$, and no $\chi_{r'}$ contains more than one interval of the form $[\max I_s, \max I_s + 1]$. The first claim now follows. (For an explicit accounting of the types of factors $\xi_r = \langle \prod_{s \in R_r} \chi_s \rangle$ see Tables 19–21.)

Regarding the second claim, note for each $r = 1, \dots, p$ that $d(r)$ equals the number of intervals among $\{\chi_s\}_{s \in R_r}$, which equals the order of R_r minus the number of singletons among $\{\chi_s\}_{s \in R_r}$. Since $\sum_{r=1}^p |R_r| = |P| = n$ and $\{\chi_s : s = 1, \dots, n\}$ contains a total of $|I| - 1$ singletons, it follows that $\sum_{r=1}^p d(r) = n + 1 - |I|$. Thus, $Y_z^* \cong D^{n+1-|I|}$. □

We wish to show, in arbitrary X_I , that attaching any Y_z to $\bigcup_{w < z} Y_w$ amounts to attaching a collection of $(n + 1 - |I|)$ -dimensional $h(z)$ -handles for some $h(z)$. Indeed,

Lemma 7.6 confirms that each Y_z^* from X_I is a compact $(n+1-|I|)$ -ball, so it remains to consider how everything is glued together. Our goal is to show that

$$(28) \quad Y_z^* \cap \bigcup_{w < z} Y_w \cong S^{h(z)-1} \times D^{n+1-|I|-h(z)}$$

and

$$(29) \quad Y_z^* \cap (Y_z \setminus Y_z^*) \subset Y_z^* \cap \bigcup_{w < z} Y_w.$$

The former implies that attaching Y_z^* to $\bigcup_{w < z} Y_w$ amounts to attaching an $(n+1-|I|)$ -dimensional $h(z)$ -handle, and the latter further implies that if we attach all the copies of Y_z^* one at a time to $\bigcup_{w < z} Y_w$, then attaching each copy amounts to attaching another $(n+1-|I|)$ -dimensional $h(z)$ -handle.

Recall that each Y_z^* has the form $\prod_{r=1}^p \xi_r(z)$. Hence,

$$\partial Y_z^* = \bigcup_{a=1}^p \left(\prod_{r=1}^{a-1} \xi_r(z) \times \partial \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z) \right).$$

We show, given arbitrary Y_z^* in X_I , that there is a subset $S(z) \subset \{1, \dots, p\}$ such that

$$(30) \quad Y_z^* \cap \bigcup_{w < z} Y_w = \bigcup_{a \in S(z)} \left(\prod_{r=1}^{a-1} \xi_r(z) \times \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z) \right).$$

Then, writing $h(z) = \sum_{r \in S(z)} \dim(\xi_r)$, we will obtain (28):

$$\begin{aligned} Y_z^* \cap \bigcup_{w < z} Y_w &= \bigcup_{a \in S(z)} \left(\prod_{r=1}^{a-1} \xi_r(z) \times \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z) \right) \\ &\cong \left(\partial \prod_{r \in S(z)} \xi_r(z) \right) \times \prod_{r \notin S(z)} \xi_r(z) \\ &\cong \partial D^{h(z)} \times D^{n+1-|I|-h(z)} = S^{h(z)-1} \times D^{n+1-|I|-h(z)}. \end{aligned}$$

Our next step is to describe the subset $S(z) \subset \{1, \dots, p\}$. To do so, we characterize each $\xi_r(z)$ as type (A) or type (B); then $S(z)$ will consist of those $r = 1, \dots, p$ for which $\xi_r(z)$ has type (A). After that, Lemmas 7.7 and 7.8 will establish (30) by double containment, implying (28), and Lemma 7.9 will establish (29).

Consider an arbitrary $Y_z^* = \prod_{r=1}^p \xi_r(z)$ from an arbitrary X_I , and classify each factor $\xi_r(z)$ into one of two classes, (A) or (B), as follows. Say that $\xi_r(z)$ is in class (B) if

- $\xi_r(z) = [i - \frac{2}{3}, i - \frac{1}{3}]$ for some $i \in I$;
- $\xi_r(z) = [i_*, i_* + \frac{1}{2}]$;

- $[\max I_s, j]$ is a factor in the expression for $\xi_r(z)$ for some s and j ; or
- some $\{i\}$ is a factor in the expression for $\xi_r(z)$ and
 - $i \in V^+$ and $i + 1 \in U^\circ \cup U^+ \cup V^+$, or $i \in U^- \cup U^\circ \cup V^-$ and $i + 1 \in V^-$;
and
 - $|V^- \cap \{i + 1, \dots, \max I_s\}|$ is even, where $i \in I_s$.

All other types of $\xi_r(z)$ are of class (A). Tables 19–21 in Appendix A list the possibilities explicitly.

Lemma 7.7 Suppose $Y_z^* = \prod_{r=1}^p \xi_r(z)$ comes from $(J, i_*, V^-, U^\circ, U^-)$. If, for some $a = 1, \dots, p$, $\xi_a(z)$ is of class (A) and

$$\vec{x} = (x_1, \dots, x_n) \in \prod_{r=1}^{a-1} \xi_r(z) \times \partial \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z),$$

then $\vec{x} \in Y_w$ for some $w < z$.

Proof Suppose first that some $\{i\}$ appears in the expression for $\xi_a(z)$, where

- $i \in I_s$;
- $i \in V^+$ and $i + 1 \in U^\circ \cup U^+ \cup V^+$, or $i \in U^- \cup U^\circ \cup V^-$ and $i + 1 \in V^-$; and
- $|V^- \cap \{i + 1, \dots, \max I_s\}|$ is odd.

Then \vec{x} is in the Y_w coming from $(J, i_*, V'^-, U^\circ, U^-)$ where V'^- is either $V^- \cup \{i\}$ or $V^- \setminus \{i + 1\}$. In either case, Proposition 7.2 implies that $V'^- < V$ and thus $w < z$.

Next, suppose that $\xi_a(z)$ has no singleton factors. There are two possibilities. If $\xi_a(z) = [i_* - 1, i_*]$ with $i_* \in J$, then \vec{x} is in some Y_w coming from $J \setminus \{i_*\} < J$. Otherwise, $\xi_a(z) = [i - 1, i - \frac{1}{2}]$ for some $i \in J \cap V^-$; in this case, $i + 1 \notin I$, and so \vec{x} is in some Y_w coming either from $J \setminus \{i\} < J$ or the from same J and i_* and $V'^- = V^- \setminus \{i\}$, where Proposition 7.2 implies that $V'^- < V^-$ because $i + 1 \notin I$.

The remaining cases follow by similar reasoning. The interested reader may find Table 21 useful for this. □

Lemma 7.8 Let $Y_z^* = \prod_{r=1}^p \xi_r(z)$ come from some $(J, i_*, V^-, U^\circ, U^-)$. If

$$\vec{x} = (x_1, \dots, x_n) \in Y_z^* \cap \bigcup_{w < z} Y_w,$$

then

$$\vec{x} \in \prod_{r=1}^{a-1} \xi_r(z) \times \partial \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z)$$

for some $a = 1, \dots, p$ such that $\xi_a(z)$ is of class (A).

Proof Let $\vec{x} = (x_1, \dots, x_n) \in Y_z^* \cap Y_{w'}$ for some $w' < z$. Choose the smallest $w < z$ such that $\vec{x} \in Y_w$, and assume that Y_w comes from some $(J', i'_*, V'^-, U'^\circ, U'^-)$ with $V'^- \subset V'$ and $U'^\circ \subset U'$, whereas Y_z comes from some $(J, i_*, V^-, U^\circ, U^-)$ with $V^- \subset V$ and $U^\circ \subset U$. Write

$$S = \left\{ a = 1, \dots, p : \vec{x} \in \prod_{r=0}^{a-1} \xi_r(z) \times \partial \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z) \right\}.$$

Assume for contradiction that $\xi_a(z)$ is of class (B) for every $a \in S$. If $S = \emptyset$, then no coordinate of \vec{x} equals i_* , so $i'_* = i_*$. Also, in that case, no coordinate of \vec{x} equals $\min I_s - 1$ for any $s \in \mathbb{Z}_m$, and so J and J' completely determine the number of coordinates that \vec{x} has in each open interval $(\min I_s - 1, \min I_{s+1} - 1)$. It follows that either $J' = J$ or $J' = T \setminus J$. If $J' = T \setminus J$, then considering the coordinates of \vec{x} in $[\min I_*, \max I_*]$ yields a contradiction. If $J' = J$, then the fact that $S = \emptyset$ implies that $V^- = V'^-$, $U^\circ = U'^\circ$, and $U^- = U'^-$, contradicting the fact that $w < z$.

Therefore, $S \neq \emptyset$. If no coordinate of \vec{x} equals i_* , then $i'_* = i_*$, so again either $J' = J$ or $J' = T \setminus J$. The latter case gives the same contradiction as before. Therefore $J' = J$, and so $V' = V$.

For each $i \in V^- \ominus V'^-$, \vec{x} has a coordinate $x_t = i - \frac{1}{2}$ (using the fact that $i'_* = i_*$ and $J' = J$). The corresponding $\xi_r(z)$ has $r \in S$, and so by assumption $\xi_r(z)$ is of class (B). Therefore, $V^- \prec V^- \ominus \{i\}$ for each $i \in V^- \ominus V'^-$. Proposition 7.4 implies that $V^- \prec V'^-$ unless $V^- = V'^-$. Since $w < z$, we must have $V^- = V'^-$.

Each $i \in U'^\circ$ must also be in U° , or else the corresponding coordinate of \vec{x} would equal $i - \frac{1}{3}$ or $i - \frac{2}{3}$, and the corresponding $\xi_a(z)$ would be of class (A) with $a \in S$, contrary to assumption. Thus, $U^\circ \subset U'^\circ$. Similarly, each $i \in U^\circ$ must also be in U'° , or else the $Y_{w'}$ coming from $(J, i_*, V, U'^\circ \cup \{i\}, U^- \setminus \{i\})$ would still contain \vec{x} but with $w' < w$, contrary to assumption. Thus, $U'^\circ = U^\circ$.

Finally, we must have $U'^- = U^-$, by Observation 7.5. This implies, contrary to assumption, that $Y_w = Y_z$. □

Lemma 7.9 Let $Y_z^* = \prod_{r=1}^p \xi_r(z)$ come from some $(J, i_*, V^-, U^\circ, U^-)$. If

$$\vec{x} = (x_1, \dots, x_n) \in Y_z^* \cap (Y_z \setminus Y_z^*),$$

then

$$\vec{x} \in \prod_{r=1}^{a-1} \xi_r(z) \times \partial \xi_a(z) \times \prod_{r=a+1}^p \xi_r(z)$$

for some $a = 1, \dots, p$ such that $\xi_a(z)$ is of class (A).

Proof This follows from a case analysis, for which the interested reader may find Tables 19–21 useful. It comes down to this. Consider two pieces $\xi_a(z)$ and $\xi_b(z)$ of Y_z^* for which the infimum $\min \xi_b(z)$ of all coordinates in $(0, k)$ among all points in $\xi_b(z)$ equals the supremum $\max \xi_a(z)$ of all coordinates in $(0, k)$ among all points in $\xi_a(z)$. Write $\max \xi_a(z) = \min \xi_b(z) = c$. Then $c \in \mathbb{Z}_k$. If c equals $i - 1$ for some $i \in T$, then $i \in J$ and $\xi_b(z)$ is of class (A). Otherwise, $c = i_*$ and $\xi_a(z)$ is of class (A). \square

7.3 Proof of the main result

The results of Sections 6 and 7.2 provide all the details we need to prove:

Theorem 7.10 For $n = 2k - 1 \in \mathbb{Z}_+$, the n -torus admits a multisection $T^n = \bigcup_{r \in \mathbb{Z}_k} X_r$ defined by

$$(1) \quad \begin{aligned} X_0 &= \{\vec{x}_\sigma : \vec{x} \in [0, 1]^2 \cdots [0, k - 1]^2 [0, k] / \sim, \sigma \in S_n\}, \\ X_i &= \{\vec{x} + (i, \dots, i) : \vec{x} \in X_0\}. \end{aligned}$$

Proof Lemma 6.8 implies that $X = \bigcup_{i \in \mathbb{Z}_k} X_i$, so it remains only to prove, for each nonempty proper subset $I \subset \mathbb{Z}_k$, that $X_I = \bigcap_{i \in I} X_i$ is an $(n + 1 - |I|)$ -dimensional submanifold of X with a spine of dimension $|I|$.

Fix some such I . Assume, without loss of generality, that I is simple. Then $X_I = (2)$, by Lemma 6.13. Decompose $X_I = \bigcup_z Y_z$ as described in Section 7.1. Lemmas 7.6 and 7.7 imply that Y_1^* is an $(n + 1 - |I|)$ -dimensional 0-handle with no pieces $\xi_r(1)$ of class (A); Lemma 7.9 and the symmetry of the construction imply further that Y_1 is a union of $(n + 1 - |I|)$ -dimensional 0-handles.

For each z , write $S(z) = \{r : \xi_r(z) \text{ is of class (A)}\}$. Lemmas 7.6–7.8 imply that attaching Y_z^* to $\bigcup_{w < z} Y_w$ amounts to attaching an $(n + 1 - |I|)$ -dimensional h -handle, where $h(z)$ is the sum of the dimensions of those $\xi_r(z)$ of class (A):

$$h(z) = \sum_{r \in S(z)} \dim(\xi_r(z)) \leq |I|.$$

Lemma 7.9 and the symmetry of the construction imply further that attaching all of Y_z to $\bigcup_{w < z} Y_w$ amounts to attaching several such handles. Thus, X_I is an $(n + 1 - |I|)$ -dimensional $|I|$ -handlebody in T^n .

It remains to check that $X_{\mathbb{Z}_k} = \bigcap_{i \in \mathbb{Z}_k} X_i$ is a closed k -manifold. We know from Lemma 6.13 that $X_{\mathbb{Z}_k}$ is given by (3).

Since $X_{\mathbb{Z}_k \setminus \{k-1\}}$ is a $(k+1)$ -manifold, it suffices to check that $X_{\mathbb{Z}_k}$ equals $\partial X_{\mathbb{Z}_k \setminus \{k-1\}}$, which is the union of those k -faces of the Y_z from the handle decomposition of $X_{\mathbb{Z}_k \setminus \{k-1\}}$ that are not glued to any other Y_w . Case analysis confirms that this union equals the expression from (3). (The reader may find Tables 19–21 useful.) \square

Alternatively, one can construct a handle decomposition of $X_{\mathbb{Z}_k}$ as follows. Cut each unit interval $[i, i + 1]$ into thirds and, for each $i_* \in \mathbb{Z}_k$, further cut $[i_* - \frac{1}{3}, i_*]$ and $[i_*, i_* + \frac{1}{3}]$ into halves. Then, for each $i_* \in \mathbb{Z}_k$, $U^\circ \subset \mathbb{Z}_k$, $U^- \subset \mathbb{Z}_k \setminus U^\circ$, and $U^* \subset (\{i_* + 1\} \cap U^-) \cup (\{i_*\} \setminus (U^\circ \cup U^-))$, define

$$\rho_i = \begin{cases} [i - \frac{2}{3}, i - \frac{1}{3}] & \text{if } i \in U^\circ, \\ [i - 1, i - \frac{2}{3}] & \text{if } i_* + 1 \neq i \in U^-, \\ [i - \frac{1}{3}, i] & \text{if } i_* \neq i \in \mathbb{Z}_k \setminus (U^\circ \cup U^-), \\ [i_*, i_* + \frac{1}{6}] & \text{if } i_* + 1 = i \in U^*, \\ [i_* + \frac{1}{6}, i_* + \frac{1}{3}] & \text{if } i_* + 1 = i \in U^- \setminus U^*, \\ [i_* - \frac{1}{6}, i_*] & \text{if } i_* = i \in U^*, \\ [i_* - \frac{1}{3}, i_* - \frac{1}{6}] & \text{if } i_* = i \in U^+ \setminus U^*, \end{cases}$$

$$X_{\mathbb{Z}_k, i_*, U^\circ, U^-, U^*} = \prod_{i \in \mathbb{Z}_k} \begin{cases} \rho_i \times \{i\} & \text{if } i \neq i_*, \\ \rho_i & \text{if } i = i_* \end{cases}.$$

Order the pieces $X_{\mathbb{Z}_k, i_*, U^\circ, U^-, U^*}$ as Y_z for $z = 1, 2, 3, \dots$ lexicographically according to the following orders on the possibilities for (i_*, U°, U^-, U^*) . Order $\{i_* \in I\}$ and $U^- \subset U^\circ$ arbitrarily. Partially order $\{U^\circ \subset \mathbb{Z}_k\}$ by inclusion, with $U^\circ < U'^\circ$ if $U^\circ \subset U'^\circ$, and extend arbitrarily to a total order. Order the possibilities for U^* the same way. Then

$$\bigcup_{i=1, \dots, k} Y_z = \bigcup_{i_* \in \mathbb{Z}_k} X_{\mathbb{Z}_k, i_*, \mathbb{Z}_k, \emptyset, \emptyset}$$

is a union of 0-handles, and to attach each $Y_z = X_{\mathbb{Z}_k, i_*, U^\circ, U^-, U^*}$ to $\bigcup_{w < z} Y_w$ is to attach a collection of $h(z)$ -handles for $h(z) = k - |U^\circ| - |U^*|$.

We leave the following question open:

Question 6 Are the multisections in Theorem 7.10 smoothable?

That is, for odd n , does T^n (under its standard smooth structure) admit a smooth multisection such that, when one passes to the unique PL structure on T^n , there is a PL homeomorphism $f : T^n \rightarrow T^n$ sending each piece of this smooth multisection to a piece of the multisection from Theorem 7.10?

8 Cubulated manifolds of odd dimension

This section extends [Theorem 7.10](#) to certain cubulated manifolds. Consider a covering space $p: M \rightarrow T^n$, where $n = 2k - 1$. Multisect $T^n = \bigcup_{i \in \mathbb{Z}_k} X_i$ as in [Theorem 7.10](#). Then, by Corollary 17 of [\[9\]](#), $M = \bigcup_{i \in \mathbb{Z}_k} p^{-1}(X_i)$ determines a PL multisection of M . In general, one expects such multisections to be less efficient than those from [Theorem 7.10](#). Also, there seems to be no reason to expect that one can extend the main construction to cubulated odd-dimensional manifolds in general. There is, however, an intermediate case to which our construction does extend.

First, we propose a modest generalization of the usual notion of a cubulation. The generalization is similar to Hatcher's Δ -complexes vis-à-vis simplicial complexes [\[3\]](#). A *cube* is a homeomorphic copy of I^n for some $n \geq 0$, with the usual cell structure; its *faces* are defined in the traditional way.

Consider an arbitrary edge of I^n joining $\vec{a} = (a_1, \dots, a_{i-1}, 0, a_{i+1}, \dots, a_n)$ and $\vec{b} = (a_1, \dots, a_{i-1}, 1, a_{i+1}, \dots, a_n)$. Orient this edge so that it runs from \vec{a} to \vec{b} . Do the same with every edge of the n -cube. Call these the *standard orientations* on the edges of the n -cube. Call a face of I^n *positive* if it contains $\vec{0}$; otherwise it is *negative*, containing $\vec{1} = (1, \dots, 1)$.

Definition 8.1 A \square -complex K is a quotient space of a collection of disjoint cubes obtained by identifying certain faces via PL homeomorphisms.¹⁴ If all of these face identifications glue a positive face of one cube to a negative face of another (not necessarily distinct) cube and respect the standard orientations on all edges, then K is a *directed* \square -complex.

Note that, by definition, a \square -complex comes equipped with a cell structure.

Definition 8.2 A *generalized cubulation* of a manifold M is a PL homeomorphism to a \square -complex. A *directed cubulation* of M is a PL homeomorphism to a directed \square -complex.

In other words, a generalized cubulation of an n -manifold M imposes a cell structure on M in which every n -cell “looks like” an n -cube, and in a directed cubulation, the n -cells are glued in a particularly nice way.

¹⁴Unlike the traditional notion of cubulation, we do not require that these identifications are between faces of *distinct* cubes.

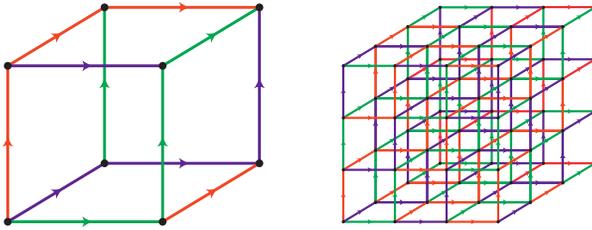


Figure 10: Face identifications (left) for the 3–manifold M from Example 8.5, and a 27 : 1 covering space $T^3 \rightarrow M$.

Example 8.3 The usual cell structure on T^n determines a generalized cubulation, and in fact a directed cubulation, but not a cubulation in the traditional sense.

Let $f : M \rightarrow K$ be a directed cubulation of an n –manifold for $n = 2k - 1$, let

$$g : I^n = [0, k]^n \rightarrow T^n = (\mathbb{R}/k\mathbb{Z})^n = [0, k]^n / \sim$$

be the quotient map, and multisect $T^n = \bigcup_{i \in \mathbb{Z}_k} X_i$ as in Theorem 7.10. Multisect M as follows. For each n –cell C in K , let $h_C : I^n \rightarrow C$ be the identification from K . For each $i \in \mathbb{Z}_k$, define

$$X'_i = \bigcup_{n\text{-cubes } C \text{ in } K} f^{-1}(h_C(g^{-1}(X_i))).$$

Proposition 8.4 With the setup above, $M = \bigcup_{i \in \mathbb{Z}_k} X'_i$ determines a multisection of M .

Proof First consider the case where $p : M \rightarrow T^n$ is a covering space. Let $I \subset \mathbb{Z}_k$ be arbitrary. Construct a handle structure on $X_I \subset T^n$, as in Section 7.1. By construction, each handle is a subset of some open cube $(a, a + k)^n \subset T^n$. Hence, the handle structure on $X_I \subset T^n$ pulls back to a handle structure on $X'_I \subset M$. The general case follows for the same reason, due to the fact that the multisection of T^n is fixed by the permutation action on the indices. \square

Remark In any multisection $M = \bigcup_{i \in \mathbb{Z}_k} X'_i$ from Proposition 8.4, all X'_i have genus $n\#(n\text{-cubes in } K)$. In particular, if $p : M \rightarrow T^n$ is an $r : 1$ covering space, then M has a multisection $M = \bigcup_{i \in \mathbb{Z}_k} X'_i$ in which each X'_i has genus nr .

Example 8.5 Consider the quotient space M obtained from I^3 by identifying the front and right faces, the left and top faces, and the bottom and back faces, all in the way that respects the standard orientations on the edges of I^n ; see Figure 10, left. The natural cell structure on M consists of one vertex, three edges, three faces, and one 3–cell. It is easy to check that the link of the vertex is a 2–sphere, and so M is a 3–manifold. Here

M is geometrically flat, since there is a $27 : 1$ covering space $T^3 \rightarrow M$; see Figure 10, right. But M is not T^3 , since $H_1(M) \cong \mathbb{Z} \oplus \mathbb{Z}_3$. Proposition 8.4 gives a genus-3 Heegaard splitting of M . Does M have an efficient (genus-2) splitting? We leave this as a puzzle for the reader.

Example 8.6 Generalizing Example 8.5, let $n = 2k - 1$ and let $\sigma \in S_n$ be an even permutation. Denote the faces of I^n by F_i^\pm , where $F_i^+ = \{(x_1, \dots, x_n) : x_i = 1\}$ and $F_i^- = \{(x_1, \dots, x_n) : x_i = 0\}$. Identify each F_i^+ with $F_{\sigma(i)}^-$ by identifying each point $(x_1, \dots, 1, \dots, x_n) \in F_i^+$ with $(x_{\sigma^{-1}(1)}, \dots, 0, \dots, x_{\sigma^{-1}(n)}) \in F_{\sigma(i)}^-$, where 1 and 0 are in the i^{th} and $\sigma^{-1}(i)^{\text{th}}$ spots, respectively.

Question 7 For what n and $\sigma \in S_n$ does the construction in Example 8.6 produce a manifold M ? When it is a manifold, is M always distinct from T^n ? Is the multisection of M from Proposition 8.4 ever efficient?

Appendix A Additional handle decomposition tables

Tables 11 and 12 explicitly detail $U_r, V_r \subset I_r$ for arbitrary I_r (following Notation 3.8). For simplicity, these tables have $I_r = I_0 = \{0, \dots, w\}$, listing U_0 and V_0 ; this is not necessarily consistent with Convention 3.9. To adapt $U_0, V_0 \subset I_0$ to the general case $U_r, V_r \subset I_r$, add $\min I_r$ in each coordinate.

Table 13 details the handle decomposition of X_I from T^9 with $I = \{0, 1, 3\} = I_1 \sqcup I_2$ for $I_1 = \{0, 1\}$ and $I_2 = \{3\}$. The interesting feature of this example is how the two blocks of indices I_1 and I_2 interact.

I_0	$0 \notin J$		$0 \in J$	
	U_0	V_0	U_0	V_0
$\{0\}$	\emptyset	\emptyset	\emptyset	$\{0\}$
$\{0,1\}$	\emptyset	$\{1\}$	\emptyset	$\{1\}$
$\{0,1,2\}$	\emptyset	$\{1,2\}$	$\{1\}$	$\{2\}$
$\{0,1,2,3\}$	\emptyset	$\{1,2,3\}$	$\{1,2\}$	$\{3\}$
$\{0,1,2,3,4\}$	\emptyset	$\{1,2,3,4\}$	$\{1,2,3\}$	$\{4\}$
$\{0, \dots, w\}$	\emptyset	$\{1, \dots, w\}$	$\{1, \dots, w - 1\}$	$\{w\}$

Table 11: The index subsets $U_0, V_0 \subset I_0$ when $i_* \notin I_0$.

I_0	U_0	V_0
$\{0\}$	\emptyset	\emptyset
$\{0,1\}$	\emptyset	\emptyset
$\{0,1,2\}$	$\left\{ \begin{array}{l} \{1\} \text{ if } i_* = 2, \\ \emptyset \text{ if } i_* \neq 2 \end{array} \right\}$	$\left\{ \begin{array}{l} \{1, 2\} \text{ if } i_* = 0, \\ \emptyset \text{ if } i_* \neq 0 \end{array} \right\}$
$\{0,1,2,3\}$	$\left\{ \begin{array}{l} \{2\} \text{ if } i_* = 0, \\ \emptyset \text{ if } i_* = 1, \\ \{1\} \text{ if } i_* = 2, \\ \{1, 2\} \text{ if } i_* = 3 \end{array} \right\}$	$\left\{ \begin{array}{l} \{i_* + 1, 3\} \text{ if } i_* \leq 1, \\ \emptyset \text{ if } i_* \geq 2 \end{array} \right\}$
$\{0, \dots, w\}$	$I_0 \setminus \{0, i_*, i_* + 1, w\}$	$\left\{ \begin{array}{l} \{i_* + 1, w\} \text{ if } i_* \leq w - 2, \\ \emptyset \text{ if } i_* \geq w - 1 \end{array} \right\}$

Table 12: The index subsets $U_0, V_0 \subset I_0$ when $i_* \in I_0$.

J	i_*	U	V	V^-	Y_z^*	h	z	glue to
\emptyset	0	\emptyset	\emptyset		$\langle \alpha 1 \beta^3 \rangle \langle 3 \delta^3 \rangle$	0	1	
	1	\emptyset	\emptyset	\emptyset	$\langle 0 \alpha \rangle \beta^3 \langle 3 \delta^3 \rangle$	1	2	1
	3	\emptyset	$\{1\}$	\emptyset	$0 \langle \alpha^+ 1 \beta^3 \rangle \delta^3$	0	3	
$\{0\}$				$\{1\}$	$\langle 0 \alpha^- \rangle \langle 1 \beta^3 \rangle \delta^3$	1	4	3
	0	\emptyset	\emptyset	\emptyset	$\langle \alpha 1 \beta^3 \rangle \langle 3 \delta^2 \rangle \varepsilon$	1	5	1,3,4
	1	\emptyset	\emptyset	\emptyset	$\langle \varepsilon 0 \alpha \rangle \beta^3 \langle 3 \delta^2 \rangle$	2	6	2,5
	3	\emptyset	$\{1\}$	\emptyset	$\langle \varepsilon 0 \rangle \langle \alpha^+ 1 \beta^3 \rangle \delta^2$	1	7	3
				$\{1\}$	$\langle \varepsilon 0 \alpha^- \rangle \langle 1 \beta^3 \rangle \delta^2$	2	8	4,7
$\{3\}$	0	\emptyset	$\{3\}$	\emptyset	$\langle \alpha 1 \beta^2 \rangle \langle \gamma^+ 3 \delta^3 \rangle$	0	9	
				$\{3\}$	$\langle \alpha 1 \beta^2 \rangle \gamma^- \langle 3 \delta^3 \rangle$	1	10	1,9
	1	\emptyset	$\{3\}$	\emptyset	$\langle 0 \alpha \rangle \beta^2 \langle \gamma^+ 3 \delta^3 \rangle$	1	11	9
				$\{3\}$	$\langle 0 \alpha \rangle \beta^2 \gamma^- \langle 3 \delta^3 \rangle$	2	12	2,10,11
	3	\emptyset	$\{1\}$	\emptyset	$0 \langle \alpha^+ 1 \beta^2 \rangle \gamma \delta^3$	1	13	2,3
$\{0,3\}$				$\{1\}$	$\langle 0 \alpha^- \rangle \langle 1 \beta^2 \rangle \gamma \delta^3$	2	14	2,4,13
	0	\emptyset	$\{3\}$	\emptyset	$\langle \alpha 1 \beta^2 \rangle \langle \gamma^+ 3 \delta^2 \rangle \varepsilon$	1	15	9,13,14
				$\{3\}$	$\langle \alpha 1 \beta^2 \rangle \gamma^- \langle 3 \delta^2 \rangle \varepsilon$	2	16	10,13,14,15
	1	\emptyset	$\{3\}$	\emptyset	$\langle \varepsilon 0 \alpha \rangle \beta^2 \langle \gamma^+ 3 \delta^2 \rangle$	2	17	11,15
				$\{3\}$	$\langle \varepsilon 0 \alpha \rangle \beta^2 \gamma^- \langle 3 \delta^2 \rangle$	3	18	6,12,16,17
	3	\emptyset	$\{1\}$	\emptyset	$\langle \varepsilon 0 \rangle \langle \alpha^+ 1 \beta^2 \rangle \gamma \delta^2$	2	19	6,7,13
				$\{1\}$	$\langle \varepsilon 0 \alpha^- \rangle \langle 1 \beta^2 \rangle \gamma \delta^2$	3	20	6,8,14,19

Table 13: A genus-9 quintisection of T^9 : X_I when $I = \{0, 1, 3\}$.

i_*	U	V	V^-	Y_z^*	h	z	glue to
0	{2}	{1,3}	\emptyset	$\alpha^-1\beta_3^{\circ}2\langle\gamma^+3\delta^3\rangle$	0	1	
				$\alpha^-\langle 1\beta_3^- \rangle 2\langle\gamma^+3\delta^3\rangle$	1	2	1
				$\alpha^-1\langle\beta_3^+2\rangle\langle\gamma^+3\delta^3\rangle$	1	3	1
			{1}	$\langle\alpha^+1\rangle\beta_3^{\circ}2\langle\gamma^+3\delta^3\rangle$	1	4	1
				$\langle\alpha^+1\beta_3^- \rangle 2\langle\gamma^+3\delta^3\rangle$	2	5	2,4
				$\langle\alpha^+1\rangle\langle\beta_3^+2\rangle\langle\gamma^+3\delta^3\rangle$	2	6	3,4
			{3}	$\alpha^-1\beta_3^{\circ}\langle 2\gamma^- \rangle \langle 3\delta^3 \rangle$	1	7	1
				$\alpha^-\langle 1\beta_3^- \rangle \langle 2\gamma^- \rangle \langle 3\delta^3 \rangle$	2	8	2,7
				$\alpha^-1\langle\beta_3^+2\gamma^- \rangle \langle 3\delta^3 \rangle$	2	9	3,7
			{1,3}	$\langle\alpha^+1\rangle\beta_3^{\circ}\langle 2\gamma^- \rangle \langle 3\delta^3 \rangle$	2	10	4,7
				$\langle\alpha^+1\beta_3^- \rangle \langle 2\gamma^- \rangle \langle 3\delta^3 \rangle$	3	11	5,8,10
				$\langle\alpha^+1\rangle\langle\beta_3^+2\gamma^- \rangle \langle 3\delta^3 \rangle$	3	12	6,9,10
1	\emptyset	{2,3}	\emptyset	$\langle 0\alpha \rangle \beta^-2\langle\gamma^+3\delta^3\rangle$	1	13	1,2
			{2}	$\langle 0\alpha \rangle \langle \beta^+2 \rangle \langle \gamma^+3\delta^3 \rangle$	2	14	1,3,13
			{3}	$\langle 0\alpha \rangle \beta^- \langle 2\gamma^- \rangle \langle 3\delta^3 \rangle$	2	15	7,8,13
			{2,3}	$\langle 0\alpha \rangle \langle \beta^+2\gamma^- \rangle \langle 3\delta^3 \rangle$	3	16	7,9,14,15
2	{1}	\emptyset	\emptyset	$0\alpha_3^{\circ}\langle 1\beta \rangle \langle \gamma 3\delta^3 \rangle$	1	17	13
				$\langle 0\alpha_3^- \rangle \langle 1\beta \rangle \langle \gamma 3\delta^3 \rangle$	2	18	13,17
				$0\langle \alpha_3^+1\beta \rangle \langle \gamma 3\delta^3 \rangle$	2	19	13,17
3	{1,2}	\emptyset	\emptyset	$0\alpha_3^{\circ}1\beta_3^{\circ}\langle 2\gamma \rangle \delta^3$	1	20	17
				$\langle 0\alpha_3^- \rangle 1\beta_3^{\circ}\langle 2\gamma \rangle \delta^3$	2	21	18,20
				$0\langle \alpha_3^+1 \rangle \beta_3^{\circ}\langle 2\gamma \rangle \delta^3$	2	22	19,20
				$0\alpha_3^{\circ}\langle 1\beta_3^- \rangle \langle 2\gamma \rangle \delta^3$	2	23	17,20
				$\langle 0\alpha_3^- \rangle \langle 1\beta_3^- \rangle \langle 2\gamma \rangle \delta^3$	3	24	18,20,23
				$0\langle \alpha_3^+1\beta_3^- \rangle \langle 2\gamma \rangle \delta^3$	3	25	19,22,23
				$0\alpha_3^{\circ}1\langle \beta_3^+2\gamma \rangle \delta^3$	2	26	17,20
				$\langle 0\alpha_3^- \rangle 1\langle \beta_3^+2\gamma \rangle \delta^3$	3	27	18,21,26
	$0\langle \alpha_3^+1 \rangle \langle \beta_3^+2\gamma \rangle \delta^3$	3	28	19,22,26			

Table 14: X_I for $I = \{0, 1, 2, 3\}$, from T^9 . Part 1: $J = \emptyset$.

Tables 14 and 15 detail the handle decomposition of X_I for $I = \{0, 1, 2, 3\}$ from the quintisection of T^9 . Note that, since $I = I_1$ consists of a single block in this example, we always have $I_1 = I_*$.

Tables 16 and 17 detail handle decompositions of X_I for $I = \{0, 1, 2, 4\}$ from the sexa-section of T^{11} . The parts of these tables with $i_* = 4$ and $0 \notin J$ feature a complication

i_*	U	V	V^-	Y_z^*	h	z	glue to
0	{2}	{1,3}	\emptyset	$\alpha^-1\beta_3^\circ 2\langle\gamma+3\delta^2\rangle\epsilon$	1	29	1,19,20
				$\alpha^-1\beta_3^- 2\langle\gamma+3\delta^2\rangle\epsilon$	2	30	2,22,23,29
				$\alpha^-1\beta_3^+ 2\langle\gamma+3\delta^2\rangle\epsilon$	2	31	3,25,26,29
			{1}	$\langle\alpha+1\rangle\beta_3^\circ 2\langle\gamma+3\delta^2\rangle\epsilon$	2	32	4,19,21
				$\langle\alpha+1\rangle\beta_3^- 2\langle\gamma+3\delta^2\rangle\epsilon$	3	33	5,22,24,30,32
				$\langle\alpha+1\rangle\beta_3^+ 2\langle\gamma+3\delta^2\rangle\epsilon$	3	34	6,25,27,31,32
			{3}	$\alpha^-1\beta_3^\circ\langle 2\gamma^- \rangle\langle 3\delta^2 \rangle\epsilon$	2	35	7,19,20,29
				$\alpha^-1\beta_3^- \langle 2\gamma^- \rangle\langle 3\delta^2 \rangle\epsilon$	3	36	8,22,23,30,35
				$\alpha^-1\beta_3^+ \langle 2\gamma^- \rangle\langle 3\delta^2 \rangle\epsilon$	3	37	9,25,26,31,35
			{1,3}	$\langle\alpha+1\rangle\beta_3^\circ\langle 2\gamma^- \rangle\langle 3\delta^2 \rangle\epsilon$	3	38	10,19,21,32,35
				$\langle\alpha+1\rangle\beta_3^- \langle 2\gamma^- \rangle\langle 3\delta^2 \rangle\epsilon$	4	39	11,22,24,33,36,38
				$\langle\alpha+1\rangle\beta_3^+ \langle 2\gamma^- \rangle\langle 3\delta^2 \rangle\epsilon$	4	40	12,25,27,34,37,38
1	\emptyset	{2,3}	\emptyset	$\langle\epsilon 0\alpha\rangle\beta^- 2\langle\gamma+3\delta^2\rangle$	2	41	13,29,30
				$\langle\epsilon 0\alpha\rangle\langle\beta+2\rangle\langle\gamma+3\delta^2\rangle$	3	42	14,29,31,41
			{3}	$\langle\epsilon 0\alpha\rangle\beta^- \langle 2\gamma^- \rangle\langle 3\delta^2 \rangle$	3	43	15,35,36,41
				$\langle\epsilon 0\alpha\rangle\langle\beta+2\rangle\langle 3\delta^2 \rangle$	4	44	16,35,37,42,43
2	{1}	\emptyset	\emptyset	$\langle\epsilon 0\rangle\alpha_3^\circ\langle 1\beta \rangle\langle\gamma 3\delta^2\rangle$	2	45	17,41,43
				$\langle\epsilon 0\alpha_3^- \rangle\langle 1\beta \rangle\langle\gamma 3\delta^2\rangle$	3	46	18,41,43,45
				$\langle\epsilon 0\rangle\langle\alpha_3^+ 1\beta \rangle\langle\gamma 3\delta^2\rangle$	3	47	19,41,43,45
3	{1,2}	\emptyset	\emptyset	$\langle\epsilon 0\rangle\alpha_3^\circ 1\beta_3^\circ\langle 2\gamma \rangle\delta^2$	2	48	20,45
				$\langle\epsilon 0\alpha_3^- \rangle 1\beta_3^\circ\langle 2\gamma \rangle\delta^2$	3	49	21,46,48
				$\langle\epsilon 0\rangle\langle\alpha_3^+ 1\rangle\beta_3^\circ\langle 2\gamma \rangle\delta^2$	3	50	22,47,48
				$\langle\epsilon 0\rangle\alpha_3^\circ\langle 1\beta_3^- \rangle\langle 2\gamma \rangle\delta^2$	3	51	23,45,48
				$\langle\epsilon 0\alpha_3^- \rangle\langle 1\beta_3^- \rangle\langle 2\gamma \rangle\delta^2$	4	52	24,46,49,51
				$\langle\epsilon 0\rangle\langle\alpha_3^+ 1\beta_3^- \rangle\langle 2\gamma \rangle\delta^2$	4	53	25,47,50,51
				$\langle\epsilon 0\rangle\alpha_3^\circ 1\langle\beta_3^+ 2\rangle\delta^2$	3	54	26,45,48
				$\langle\epsilon 0\alpha_3^- \rangle 1\langle\beta_3^+ 2\rangle\delta^2$	4	55	27,46,49,54
				$\langle\epsilon 0\rangle\langle\alpha_3^+ 1\rangle\langle\beta_3^+ 2\rangle\delta^2$	4	56	28,47,50,54

Table 15: X_I for $I = \{0, 1, 2, 3\}$, from T^9 . Part 2: $J = \{0\}$.

that does not appear in dimensions $n \leq 9$. Also see Tables 9 and 18 for more complicated examples of this pattern.

Table 18 details the start of the handle decomposition of X_I from T^{15} with $I = \{0, 1, 2, 3, 4, 6\}$, focusing on the first few pieces Y_z . Those pieces have $J = \emptyset$, $i_* = 6$,

J	i_*	U	V	V^-	Y_z^*	h	z	glue to
\emptyset	4	\emptyset	{1,2}	\emptyset	$0\langle\alpha^+1\rangle\langle\beta^+2\gamma^3\rangle\epsilon^3$	0	1	
				{1}	$\langle 0\alpha^- \rangle 1\langle\beta^+2\gamma^3\rangle\epsilon^3$	1	2	1
				{1,2}	$\langle 0\alpha^- \rangle \langle 1\beta^- \rangle \langle 2\gamma^3 \rangle \epsilon^3$	1	3	2
				{2}	$0\langle\alpha^+1\beta^- \rangle \langle 2\gamma^3 \rangle \epsilon^3$	2	4	1,3
	0	\emptyset	{1,2}	\emptyset	$\alpha^-1\langle\beta^+2\gamma^3\rangle\langle 4\epsilon^3 \rangle$	0	5	
				{1}	$\langle\alpha^+1\rangle\langle\beta^+2\gamma^3\rangle\langle 4\epsilon^3 \rangle$	1	6	5
				{2}	$\alpha^- \langle 1\beta^- \rangle \langle 2\gamma^3 \rangle \langle 4\epsilon^3 \rangle$	1	7	5
				{1,2}	$\langle\alpha^+1\beta^- \rangle \langle 2\gamma^3 \rangle \langle 4\epsilon^3 \rangle$	2	8	5,6
	1	\emptyset	\emptyset	\emptyset	$\langle 0\alpha \rangle \langle \beta 2\gamma^3 \rangle \langle 4\epsilon^3 \rangle$	1	9	5,7
	2	{1}	\emptyset	\emptyset	$0\alpha_3^\circ \langle 1\beta \rangle \gamma^3 \langle 4\epsilon^3 \rangle$	1	10	9
					$\langle 0\alpha_3^- \rangle \langle 1\beta \rangle \gamma^3 \langle 4\epsilon^3 \rangle$	2	11	9,10
					$0\langle\alpha_3^+1\beta \rangle \gamma^3 \langle 4\epsilon^3 \rangle$	2	12	9,10
{4}	4	\emptyset	{1,2}	\emptyset	$0\langle\alpha^+1\rangle\langle\beta^+2\gamma^2\rangle\delta\epsilon^3$	1	13	1,10,12
				{1}	$\langle 0\alpha^- \rangle 1\langle\beta^+2\gamma^2\rangle\delta\epsilon^3$	2	14	2,10,11,13
				{1,2}	$\langle 0\alpha^- \rangle \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta\epsilon^3$	2	15	3,10,11,14
				{2}	$0\langle\alpha^+1\beta^- \rangle \langle 2\gamma^2 \rangle \delta\epsilon^3$	3	16	4,10,12,13,15
	0	\emptyset	{1,2}	\emptyset	$\alpha^-1\langle\beta^+2\gamma^2\rangle\langle\delta^+4\epsilon^3\rangle$	0	17	
					$\alpha^-1\langle\beta^+2\gamma^2\rangle\delta^- \langle 4\epsilon^3 \rangle$	1	18	5,17
				{1}	$\langle\alpha^+1\rangle\langle\beta^+2\gamma^2\rangle\langle\delta^+4\epsilon^3\rangle$	1	19	17
					$\langle\alpha^+1\rangle\langle\beta^+2\gamma^2\rangle\delta^- \langle 4\epsilon^3 \rangle$	2	20	6,18,19
				{2}	$\alpha^- \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \langle \delta^+ 4\epsilon^3 \rangle$	1	21	19
					$\alpha^- \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta^- \langle 4\epsilon^3 \rangle$	2	22	7,20,21
				{1,2}	$\langle\alpha^+1\beta^- \rangle \langle 2\gamma^2 \rangle \langle\delta^+4\epsilon^3\rangle$	2	23	19,21
					$\langle\alpha^+1\beta^- \rangle \langle 2\gamma^2 \rangle \delta^- \langle 4\epsilon^3 \rangle$	3	24	8,20,22,23
	1	\emptyset	{4}	\emptyset	$\langle 0\alpha \rangle \langle \beta 2\gamma^2 \rangle \langle \delta^+ 4\epsilon^3 \rangle$	1	25	17,21
				{4}	$\langle 0\alpha \rangle \langle \beta 2\gamma^2 \rangle \delta^- \langle 4\epsilon^3 \rangle$	2	26	9,18,22,25
	2	{1}	{4}	\emptyset	$0\alpha_3^\circ \langle 1\beta \rangle \gamma^2 \langle \delta^+ 4\epsilon^3 \rangle$	1	27	25
					$\langle 0\alpha_3^- \rangle \langle 1\beta \rangle \gamma^2 \langle \delta^+ 4\epsilon^3 \rangle$	2	28	25,27
				$0\langle\alpha_3^+1\beta \rangle \gamma^2 \langle \delta^+ 4\epsilon^3 \rangle$	2	29	25,27	
{4}				$0\alpha_3^\circ \langle 1\beta \rangle \gamma^2 \delta^- \langle 4\epsilon^3 \rangle$	2	30	10,26,27	
				$\langle 0\alpha_3^- \rangle \langle 1\beta \rangle \gamma^2 \delta^- \langle 4\epsilon^3 \rangle$	3	31	11,26,28,30	
				$0\langle\alpha_3^+1\beta \rangle \gamma^2 \delta^- \langle 4\epsilon^3 \rangle$	3	32	12,26,29,30	

Table 16: Part 1 of X_I for $I = \{0, 1, 2, 4\}$, from T^{11} .

J	i_*	U	V	V^-	Y_z^*	h	z	glue to				
{0}	4	{1}	{2}	\emptyset	$\langle \zeta 0 \rangle \alpha_3^0 1 \langle \beta + 2\gamma^3 \rangle \varepsilon^2$	1	33	1,2				
					$\langle \zeta 0 \rangle \langle \alpha_3^+ 1 \rangle \langle \beta + 2\gamma^3 \rangle \varepsilon^2$	2	34	1,33				
					$\langle \zeta 0 \alpha_3^- \rangle 1 \langle \beta + 2\gamma^3 \rangle \varepsilon^2$	2	35	2,33				
				{2}	$\langle \zeta 0 \rangle \alpha_3^0 \langle 1\beta^- \rangle \langle 2\gamma^3 \rangle \varepsilon^2$	2	36	3,4,33				
					$\langle \zeta 0 \rangle \langle \alpha_3^+ 1\beta^- \rangle \langle 2\gamma^3 \rangle \varepsilon^2$	3	37	3,34,36				
					$\langle \zeta 0 \alpha_3^- \rangle \langle 1\beta^- \rangle \langle 2\gamma^3 \rangle \varepsilon^2$	3	38	4,35,36				
	0	\emptyset	{1,2}	\emptyset	$\zeta \alpha^- 1 \langle \beta + 2\gamma^3 \rangle \langle 4\varepsilon^2 \rangle$	1	39	2,5				
					$\zeta \langle \alpha^+ 1 \rangle \langle \beta + 2\gamma^3 \rangle \langle 4\varepsilon^2 \rangle$	2	40	1,6,39				
					$\zeta \alpha^- \langle 1\beta^- \rangle \langle 2\gamma^3 \rangle \langle 4\varepsilon^2 \rangle$	2	41	3,7,39				
				{1,2}	$\zeta \langle \alpha^+ 1\beta^- \rangle \langle 2\gamma^3 \rangle \langle 4\varepsilon^2 \rangle$	3	42	4,8,40				
					$\langle \zeta 0 \alpha \rangle \langle \beta 2\gamma^3 \rangle \langle 4\varepsilon^2 \rangle$	2	43	9,39,41				
					$\langle \zeta 0 \rangle \alpha_3^0 \langle 1\beta \rangle \gamma^3 \langle 4\varepsilon^2 \rangle$	2	44	10,43				
	1	\emptyset	\emptyset	\emptyset	$\langle \zeta 0 \alpha_3^- \rangle \langle 1\beta \rangle \gamma^3 \langle 4\varepsilon^2 \rangle$	3	45	11,43,44				
					$\langle \zeta 0 \rangle \langle \alpha_3^+ 1\beta \rangle \gamma^3 \langle 4\varepsilon^2 \rangle$	3	46	12,43,44				
					2	{1}	\emptyset	\emptyset	$\langle \zeta 0 \rangle \alpha_3^0 \langle 1\beta \rangle \gamma^3 \langle 4\varepsilon^2 \rangle$	2	44	10,43
									$\langle \zeta 0 \alpha_3^- \rangle \langle 1\beta \rangle \gamma^3 \langle 4\varepsilon^2 \rangle$	3	45	11,43,44
									$\langle \zeta 0 \rangle \langle \alpha_3^+ 1\beta \rangle \gamma^3 \langle 4\varepsilon^2 \rangle$	3	46	12,43,44
									{0, 4}	4	{1}	{2}
$\langle \zeta 0 \rangle \langle \alpha_3^+ 1 \rangle \langle \beta + 2\gamma^2 \rangle \delta \varepsilon^2$	3	48	13,34,45,47									
$\langle \zeta 0 \alpha_3^- \rangle 1 \langle \beta + 2\gamma^2 \rangle \delta \varepsilon^2$	3	49	14,35,46,47									
{2}	$\langle \zeta 0 \rangle \alpha_3^0 \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta \varepsilon^2$	3	50	15,16,36,44,47								
	$\langle \zeta 0 \rangle \langle \alpha_3^+ 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta \varepsilon^2$	4	51	16,37,45,48,50								
	$\langle \zeta 0 \alpha_3^- \rangle \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta \varepsilon^2$	4	52	15,38,46,49,50								
0	\emptyset	{1,2}	\emptyset	$\zeta \alpha^- 1 \langle \beta + 2\gamma^2 \rangle \langle \delta + 4\varepsilon^2 \rangle$	1	53	14,17					
				$\zeta \alpha^- 1 \langle \beta + 2\gamma^2 \rangle \delta^- \langle 4\varepsilon^2 \rangle$	2	54	14,18,39,53					
				$\zeta \langle \alpha^+ 1 \rangle \langle \beta + 2\gamma^2 \rangle \langle \delta + 4\varepsilon^2 \rangle$	2	55	13,19,53					
			{1}	$\zeta \langle \alpha^+ 1 \rangle \langle \beta + 2\gamma^2 \rangle \delta^- \langle 4\varepsilon^2 \rangle$	3	56	13,20,40,54,55					
				$\zeta \alpha^- \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \langle \delta + 4\varepsilon^2 \rangle$	2	57	15,21,53					
				$\zeta \alpha^- \langle 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta^- \langle 4\varepsilon^2 \rangle$	3	58	15,22,41,54,57					
1	\emptyset	{4}	\emptyset	$\zeta \langle \alpha^+ 1\beta^- \rangle \langle 2\gamma^2 \rangle \langle \delta + 4\varepsilon^2 \rangle$	3	59	16,23,55,57					
				$\zeta \langle \alpha^+ 1\beta^- \rangle \langle 2\gamma^2 \rangle \delta^- \langle 4\varepsilon^2 \rangle$	4	60	16,24,42,56,58,59					
				$\langle \zeta 0 \alpha \rangle \langle \beta 2\gamma^2 \rangle \langle \delta + 4\varepsilon^2 \rangle$	2	61	25,53,57					
				$\langle \zeta 0 \alpha \rangle \langle \beta 2\gamma^2 \rangle \delta^- \langle 4\varepsilon^2 \rangle$	3	62	26,43,54,58,61					
				2	{1}	{4}	\emptyset	$\langle \zeta 0 \rangle \alpha_3^0 \langle 1\beta \rangle \gamma^2 \langle \delta + 4\varepsilon^2 \rangle$		2	63	27,61
								$\langle \zeta 0 \alpha_3^- \rangle \langle 1\beta \rangle \gamma^2 \langle \delta + 4\varepsilon^2 \rangle$		3	64	28,61,63
$\langle \zeta 0 \rangle \langle \alpha_3^+ 1\beta \rangle \gamma^2 \langle \delta + 4\varepsilon^2 \rangle$	3	65	29,61,63									
{4}	$\langle \zeta 0 \rangle \alpha_3^0 \langle 1\beta \rangle \gamma^2 \delta^- \langle 4\varepsilon^2 \rangle$	3	66					30,44,62,63				
	$\langle \zeta 0 \alpha_3^- \rangle \langle 1\beta \rangle \gamma^2 \delta^- \langle 4\varepsilon^2 \rangle$	4	67					31,44,62,64,66				
	$\langle \zeta 0 \rangle \langle \alpha_3^+ 1\beta \rangle \gamma^2 \delta^- \langle 4\varepsilon^2 \rangle$	4	68					32,45,62,65,66				

Table 17: Part 2 of X_I for $I = \{0, 1, 2, 4\}$, from T^{11} .

V^-	Y_z^*	h	z	glue to
\emptyset	$0\langle\alpha^+1\rangle\langle\beta^+2\rangle\langle\gamma^+3\rangle\langle\delta^+4\varepsilon^3\rangle\eta^3$	0	1	
$\{1\}$	$\langle 0\alpha^- \rangle 1\langle\beta^+2\rangle\langle\gamma^+3\rangle\langle\delta^+4\varepsilon^3\rangle\eta^3$	1	2	1
$\{1, 2\}$	$\langle 0\alpha^- \rangle 1\langle\beta^-2\rangle\langle\gamma^+3\rangle\langle\delta^+4\varepsilon^3\rangle\eta^3$	1	3	2
$\{2\}$	$0\langle\alpha^+1\beta^- \rangle 2\langle\gamma^+3\rangle\langle\delta^+4\varepsilon^3\rangle\eta^3$	2	4	1,3
$\{2, 3\}$	$0\langle\alpha^+1\beta^- \rangle 2\langle\gamma^-3\rangle\langle\delta^+4\varepsilon^3\rangle\eta^3$	1	5	4
$\{1, 2, 3\}$	$\langle 0\alpha^- \rangle 1\langle\beta^-2\rangle\langle\gamma^-3\rangle\langle\delta^+4\varepsilon^3\rangle\eta^3$	2	6	3,5
$\{1, 3\}$	$\langle 0\alpha^- \rangle 1\langle\beta^+2\gamma^- \rangle 3\langle\delta^+4\varepsilon^3\rangle\eta^3$	2	7	2,6
$\{3\}$	$0\langle\alpha^+1\rangle\langle\beta^+2\gamma^- \rangle 3\langle\delta^+4\varepsilon^3\rangle\eta^3$	3	8	1,5,7
$\{3, 4\}$	$0\langle\alpha^+1\rangle\langle\beta^+2\gamma^- \rangle 3\langle\delta^-4\varepsilon^3\rangle\eta^3$	1	9	8
$\{1, 3, 4\}$	$\langle 0\alpha^- \rangle 1\langle\beta^+2\gamma^- \rangle 3\langle\delta^-4\varepsilon^3\rangle\eta^3$	2	10	7,9
$\{1, 2, 3, 4\}$	$\langle 0\alpha^- \rangle 1\langle\beta^-2\rangle\langle\gamma^-3\rangle\langle\delta^-4\varepsilon^3\rangle\eta^3$	2	11	6,10
$\{2, 3, 4\}$	$0\langle\alpha^+1\beta^- \rangle 2\langle\gamma^-3\rangle\langle\delta^-4\varepsilon^3\rangle\eta^3$	3	12	5,9,11
$\{2, 4\}$	$0\langle\alpha^+1\beta^- \rangle 2\langle\gamma^+3\delta^- \rangle 4\varepsilon^3\eta^3$	2	13	4,12
$\{1, 2, 4\}$	$\langle 0\alpha^- \rangle 1\langle\beta^-2\rangle\langle\gamma^+3\delta^- \rangle 4\varepsilon^3\eta^3$	3	14	3,11,13
$\{1, 4\}$	$\langle 0\alpha^- \rangle 1\langle\beta^+2\rangle\langle\gamma^+3\delta^- \rangle 4\varepsilon^3\eta^3$	3	15	2,10,14
$\{4\}$	$0\langle\alpha^+1\rangle\langle\beta^+2\rangle\langle\gamma^+3\delta^- \rangle 4\varepsilon^3\eta^3$	4	16	1,9,13,15

Table 18: Start of the handle decomposition from T^{15} with $I = \{0, 1, 2, 3, 4, 6\}$, $J = \emptyset$, $i_* = 6$, $U = \emptyset$, and $V = \{1, 2, 3, 4\}$.

$U = \emptyset$, and $V = \{1, 2, 3, 4\}$. The interesting feature of this example is the ordering of these pieces. Compare to (23) and Tables 9, 16, and 17.

Tables 19, 20, and 21 list the possible forms for $\xi_r(z)$. Table 19 lists those with no singleton factor. Table 20 lists those with a singleton factor $\{i\}$, where $i \in V^+$ and $i + 1 \in U^\circ \cup U^+ \cup V^+$, or $i \in U^- \cup U^\circ \cup V^-$ and $i + 1 \in V^-$; the class of this case

class	$\xi_r(z)$	conditions
(A)	$[i_* - 1, i_*]$	$i_* \in J$
(A)	$[i - 1, i - \frac{1}{2}]$	$i \in J \cap V^- \implies i \neq i_*, i + 1 \notin I$
(B)	$[i_*, i_* + \frac{1}{2}]$	$a \leq i_* \leq b - 2, i_* + 1 \in V^-$
(B)	$[i - \frac{2}{3}, i - \frac{1}{3}]$	$i \in U^\circ$
(B)	$\prod_{j=i_*+1}^{c-1} [i_*, j]^2$	$i_* = b, c \in J$
(B)	$\prod_{j=i_*+1}^{c-2} [i_*, j]^2 [i_*, c - 1]^3$	$i_* = b, c \notin J$

Table 19: The possible forms for $\xi_r(z)$ with no singleton factor, where $i_* \in I_s$, $a = \min I_s$, $b = \max I_s$, and $c = \min I_{s+1}$.

class	$\xi_r(z)$	conditions on i	conditions on $i + 1$	parity
(A)	$[i - \frac{1}{2}, i]\{i\}$	$i \in V^+$	$i + 1 \in U^\circ \cup U^+ \cup V^+$	odd
(A)	$\{i\}[i, i + \frac{1}{2}]$	$i \in U^- \cup U^\circ \cup V^-$	$i + 1 \in V^-$	odd
(A)	$[i - \frac{1}{2}, i]\{i\}[i, i + \frac{1}{2}]$	$i \in V^+$	$i + 1 \in V^-$	odd
(B)	$[i - \frac{1}{2}, i]\{i\}$	$i \in V^+$	$i + 1 \in U^\circ \cup U^+ \cup V^+$	even
(B)	$\{i\}[i, i + \frac{1}{2}]$	$i \in U^- \cup U^\circ \cup V^-$	$i + 1 \in V^-$	even
(B)	$[i - \frac{1}{2}, i]\{i\}[i, i + \frac{1}{2}]$	$i \in V^+$	$i + 1 \in V^-$	even

Table 20: The possible forms for each $\xi_r(z)$ containing a singleton factor $\{i\}$, where $i \in V^+$ and $i + 1 \in U^\circ \cup U^+ \cup V^+$, or $i \in U^- \cup U^\circ \cup V^-$ and $i + 1 \in V^-$; the class depends on the parity of $\#(V^- \cap \{i + 1, \dots, \max I_s\})$, where $i \in I_s$.

depends on the parity of $\#(V^- \cap \{i + 1, \dots, \max I_s\})$, where $i \in I_s$. Table 21 lists the remaining possibilities for $\xi_r(z)$.

class	$\xi_r(z)$	conditions on i
(A)	$[i - 1, i]\{i\}$	$i \in J, i + 1 \in U^\circ \cup U^+ \cup V^+$
(A)	$[i - 1, i]\{i\}[i, i + 1]$	$i \in J, i_* = i + 1$
(A)	$[i - 1, i]\{i\}[i, i + \frac{1}{3}]$	$i \in J, i + 1 \in U^-$
(A)	$[i - 1, i]\{i\}[i, i + \frac{1}{2}]$	$i \in J, i + 1 \in V^-$
(A)	$[i - \frac{1}{3}, i]\{i\}$	$i \in U^+, i + 1 \in U^\circ \cup U^+ \cup V^+$
(A)	$\{i\}[i, i + \frac{1}{3}]$	$i + 1 \in U^-, i \in U^- \cup U^\circ \cup V^-$
(A)	$[i - \frac{1}{3}, i]\{i\}[i, i + \frac{1}{3}]$	$i \in U^+, i + 1 \in U^-$
(A)	$[i - \frac{1}{3}, i]\{i\}[i, i + \frac{1}{2}]$,	$i \in U^+, i + 1 \in V^-$ $\implies i + 1 = \max I_s \neq i_*$
(A)	$[i - \frac{1}{2}, i]\{i\}[i, i + \frac{1}{3}]$,	$i \in V^+, i + 1 \in U^-$ $\implies i = i_* + 1 \leq \max I_s - 1$
(A)	$\{i\}$	$i \in (T \setminus J) \cup U^- \cup U^\circ \cup V^-$, $i + 1 \in U^\circ \cup U^+ \cup V^+$
(B)	$[i - 1, i]\{i\} \prod_{j=i+1}^{c-2} [i, j]^2 [i, c - 1]^q$	$i_* = \min I_s = i - 1 = \max I_s - 1$
(B)	$[i - \frac{1}{2}, i]\{i\} \prod_{j=i+1}^{c-2} [i, j]^2 [i, c - 1]^q$	$i = \max I_s \in V^+$
(B)	$\{i\} \prod_{j=i+1}^{c-2} [i, j]^2 [i, c - 1]^q$	$i = \max I_s \in V^-$

Table 21: The possible forms for each $\xi_r(z)$ not listed in Table 19 or 20. Each contains a singleton factor $\{i\}$ for $i_* \neq i \in I_s$ with $s \in \mathbb{Z}_m$. Write $c = \min I_{s+1}$ with $q \in \{2, 3\}$.

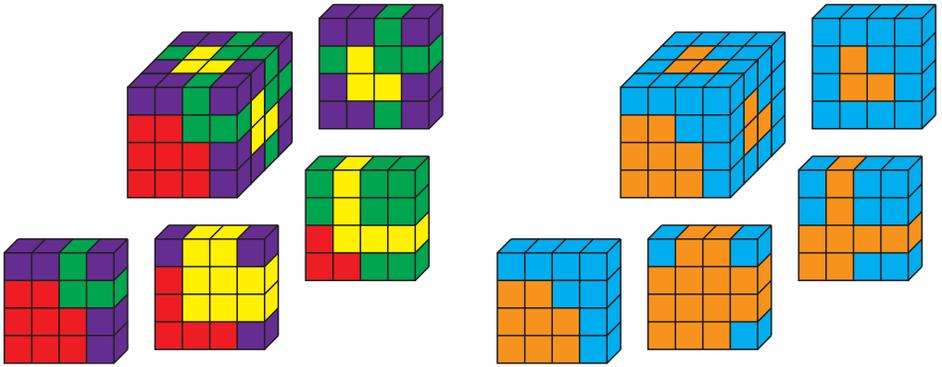


Figure 11: Another construction of the minimal-genus Heegaard splitting of S^3 .

Appendix B Four other attempts to multisect T^n for n odd

From the handle decomposition

The n -torus has a natural handle decomposition, with $\binom{n}{h}$ h -handles for each $h = 0, \dots, n$, which one can construct as follows. View T^n as $(\mathbb{R}/2\mathbb{Z})^n$, and decompose it into the 2^n subcubes with vertices in $(\mathbb{Z}/2\mathbb{Z})^n$. Then, using Notation 3.4, for each $h = 0, \dots, n$, the h -handles are the subcubes which are permutations of $\alpha^{n-h} \beta^h$.¹⁵

One might hope that $X_i = \langle \alpha^{n-i} \beta^i \rangle \cup \langle \alpha^{n+1-i} \beta^{i-1} \rangle$ determines a multisection.¹⁶ Indeed, in dimension 3 this is the Heegaard splitting shown in Figure 1. Yet, the construction does not work beyond dimension 3, as one can see by noting, eg, that $X_0 \cap X_{k-1} = \bigcup_{r=0}^{n-2} \langle \alpha \beta 0^r 1^{n-2-r} \rangle$ is always 2-dimensional.

By gluing pairs of balls

Instead, one might attempt to generalize the following construction; see Figure 11. View T^n as $(\mathbb{R}/2k\mathbb{Z})^n = [0, 2k]^n / \sim$. Partition the $(2k)^n$ unit cubes with vertices in the lattice $(\mathbb{Z}/2k\mathbb{Z})^n$ so as to form V_0, \dots, V_n subject to the following conditions:¹⁷

- If $\vec{x} \in V_0$, then $\vec{x} + (r, \dots, r) \in V_r$.
- The permutation action on the indices fixes each V_r .
- V_0 contains $[0, 1]^n$, is star-shaped about $(0, \dots, 0)$, and contains no points with any coordinate in $(n - 1, n)$.

¹⁵This handle decomposition is optimal in the sense that it has the minimum possible number of handles of each index, since $H_h(T^n)$ has rank $\binom{n}{h}$.

¹⁶Note that n is odd throughout Appendix B.

¹⁷These conditions uniquely determine V_0, \dots, V_n .

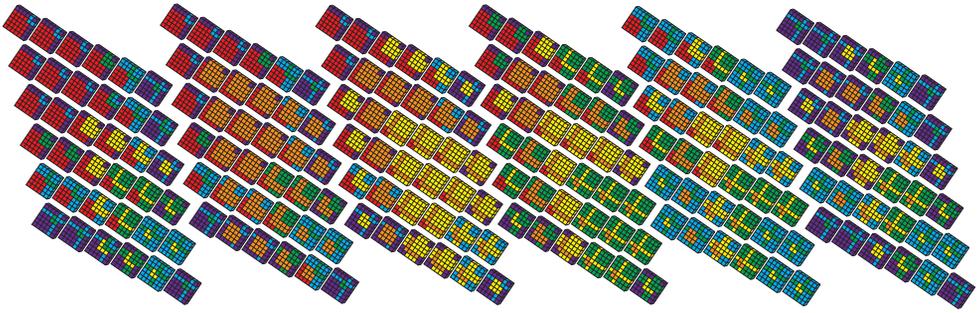


Figure 12: Decomposing $T^5 = [0, 6]^5/\sim$ as $V_0 \cup \dots \cup V_5$. Does $(V_0 \cup V_1, V_2 \cup V_3, V_4 \cup V_5)$ determine a trisection?

Then, for $i = 0, \dots, k = \frac{1}{2}(n + 1)$, let $X_i = V_{2i} \cup V_{2i+1}$. Figure 11 shows that this construction does in fact give a genus-3 Heegaard splitting of T^3 .

In higher dimensions this construction is promising for many of the same reasons as the construction behind Theorem 7.10. This construction has at least one additional advantage, namely that each V_i is a ball. This makes it easy to check that each X_i is indeed an n -dimensional handlebody of genus n . Unfortunately, the complexity of this construction grows much more rapidly than the construction behind Theorem 7.10, making it hard to check the other details, even in dimension 5. Indeed, see Figure 12.

Question 8 Does this construction also give a trisection of T^5 ? Does it give a multisection of T^n for arbitrary $n = 2k - 1$?

By summing coordinates

As shown in Figure 13, the genus-3 Heegaard splitting of $T^3 = [0, 2]^3/\sim$ can be constructed as $T^3 = X_0 \cup X_1$, where

$$X_i = \{(x_1, x_2, x_3) : 3i \leq x_1 + x_2 + x_3 \leq 3(i + 1)\}/\sim.$$

The splitting surface consists of the hexagon $\{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 3\}/\sim$ together with three other hexagons. One is $\{(0, x_2, x_3) : 1 \leq x_2 + x_3 \leq 5\}/\sim$, and the others are obtained from this one by permuting coordinates. A cocore of one 1-handle in X_0 is the triangle $\{(0, x_2, x_3) : x_2 + x_3 \leq 1\}/\sim$, and a cocore of a 1-handle in X_0 is the triangle $\{(0, x_2, x_3) : 5 \leq x_2 + x_3\}/\sim$; the other 1-handles of X_0 and X_1 are related to these by permuting coordinates.

One might attempt to trisect $T^5 = [0, 3]^5/\sim$ as $T^5 = X_0 \cup X_1 \cup X_2$ with

$$X_i = \{(x_1, \dots, x_5) : 5i \leq x_1 + \dots + x_5 \leq 5(i + 1)\}/\sim.$$

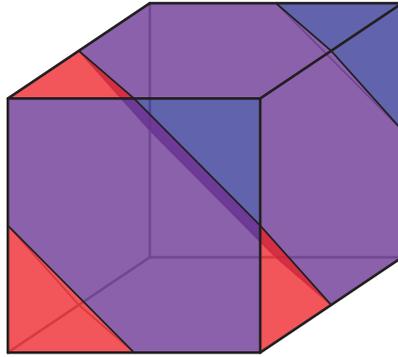


Figure 13: The efficient Heegaard splitting $T^3 = X_0 \cup X_1$ constructed by summing coordinates. Four purple hexagons form the splitting surface. Red and blue triangles are cocores of the 1–handles in X_0 and X_1 , respectively.

Then each X_i is in fact a 4–dimensional 1–handlebody of genus 4: a cocore of a 1–handle of X_0 is the 4–simplex $\{(0, x_2, x_3, x_4, x_5) : x_2 + x_3 + x_4 + x_5 \leq 2\}/\sim$, a cocore of a 1–handle of X_1 is $\{(0, x_2, x_3, x_4, x_5) : 5 \leq x_2 + x_3 + x_4 + x_5 \leq 7\}/\sim$, and a cocore of a 1–handle of X_2 is $\{(0, x_2, x_3, x_4, x_5) : 12 \leq x_2 + x_3 + x_4 + x_5\}/\sim$. The other 1–handles of X_0, X_1 , and X_2 are related to these by permuting coordinates.

Yet, this is not a trisection, because

$$X_0 \cap X_2 = \{(x_1, x_2, x_3, 0, 0)_\sigma : 4 \leq x_1 + x_2 + x_3 \leq 5, \sigma \in S_5\}/\sim$$

is 3–dimensional, not 4–dimensional.

To fix this problem, one could choose $0 = a_0 < a_1 < a_2 < a_3 = 15$ differently and define

$$X_i = \{(x_1, \dots, x_5) : a_i \leq x_1 + \dots + x_5 \leq a_{i+1}\}.$$

Then $X_0 \cap X_2$ will be 4–dimensional if and only if $a_2 - a_1 < 3$. This creates a new problem: if $a_2 - a_1 < 3$, then X_1 is contractible, and hence a 5–ball. It now follows from Proposition 2.6 that no choice of a_1 and a_2 produces a trisection of T^5 . The same difficulty prevails in all other dimensions $n > 3$ (including even dimensions).

Using the symmetric space T^n/S_n

Given a triangulation K of an n –manifold X , Rubinstein and Tillmann multisect X by mapping each n –simplex of K to the standard $(k-1)$ –simplex

$$(31) \quad \Delta_{k-1} = [\vec{v}_0, \dots, \vec{v}_{k-1}] = \left\{ \sum_{j \in \mathbb{Z}_k} a_j \vec{v}_j : 0 \leq a_j, \sum_{j \in \mathbb{Z}_k} a_j = 1 \right\},$$

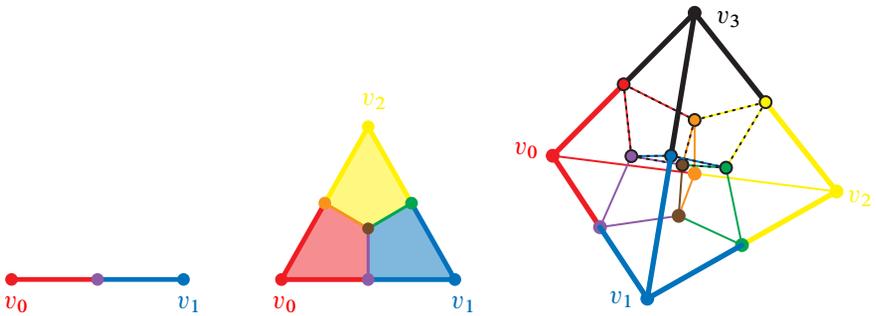


Figure 14: The decompositions $\Delta_{k-1} = \bigcup_{i \in \mathbb{Z}_k} Z_i$ of the 1-, 2-, and 3-simplices following Rubinstein and Tillmann.

decomposing $\Delta_{k-1} = \bigcup_{i \in \mathbb{Z}_k} Z_i$, where each

$$(32) \quad Z_i = \{\vec{x} \in \Delta_{k-1} : |\vec{x} - \vec{v}_i| \leq |\vec{x} - \vec{v}_j| \text{ for all } j \in \mathbb{Z}_k\}$$

(see Figure 14), and pulling back. Their maps from the n -simplices of K to Δ_{k-1} are simplest to construct in odd dimension $n = 2k - 1$. Namely,

- map the barycenter of each r -face to $\vec{v}_j \in \Delta_{k-1}$ for $j = 2r, 2r + 1$; and
- extend linearly in the first barycentric subdivision of K .

The even-dimensional case is similar, but with an extra move.

For example, the triangulation of S^3 with two 3-simplices gives a genus-3 Heegaard splitting, as shown in Figure 15.

Following Rubinstein and Tillmann, one might try to construct a, say PL, multisection of T^n using the symmetric space T^n/S_n , which is homeomorphic to a disk-bundle over the circle; this bundle is twisted when n is even and untwisted when n is odd.

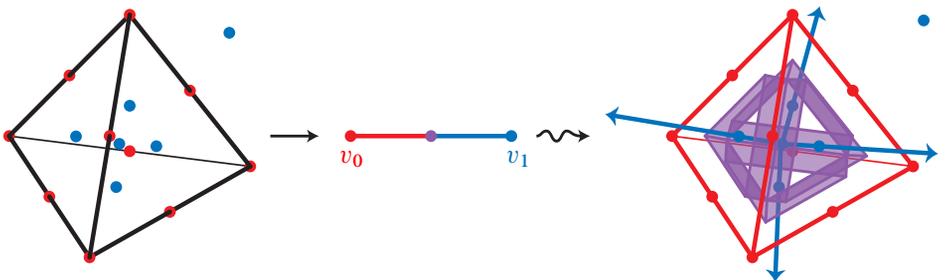


Figure 15: A genus-3 Heegaard splitting (right) of S^3 , following Rubinstein and Tillmann’s construction.

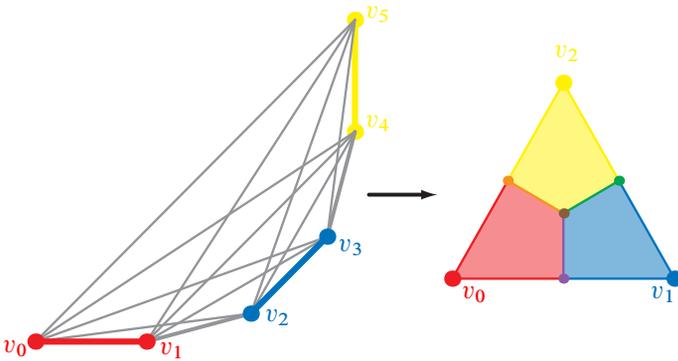


Figure 16: Try viewing T^n/S_n as Δ_n/\sim and Δ_n as an iterated join of k intervals. Then map $\Delta_n \rightarrow \Delta_{k-1}$, decompose Δ_{k-1} , and pull back. It fails, even for $n = 5$, shown.

One can also view the symmetric space T^n/S_n as an n -simplex $\Delta_n = [\vec{v}_0, \dots, \vec{v}_n]$ with certain faces identified. When $n = 2k - 1$, one can also view Δ_n as an iterated join of intervals,

$$\Delta_n = [\vec{v}_0, \vec{v}_1] * \dots * [\vec{v}_{2(k-1)}, \vec{v}_{2k-1}].$$

Hence, there is a map $\phi: \Delta_n \rightarrow \Delta_{k-1} = [\vec{v}_0, \dots, \vec{v}_n]$ given by

$$\phi: \vec{x} = \sum_{i=0}^{k-1} w_i (c_i \vec{v}_{2i} + (1 - c_i) \vec{v}_{2i+1}) \mapsto \sum_{i=0}^{k-1} w_i \vec{v}_i.$$

One can then decompose Δ_{k-1} symmetrically into k pieces using barycentric coordinates as in (32) and Figure 16. Following Rubinstein and Tillmann’s construction of PL multisections from triangulations [9], one might attempt to construct a multisection of T^n by pulling back each X_i via ϕ , mapping forward by the quotient map $\Delta_n \rightarrow T^n/S_n$, and pulling back by the quotient map $T^n \rightarrow T^n/S_n$.

This construction works for T^3 and cuts any T^n into k 1-handlebodies of genus n . Unfortunately, the needed intersection properties fail, even for T^5 , so the decomposition is not a multisection. Note that by writing

$$\Delta_n = [\vec{v}_0, \vec{v}_1] * \dots * [\vec{v}_{2(k-1)}, \vec{v}_{2k-1}]$$

we made an asymmetric choice, and that the resulting decomposition is generally different than the one obtained by writing

$$\Delta_n = [\vec{v}_{\sigma(0)}, \vec{v}_{\sigma(1)}] * \dots * [\vec{v}_{\sigma(2k-2)}, \vec{v}_{\sigma(2k-1)}]$$

for arbitrary $\sigma \in S_n$ and then following the same procedure.

References

- [1] **C Frohman**, *Minimal surfaces and Heegaard splittings of the three-torus*, Pacific J. Math. 124 (1986) 119–130 [MR](#) [Zbl](#)
- [2] **D Gay**, **R Kirby**, *Trisecting 4-manifolds*, Geom. Topol. 20 (2016) 3097–3132 [MR](#) [Zbl](#)
- [3] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) [MR](#) [Zbl](#)
- [4] **D Koenig**, *Trisections of 3-manifold bundles over S^1* , Algebr. Geom. Topol. 21 (2021) 2677–2702 [MR](#) [Zbl](#)
- [5] **P Lambert-Cole**, **J Meier**, *Bridge trisections in rational surfaces*, J. Topol. Anal. 14 (2022) 655–708 [MR](#) [Zbl](#)
- [6] **P Lambert-Cole**, **M Miller**, *Trisections of 5-manifolds*, from “2019–20 MATRIX annals” (DR Wood, J de Gier, CE Praeger, T Tao, editors), MATRIX Book Ser. 4, Springer (2021) 117–134 [MR](#) [Zbl](#)
- [7] **J Meier**, **T Schirmer**, **A Zupan**, *Classification of trisections and the generalized property R conjecture*, Proc. Amer. Math. Soc. 144 (2016) 4983–4997 [MR](#) [Zbl](#)
- [8] **C P Rourke**, **B J Sanderson**, *Introduction to piecewise-linear topology*, Ergebnisse der Math. 69, Springer (1972) [MR](#) [Zbl](#)
- [9] **JH Rubinstein**, **S Tillmann**, *Multisections of piecewise linear manifolds*, Indiana Univ. Math. J. 69 (2020) 2209–2239 [MR](#) [Zbl](#)
- [10] **NJA Sloane**, et al., *The on-line encyclopedia of integer sequences* Available at <http://oeis.org>
- [11] **M Williams**, *Trisections of flat surface bundles over surfaces*, PhD thesis, University of Nebraska–Lincoln (2020) [MR](#) Available at <https://www.proquest.com/docview/2437426769>

Department of Mathematics, Wake Forest University
Winston-Salem, NC, United States

kindret@wfu.edu

<https://www.thomaskindred.com>

Received: 5 November 2020 Revised: 9 March 2022

Bigrading the symplectic Khovanov cohomology

ZHECHI CHENG

We construct a well-defined relative second grading on symplectic Khovanov cohomology from holomorphic disc counting. We use a version of symplectic Khovanov cohomology defined for bridge diagrams rather than braids. We show that our second grading recovers the Jones grading of Khovanov homology up to an overall grading shift over any characteristic-zero field, through proving that the isomorphism of Abouzaid and Smith can be refined as an isomorphism between bigraded cohomology theories. The central idea of the proof is to construct an exact triangle for symplectic Khovanov cohomology that behaves similarly to the unoriented skein exact triangle for Khovanov homology.

53D40, 57K18, 57K10, 57R58

1 Introduction

In [11], Seidel and Smith defined a singly graded link invariant, *symplectic Khovanov cohomology* $\text{Kh}_{\text{symp}}^*(L)$. It is the Lagrangian intersection Floer cohomology of two Lagrangians in a symplectic manifold \mathcal{Y}_n . The manifold \mathcal{Y}_n is built through taking a fiber of the restriction of the adjoint quotient map $\chi: \mathfrak{sl}_{2n}(\mathbb{C}) \rightarrow \text{Conf}_{2n}^0(\mathbb{C})$ to a nilpotent slice \mathcal{S}_n . The Lagrangians are determined by the link L as follows. A given link L in S^3 can be realized as a braid closure of $\beta_L \in \text{Br}_n$ for some n depending on L . The braid $\beta_L \times \text{id} \in \text{Br}_{2n}$ gives a path in the configuration space $\text{Conf}_{2n}^0(\mathbb{C})$. Parallel transport along $\beta_L \times \text{id}$ induces a symplectomorphism of \mathcal{Y}_n to itself (precisely speaking, from an arbitrarily large compact subset of \mathcal{Y}_n to itself). There is a distinguished Lagrangian submanifold \mathcal{K} given by iterated vanishing cycles. Let $(\beta_L \times \text{id})(\mathcal{K})$ be its image under the parallel transport. $\text{Kh}_{\text{symp}}^*(L)$ is defined to be the Floer cohomology group

$$(1-1) \quad \text{Kh}_{\text{symp}}^*(L) = \text{HF}^{*+n+w}(\mathcal{K}, (\beta_L \times \text{id})(\mathcal{K})),$$

where w is the writhe of β_L . There is a conjectural relation between $\text{Kh}_{\text{symp}}^*$ and the original Khovanov homology $\text{Kh}^{*,*}$ defined by Khovanov in [5].

Conjecture 1.1 [11, Conjecture 2] For any link $L \subset S^3$,

$$\mathrm{Kh}_{\mathrm{symp}}^k(L) \cong \bigoplus_{i-j=k} \mathrm{Kh}^{i,j}(L).$$

This conjecture is true over any characteristic-zero field, proved by Abouzaid and Smith in [2, Theorem 7.6]. For the cases of characteristic-nonzero fields, in the very few examples that have been computed, such as the case of the trefoil in [11, Proposition 55], the theories are isomorphic.

Our results rely on the theorem of Abouzaid and Smith that [Conjecture 1.1](#) is true over any characteristic-zero field, so we assume the characteristic of the base field \mathbb{k} to be zero unless noted otherwise.

Note that both Khovanov homology and symplectic Khovanov cohomology are the homology of graded or relatively graded cochain complexes in which the differential raises the grading by one.

We work with the framework of Manolescu's Hilbert scheme reformulation of $\mathrm{Kh}_{\mathrm{symp}}^*$, in which \mathcal{Y}_n is viewed as a symplectically embedded open subscheme of $\mathrm{Hilb}^n(A_{2n-1})$, the n^{th} Hilbert scheme of the Milnor fiber of A_{2n-1} -surface singularity.

One of the advantages of Manolescu's reformulation [7] is that we can work with *bridge diagrams* instead of braid closures. A bridge diagram is a decorated link diagram obtained by breaking the link diagrams into n pairwise disjoint α arcs and n pairwise disjoint β arcs so that β arcs overcross α arcs at any intersection. These arcs give rise to two Lagrangians \mathcal{K}_α and \mathcal{K}_β in $\mathcal{Y}_n \subset \mathrm{Hilb}^n(A_{2n-1})$. The main theorem of [7] implies that $\mathrm{Kh}_{\mathrm{symp}}^*(L) \cong \mathrm{HF}^{*+n+w}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ for a specific type of bridge diagram, called a *flattened braid diagram*, where n is the number of strands and w is the writhe of the corresponding braid. All braids can give rise to flattened braid diagrams, but not all bridge diagrams are isotopic to flattened braid diagrams.

Attempts were made to generalize symplectic Khovanov cohomology to arbitrary bridge diagrams, an \mathbb{F}_2 -coefficients version by Hendricks, Lipshitz and Sarkar [4, Section 7], and a relatively graded version by Waldron [13, Section 6]. In this paper, we give an absolute grading for Waldron's construction. Recall that the writhe w of the diagram is the number of positive crossings minus the number of negative crossings, and the rotation number rot of the diagram is the number of counterclockwise minus the number of clockwise Seifert circles. We can conclude:

Theorem 1.2 For any oriented bridge diagram representing a link L , let x_0 be the generator whose coordinates are the starting point of each β arc. Then the Floer cohomology groups

$$(1-2) \quad \text{Kh}_{\text{symp}}^*(L) = \text{HF}^{*+\text{gr}(x_0)+w+\text{rot}}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$$

are link invariants.

As Waldron proved in the relative case [13, Theorem 1.1], the absolutely graded invariant defined with bridge diagram is also canonical, ie the Floer groups from two equivalent bridge diagrams are related through a canonical isomorphism.

The orientation of the diagram, especially for a link diagram, is crucial in computing the correction terms and locating x_0 . Throughout this paper, we always assume that our bridge diagrams are oriented.

Abouzaid and Smith [1, Equation 2.31] constructed an endomorphism ϕ of $\text{CF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ from the linear part of an nc -vector field, which was introduced in [1, Definition 2.3]. This endomorphism is a chain map and thus induces a generalized eigenspace decomposition of $\text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$. The eigenvalues will equip the Floer cohomology group $\text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ with an additional grading, called the *weight grading*.

We will only use a relative version of the weight grading because an absolute grading relies on choices of auxiliary data for each Lagrangian, called *equivariant structures*. Equivariant structures are defined in [1, Definition 2.10] and in Definition 3.3 of this paper; the choice of equivariant structure changes the grading by a shift, as in [1, Lemma 2.1.2(2)], so we may obtain our relative grading without addressing this choice. As a result, we will skip the discussion of equivariant structures in this paper. We prove that for any bridge diagram, the relative weight grading recovers the Jones grading (called the quantum grading in some contexts) of Khovanov homology.

Theorem 1.3 Symplectic Khovanov cohomology, graded by (gr, wt) , and Khovanov homology, graded by (i, j) , are isomorphic as bigraded vector spaces over any field of characteristic zero, where the gradings are related by $\text{gr} = i - j$ and $\text{wt} = -j + c$, where c is a correction term for the relative weight grading.

By definition, the weight grading lives in $\overline{\mathbb{k}}$, the algebraic closure of the base field \mathbb{k} . The theorem implies that the weight grading is integral. Defining an absolute weight grading requires us to make a specific choice on the equivariant structures on the

Lagrangians depending on the writhe, crossing number, and other properties of the bridge diagram. It is natural for us to ask the following question.

Question 1.4 How do we specify the choice of the equivariant structure such that the weight grading is an absolute grading?

As for the purpose of this paper, we are satisfied with a relative grading. Thus, we will only fix an arbitrary choice of equivariant structures throughout this paper.

To prove [Theorem 1.3](#), we show that Abouzaid and Smith's long exact sequence of the unoriented skein relation for symplectic Khovanov cohomology groups [2, Equation 7.9] decomposes with respect to the weight grading. In other words, if we choose any weight grading wt_1 of the first group, there exists a weight grading wt_2 of the second group and wt_3 of the third group such that the map connecting the first and second group changes the weight grading by a constant $\text{wt}_2 - \text{wt}_1$, and constant weight grading changes $\text{wt}_3 - \text{wt}_2$ and $\text{wt}_1 - \text{wt}_3$ for the other two maps:

$$(1-3) \quad \cdots \rightarrow \text{HF}^{*,\text{wt}_1}(L_+) \rightarrow \text{HF}^{*,\text{wt}_2}(L_0) \rightarrow \text{HF}^{*+2,\text{wt}_3}(L_\infty) \\ \rightarrow \text{HF}^{*+1,\text{wt}_1}(L_+) \rightarrow \cdots,$$

where L_+ is a link with a positive crossing at p , and L_0 and L_∞ are the two resolutions at p .

In the singly graded case, the isomorphism of Abouzaid and Smith between symplectic Khovanov cohomology and Khovanov homology can be used to build a commutative diagram between the exact triangle above and the exact triangle for the unoriented skein relation in Khovanov homology. We show the maps between the exact triangles given by the isomorphism of Abouzaid and Smith are bigraded by induction on the number of crossings. A purity result of Abouzaid and Smith [1, Theorem 1.1] leads to a computation for crossingless diagrams of an unlink that every element $x \in \text{Kh}_{\text{symp}}^k(L)$ will satisfy $\text{wt}(x) = k$ for some choice of equivariant structures.

This finishes the proof of [Theorem 1.3](#). Our argument so far is diagrammatic: we have not proved that the relative weight grading is independent of bridge diagrams yet. Now that we know the relative weight grading recovers the Jones grading up to an overall grading shift for any diagram, and with the fact that Jones grading is independent of link diagrams, we prove a conjecture of Abouzaid and Smith.

Theorem 1.5 *The relative weight grading on $\text{Kh}_{\text{symp}}^*(L)$ is independent of the choice of link diagram.*

It is worth noting that our proof of [Theorem 1.5](#) is not internal to symplectic geometry, and the invariance of the relative weight grading relies on the well-definedness of the Jones grading in combinatorial Khovanov homology. We conclude this introduction with the following question.

Question 1.6 Is there a proof of invariance of the weight grading from pure symplectic geometry?

Organization

The paper is organized as follows. In [Section 2](#), we review the definition of symplectic Khovanov cohomology and construct an absolute grading on the symplectic Khovanov cohomology for bridge diagrams. In [Section 3](#), we give a precise definition of the weight grading and construct a bigraded unoriented skein exact triangle of symplectic Khovanov cohomology. In [Section 4](#), we prove the main theorem by showing that Abouzaid and Smith's isomorphism between symplectic Khovanov cohomology and combinatorial Khovanov homology preserves the second grading.

Acknowledgements

The author would like to thank his thesis advisor Mohammed Abouzaid for his guidance throughout this project. He also thanks Kristen Hendricks, Mikhail Khovanov, Francesco Lin, Robert Lipshitz and Ivan Smith for many helpful discussions on the subject. He would like to thank the editor and referees for their constructive comments on the manuscript. Finally, he thanks Columbia University for its supportive studying environment during the early stage of writing the paper, Institut Mittag-Leffler for its hospitality during the later stage of preparation of this paper, and Wuhan University for its productive research environment during the publication phase of this paper. The author was supported by NSFC grant 12126101, NSF grants DMS-1609148 and DMS-1564172, and the Swedish Research Council under grant 2016-06596.

2 A review of symplectic Khovanov cohomology $\text{Kh}_{\text{symp}}^*(L)$

We will briefly review the original definition of symplectic Khovanov cohomology and give a formal definition of symplectic Khovanov cohomology of a bridge diagram in [Section 2.1](#). We will discuss the homological grading in [Section 2.2](#). The construction of this section is not restricted to characteristic-zero fields.

2.1 Symplectic Khovanov cohomology for bridge diagrams

The link invariant $\text{Kh}_{\text{symp}}^*(L)$ was first introduced by Seidel and Smith in [11] as the Lagrangian intersection Floer cohomology of two Lagrangians in \mathcal{Y}_n , constructed as a nilpotent slice in $\mathfrak{sl}_{2n}(\mathbb{C})$. In this original formulation, the first Lagrangian is created by a technique called the *iterated vanishing cycle*, while the second Lagrangian is the image of the first Lagrangian under the symplectomorphism induced by the braid whose closure is the link L . Manolescu in [7] introduced a reformulation using Hilbert schemes, which is easier to visualize and more similar to other low-dimensional invariants, such as Heegaard Floer homology. In this subsection, Lagrangian Floer cohomology groups are relatively graded; see Remark 2.4.

Let us start with the complex surface of Milnor fiber A_{2n-1} . Consider the complex surface

$$(2-1) \quad A_{2n-1} = \{(u, v, z) \in \mathbb{C}^3 \mid u^2 + v^2 + p(z) = 0\} \in \mathbb{C}^3,$$

where $p(z) = (z - p_1) \cdots (z - p_{2n})$.

Recall that the Hilbert scheme of n points on A_{2n-1} , $\text{Hilb}^n(A_{2n-1})$, is the space of closed zero-dimensional subschemes of A_{2n-1} of length n . The Hilbert scheme $\text{Hilb}^n(A_{2n-1})$ is closely related to the n -fold symmetric product $\text{Sym}^n(A_{2n-1})$ of A_{2n-1} , through the *Hilbert–Chow morphism*:

Proposition 2.1 [8, Chapter 9] *The Hilbert–Chow morphism π is a natural morphism from $\text{Hilb}^n(X)$ to $\text{Sym}^n(X)$ such that*

$$(2-2) \quad \pi(Z) = \sum_{x \in X} \text{length}(Z_x)[x].$$

Moreover, if X is complex one-dimensional, then π is an isomorphism. If X is smooth and complex two-dimensional, π is a crepant resolution of singularities and $\text{Hilb}^n(X)$ is smooth.

Let $\Delta = \{(x_1, \dots, x_n) \in \text{Sym}^n(A_{2n-1}) \mid x_i = x_j \text{ for some } i \neq j\}$ be the *diagonal* of the symmetric product. The proposition above implies that π is one-to-one away from the diagonal Δ , so we can think of $\text{Hilb}^n(A_{2n-1})$ as $\text{Sym}^n(A_{2n-1})$ when away from Δ . Explicitly, a point in $\text{Hilb}^n(A_{2n-1})$ corresponds to an ideal

$$I \subset \mathcal{O} = \mathbb{C}[u, v, z]/(u^2 + v^2 + p(z))$$

with $\dim_{\mathbb{C}}(\mathcal{O}/I) = n$. Thus, given n distinct points (u_i, v_i, z_i) in A_{2n-1} , we can construct an ideal by taking the product of $(u - u_i, v - v_i, z - z_i)$, which gives a point in $\text{Hilb}^n(A_{2n-1})$.

Now we proceed to the definition of the manifold \mathcal{Y}_n . First, we consider a projection $i: A_{2n-1} \rightarrow \mathbb{C}$ such that $i(u, v, z) = z$. It induces a map $\mathbb{C}[z] \hookrightarrow \mathbb{C}[u, v, z] \rightarrow \mathcal{O}$. As $\mathbb{C}[z]$ intersects with the ideal $(u^2 + v^2 + p(z))$ in $\mathbb{C}[u, v, z]$ trivially, if we denote by R_1 the image of $\mathbb{C}[z]$ in \mathcal{O} , we have $R_1 \cong \mathbb{C}[z]$. For any ideal $I \in \text{Hilb}^n(\mathbb{C})$, we then define the projection $i(I) = I \cap R_1$ and the manifold

$$(2-3) \quad \mathcal{Y}_n = \{I \in \text{Hilb}^n(A_{2n-1}) \mid i(I) \text{ has length } n\}.$$

The complement of \mathcal{Y}_n defines a divisor in the Hilbert scheme $\text{Hilb}^n(A_{2n-1})$,

$$(2-4) \quad D_r = \{I \in \text{Hilb}^n(A_{2n-1}) \mid i(I) \text{ has at most length } n - 1\}.$$

Manolescu proved [7, Proposition 2.7] that \mathcal{Y}_n is biholomorphic to the space Seidel and Smith considered in [11].

Now, we construct the Lagrangians from a link diagram. A *bridge diagram* D for a link L is a triple $(\vec{\alpha}, \vec{\beta}, \vec{p})$, where $\vec{p} = (p_1, p_2, \dots, p_{2n})$ are $2n$ distinct points in \mathbb{R}^2 , $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$ are n pairwise disjoint embedded arcs and $\vec{\beta} = (\beta_1, \beta_2, \dots, \beta_n)$ are also pairwise disjoint embedded arcs such that $\partial(\bigcup \alpha_i) = \partial(\bigcup \beta_i) = \{p_1, \dots, p_{2n}\}$ and if we let the β arcs overcross the α arcs at the intersections in \mathbb{R}^2 , we obtain a diagram for L .

For each arc α_i or β_i , we can associate a Lagrangian sphere Σ_{α_i} or Σ_{β_i} in A_{2n-1} through the equations

$$(2-5) \quad \Sigma_{\alpha_i} = \{(u, v, z) \in A_{2n-1} \mid z \in \alpha_i, u, v \in \sqrt{-p(z)}\mathbb{R}\},$$

$$(2-6) \quad \Sigma_{\beta_i} = \{(u, v, z) \in A_{2n-1} \mid z \in \beta_i, u, v \in \sqrt{-p(z)}\mathbb{R}\}.$$

There are natural projection maps $\Sigma_{\alpha_i} \rightarrow \alpha_i$ and $\Sigma_{\beta_i} \rightarrow \beta_i$. Each interior point of the arc lifts to a copy of the circle, whereas each endpoint lifts to a single point. Therefore Σ_{α_i} and Σ_{β_i} are copies of S^2 . With appropriate choice of Kähler form, they are Lagrangian spheres; see [7, Section 4].

The spheres Σ_{α_i} are pairwise disjoint in A_{2n-1} because their projection arcs α_i are pairwise disjoint in \mathbb{C} . Similarly, the spheres Σ_{β_i} are also pairwise disjoint. With these

spheres, we can build two Lagrangians in $\text{Hilb}^n(A_{2n-1})$ and, in fact, in \mathcal{Y}_n by

$$(2-7) \quad \mathcal{K}_\alpha = \Sigma_{\alpha_1} \times \Sigma_{\alpha_2} \times \cdots \times \Sigma_{\alpha_n},$$

$$(2-8) \quad \mathcal{K}_\beta = \Sigma_{\beta_1} \times \Sigma_{\beta_2} \times \cdots \times \Sigma_{\beta_n}.$$

It is worth pointing out that \mathcal{K}_α and \mathcal{K}_β do not intersect transversely. The intersections of the spheres Σ_{α_i} and Σ_{β_j} in A_{2n-1} are points in the fibers of the endpoints of the α arcs and β arcs and circles S^1 in the fibers of interior intersections of those arcs. Thus, we could have some tori in $\mathcal{K}_\alpha \cap \mathcal{K}_\beta$. To deal with this, we can either perturb one of the Lagrangians (see [7, Section 6.1]), or use Floer theory with clean intersections, as in [9]. We can use either method when we compute Floer cohomology groups, because both methods result in isomorphic homology groups.

For example, we can perturb \mathcal{K}_β as follows. As \mathcal{K}_β is a product $\Sigma_{\beta_1} \times \Sigma_{\beta_2} \times \cdots \times \Sigma_{\beta_n}$, it will be easier to perturb each Σ_{β_i} in the complex surface A_{2n-1} . Let N be a circle in the intersection $\Sigma_{\alpha_j} \cap \Sigma_{\beta_i} \subset A_{2n-1}$ and V be a small neighborhood of N such that V intersects no other Lagrangian spheres. Following Weinstein [14, Theorem 6.5], we use the standard height function on N to isotope Σ_{β_i} into Σ'_{β_i} so that Σ'_{β_i} is identical to Σ_{β_i} outside of V and Σ'_{β_i} intersects $\Sigma_{\beta_i} \cap V$ only at the maximum and minimum of the height function. We repeat this process for all the S^1 intersections on Σ_{β_i} and for all the β Lagrangian spheres Σ_{β_i} . With a slight abuse of notation, we call the resulting Lagrangian spheres Σ'_{β_i} and let $\mathcal{K}'_\beta = \Sigma'_{\beta_1} \times \Sigma'_{\beta_2} \times \cdots \times \Sigma'_{\beta_n}$. The resulting Floer cochain complex $\text{CF}^*(\mathcal{K}_\alpha, \mathcal{K}'_\beta)$ will be quasi-isomorphic to $\text{CF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$.

For simplicity, we treat \mathcal{K}_β as the original Lagrangian perturbed, unless mentioned otherwise, so that it intersects transversely with \mathcal{K}_α at isolated points. Each intersection at the interior of arcs α_i and β_j now gives the intersection of Σ_{α_i} and Σ_{β_j} at two points instead of a circle.

Proposition 2.2 [7, Theorem 1.2; 13, Theorem 4.12] *For any bridge diagram D representing a link L , the Floer cohomology $\text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ in $\mathcal{Y}_n = \text{Hilb}^n(A_{2n-1}) \setminus D_r$ is canonically isomorphic to Seidel and Smith’s symplectic Khovanov homology $\text{Kh}^*_{\text{symp}}(L)$.*

With Proposition 2.2 in mind, we finally define:

Definition 2.3 The symplectic Khovanov cohomology $\text{Kh}^*_{\text{symp}}(D)$ of a bridge diagram D is defined to be $\text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ in $\mathcal{Y}_n = \text{Hilb}^n(A_{2n-1}) \setminus D_r$.

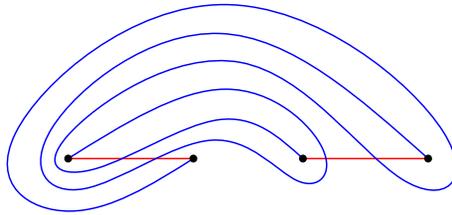


Figure 1: A flattened braid diagram for the trefoil.

Remark 2.4 We have not yet given an absolute grading for this definition. In fact, Proposition 2.2 is proved in the relatively graded case. Manolescu discussed an absolute grading in [7, Section 6.2], but it only applies in the case of flattened braid diagrams, where explicit choices can be made on Lagrangians to construct an absolute Maslov grading. Waldron’s construction in [13] works for any bridge diagram, but only in the relatively graded case.

In the rest of the paper, whenever we mention $\text{Kh}_{\text{symp}}^*(L)$, we will always be working with Definition 2.3 rather than the original definition of Seidel and Smith, unless noted otherwise.

2.2 An absolute grading for bridge diagrams

In this subsection, we prove Theorem 1.2 by showing that the homological grading shifted by $\text{gr}(x_0) + \text{rot} + w$ is invariant under isotopy, handlesliding and stabilization. We reiterate our conventions here that are crucial in defining the absolute grading. Our link is oriented. At each intersection, the β arc overcrosses the α arc. The distinguished generator x_0 has the coordinates at the starting points of all β arcs.

The idea of this correction term is from Droz and Wagner in [3], in which the authors considered grid diagrams obtained from deforming flattened braid diagrams. A flattened braid diagram of a braid $b \in \text{Br}_n$ is a special bridge diagram associated to the braid b whose marked points p_i are placed on the real line ordered by their indices, the α arcs are segments on the real line connecting p_{2i-1} and p_{2i} , and the β arcs are the images of α arcs after applying the braid b action on the odd-numbered strands.

Remark 2.5 An interesting example is that if we take a bridge diagram representing a braid closure, the rotation number is exactly the number of strands (because all Seifert circles are going counterclockwise), and the writhe of the diagram matches the writhe

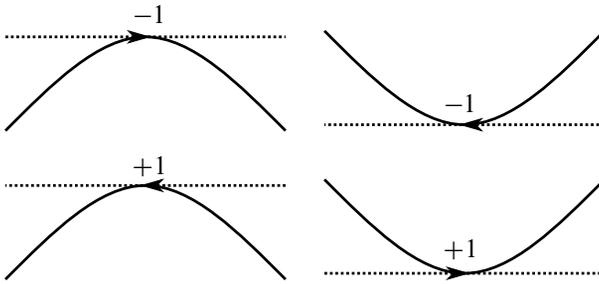


Figure 2: Local picture near the marked points with horizontal tangent lines and their sign assignments.

of the braid. The distinguished generator has homological grading 0 in any flattened braid diagram. Thus the correction term $gr(x_0) + rot + w$ for the braid closure diagram agrees with Manolescu’s correction term $n + w$ of the flattened braid diagram, where these two diagrams are isotopic as link diagrams.

Equipping the Lagrangian Floer homology with an absolute grading requires us to grade our Lagrangians by choosing some sections of the canonical bundle. In the braid closure setup, the two Lagrangians are related by some fibered Dehn twists and thus the choice on the second Lagrangian can be induced from the first one canonically. In the general bridge diagram case (not a flattened braid diagram), there is no easy way to assign a choice to the second Lagrangian, so instead, we assign to the generator x_0 the grading 0.

Manolescu in [7, Section 6.2] pointed out a combinatorial method to compute the relative homological grading. In short, we replace each β arc β_j with an oriented figure-eight γ_j (and the orientation does not matter) in a small neighborhood of β_j . We arrange our diagram so that each α arc is horizontal and each figure-eight is vertical wherever they intersect. Each intersection of Lagrangian spheres Σ_{α_i} and Σ_{β_j} corresponds to an intersection of the arc α_i and the figure-eight γ_j . We travel along the figure-eight γ_j and mark the points that have horizontal tangent lines. We assign $+1$ to those marked points if the part of the figure-eight near the marked point is locally oriented counterclockwise, and -1 if clockwise; see Figure 2.

Now, we label each point on γ_j with a number, starting with the quantity 0 at the intersection representing the starting point of β_j . Note that this is an arbitrary choice, but we will correct for it later. We move along the figure-eight γ_j following its orientation. We change the quantity we assign only when we move past the marked

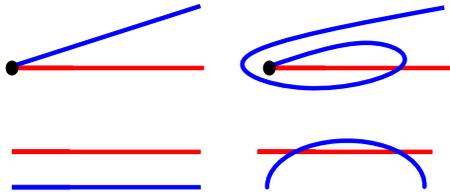


Figure 3: Two possible bridge diagram isotopies corresponding to Reidemeister I and II moves.

points with horizontal tangent lines. The new quantity will be the sum of the previous one and the number we put on the marked points, ie $+1$ or -1 ; see Figure 4 for an illustration of the computation. An easy combinatorial argument implies that the sum of the numbers on such marked points is 0 and thus we get back to 0 when we reach back to the starting point. Each of the generators of the Floer cohomology group corresponds to n intersection points and the sum of the labeled numbers will be its relative grading. In the construction, we made a choice that the coordinate at each starting point of the β arc has grading 0, and thus the distinguished generator x_0 has grading 0. In other words, the grading shifted by $\text{gr}(x_0)$ will be independent of the grading we assign to the starting points of the β arc. We now prove the following proposition about the absolute grading.

Proposition 2.6 $\text{HF}^{*+\text{gr}(x_0)+\text{rot}+w}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ is invariant under bridge diagram isotopy.

Proof An isotopy of a bridge diagram induces Hamiltonian-isotopic Lagrangians and thus keeps the relatively graded group $\text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ unchanged. If any crossing arises, it must come from one of the cases shown in Figure 3, which correspond to a Reidemeister I move and a Reidemeister II move (whereas the Reidemeister move III is from a combination of (de-)stabilizations and handleslides).

In the first case (see top row of Figure 3), there are two subcases: the α arc oriented to the left, and to the right. The easier subcase is the right-going α curve. The crossing is a positive crossing and the bigon region gives an additional clockwise Seifert circle. The distinguished generator x_0 has no coordinate in this region and thus its grading does not change. The total change of $\text{gr}(x_0) + \text{rot} + w$ is 0.

The other subcase is more subtle. The crossing is still positive but the Seifert circle is now counterclockwise, which increases the correction term by 2. The distinguished generator x_0 indeed takes a coordinate at the black dot in the top row of Figure 3.

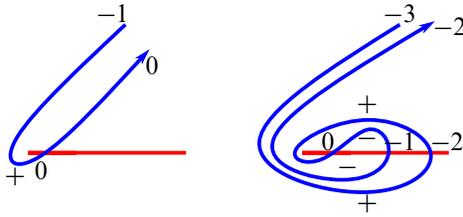


Figure 4: Isotopy corresponding to Reidemeister I move. We label each intersection and the endpoint of the figure-eight with its grading and we abbreviate the ± 1 labels on the marked points with horizontal tangent lines to \pm signs to avoid confusion.

With the illustration of figure-eights in Figure 4, we can see that the relative grading is changed by 2, which cancels the contribution of the other two terms to $\text{gr}(x_0) + \text{rot} + w$. In fact, we can explicitly describe the new cochain complex. The bigon region in the complex plane lifts to topological disks that have holomorphic representatives in the complex surface A_{2n-1} and thus in \mathcal{Y}_n when multiplied with constant discs, connecting one of the two generators at the interior intersection to the endpoint. One of the moduli spaces has dimension one and contributes to differentials which cancel pairs of generators having identical coordinates except one coordinate in the picture. At the cochain level, the isotopy created three copies of the original cochain complex with generators that have a coordinate at the black dot, but two of the three copies are canceled on the homology level. The surviving copy should have the same grading as the original cochain. It is not hard to see that the other moduli space of the bigon region has dimension two; thus, locally changing coordinate from the lower grading coordinate at the intersection to the black dot decreases the grading by 2. So in the new diagram, the grading of the distinguished generator decreases by 2 relative to other generators, which cancels the contribution by the sum of writhe and rotation number.

In the second case which corresponds to a Reidemeister II move, the two new crossings always have different signs so the writhe remains unchanged regardless of the orientation. If the two strands are oriented in the same direction, the Seifert surface remains unchanged and thus the rotation number is also the same. If the two strands are oriented in the opposite direction, there will be two subcases, whether the two strands are from the same Seifert circle or not. In the first subcase (the top row of Figure 5), locally one Seifert circle breaks into three circles, but the middle one is of the opposite direction. In the second subcase (the bottom row of Figure 5), there are two Seifert circles of the same direction in each picture. In either case, the rotation number remains the same as

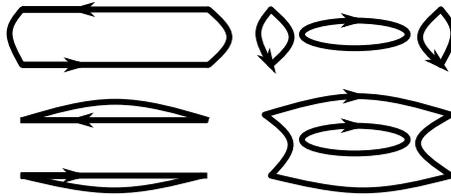


Figure 5: Two cases for Reidemeister II isotopy when two strands are oriented in the opposite direction. The first row corresponds to the case when both strands are oriented in the same direction and the second row corresponds to the case when strands are oriented in the opposite direction.

well. Apart from the two new intersections in the new diagram, all other generators are from the diagram before the isotopy and it is not hard to see that the gradings of such generators remain unchanged. Thus $\text{gr}(x_0)$ remains unchanged as well, relative to other generators. \square

Proposition 2.7 $\text{HF}^{*\text{+gr}(x_0)\text{+rot}+w}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ is invariant under stabilization.

Proof Stabilizations will not change the writhe or the rotation number of the diagram. It suffices to show that the grading change of the homology is the same as the grading change of the distinguished generator x_0 , when we apply stabilizations. This follows because Waldron’s proof of invariance of the relative grading under stabilization establishes an isomorphism on the cochain level; see [13, Lemma 5.19]. \square

Handleslide invariance is the hardest among the three. Before the proof, we make some topological observations. With isotopy invariance, we should arrange our diagram to be as simple as possible.

We explicitly describe the process of the handleslide of arc α_2 over α_1 , as shown in Figure 6. First, we choose the boundary of a tubular neighborhood of α_1 , so that each midpoint intersection of any β arc locally intersects the circle exactly twice, and the β arcs connecting the endpoints of α_1 locally intersect the circle exactly once. Then pick any path γ from α_2 to the circle that does not intersect any α arc in its interior, and perform a connected sum of α_2 and the circle along γ so that each intersection between the path and the β arc creates exactly two intersections on the connected sum.

Proposition 2.8 $\text{HF}^{*\text{+gr}(x_0)\text{+rot}+w}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ is invariant under handleslides.

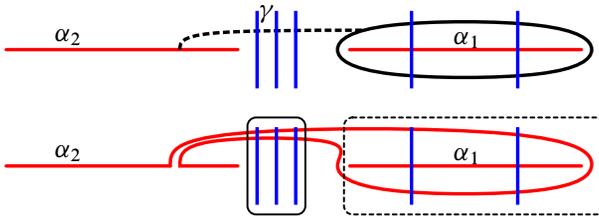


Figure 6: Handlesliding α_2 over α_1 along the arc γ . In the diagram after handlesliding, as shown in the second row, the left (solid) box corresponds to Figure 7 and the right (dashed) box corresponds to Figure 8.

Proof It is clear that new intersections are created in pairs, and each pair contains exactly one negative and one positive crossing. Thus, the writhe of the diagram remains unchanged.

Next, we show that the rotation number remains unchanged as well. The proof breaks into three steps.

First, we study the intersections created around the path γ , ie the intersections in the solid box in Figure 6. Locally, there is a series of parallel β arcs intersecting two α arcs. The two α arcs are parallel to the path γ and always oriented in the opposite directions.

Let us label a β arc with positive sign if it goes from bottom to top, and with negative sign if it goes from top to bottom. For convenience, we order those parallel β arcs from left to right. We study the local picture with only two adjacent β arcs. There are four cases in total shown in Figure 7. In the case $++$, the Seifert circle containing the left arc remains unchanged. The Seifert circle containing the right arc goes to the local picture involving the right β arc and the next one to the right outside the figure. The Seifert circle will either be unchanged, if the next β arc is also $+$, or it will be discussed in the case $+ -$ below, if the next β arc is $-$. In the case $--$, the Seifert circle containing the right arc remains the same and the other Seifert circle will either be unchanged or will be discussed in the case $+ -$ below.

In the case $+ -$, there are two subcases, whether the two β arcs belong to the same Seifert circle or different ones, similar to the Reidemeister II move argument as shown in Figure 5. In this case either a clockwise circle is broken into two or two clockwise circles are merged into one. In either case, the rotation number is reduced by 1.

In the case $- +$, one counterclockwise circle is created. The other two Seifert circles are studied already in other cases. Thus, the rotation number is increased by 1.

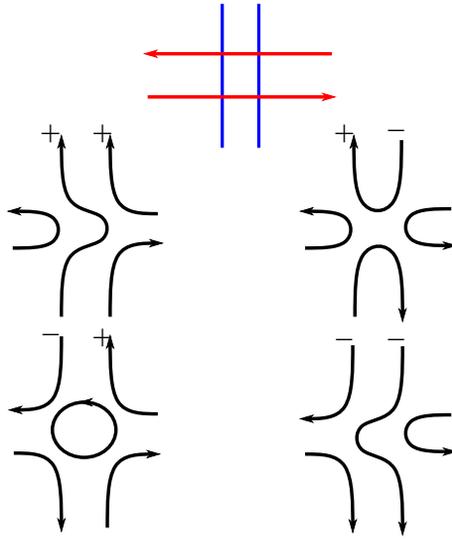


Figure 7: All four possible Seifert circles at the intersections near the path γ , depending on the orientation of β arcs.

To sum this up, each time the sign changes from $+$ to $-$, the rotation number decreases by 1, and each time the sign changes from $-$ to $+$, the rotation number increases by 1. At the intersection of the path γ and α_1 , α_1 is oriented to the right to match the orientation of the α arcs in the handleslide picture, and the circle will remain the same if the first β is positive, and decreases the rotation number by 1 if the first β is negative. Thus, if the last β arc is positive, the rotation number remains the same, and if the last one is negative, the rotation number is reduced by 1. The last β arc is also related to the last step of the argument.

The second step is to compute the contribution of new intersections created at the circle near α_1 , ie the intersections in the dashed box of Figure 6. Without loss of generality, let us assume α_1 is oriented from left to right. The part in the dashed box of Figure 8 is exactly the same as the picture before the handleslide. The rest of the region is two parallel α arcs oriented in the opposite direction, intersecting a series of parallel β arcs. The study of these new intersections is similar to the first half of the argument and the conclusion is similar, except now the sign depends on the first β arc. We claim that the rotation number remains the same if the first arc is positive, and increases by 1 if the first arc is negative.

Combing the results from the first two steps, the change of the rotation number only depends on the direction of the last β arc in Figure 7 and the first β arc in Figure 8.

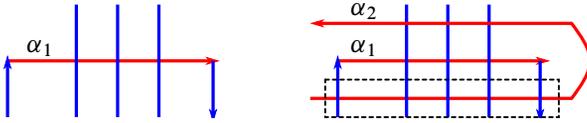


Figure 8: Diagrams near α_1 . The left one is before the handleslide and the right one is after the handleslide.

If both are of the same direction, the total contribution to the rotation number is 0. If the β from the first half is $-$ and the β from the second half is $+$, then we have a counterclockwise Seifert circle while all the other Seifert circles remain unchanged; see Figure 9. This Seifert circle cancels the effect of the second arc being negative, ie increases the rotation number by 1. In the symmetric case, the creation of a clockwise Seifert circle between the two arcs cancels the effect of the first β arc being negative, ie decreases the rotation number by 1. In any of the cases above, the total change of rotation number is 0.

Lastly, we need to check the relative grading of the homology remains the same relative to the distinguished generator x_0 . First of all, the isomorphism between symplectic Khovanov homology groups induced by handlesliding is induced by a continuation map on the cochain level. It is also easy to see that there is an injection from the generators of the original cochain complex to the ones of the cochain complex after the handleslide. It is easy to see that the homological gradings of the corresponding generators are not changed by considering the formulation of the relative grading in terms of tangencies of figure-eights, because we only changed one of the arcs and we essentially only applied an isotopy to one of the α arcs if we allow the isotopy to move across other marked points and α arcs. Moreover, the distinguished generator is obviously one of the generators that are preserved in the correspondence, and thus the homological grading relative to the distinguished generator x_0 remains the same. \square

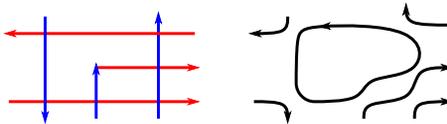


Figure 9: The figure on the left is the region connecting Figures 7 and 8 with the case $-+$. The left β arc is the rightmost β arc of Figure 7, the middle short β arc is the leftmost β arc of Figure 8, and the right β arc is the second leftmost β arc of Figure 8. The figure on the right is its Seifert circles and there is a counterclockwise Seifert circle in the middle.

Combining the propositions in this subsection, we complete a proof for [Theorem 1.2](#), that we obtain an absolutely graded version of symplectic Khovanov cohomology for bridge diagrams.

3 Weight grading on $\text{Kh}_{\text{symp}}^*(L)$

In this section, we discuss the construction of the weight grading in [Section 3.1](#), following the idea of Abouzaid and Smith in [\[1\]](#). We then turn to its behavior under certain Floer products in [Section 3.2](#). Lastly, we construct a long exact sequence of the unoriented skein relation for bigraded symplectic Khovanov cohomology groups in [Section 3.3](#).

3.1 Construction of weight grading

In this subsection, we define the weight grading wt on $\text{Kh}_{\text{symp}}^*$ following the idea of Abouzaid and Smith in [\[1, Section 3\]](#). The key point of constructing such a grading is to build an automorphism, more precisely, the linear term of a *noncommutative vector field*, or an *nc-vector field* [\[1, Definition 2.3\]](#),

$$(3-1) \quad \phi: \text{CF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta) \rightarrow \text{CF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta),$$

which preserves the homological grading and commutes with the differential. Thus, it induces an automorphism Φ on homology. If x is an eigenvector of eigenvalue λ , we define the weight grading $\text{wt}(x) = \lambda$. Most results of this section hold for fields of any characteristic but we will restrict our discussion to fields of characteristic zero; see [Remark 3.4](#) for more detail. In this subsection only, we use L for Lagrangians rather than links, as we will not use link diagrams in this subsection.

The idea of defining the map ϕ is to study certain moduli spaces of holomorphic maps in a partially compactified space $\bar{\mathcal{Y}}_n$ of \mathcal{Y}_n . We will begin by introducing the partial compactification $\bar{\mathcal{Y}}_n$ before we set up the moduli spaces. We leave some of the technical details to [\[1\]](#). Abouzaid and Smith give a hypothesis under which it is possible to define an nc-vector field on an exact Fukaya category [\[1, Hypothesis 3.1\]](#), and show that a certain partial compactification of \mathcal{Y}_n satisfies their hypothesis [\[1, Section 6\]](#).

Remark 3.1 It is worth noting that for some geometric hypotheses to hold — see [\[1, Lemma 6.5\]](#), for example — Abouzaid and Smith considered Lagrangians in fibered position, which are the Lagrangians that are fibered over α or β arcs, throughout their

papers [1; 2]. This will not hold if we perturb one of the Lagrangians to make the Lagrangians intersect transversely. As a result, Abouzaid and Smith worked with Lagrangian Floer theory with clean intersections. From now on, we no longer perturb our Lagrangians but assume that the Lagrangians intersect cleanly. We do so without further comment.

Recall that the Milnor fiber A_{2n-1} admits a Lefschetz fibration structure over \mathbb{C} , with regular fibers cones and singular fibers cylinders. We partially compactify A_{2n-1} into \bar{A}_{2n-1} by adding two sections so that each regular fiber is compactified to two 2-spheres, while each singular fiber is compactified to a 2-sphere. Then we can partially compactify \mathcal{Y}_n as $\bar{\mathcal{Y}}_n = \text{Hilb}^n(\bar{A}_{2n-1}) \setminus D_r$, where D_r is defined analogously to equation (2-4) for the partially compactified surface. Following the notation of Abouzaid and Smith, we let s_0 and s_∞ be the two sections we add to A_{2n-1} . Let D_0 be the divisor of subschemes whose support meets $s_0 \cup s_\infty$.

Now we turn to the moduli space. Let $\mathcal{R}_{(0,1)}^{k+1}$ be the moduli space of holomorphic classes of closed unit discs with the additional data of

- two interior marked points $z_0 = 0$ and $z_1 \in (0, 1)$;
- $k + 1$ boundary punctures at $q_0 = 1$, and k others q_1, \dots, q_k placed counterclockwise.

Following [1, Section 3.7], we define

$$(3-2) \quad \mathcal{R}_{(0,1)}^{k+1}(x_0; x_k, \dots, x_1)$$

as the moduli space of finite-energy holomorphic maps $u: \mathbb{D} \rightarrow \bar{\mathcal{Y}}_n$ such that

- (M-1) $u^{-1}(D_r) = \emptyset$,
- (M-2) $u^{-1}(D_0) = z_0$,
- (M-3) $u^{-1}(D'_0) = z_1$, where D'_0 is a divisor linearly equivalent to D_0 but shares no irreducible component with D_0 ,
- (M-4) the relative homology class satisfies $[u] \cdot [D_0] = 1$,
- (M-5) $u(q_i) = x_i$,
- (M-6) the boundary segment between q_i and q_{i+1} maps to L_i .

This is the moduli space originally used by Seidel and Solomon in [12] for the q -intersection number, and it was later used by Abouzaid and Smith in [1, Section 3] for the nc-vector field.

Now we are ready to introduce our automorphism Φ , which only involves the case of $k = 1$. When $k = 1$, the virtual dimension of $\mathcal{R}_{(0,1)}^2(x, y)$ is the difference of the homological grading $\text{gr}(x) - \text{gr}(y)$; see [1, Lemma 3.16]. It makes sense to consider a map of degree 0 that counts holomorphic discs in $\mathcal{R}_{(0,1)}^2(x, y)$:

$$(3-3) \quad b^1(x) = \sum_{y \mid \text{gr}(y) = \text{gr}(x)} \# \mathcal{R}_{(0,1)}^2(x, y)y.$$

In fact, b^1 is the linear part of a Hochschild cochain $b \in \text{CC}^*(\mathcal{F}(\mathcal{Y}_n), \mathcal{F}(\mathcal{Y}_n))$ if we allow multiple inputs instead of one single input. A precise construction of the Hochschild cochain b can be found at [1, Equation 3.89].

Remark 3.2 The construction of an nc-vector field requires a closed Hochschild cochain, but the cochain b is not closed; see [1, Proposition 3.20]. However, it can be made closed by adding some terms involving sphere bubbles; see [1, Equation 3.90]. To keep this section short, we will not introduce these additional terms here, because the degenerations involving sphere bubbles can be excluded analogously to the argument in last three paragraphs of [1, Section 3.9].

We can also denote by b^k the part of b with k inputs, which can be defined by counting discs in $\mathcal{R}_{(0,1)}^{k+1}(x_0; x_k, \dots, x_1)$, similarly to the definition of b^1 . The terms b^k for general k are not needed in this paper, except for the case $k = 0$, which will only be used in the definition below about equivariant structures. As we do not intend to expand the discussion of the equivariant structures, we refer the readers to [1, Section 3.4] for more details. The term b^0 is defined by counting holomorphic discs in $\mathcal{R}_{(0,1)}^1(x)$ with one output and no input such that the entire boundary is mapped to some Lagrangian L . The virtual dimension of $\mathcal{R}_{(0,1)}^1(y)$ is $\text{gr}(y) - 1$ and thus $b^0|_L \in \text{CF}^1(L, L)$ for each Lagrangian submanifold L .

Definition 3.3 An equivariant object is a pair (L, c) , with $L \in \text{Ob}(\mathcal{F}(M))$ and $c \in \text{CF}^0(L, L)$, with $dc = b^0|_L$.

By definition, given an exact Lagrangian L , the obstruction to the existence of an equivariant structure c is given by $b^0|_L \in \text{HF}^1(L, L) \cong H^1(L)$, and the set of choices when this vanishes is an affine space equivalent to $H^0(L)$. In our case, the Lagrangians are products of spheres, thus $H^1(L) \cong \{0\}$ and $H^0(L) \cong \mathbb{k}$.

Since we do not assume that b^0 vanishes, for a cocycle $b \in CC^1(\mathcal{F}(M), \mathcal{F}(M))$ and equivariant objects $(\mathcal{K}_\alpha, c_\alpha)$ and $(\mathcal{K}_\beta, c_\beta)$, the linear term b^1 is not always a chain map. However, we can define a chain map following [1, Equation 2.31],

$$(3-4) \quad \phi(x) = b^1(x) - \mu^2(c_\alpha, x) + \mu^2(x, c_\beta).$$

It induces an endomorphism Φ on $HF^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$. If we consider the generalized eigenspace decomposition, the eigenvalue of the generalized eigenvector x will be its weight grading, denoted as $wt(x)$.

Remark 3.4 The construction above applies to fields \mathbb{k} of any characteristic. The weight grading is a priori indexed by elements of the algebraic closure $\bar{\mathbb{k}}$. If $\text{char}(\mathbb{k}) = 0$, there exist additive subgroups \mathbb{Z} of \mathbb{k} , and thus of its algebraic closure $\bar{\mathbb{k}}$. As for finite fields, if $\mathbb{k} = \mathbb{F}_p$ with a prime number p for simplicity, we can then compute $\bar{\mathbb{k}} = \bigcup_n \mathbb{F}_{p^n}$; it is no longer possible to find an additive subgroup \mathbb{Z} of \mathbb{F}_p or of its algebraic closure. As a result, working with a characteristic-zero field will not only enable us to use the result of Abouzaid and Smith that identifies symplectic Khovanov cohomology and Khovanov homology, but also make it easier to obtain an integral weight grading, in contrast to finite fields.

We learn from [1, Lemma 2.12] that since $HF^0(\mathcal{K}, \mathcal{K}) \cong \mathbb{k}$, changing equivariant structures shifts the overall weight gradings (ie the eigenvalues) by a constant

$$(3-5) \quad \Phi_{(\mathcal{K}_\alpha, c_\alpha), (\mathcal{K}_\beta, c_\beta)} = \Phi_{(\mathcal{K}_\alpha, c_\alpha + s_\alpha), (\mathcal{K}_\beta, c_\beta + s_\beta)} + (s_\alpha - s_\beta) \text{id}.$$

So we have the following definition of the relative weight grading:

Definition 3.5 Let Φ be the endomorphism constructed above on $HF^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ and x be an eigenvector of Φ . The relative weight grading $wt(x)$ is defined to be the eigenvalue of x . This construction relies on auxiliary choices of equivariant structures on \mathcal{K}_α and \mathcal{K}_β , but different choices of such structures will only change all gradings by a fixed number. Thus $wt(x)$ as a relative grading is independent of choices of equivariant structures.

Remark 3.6 With given equivariant structures on \mathcal{K}_α and \mathcal{K}_β , we can compute an absolute weight grading. But at the time of writing this paper, the author does not know the choices that would give a correction term independent of the link diagram.

Then we can rephrase Abouzaid and Smith’s purity result as follows.

Proposition 3.7 [1, Proposition 6.11] *Let D be a bridge diagram without any crossings. Then we can choose c_α on \mathcal{K}_α and c_β on \mathcal{K}_β such that for any $x \in \text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta)$, we have $\text{wt}(x) = \text{gr}(x)$.*

Recall the Khovanov homology of an unlink of k components U_k is

$$(3-6) \quad \text{Kh}^{*,*}(U_k) = \bigotimes^k (\mathbb{k}_{(0,1)} \oplus \mathbb{k}_{(0,-1)}).$$

With the choice of the equivariant structures by Abouzaid and Smith, we have

$$(3-7) \quad \text{Kh}_{\text{symp}}^{*,*}(U_k) = \bigotimes^k (\mathbb{k}_{(1,1)} \oplus \mathbb{k}_{(-1,-1)}).$$

This proposition is a special case of our main theorem with crossing number equal to 0, if we relate gradings (gr, wt) on symplectic Khovanov cohomology and (i, j) on Khovanov homology with the formula

$$(3-8) \quad i = \text{gr} - \text{wt},$$

$$(3-9) \quad j = -\text{wt}.$$

3.2 Floer product and weight grading

In [1, Section 3], Abouzaid and Smith pointed out an important fact that the weight grading is compatible with Floer products, without actually phrasing and proving the precise statement. Also in [12, Equation 4.9], Seidel and Solomon discussed the derivation property in a similar setup. We prove the following proposition:

Proposition 3.8 *Let $\mathcal{K}_0, \mathcal{K}_1$ and \mathcal{K}_2 be compact Lagrangians given by crossingless matchings in \mathcal{Y}_n . Then for any eigenvector $\alpha \in \text{HF}^*(\mathcal{K}_1, \mathcal{K}_2)$ and $\beta \in \text{HF}^*(\mathcal{K}_0, \mathcal{K}_1)$,*

$$(3-10) \quad \text{wt}(\mu^2(\alpha, \beta)) = \text{wt}(\alpha) + \text{wt}(\beta).$$

Proof Consider the boundary strata of the moduli space $\overline{\mathcal{R}}^3_{(0,1)}(x_0; x_1, x_2)$, where we have three boundary marked points and two interior marked points. As before, we exclude sphere bubbles. We can also exclude the cases when two interior marked points are split into two different components by the assumption (M-4). This is because both components will have positive intersection with D_0 and thus $[\mu] \cdot [D_0] \geq 2$, which contradicts (M-4).

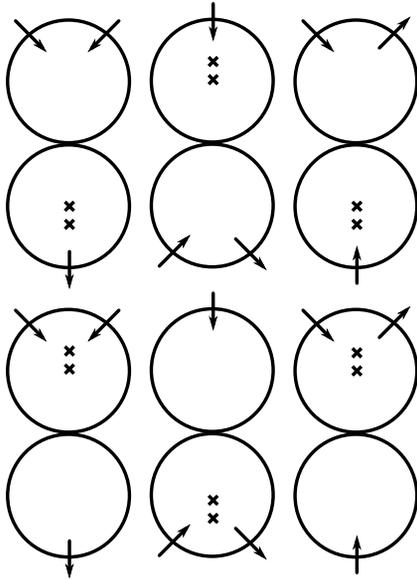


Figure 10: Six possible boundary bubbles. The incoming arrows are inputs and the outgoing ones are outputs.

With sphere bubbles excluded as discussed in Remark 3.2, there are still six possible types of degeneration; see Figure 10. Three degenerations shown in the first row compute

$$(3-11) \quad \phi(\mu^2(x_1, x_2)),$$

$$(3-12) \quad \mu^2(x_1, \phi(x_2)),$$

$$(3-13) \quad \mu^2(\phi(x_1), x_2).$$

The three degenerations shown in the second row compute

$$(3-14) \quad \mu^1\phi^2(x_1, x_2),$$

$$(3-15) \quad \phi^2(\mu^1(x_1), x_2),$$

$$(3-16) \quad \phi^2(x_1, \mu^1(x_2)).$$

If we pass to homology, those terms with μ^1 will vanish and thus we have the following relation by counting all the boundary components of the one-dimensional moduli space $\overline{\mathcal{R}}_{(0,1)}^2(x_0; x_1, x_2)$,

$$(3-17) \quad \phi(\mu^2(x_1, x_2)) = \mu^2(x_1, \phi(x_2)) + \mu^2(\phi(x_1), x_2).$$

This is equivalent to

$$(3-18) \quad \text{wt}(\mu^2(x_1, x_2))\mu^2(x_1, x_2) = \mu^2(x_1, \text{wt}(x_2)x_2) + \mu^2(\text{wt}(x_1)x_1, x_2) \\ = (\text{wt}(x_1) + \text{wt}(x_2))\mu^2(x_1, x_2),$$

which proves the result. □

Remark 3.9 By studying a similar setup with more boundary marked points, we can generalize the result above to higher Floer products. Specifically, when we consider the product of three elements, we have $\text{wt}(\mu^3(x_1, x_2, x_3)) = \text{wt}(x_1) + \text{wt}(x_2) + \text{wt}(x_3)$.

3.3 A long exact sequence of $\text{Kh}_{\text{symp}}^{*,*}(L)$

Abouzaid and Smith [2, Equation 7.9] constructed a long exact sequence from an exact triangle of bimodules over the Fukaya category of \mathcal{Y}_n ,

$$(3-19) \quad \dots \rightarrow \text{Kh}_{\text{symp}}^*(L_+) \rightarrow \text{Kh}_{\text{symp}}^*(L_0) \rightarrow \text{Kh}_{\text{symp}}^{*+2}(L_\infty) \rightarrow \dots,$$

where L_+ is a link diagram with a positive crossing and L_0 and L_∞ are diagrams given by 0 or ∞ resolutions at the positive crossing. The goal of this chapter is to give an explicit construction of such a long exact sequence within the framework of bridge diagrams that preserves the weight grading, just like the combinatorial Khovanov homology.

We use the following local diagrams for the computation: we name the blue curves β , the green curves γ and the yellow curves δ , respectively, in Figure 11, and the other components of γ and δ are small isotopies of the components of β such that there are no intersections in the interiors of the arcs. Thus, the Lagrangians $\mathcal{K}_\beta, \mathcal{K}_\gamma$ and \mathcal{K}_δ intersect pairwise transversely. (Here we moved the actual arcs rather than perturbing the Lagrangians to avoid issues we discussed in Remark 3.2.) Pairing α with β gives L_+ , pairing α with γ gives L_0 , and pairing α with δ gives L_∞ . We have the following exact sequence:

Proposition 3.10 [2, Proposition 7.4] *If we have α, β, γ and δ curves presented locally as in Figure 11 and β, γ and δ are the same apart from in this region, then we have the following exact sequence*

$$(3-20) \quad \dots \xrightarrow{c_1} \text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\beta) \xrightarrow{c_2} \text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\gamma) \xrightarrow{c_3} \text{HF}^{*+2}(\mathcal{K}_\alpha, \mathcal{K}_\delta) \xrightarrow{c_1} \dots$$

In particular, there are elements

$$c_1 \in \text{CF}^*(\mathcal{K}_\beta, \mathcal{K}_\delta), \quad c_2 \in \text{CF}^*(\mathcal{K}_\gamma, \mathcal{K}_\beta) \quad \text{and} \quad c_3 \in \text{CF}^*(\mathcal{K}_\delta, \mathcal{K}_\gamma)$$

such that the maps above are Floer products with the corresponding elements.

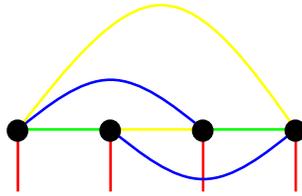


Figure 11: Exact triangle of Lagrangians. The red, blue, green and yellow curves are α , β , γ and δ curves, respectively.

Proof Abouzaid and Smith proved that there is an exact triangle of bimodules of the Fukaya category of \mathcal{Y}_n among the identity bimodule, the cup–cap bimodule and the bimodule representing a half-twist τ . Evaluating these bimodules at \mathcal{K}_γ as the second object, we have an exact triangle of one-sided modules between \mathcal{K}_γ , \mathcal{K}_δ and a one-sided module that is equivalent to $\text{HF}^*(\bullet, \mathcal{K}_\beta)$, such that the maps connecting those terms are Floer products with some elements $c_1 \in \text{CF}^*(\mathcal{K}_\beta, \mathcal{K}_\delta)$, $c_2 \in \text{CF}^*(\mathcal{K}_\gamma, \mathcal{K}_\beta)$ and $c_3 \in \text{CF}^*(\mathcal{K}_\delta, \mathcal{K}_\gamma)$. Evaluating these one-sided modules at \mathcal{K}_α , we have the desired long exact sequence. □

Now we need to show that this long exact sequence preserves the relative weight grading in the sense that each element summing to c_i has the same weight grading so that c_i has a well-defined weight grading, and moreover the weight gradings of c_1 , c_2 and c_3 sum to 0. With a closer look into the diagram, we have the following observation:

Lemma 3.11 *Each pair of β , γ and δ forms a bridge diagram for an unlink of $n - 1$ components without any crossings.*

Proof In the region shown in Figure 11, each pair of β , γ and δ forms a crossingless unknot. In the other regions not shown in the figure, each pair of arcs forms a crossingless unknot component as well. Thus we have one crossingless unknot component in Figure 11 and $n - 2$ crossingless unknot components outside that figure. Together, each pair of β , γ and δ forms a bridge diagram for an unlink of $n - 1$ components without any crossings. □

From the computations of [1, Proposition 5.12], we have:

Lemma 3.12 *With some grading shifts, we compute*

$$(3-21) \quad \text{CF}^*(\mathcal{K}_\beta, \mathcal{K}_\gamma) \cong \text{CF}^*(\mathcal{K}_\gamma, \mathcal{K}_\delta) \cong \text{CF}^*(\mathcal{K}_\delta, \mathcal{K}_\beta) \cong \bigotimes^{n-1} H^*(S^2),$$

and thus all generators are cocycles.

Now, we prove the following proposition about the weight grading on those groups:

Proposition 3.13 *For a fixed choice of equivariant structures on the Lagrangians \mathcal{K}_β , \mathcal{K}_γ and \mathcal{K}_δ , there are well-defined weight gradings for the elements $c_1 \in \text{CF}^*(\mathcal{K}_\beta, \mathcal{K}_\delta)$, $c_2 \in \text{CF}^*(\mathcal{K}_\gamma, \mathcal{K}_\beta)$ and $c_3 \in \text{CF}^*(\mathcal{K}_\delta, \mathcal{K}_\gamma)$.*

Proof From Lemma 3.12, we know that each of the Floer cochain complexes and the Floer cohomology groups above can be made so that weight grading equals homological grading. Together with the observation that choices of equivariant structures will only apply overall grading shifts to all weight gradings, thus we know the weight grading of any element in CF^n must be the same. If we fix a set of equivariant structures on \mathcal{K}_β , \mathcal{K}_γ and \mathcal{K}_δ , we have a well-defined weight grading for each c_i from its homological grading plus the effect a grading shift from changing the equivariant structure from the standard one. □

Before we prove $\text{wt}(c_1) + \text{wt}(c_2) + \text{wt}(c_3) = 0$, recall the following lemma from Seidel:

Lemma 3.14 [10, Lemma 3.7] *A triple $\mathcal{K}_\beta, \mathcal{K}_\gamma$ and \mathcal{K}_δ forms an exact triangle of Lagrangians if and only if there exist*

$$c_1 \in \text{CF}^1(\mathcal{K}_\beta, \mathcal{K}_\delta), \quad c_2 \in \text{CF}^0(\mathcal{K}_\gamma, \mathcal{K}_\beta), \quad c_3 \in \text{CF}^0(\mathcal{K}_\delta, \mathcal{K}_\gamma),$$

$$h_1 \in \text{HF}^0(\mathcal{K}_\gamma, \mathcal{K}_\beta), \quad h_2 \in \text{HF}^0(\mathcal{K}_\beta, \mathcal{K}_\delta), \quad k \in \text{HF}^{-1}(\mathcal{K}_\beta, \mathcal{K}_\beta)$$

such that

$$(3-22) \quad \mu^1(h_1) = \mu^2(c_3, c_2),$$

$$(3-23) \quad \mu^1(h_2) = -\mu^2(c_1, c_3),$$

$$(3-24) \quad \mu^1(k) = -\mu^2(c_1, h_1) + \mu^2(h_2, c_2) + \mu^3(c_1, c_3, c_2) - e_{\mathcal{K}_\beta},$$

where $e_{\mathcal{K}_\beta} \in \text{HF}^0(\mathcal{K}_\beta, \mathcal{K}_\beta)$ is the identity element of the Floer product

$$\mu^2: \text{HF}^*(\mathcal{K}_\beta, \mathcal{K}) \otimes \text{HF}^*(\mathcal{K}_\beta, \mathcal{K}_\beta) \rightarrow \text{HF}^*(\mathcal{K}_\beta, \mathcal{K}).$$

Lemma 3.15 *For any choice of equivariant structures on Lagrangians, the sum of weight gradings satisfies $\text{wt}(c_1) + \text{wt}(c_2) + \text{wt}(c_3) = 0$.*

Proof We know μ^1 vanishes on all the Floer groups above. If we prove that $h_1 = 0$ and $h_2 = 0$, then the equation (3-23) becomes $\mu^3(c_1, c_3, c_2) = e_{\mathcal{K}_\beta}$. Together with the fact that weight grading is compatible with Floer product, see Remark 3.9, we know that $\text{wt}(c_1) + \text{wt}(c_2) + \text{wt}(c_3) = \text{wt}(e_{\mathcal{K}_\beta})$. From [1, Lemma 4.10], no matter which equivariant structure we choose on \mathcal{K}_β , the identity always has the weight grading 0.

To see that $h_1 = 0$ and $h_2 = 0$, we need to look into the absolute grading. We claim that $\text{HF}^*(\mathcal{K}_\gamma, \mathcal{K}_\beta)$ and $\text{HF}^*(\mathcal{K}_\beta, \mathcal{K}_\delta)$ are supported in odd degrees. This is because pairing γ with β gives a flattened braid diagram for a braid σ with writhe 1. The number of strands is even and thus this group is supported in odd degrees. Similarly, pairing β with δ gives σ^{-1} , which also gives an odd-degree supported group. Notice that the third map c_3 should have degree 2 in the absolute grading case, thus the exact triangle is actually between $\mathcal{K}_\beta, \mathcal{K}_\gamma$ and \mathcal{K}_δ [2]. But shifting the grading of Floer groups by 2 does not change the fact that h_1 and h_2 have even degrees. Thus, they must be 0. \square

Thus we conclude the following:

Proposition 3.16 *The long exact sequence of equation (3-20),*

$$(3-25) \quad \dots \xrightarrow{c_1} \text{HF}^{*, \text{wt}_1}(\mathcal{K}_\alpha, \mathcal{K}_\beta) \xrightarrow{c_2} \text{HF}^{*, \text{wt}_2}(\mathcal{K}_\alpha, \mathcal{K}_\gamma) \xrightarrow{c_3} \text{HF}^{*+2, \text{wt}_3}(\mathcal{K}_\alpha, \mathcal{K}_\delta) \xrightarrow{c_1} \dots,$$

decomposes with respect to weight gradings.

Proof From Proposition 3.13, we know that there is a well-defined weight grading for c_1, c_2 and c_3 . If we start with the first group $\text{HF}^{*, \text{wt}_1}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$, the only nontrivial map will be at weight grading $\text{wt}_1 + \text{wt}(c_2)$ because the weight grading is compatible with Floer products. The same goes for $\text{wt}_3 = \text{wt}_1 + \text{wt}(c_3) + \text{wt}(c_2)$. The next weight grading should be $\text{wt}_1 + \text{wt}(c_2) + \text{wt}(c_3) + \text{wt}(c_1) = \text{wt}_1$, which is exactly where we started with the first group. \square

4 Proof of the main theorem via a bigraded isomorphism

In this section, we prove our main theorem, Theorem 1.3, through showing that our isomorphism is a bigraded refinement of the isomorphism in [2]. We start this section by rephrasing the main theorem of [2] as follows:

Proposition 4.1 [2, Theorem 7.5] *For any bridge diagram L , we have an isomorphism H between symplectic Khovanov cohomology and Khovanov homology,*

$$(4-1) \quad H: \text{Kh}_{\text{symp}}^*(L) \rightarrow \text{Kh}^*(L).$$

In [2], we can only conclude from the original argument of Abouzaid and Smith that H is canonical for braid closures. With Proposition 2.2, that symplectic Khovanov cohomology is defined canonically (and the same for combinatorial Khovanov homology), we can claim that H is canonical for any link diagram. This is an isomorphism with only information on the homological grading. But Abouzaid and Smith’s proof of the isomorphism between symplectic and ordinary Khovanov homology implies

that the long exact sequence in equation (3-20) commutes with the corresponding long exact sequence for Khovanov homology, as illustrated in the following lemma.

Lemma 4.2 *Fix a link diagram L_+ and its unoriented resolutions L_0 and L_∞ at one of the crossings. We represent their corresponding bridge diagrams with (α, β) , (α, γ) and (α, δ) arcs respectively such that α, β, γ and δ are locally shown in Figure 11. The isomorphism H is compatible with the exact sequence (3-20), ie the following diagram commutes:*

$$\begin{array}{ccccccc}
 \text{HF}^*(\mathcal{K}_\alpha, \mathcal{K}_\gamma) & \longrightarrow & \text{HF}^{*+2}(\mathcal{K}_\alpha, \mathcal{K}_\delta) & \longrightarrow & \text{HF}^{*+1}(\mathcal{K}_\alpha, \mathcal{K}_\beta) & & \\
 \downarrow H & & \downarrow H & & \downarrow H & & \\
 \text{Kh}^*(L_0) & \longrightarrow & \text{Kh}^{*+2}(L_\infty) & \longrightarrow & \text{Kh}^{*+1}(L_+) & & \\
 & & & & & \longrightarrow & \text{HF}^{*+1}(\mathcal{K}_\alpha, \mathcal{K}_\gamma) \longrightarrow \text{HF}^{*+3}(\mathcal{K}_\alpha, \mathcal{K}_\delta) \\
 & & & & & \downarrow H & \downarrow H \\
 & & & & & \longrightarrow & \text{Kh}^{*+1}(L_0) \longrightarrow \text{Kh}^{*+3}(L_\infty)
 \end{array}$$

where the upper row is the exact sequence of equation (3-20), and the lower row is the exact sequence for combinatorial Khovanov homology with grading $i - j$.

Proof In the proof of the Abouzaid–Smith isomorphism, they show that the cup functors of Khovanov and symplectic Khovanov are identified with the isomorphism in the arc algebra; see [2, Corollary 6.16]. Moreover, we know that the cap functors in both cases are adjoint to the corresponding cup functors from [2, Proposition 7.4]. The horizontal maps in the first squares are given by applying the same cap–cup functor in each case, and thus they commute with the isomorphisms in the first square.

For the other squares, the diagram naturally commutes if we replace the third group of each row with the mapping cone of the other two, given the fact that the first square is already commutative. Moreover, in the proof of [2, Theorem 7.6], the isomorphism H and the long exact sequences are constructed via the mapping cones, and thus factor through the cones of horizontal maps of the first square. Together with the fact that the third group of each row is isomorphic to a mapping cone of the other two (see [2, Proposition 7.4]), we conclude that the second and third squares are also commutative. □

Corollary 4.3 *The diagram involving the exact sequences of resolving a negative crossing is also commutative.*

Recall that we compared the bigradings of two theories at the end of Section 3.1 for unlinks.

Proposition 4.4 [1, Proposition 6.11] *The isomorphism H preserves the weight grading if D is a crossingless diagram, with grading correspondence $gr = i - j$ and $wt = -j$.*

We are now ready to prove the main theorem.

Proof of Theorem 1.3 We only need to prove that for any fixed Jones grading j_0 , H is also an isomorphism with $wt = -j_0 + c$ with some grading shift c ,

$$(4-3) \quad H: \text{Kh}_{\text{symp}}^{*, -j_0}(L) \rightarrow \text{Kh}^{*, j_0}(L).$$

We prove that this statement is true for any link (bridge) diagram by induction on the number of crossings. The base case for unlinks is proved with Proposition 4.4.

Now we assume that L_+ is a link diagram with n crossings, whereas its resolutions L_0 and L_∞ have $n - 1$ crossings. Let us assume we are resolving at a positive crossing. If all crossings are negative, a similar argument can be applied for the exact sequences induced by resolving at a negative crossing. By the inductive hypothesis, the maps H are bigraded isomorphisms for L_0 and L_∞ . Let us fix a Jones grading j_1 on $\text{Kh}(L_0)$. The maps in the long exact sequence will be trivial unless the Jones grading j_2 on $\text{Kh}(L_\infty)$ is $j_1 - 3v - 2$, where v is the signed count of crossings between the arc that ends at the left-most endpoint in Figure 11 and other components, and j_3 on $\text{Kh}(L_+)$ is $j_1 + 1$. Thus we can decompose our commutative diagram with respect to the Jones grading:

$$(4-4) \quad \begin{array}{ccccc} \text{HF}^{*, -j_1}(\mathcal{K}_\alpha, \mathcal{K}_\gamma) & \xrightarrow{c_2} & \text{HF}^{*+2, -j_2}(\mathcal{K}_\alpha, \mathcal{K}_\delta) & \xrightarrow{c_3} & \text{HF}^{*+1, j'_3}(\mathcal{K}_\alpha, \mathcal{K}_\beta) \\ \downarrow H & & \downarrow H & & \downarrow H \\ \text{Kh}^{*, j_1}(L_0) & \longrightarrow & \text{Kh}^{*+2, j_2}(L_\infty) & \longrightarrow & \text{Kh}^{*+1, j_3}(L_+) \\ & & & & \xrightarrow{c_1} \text{HF}^{*+1, -j_1}(\mathcal{K}_\alpha, \mathcal{K}_\gamma) \xrightarrow{c_2} \text{HF}^{*+3, -j_2}(\mathcal{K}_\alpha, \mathcal{K}_\delta) \\ & & & & \downarrow H \qquad \qquad \downarrow H \\ & & & & \longrightarrow \text{Kh}^{*+1, j_1}(L_0) \longrightarrow \text{Kh}^{*+3, j_2}(L_\infty) \end{array}$$

where the weight grading j'_3 is given by $-j_2 + \text{wt}(c_3)$. The first, second, fourth and fifth columns are all isomorphisms, so by the five lemma, we conclude that the third column is also an isomorphism and thus we know that the map H is also a bigraded

isomorphism for L_+ . As for the grading correspondence, because the c_i all have fixed weight grading after specifying the choice of equivariant structures, if we change j_1 by any number k , we change j_3 also by k . As for the first row, if $-j_1$ is changed into $-j_1 - k$, this will result in j'_3 shifting by $-k$ as well. This is enough to show that the weight grading recovers the Jones grading as relative gradings. \square

As a corollary of the theorem above, we also conclude [Theorem 1.5](#), that the relative weight grading is independent of the choice of link diagrams.

Proof of Theorem 1.5 For any two bridge diagrams D and D' representing a link L , relative weight gradings wt and wt' can be defined on D , and respectively D' . [Theorem 1.3](#) indicates that both wt and wt' coincide with $-j$ with as relative gradings. Thus wt and wt' are the same as relative gradings. \square

Lastly, we provide some insight into [Question 1.6](#). At the writing of this paper, the author cannot provide a pure symplectic proof of [Theorem 1.5](#) without referring to the invariance of the bigrading on Khovanov homology. Such a proof would consist of the invariance of weight grading under isotopy, handleslide and stabilization.

Isotopy and handleslide invariance could be potentially proved via [Proposition 3.8](#), if we look closely enough into the bridge diagrams and the Floer products. Isotopy and handleslide invariance can also be deduced from Hamiltonian isotopy invariance. But general Hamiltonian isotopy invariance will require virtual perturbations (see [[1](#), Remark 3.21]) because some of the transversality assumptions will not be preserved under general isotopies — the Lagrangian might bound some Maslov zero discs after isotopy, say.

The proof of stabilization invariance will be different from the other two invariance proofs. It requires some degeneration arguments in the Hilbert scheme setup, relating holomorphic discs in $\bar{\mathcal{Y}}_n$ and discs in $\bar{\mathcal{Y}}_{n-1}$ so that differentials in the stabilized diagram can be identified with differentials in the original diagram. One could extend the existing degeneration arguments for \mathcal{Y}_n before the partial compactification. In the original nilpotent slice setup, Seidel and Smith gave such an argument in [[11](#), Section 5.4]. However, the weight grading is defined using Hilbert schemes, but a stabilization invariance is never fully established in the Hilbert scheme setup. The author expects that one could potentially adapt the recent work of Mak and Smith [[6](#)] into the degeneration argument (or the so-called *neck-stretching argument*) of Hendricks, Lipshitz and Sarkar in [[4](#), Section 7.4.1], fixing the issue mentioned in their correction, and then generalize the argument to the partial compactification $\bar{\mathcal{Y}}_n$.

References

- [1] **M Abouzaid, I Smith**, *The symplectic arc algebra is formal*, Duke Math. J. 165 (2016) 985–1060 [MR](#) [Zbl](#)
- [2] **M Abouzaid, I Smith**, *Khovanov homology from Floer cohomology*, J. Amer. Math. Soc. 32 (2019) 1–79 [MR](#) [Zbl](#)
- [3] **J-M Droz, E Wagner**, *Grid diagrams and Khovanov homology*, Algebr. Geom. Topol. 9 (2009) 1275–1297 [MR](#) [Zbl](#)
- [4] **K Hendricks, R Lipshitz, S Sarkar**, *A flexible construction of equivariant Floer homology and applications*, J. Topol. 9 (2016) 1153–1236 [MR](#) [Zbl](#) Correction in J. Topol. 13 (2020) 1317–1331
- [5] **M Khovanov**, *A categorification of the Jones polynomial*, Duke Math. J. 101 (2000) 359–426 [MR](#) [Zbl](#)
- [6] **C Y Mak, I Smith**, *Fukaya–Seidel categories of Hilbert schemes and parabolic category \mathcal{O}* , J. Eur. Math. Soc. 24 (2022) 3215–3332 [MR](#) [Zbl](#)
- [7] **C Manolescu**, *Nilpotent slices, Hilbert schemes, and the Jones polynomial*, Duke Math. J. 132 (2006) 311–369 [MR](#) [Zbl](#)
- [8] **H Nakajima**, *Lectures on Hilbert schemes of points on surfaces*, University Lecture Series 18, Amer. Math. Soc., Providence, RI (1999) [MR](#) [Zbl](#)
- [9] **M Poźniak**, *Floer homology, Novikov rings and clean intersections*, from “Northern California symplectic geometry seminar” (Y Eliashberg, D Fuchs, T Ratiu, A Weinstein, editors), Amer. Math. Soc. Transl. Ser. 2 196, Amer. Math. Soc., Providence, RI (1999) 119–181 [MR](#) [Zbl](#)
- [10] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) [MR](#) [Zbl](#)
- [11] **P Seidel, I Smith**, *A link invariant from the symplectic geometry of nilpotent slices*, Duke Math. J. 134 (2006) 453–514 [MR](#) [Zbl](#)
- [12] **P Seidel, J P Solomon**, *Symplectic cohomology and q -intersection numbers*, Geom. Funct. Anal. 22 (2012) 443–477 [MR](#) [Zbl](#)
- [13] **J W Waldron**, *An invariant of link cobordisms from symplectic Khovanov homology*, preprint (2009) [arXiv 0912.5067](#)
- [14] **A Weinstein**, *Lagrangian submanifolds and Hamiltonian systems*, Ann. of Math. 98 (1973) 377–410 [MR](#) [Zbl](#)

School of Mathematics and Statistics, Wuhan University
Wuhan, Hubei, China

zcheng@whu.edu.cn

Received: 1 February 2021 Revised: 16 May 2022

Fibrations of 3–manifolds and asymptotic translation length in the arc complex

BALÁZS STRENNER

Given a 3–manifold M fibering over the circle, we investigate how the asymptotic translation lengths of pseudo-Anosov monodromies in the arc complex vary as we vary the fibration. We formalize this problem by defining normalized asymptotic translation length functions μ_d for every integer $d \geq 1$ on the rational points of a fibered face of the unit ball of the Thurston norm on $H^1(M; \mathbb{R})$. We show that, even though the functions μ_d themselves are typically nowhere continuous, the sets of accumulation points of their graphs on d –dimensional slices of the fibered face are rather nice and in a way reminiscent of Fried’s convex and continuous normalized entropy function. We also show that these sets of accumulation points depend only on the shape of the corresponding slice. We obtain a particularly concrete description of these sets when the slice is a simplex. We also compute μ_1 at infinitely many points for the mapping torus of the simplest pseudo-Anosov braid to show that the values of μ_1 are rather arbitrary. This suggests that giving a formula for the functions μ_d seems very difficult even in the simplest cases.

57M10, 57M50, 57M60; 11H06, 11P21

1. Introduction	4088
2. Background	4094
3. Asymptotic translation length via cycles in graphs	4097
4. Estimating the stashing sets	4107
5. Lemmas on cones, lattices and volumes	4115
6. Proof of the main theorem	4123
7. An example	4128
References	4141

1 Introduction

To every fibration $M \rightarrow S^1$ of a 3-manifold M over the circle, there is an associated element of $H^1(M; \mathbb{Z})$, the pullback of a generator of $H^1(S^1; \mathbb{Z}) \cong \mathbb{Z}$. The integral cohomology classes that correspond to fibrations of M are organized by the faces of the unit ball of the Thurston norm $\|\cdot\|$ on $H^1(M; \mathbb{R})$ [16]: a face \mathcal{F} can be *fibred*, in which case every integral point in the interior of the cone $\mathbb{R}_+\mathcal{F}$ corresponds to a fibration, or not fibred, in which case no integral point in $\mathbb{R}_+\mathcal{F}$ corresponds to a fibration.

An element $\phi \in H^1(M; \mathbb{Z})$ is *primitive* if it cannot be written in the form $k\phi'$ for some $\phi' \in H^1(M; \mathbb{Z})$ and integer $k \geq 2$. If an element $\phi \in H^1(M; \mathbb{Z})$ corresponds to a fibration, then it is primitive if and only if the fibers are connected. For any $\phi \in H^1(M; \mathbb{Q})$, denote by $\bar{\phi} \in H^1(M; \mathbb{Z})$ the unique primitive integral point on the ray $\mathbb{R}_+\phi$.

We can now state a classical result of Fried from 1982, which was the main motivation for this work. If M admits a complete finite-volume hyperbolic metric, then the monodromies of the fibrations are pseudo-Anosov mapping classes by Thurston's hyperbolization theorem; see Otal [13]. For a fibred face \mathcal{F} of M , define the *normalized entropy function*

$$\xi: \text{int}(\mathcal{F}) \cap H^1(M; \mathbb{Q}) \rightarrow \mathbb{R}_+$$

by the formula

$$(1-1) \quad \xi(\phi) = \|\bar{\phi}\| \cdot \log \lambda(\bar{\phi}),$$

where $\lambda(\bar{\phi})$ denotes the stretch factor of the pseudo-Anosov monodromy corresponding to $\bar{\phi}$. Fried [5, Theorem E] proves that the function ξ extends to a convex, continuous function to the interior of \mathcal{F} and $\xi(\phi) \rightarrow \infty$ as $\phi \rightarrow \partial\mathcal{F}$. The goal of this paper is to investigate analogous functions on the rational points of the fibred faces that are defined not in terms of the stretch factor but via another numerical invariant of pseudo-Anosov maps, the asymptotic translation length in the arc complex.

The *arc complex* $\mathcal{A}(S)$ of a connected punctured surface S is a simplicial complex whose vertices are isotopy classes of properly embedded essential arcs in S and whose simplices correspond to collections of disjoint arcs. For two vertices α and β of $\mathcal{A}(S)$, their distance $d_{\mathcal{A}}(\alpha, \beta)$ is defined as the minimal number of edges of a path in the 1-skeleton of $\mathcal{A}(S)$ that starts at α and ends at β . The *asymptotic translation length* of

a mapping class f in the arc complex is defined as

$$\ell_{\mathcal{A}}(f) = \liminf_{n \rightarrow \infty} \frac{d_{\mathcal{A}}(\alpha, f^n(\alpha))}{n},$$

where α is any arc. The number $\ell_{\mathcal{A}}(f)$ is a natural invariant encoding geometric information about the 3-manifold M : Futer and Schleimer [7] showed that it is proportional to the height and area of the boundary of the maximal cusp in M .

Based on work of Baik, Shin and Wu [3], we define the d -adic normalized asymptotic translation length function

$$\mu_d : \text{int}(\mathcal{F}) \cap H^1(M; \mathbb{Q}) \rightarrow \mathbb{R}_+$$

by the formula

$$(1-2) \quad \mu_d(\phi) = \|\bar{\phi}\|^{1+1/d} \cdot \ell_{\mathcal{A}}(\bar{\phi}),$$

where $\ell_{\mathcal{A}}(\bar{\phi})$ is defined as $\ell_{\mathcal{A}}(f)$, where f is the monodromy of the connected fiber corresponding to $\bar{\phi}$. In order for μ_d to be defined, the fibers of M have to be punctured, so M has to be a cusped 3-manifold. We will work under the stronger hypothesis that the fibered face \mathcal{F} is *fully punctured*, meaning that the singular set of every pseudo-Anosov monodromy in $\mathbb{R}_+\mathcal{F}$ is contained in the set of punctures of the fiber. (If this condition holds for one monodromy in $\mathbb{R}_+\mathcal{F}$, then it holds for all.)

A d -dimensional slice of a fibered face \mathcal{F} is an intersection $\mathcal{F} \cap \Sigma$, where Σ is a $(d+1)$ -dimensional linear subspace of $H^1(M; \mathbb{R})$ intersecting the interior of \mathcal{F} . The slice is *rational* if $\Sigma \cap H^1(M; \mathbb{Q})$ is dense in Σ .

Theorem 1.1 *Let M be a connected cusped 3-manifold that admits a complete finite-volume hyperbolic metric. Let \mathcal{F} be a fully punctured fibered face of the unit ball of the Thurston norm on $H^1(M; \mathbb{R})$. Suppose that $1 \leq d \leq \dim(H^1(M; \mathbb{R})) - 1$ and let Ω be a rational d -dimensional slice of \mathcal{F} . Consider $\text{Graph}(\mu_d|_{\Omega}) \subset \Omega \times \mathbb{R}$, the graph of the normalized asymptotic translation length function μ_d restricted to Ω .*

There is a continuous function $g : \text{int}(\Omega) \rightarrow \mathbb{R}_+$ such that $g(\phi) \rightarrow \infty$ as $\phi \rightarrow \partial\Omega$ and the set of accumulation points of $\text{Graph}(\mu_d|_{\Omega})$ is

$$\{(\omega, g(\omega)) : \omega \in \text{int}(\Omega)\}$$

if $d = 1$ and

$$\{(\omega, r) : \omega \in \text{int}(\Omega), 0 \leq r \leq g(\omega)\} \cup (\partial\Omega \times [0, \infty))$$

if $d \geq 2$.

In words, the set of accumulation points is the graph of g if $d = 1$ and the closure of the region under the graph of g if $d \geq 2$.

As an immediate corollary, we have:

Corollary 1.2 *If $M, \mathcal{F}, d \geq 2$ and Ω are as in Theorem 1.1, then $\mu_d|_\Omega$ is a nowhere-continuous function.*

In this sense, the functions μ_d are therefore very different from Fried’s function ξ , which is always continuous. Nevertheless, the properties of continuity and blowing up at the boundary still make an appearance in Theorem 1.1 for the bounding function g .

We derive a formula for g in Theorem 6.1. However, it is not clear from this formula whether g is always convex.

Question 1.3 Is the function g in Theorem 1.1 convex?

When Ω is a simplex, we are able to describe the function g explicitly. We will show in Lemma 6.2 that convexity holds in this case.

Theorem 1.4 *Let $M, \mathcal{F}, d, \Omega$ and g be as in Theorem 1.1. Suppose Ω is a simplex with vertices $\omega_1, \dots, \omega_{d+1}$ and define the reparametrization*

$$g^*(\alpha_1, \dots, \alpha_{d+1}) = g\left(\sum_{i=1}^{d+1} \alpha_i \omega_i\right)$$

of the function g by

$$\left\{(\alpha_1, \dots, \alpha_{d+1}) \mid \alpha_i > 0, \sum_{i=1}^{d+1} \alpha_i = 1\right\},$$

the interior of the standard simplex. Let Σ be the subspace spanned by Ω , let $\Lambda = \Sigma \cap H^1(M; \mathbb{Z})$ be the integral lattice in Σ and let vol_Λ be the translation-invariant volume form on Σ with respect to which Λ has covolume 1. Then

$$g^*(\alpha_1, \dots, \alpha_{d+1}) = \sqrt[d]{\frac{1}{O_d \cdot d! \cdot \text{vol}_\Lambda(\Sigma / \langle \omega_1, \dots, \omega_{d+1} \rangle_{\mathbb{Z}}) \cdot \prod_{i=1}^{d+1} \alpha_i}},$$

where O_d is a constant depending only on d .

In the case $d = 1$, we have $O_1 = 1$; therefore,

$$g^*(\alpha, 1 - \alpha) = \frac{1}{\text{vol}_\Lambda(\Sigma / \langle \omega_1, \omega_2 \rangle_{\mathbb{Z}}) \cdot \alpha(1 - \alpha)}.$$

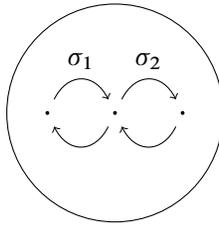


Figure 1: The half-twists σ_1 and σ_2 .

The constant O_d has a concrete interpretation: it is the smallest possible volume for a d -dimensional simplex σ in \mathbb{R}^d with the property that each larger scaled and translated copy of σ ($a\sigma + b$ with $a, b \in \mathbb{R}$ and $a > 1$) contains a point of \mathbb{Z}^d in its interior. Although determining the value of O_d for $d \geq 2$ seems to be an elementary lattice geometry question, we do not even know the value of O_2 .

Question 1.5 What is the value of the constant O_d for $d \geq 2$?

To shed some light on the exact values of the functions μ_d in addition to the accumulation points of their graphs, we compute μ_1 at infinitely many points for the mapping torus of the simplest hyperbolic braid. Both the answer and the proof are rather ad hoc, suggesting that it is very difficult to elegantly describe μ_d even in the simplest cases.

Theorem 1.6 Let M be the mapping torus of the pseudo-Anosov braid $f = \sigma_1\sigma_2^{-1}$ (read in either order) on three strands; see Figure 1. The fibered face \mathcal{F} containing f is 1-dimensional and f corresponds to the midpoint of \mathcal{F} . By choosing a linear identification of \mathcal{F} with $[-1, 1]$, we have $\mu_1(0) = \frac{8}{3}$ and

$$\mu_1(t) = \begin{cases} \frac{8}{3} & \text{if } t = \pm \frac{1}{2}, \\ 4 & \text{if } t = \pm \frac{1}{3}, \\ \frac{64}{13} & \text{if } t = \pm \frac{1}{4}, \\ 8/(1 + |t|)^2 & \text{if } t = \pm 1/k \text{ when } k \geq 5 \text{ is odd,} \\ 8/(1 + 2|t| - t^2) & \text{if } t = \pm 1/k \text{ when } k \geq 6 \text{ is even.} \end{cases}$$

Moreover,

$$\lim_{\mathbb{Q} \ni u \rightarrow t} \mu_1(u) = \frac{8}{1 - t^2}$$

for all $t \in (-1, 1)$. Therefore,

$$\mu_1(t) < \lim_{\mathbb{Q} \ni u \rightarrow t} \mu_1(u)$$

for $t = 0$ and all $t = \pm 1/k$ with $k \in \mathbb{Z}$ and $k \geq 2$, and μ_1 is discontinuous at all of these points.

In other words, the function μ_1 defined on the 1–dimensional fibered face in [Theorem 1.6](#) is discontinuous at every point where we have computed its value. We wonder if μ_1 is discontinuous at every rational point of every 1–dimensional slice. More generally:

Question 1.7 Suppose $M, \mathcal{F}, d, \Omega$ and $g: \text{int}(\Omega) \rightarrow \mathbb{R}_+$ are as in [Theorem 1.1](#). Does

$$\mu_d(x) < g(x)$$

hold for every rational point x in the interior of Ω ?

It would be interesting to generalize [Theorem 1.1](#) in various directions. For example, one could try to drop the hypothesis that the fibered face \mathcal{F} is fully punctured. Instead of the arc complex, one could also consider the curve complex and define the normalized asymptotic translation length functions analogously. Our proof has two key ingredients that are specific to the arc complex:

- Agol’s veering triangulation of 3–manifolds [\[1\]](#), and
- a theorem of Minsky and Taylor stating that there is a 1–Lipschitz retraction from the arc complex $\mathcal{A}(S)$ to the edges of the veering triangulation [\[12\]](#).

Generalizing [Theorem 1.1](#) to other cases would require replacing these technical tools with tools suitable in the other cases. Of [Sections 3, 4, 5](#) and [6](#), containing the proof of [Theorem 1.1](#), only [Section 3](#) relies crucially on veering triangulations. We use veering triangulations also for proving [Proposition 4.5](#), but, as we remark there, alternative approaches to analogous results already exist. The remaining parts of [Sections 4, 5](#) and [6](#) should generalize to other cases essentially without modifications.

Dependence only on shape

One interesting property of the functions μ_d is that, up to a constant factor, their bounding function g on any d –dimensional slice Ω only depends on the shape of Ω . This is in sharp contrast to Fried’s normalized entropy function ξ , which can take different forms even on 1–dimensional fibered faces.

Theorem 1.8 For $i = 1, 2$, let M_i be 3–manifolds as in [Theorem 1.1](#). Suppose $\mathcal{F}_i \subset H^1(M_i; \mathbb{R})$ are fibered faces of M_i and $\Omega_i \subset \mathcal{F}_i$ are d –dimensional slices for some integer $d \geq 1$. Let Σ_i be the span of Ω_i in $H^1(M_i; \mathbb{R})$ and consider the lattice $\Lambda_i = \Sigma_i \cap H^1(M_i; \mathbb{Z})$ in Σ_i . Let $g_i: \text{int}(\Omega_i) \rightarrow \mathbb{R}_+$ be the bounding functions for the functions $\mu_d^{\mathcal{F}_i}|_{\Omega_i}$ as in [Theorem 1.1](#).

If there is a linear isomorphism $i : \Sigma_1 \rightarrow \Sigma_2$ such that $i(\Omega_1) = \Omega_2$, then

$$g_2(i(\phi_1)) = \theta^{1/d} g_1(\phi_1)$$

holds for all $\phi_1 \in \text{int}(\Omega_1)$ for

$$\theta = \frac{\text{vol}(\Sigma_2/\Lambda_2)}{\text{vol}(\Sigma_2/i(\Lambda_1))},$$

where vol is any translation-invariant volume form on Σ_2 .

Related results

Kin and Shin [10] have shown that the function μ_1^C , defined analogously to μ_1 using the curve complex instead of the arc complex, is bounded from above on infinite subsets of slices arising from projecting an arithmetic progression in $H^1(M; \mathbb{Z})$ onto \mathcal{F} . Using this, they improved the upper bound of a result of Gadre and Tsai [8], stating that the minimal asymptotic translation length in the curve complex for pseudo-Anosov maps on the closed surface S_g of genus g is between C_1/g^2 and C_2/g^2 for some C_1 and C_2 . Using the bounds on μ_1^C , Kin and Shin [10] also provided upper bounds for the minimal asymptotic translation length in the curve complex for certain sequences of punctured surfaces (for this, see also Valdivia [17]), handlebody groups and hyperelliptic handlebody groups. It would be interesting to investigate the implications of our more explicit description of the function μ_1 on similar questions.

Baik, Shin and Wu [3] have studied the function μ_d^C , defined analogously to μ_d using the curve complex instead of the arc complex. They proved that the function μ_d^C is bounded from above on compact d -dimensional polytopes contained in the interior of \mathcal{F} . In Conjecture 1 of their paper, they conjecture that their bound is sharp in the sense that, for each $d \geq 2$, there exist M, \mathcal{F} and Ω such that the function μ_d^C is bounded away from 0 on an infinite subset of Ω . Although in the arc complex instead of the curve complex, our Theorem 1.1 verifies the stronger statement that μ_d is bounded away from zero on an infinite subset of Ω for all choices of M, \mathcal{F} and Ω . Moreover, in addition to showing that the values are bounded away from zero, Theorem 1.1 precisely specifies the values μ_d can approach along accumulating sequences in Ω .

For more related results involving the curve complex instead of the arc complex, see Baik, Kin, Shin and Wu [2].

Acknowledgements

We thank the referees for their thorough review and for pointing out an error in an earlier version of the paper.

2 Background

2.1 Fibrations over the circle

Let $\pi: M \rightarrow S^1$ be a fibering of M over the circle with fiber $S = \pi^{-1}(0)$. Let $\phi \in H^1(M)$ be the pullback of one of the two generators of $H^1(S^1) \cong \mathbb{Z}$. There is an infinite cyclic cover $S \times \mathbb{R} \rightarrow M$ corresponding to the homomorphism $\phi: \pi_1(M) \rightarrow \mathbb{Z}$. Let $h: S \times \mathbb{R} \rightarrow S \times \mathbb{R}$ be the element of the deck group that maps $S \times \{1\}$ to $S \times \{0\}$. The composition

$$S \times \{0\} \rightarrow S \times \{1\} \xrightarrow{h} S \times \{0\},$$

where the first map is the isotopy map $(x, 0) \mapsto (x, 1)$ in the product $S \times \mathbb{R}$, yields a homeomorphism of S . The mapping class of this homeomorphism is the *monodromy* f of the fibration. The monodromy depends on which of the two generators we pick for S^1 .

In other words, we can present the 3-manifold M as the quotient

$$M = (S \times \mathbb{R}) / \langle (x, t) \sim (\psi(x), t - 1) \rangle$$

for any homeomorphism $\psi: S \rightarrow S$ representing the mapping class f .

The map $((x, t), s) \mapsto (x, t + s)$ defines a flow $(S \times \mathbb{R}) \times \mathbb{R} \rightarrow S \times \mathbb{R}$ which is h -equivariant; therefore, it descends to a map $M \times \mathbb{R} \rightarrow M$, defining the *suspension flow* on M .

Finally, we will use the following conventions to make the discussions more intuitive. We picture the product $S \times \mathbb{R}$ so that the \mathbb{R} -coordinate axis is vertical and where ∞ is up and $-\infty$ is down. So, if $t_1 < t_2$, then we will say that the slice $S \times \{t_2\}$ is *above* $S \times \{t_1\}$ and $S \times \{t_1\}$ is *below* $S \times \{t_2\}$.

2.2 Pseudo-Anosov monodromies

When $\pi: M \rightarrow S^1$ is a fibration and M is hyperbolic, the monodromy f is pseudo-Anosov by Thurston's hyperbolization theorem. Let $\lambda^\pm \subset S$ be the invariant singular (unmeasured) foliations of f . We will refer to the foliation whose leaves are expanded by f as the *horizontal foliation* and the foliation whose leaves are contracted as the *vertical foliation*.

Since λ^\pm are invariant under the monodromy, their orbits under the suspension flow are singular 2-dimensional foliations Λ^\pm in M , transverse to the fibers, whose singular set is the suspension of the singular points of λ^\pm . Conversely, the foliations λ^\pm can be obtained from Λ^\pm by taking the intersection of Λ^\pm with the fibers.

2.3 Fully punctured fibered faces

Let M be a hyperbolic 3-manifold and let \mathcal{F} be a fibered face of the Thurston norm ball of $H^1(M; \mathbb{R})$. Every integral cohomology class in the interior of the cone $\mathbb{R}_+\mathcal{F}$ corresponds to a fibration of M over the circle.

Fried [5] (see also McMullen [11, Corollary 3.2]) showed that the suspension foliations Λ^\pm constructed from any two fibrations in this fibered cone are the same (up to isotopy) when the singular points of λ^\pm are all at punctures of the fiber for *some* fibration. It follows that in this case Λ^\pm do not have any singular points and therefore the singular points of λ^\pm are all at punctures of the fiber for *all* fibrations in this fibered cone. Such a fibered face \mathcal{F} is called *fully punctured*.

2.4 Relating different fibrations

From now on, suppose that \mathcal{F} is a fully punctured fibered face. The *maximal abelian cover* \tilde{M} of M is the cover corresponding to the natural homomorphism

$$\pi_1(M) \rightarrow G = H_1(M; \mathbb{Z})/\text{torsion}.$$

The foliations Λ^\pm in M lift to foliations $\tilde{\Lambda}^\pm$ in \tilde{M} . The suspension flow on M also lifts to a flow on \tilde{M} , leaving invariant the foliations $\tilde{\Lambda}^\pm$. The leaf space of this flow is homeomorphic to a surface \tilde{S} and the foliations $\tilde{\Lambda}^\pm$ define foliations $\tilde{\lambda}^\pm$ on \tilde{S} .

Every fiber of every fibration in the cone $\mathbb{R}_+\mathcal{F}$ is a quotient of the foliated surface $(\tilde{S}, \tilde{\lambda}^\pm)$ by a covering map. One can see this as follows. Let $\phi \in \mathbb{R}_+\mathcal{F}$ be a primitive integral point with monodromy f and fiber S with stable and unstable foliations λ^\pm . The covering $\tilde{M} \rightarrow M$ factors through the infinite cyclic covering $S \times \mathbb{R} \rightarrow M$ induced by $\phi: \pi_1(M) \rightarrow \mathbb{Z}$. The lift of S to $S \times \mathbb{R}$ is an infinite collection of parallel copies of S . Under the covering $\tilde{M} \rightarrow S \times \mathbb{R}$, each copy lifts to a surface intersecting every flowline of \tilde{M} exactly once. This gives rise to a foliation-preserving covering $(\tilde{S}, \tilde{\lambda}^\pm) \rightarrow (S, \lambda^\pm)$ whose deck group is the kernel of the homomorphism $G \rightarrow \mathbb{Z}$ induced by ϕ .

2.5 Veering triangulations

This section recalls some facts about veering triangulations of hyperbolic 3-manifolds defined by Agol [1], refined by Guéritaud [9] and further studied by Minsky and Taylor [12].

When the foliations λ^\pm for some fibration are endowed with the measure invariant under the pseudo-Anosov monodromy and this measure is lifted to $\tilde{\lambda}^\pm$, the measured foliations $\tilde{\lambda}^\pm$ endow the surface \tilde{S} with a singular Euclidean metric. Denote the metric completion of \tilde{S} by \hat{S} . Each completion point in $\hat{S} - \tilde{S}$ is an isolated point whose small neighborhood minus the completion point covers the neighborhood of a puncture in any fiber S . The metric on \tilde{S} depends on the fibration chosen in the construction, but the topology of \hat{S} does not. In the future, we will ignore the metric and consider \hat{S} together with the *unmeasured* foliations $\hat{\lambda}^\pm$ obtained from $\tilde{\lambda}^\pm$ by extending to the completion points.

A *singularity-free rectangle* in $(\hat{S}, \hat{\lambda}^\pm)$ is an immersion $[0, 1]^2 \rightarrow \hat{S}$ such that the vertical and horizontal foliations of $[0, 1]^2$ map to $\hat{\lambda}^\pm$ and the interior of the rectangle does not contain any completion point of $\hat{S} - \tilde{S}$. By the *interior* of the rectangle, we mean the image of $(0, 1)^2$ under the immersion. Similarly, by the *boundary* of the rectangle, we mean the image of the boundary of $[0, 1]^2$ under the immersion.

A singularity-free rectangle is *maximal* if all four sides of $[0, 1]^2$ contain the preimage of a completion point in their interior under the immersion map. (Each side may contain only one completion point, since the horizontal and vertical foliations, being invariant foliations of a pseudo-Anosov map, cannot have saddle connections.) By connecting each pair of the four points with an arc inside $[0, 1]^2$ and considering the image under the immersion map, we obtain six arcs in \hat{S} , forming a flattened tetrahedron in \hat{S} .

These arcs are defined only up to isotopy. To make the choice of the arcs canonical, we choose a fibration in our fibered cone and — as we have seen above — this choice endows \hat{S} with a singular Euclidean metric. We choose the arcs to be the unique geodesics in their isotopy class in this metric.

We think of the arc connecting the horizontal sides to be *above* the arc connecting the vertical sides. So the two triangles containing the arc connecting the vertical sides are the two bottom triangles and the remaining two triangles are the two top triangles of the tetrahedron.

Consider *all* maximal singularity-free rectangles in \hat{S} and all the arcs, triangles and tetrahedra they define through this process. For each triangle the smallest singularity-free rectangle containing it can be enlarged in two ways to a maximal singularity-free rectangle: we can enlarge the rectangle horizontally or vertically. In the former case, we obtain a tetrahedron that contains our triangle as one of the two top triangles. In the

latter case, we obtain a tetrahedron that contains our triangle as one of the two bottom triangles. Therefore the tetrahedra glue together in a layered fashion.

The links of the triangulation around the vertices are not spheres. Instead we glue up the *ideal* tetrahedra that do not include the vertices. With more work (see [9]), one can check that the links of the edges are circles, so the ideal tetrahedra glue up to a 3-manifold. Moreover, this 3-manifold is homeomorphic to $\widetilde{M} \cong \widetilde{S} \times \mathbb{R}$ and the ideal triangulation is called the *veering triangulation* of \widetilde{M} . The veering triangulation is invariant under the G -action, and the quotient is the veering triangulation of M .

We conclude by comparing the conventions regarding *above* and *below* introduced in this section versus the conventions introduced earlier. Recall from Section 2.1 that, for any fibration in our fibered cone, the deck transformation $h: S \times \mathbb{R} \rightarrow S \times \mathbb{R}$ satisfies $h(x, t) = (\psi(x), t - 1)$, where ψ is a pseudo-Anosov homeomorphism representing the monodromy mapping class. Recall also that our convention is that ψ expands horizontally and contracts vertically. Therefore the tetrahedra and the corresponding maximal singularity-free rectangles become wider and shorter as we go down in the product $\widetilde{S} \times \mathbb{R}$. This is consistent with the convention that the top edge of each tetrahedron, connecting the horizontal sides of the corresponding rectangle, has larger slope than the bottom edge, connecting the vertical sides.

3 Asymptotic translation length via cycles in graphs

3.1 Intersecting edges of the veering triangulation

Given an edge of the veering triangulation of $\widetilde{M} \cong \widetilde{S} \times \mathbb{R}$, its projection onto \widetilde{S} is an arc in \widetilde{S} . We say that two edges *intersect* if their projections intersect in \widetilde{S} . Otherwise we say that the two edges are *disjoint*. Recall that we have chosen these arcs to be geodesics in a singular Euclidean metric, so the arcs are automatically in minimal position and we do not need to be concerned about isotopies. Recall also that the edges do not have endpoints, so, if they intersect, they have to intersect in their interiors.

For our applications, it will be important to keep track of which pairs of edges of the veering triangulation of \widetilde{M} intersect and which two are disjoint. We can organize this information as follows.

Let E be the set of edges of the veering triangulation of M . The set E is finite, which follows from Agol's construction of the veering triangulations by periodic train track

sequences [1]. For each edge $e \in E$, choose a lift \tilde{e} in the veering triangulation of \tilde{M} . Denote the set of these lifts by \tilde{E} . Each edge of the veering triangulation of \tilde{M} can be uniquely written in the form $g\tilde{e}$ for some $g \in G$ and $\tilde{e} \in \tilde{E}$.

When two edges $g\tilde{e}$ and $g'\tilde{e}'$ intersect, one of the edges is *above* the other with respect to the pseudo-Anosov flow. If $g\tilde{e}$ is above $g'\tilde{e}'$, we write $g\tilde{e} > g'\tilde{e}'$. By our conventions, $g\tilde{e}$ is above $g'\tilde{e}'$ if they intersect and the smallest singularity-free rectangle containing $g'\tilde{e}'$ is wider and shorter than the smallest rectangle containing $g\tilde{e}$.

Definition 3.1 (stashing set) For any $e, e' \in E$, introduce the notation

$$\text{Stash}(e, e') = \{g \in G : \tilde{e} > g\tilde{e}'\}.$$

In words, $\text{Stash}(e, e')$ is the set of deck transformations in G that send (or *stash*) \tilde{e}' below \tilde{e} ; hence, we call $\text{Stash}(e, e')$ the *stashing set* of \tilde{e}' with respect to \tilde{e} .

The knowledge of the sets $\text{Stash}(e, e')$ for all pairs $e, e' \in E$ contains all disjointness information, since $g\tilde{e} > g'\tilde{e}'$ if and only if $g^{-1}g' \in \text{Stash}(e, e')$.

3.2 Frobenius numbers

We define the *Frobenius number* of a function $\beta: A \rightarrow \mathbb{Z}$ as

$$(3-1) \quad \text{Frob}(\beta) = \max(\mathbb{Z} - \beta(A))$$

if the maximum exists.

We remark that this notion is closely related to the Frobenius coin problem [15], which, given relatively prime positive integers a_1, \dots, a_n , asks for the largest integer that cannot be written as a linear combination of a_1, \dots, a_n with nonnegative integer coefficients. Indeed, let H be the free commutative monoid generated by x_1, \dots, x_n and let $\beta: H \rightarrow \mathbb{Z}$ be a homomorphism such that $\beta(x_1), \dots, \beta(x_n)$ are positive. Then the Frobenius number of β , as defined in (3-1), is the largest integer that cannot be written as a nonnegative integral linear combination of $\beta(x_1), \dots, \beta(x_n)$.

3.3 Translation length in the arc complex via graphs

To every primitive integral class ϕ in the interior of $\mathbb{R}_+\mathcal{F}$, we associate a weighted directed graph $W(\phi)$ on the vertex set E . There is an edge from e to e' if and only if there is at least one integer that is *not* contained in the subset

$$-\phi(\text{Stash}(e, e')) \cup \phi(\text{Stash}(e', e))$$

of \mathbb{Z} . Here ϕ stands for the surjective linear functional $G \rightarrow \mathbb{Z}$ associated to ϕ . If there is an edge from e to e' , then its weight $w(ee')$ is defined as the largest integer not contained in the subset $-\phi(\text{Stash}(e, e'))$ of \mathbb{Z} . Alternatively,

$$(3-2) \quad w(ee') = \text{Frob}(\phi|_{-\text{Stash}(e, e')}).$$

In Corollary 3.4, we will see that this largest integer always exists and therefore $w(ee')$ is always well defined.

Lemma 3.3 below will explain the information contained by the weighted graph $W(\phi)$. First we need the following lemma:

Lemma 3.2 *Any element of $-\phi(\text{Stash}(e, e'))$ is larger than any one of $\phi(\text{Stash}(e', e))$. In addition, if $-\phi(\text{Stash}(e, e')) \cup \phi(\text{Stash}(e', e))$ is not all of \mathbb{Z} , then the difference between the smallest element of $-\phi(\text{Stash}(e, e'))$ and the largest one of $\phi(\text{Stash}(e', e))$ is at least 2.*

Proof To prove the first statement, let $g_1, g_2 \in G$ be such that $\tilde{e} > g_1\tilde{e}'$ and $\tilde{e}' > g_2\tilde{e}$. Since the relation $>$ is transitive, we have $\tilde{e}' > g_1g_2\tilde{e}'$. One should think of ϕ as a height function: since $g_1g_2\tilde{e}'$ is below \tilde{e}' , we have $\phi(g_1g_2) = \phi(g_1) + \phi(g_2) < 0$. So $-\phi(g_1)$ is indeed larger than $\phi(g_2)$.

Assume that the difference is between the smallest element of $-\phi(\text{Stash}(e, e'))$ and the largest element of $\phi(\text{Stash}(e', e))$ is 1. Then there are g_1, g_2 such that $\tilde{e} > g_1\tilde{e}'$ and $\tilde{e}' > g_2\tilde{e}$ and $\phi(g_1g_2) = -1$. As before, we have $\tilde{e}' > g_1g_2\tilde{e}'$. So

$$\tilde{e} > g_1\tilde{e}' > g_1(g_1g_2)\tilde{e}' > g_1(g_1g_2)^2\tilde{e}' > \dots,$$

which means that every integer at least $-\phi(g_1)$ is contained in $-\phi(\text{Stash}(e, e'))$. Similarly, we obtain that every integer at most $\phi(g_2)$ is contained in $\phi(\text{Stash}(e', e))$. Since the gap between $-\phi(g_1)$ and $\phi(g_2)$ is 1, we have $-\phi(\text{Stash}(e, e')) \cup \phi(\text{Stash}(e', e)) = \mathbb{Z}$. This proves the second statement. \square

In the following lemma, S is the fiber of the fibration corresponding to ϕ , f is the monodromy and p_ϕ is the composition $\tilde{M} \cong \tilde{S} \times \mathbb{R} \rightarrow \tilde{S} \rightarrow S$.

Lemma 3.3 *There is an edge from e to e' in $W(\phi)$ if and only if there exists an integer k such that $p_\phi(\tilde{e})$ and $f^k(p_\phi(\tilde{e}'))$ are disjoint in S . Moreover, if there is an edge from e to e' , then its weight $w(ee')$ is the largest integer k such that $p_\phi(\tilde{e})$ and $f^k(p_\phi(\tilde{e}'))$ are disjoint in S .*

Proof One can check step by step that following are equivalent for any integer k :

- (1) The arcs $p_\phi(\tilde{e})$ and $f^k(p_\phi(\tilde{e}'))$ are disjoint in S .
- (2) The edge \tilde{e} is disjoint from all lifts of $f^k(p_\phi(\tilde{e}'))$ to \tilde{S} .
- (3) The edge \tilde{e} is disjoint from $g\tilde{e}'$ for all $g \in G$ with $\phi(g) = -k$.
- (4) $\tilde{e} \not\prec g\tilde{e}'$ and $\tilde{e} \not\prec g\tilde{e}'$ for all $g \in G$ with $\phi(g) = -k$.
- (5) $g \notin \text{Stash}(e, e')$ and $g^{-1} \notin \text{Stash}(e', e)$ for all $g \in G$ with $\phi(g) = -k$.
- (6) $-k \notin \phi(\text{Stash}(e, e'))$ and $k \notin \phi(\text{Stash}(e', e))$.
- (7) $k \notin -\phi(\text{Stash}(e, e')) \cup \phi(\text{Stash}(e', e))$.

By definition, there is an edge from e to e' in $W(\phi)$ if an integer k satisfying the last statement exists. The first statement of the lemma follows.

Using the equivalences again, the largest integer k such that $p_\phi(\tilde{e})$ and $f^k(p_\phi(\tilde{e}'))$ are disjoint in S is the largest k that is not contained in either $-\phi(\text{Stash}(e, e'))$ or $\phi(\text{Stash}(e', e))$. But, if there exists such a k , then by Lemma 3.2 it is the largest k that is not contained in $-\phi(\text{Stash}(e, e'))$. The second statement now follows from the definition of $w(ee')$. □

Corollary 3.4 For any pair $e, e' \in E$ and primitive integral class ϕ in the interior of $\mathbb{R}_+\mathcal{F}$, there exists some integer N such that $n \in -\phi(\text{Stash}(e, e'))$ for all $n > N$.

Proof We will show that there exists some N such that, for all $n > N$, the arcs $p_\phi(\tilde{e})$ and $f^n(p_\phi(\tilde{e}'))$ are not disjoint. Using the equivalences in the proof of Lemma 3.3, the statement will follow.

Let R and R' be the rectangles with horizontal and vertical sides whose diagonals are $p_\phi(\tilde{e})$ and $p_\phi(\tilde{e}')$, respectively. Let s' be a horizontal side of R' . The side s' is a starting segment of a horizontal separatrix emanating from a singularity. For any n , the segment $f^n(s')$ is the starting segment of a horizontal separatrix. There are finitely many horizontal separatrices and each such separatrix is dense in the surface. The map f stretches the surface horizontally by the stretch factor, so, if n is large enough, then $f^n(s')$ intersects the interior of R and consequently both horizontal sides of $f^n(R')$ intersect the interior of R and therefore $f^n(p_\phi(\tilde{e}'))$, the diagonal of $f^n(R')$, intersects $p_\phi(\tilde{e})$, the diagonal of R . □

We define the average weight of a cycle $\gamma = e_1 \dots e_n e_1$ in $W(\phi)$ as

$$\bar{w}(\gamma) = \frac{w(e_1 e_2) + \dots + w(e_{n-1} e_n) + w(e_n e_1)}{n}.$$

Proposition 3.5 (asymptotic translation length via weighted graphs) *For any primitive integral class ϕ in the interior of the cone $\mathbb{R}_+ F$, the asymptotic translation length in the arc complex of the pseudo-Anosov monodromy corresponding to ϕ is*

$$\ell_{\mathcal{A}}(\phi) = \frac{1}{\max\{\bar{w}(\gamma) : \gamma \text{ is a cycle in } W(\phi)\}}.$$

Proof First note that the maximum is indeed realized, since any cycle decomposes to minimal cycles and the average weight of the cycle is at most the average weight of the minimal cycle with the largest weight. The fact that there is at least one cycle in $W(\phi)$ will follow from the rest of the proof.

For any cycle $\gamma = e_1 \dots e_d e_1$, extend the sequence e_i for $i \geq d + 1$ such that $e_{i+d} = e_i$ for each integer $i \geq 1$. Consider the sequence

$$p_\phi(\tilde{e}_1), f^{w(e_1 e_2)}(p_\phi(\tilde{e}_2)), f^{w(e_1 e_2) + w(e_2 e_3)}(p_\phi(\tilde{e}_3)), \dots$$

of arcs in S . By Lemma 3.3, consecutive arcs are disjoint. Therefore, we have

$$d_{\mathcal{A}}(p_\phi(\tilde{e}_1), f^{\sum_{i=1}^{nd} w(e_i e_{i+1})}(p_\phi(\tilde{e}_1))) \leq nd$$

for any integer $n \geq 1$. This demonstrates that

$$\begin{aligned} \ell_{\mathcal{A}}(\phi) &= \lim_{n \rightarrow \infty} \frac{d_{\mathcal{A}}(p_\phi(\tilde{e}_1), f^{\sum_{i=1}^{nd} w(e_i e_{i+1})}(p_\phi(\tilde{e}_1)))}{\sum_{i=1}^{nd} w(e_i e_{i+1})} \\ &\leq \lim_{n \rightarrow \infty} \frac{nd}{\sum_{i=1}^{nd} w(e_i e_{i+1})} = \frac{1}{\bar{w}(\gamma)}. \end{aligned}$$

Since this inequality holds for any cycle γ , it follows that the left-hand side in the proposition is bounded from above by the right-hand side.

A key ingredient for the inequality in the reverse direction is a result of Minsky and Taylor [12, Theorem 1.4], which states that there is a 1-Lipschitz retraction from the arc complex $\mathcal{A}(S)$ to the set of arcs that are projections of the edges of the veering triangulation of \tilde{M} under p_ϕ . In particular, any two arcs $f^k(p_\phi(\tilde{e}))$ and $f^{k'}(p_\phi(\tilde{e}'))$ in S are joined by a geodesic in the arc complex $\mathcal{A}(S)$ whose vertices are all of the form $f^{k''}(p_\phi(\tilde{e}''))$.

Fix $\tilde{e}_1 \in \tilde{E}$ and let n be a positive integer. Denote the distance $d_{\mathcal{A}}(p_\phi(\tilde{e}_1), f^n(p_\phi(\tilde{e}_1)))$ by d_n ; then there is a sequence of arcs

$$f^{k_1}(p_\phi(\tilde{e}_1)), f^{k_2}(p_\phi(\tilde{e}_2)), \dots, f^{k_{d_n+1}}(p_\phi(\tilde{e}_{d_n+1}))$$

such that consecutive arcs are disjoint in S , $\tilde{e}_{d_n+1} = \tilde{e}_1$, the k_i are integers and $k_{d_n+1} - k_1 = n$. By Lemma 3.3, there is an edge from e_i to e_{i+1} in $W(\phi)$ and we have $k_{i+1} - k_i \leq w(e_i e_{i+1})$ for all $i = 1, \dots, d_n$. Summing these inequalities, we obtain $n \leq \sum_{i=1}^{d_n} w(e_i e_{i+1})$. After dividing both sides by d_n and taking reciprocals, we have

$$\frac{d_n}{n} \geq \frac{1}{\bar{w}(\gamma_n)},$$

where γ_n denotes the cycle $e_1 \dots e_{d_n} e_1$. In particular, this shows that there is at least one cycle in $W(\phi)$.

Note that

$$\ell_{\mathcal{A}}(\phi) = \lim_{n \rightarrow \infty} \frac{d_{\mathcal{A}}(p_{\phi}(\tilde{e}_1), f^n(p_{\phi}(\tilde{e}_1)))}{n} = \lim_{n \rightarrow \infty} \frac{d_n}{n} \geq \liminf_{n \rightarrow \infty} \frac{1}{\bar{w}(\gamma_n)}.$$

The right-hand side in the proposition is a lower bound for $1/\bar{w}(\gamma)$ for any cycle γ in $W(\phi)$; therefore, it is also a lower bound for $\ell_{\mathcal{A}}(\phi)$. This completes the proof of the reverse inequality. □

3.4 The graph Δ

In this section, we introduce a digraph Δ that serves as a model for the veering triangulation. We will use this graph to compute the weighted graphs $W(\phi)$ discussed in the previous section.

The vertices and edges of Δ correspond to the tetrahedra and the triangles, respectively, of the veering triangulation of M . The edge corresponding to a triangle t starts at the tetrahedron that has t as one of its two bottom triangles and ends at the tetrahedron that has t as one of its two top triangles. Note that every vertex has exactly two outgoing and two incoming edges.

There is a one-to-one correspondence from the tetrahedra to the edges of the veering triangulation that assigns to each tetrahedron its bottom edge. Using this correspondence, we can alternatively think about the vertices of Δ as edges of the veering triangulation. For an edge $e \in E$, Figure 2 illustrates the two other edges $e_1, e_2 \in E$ such that there is an edge of Δ from e_i to e . We can describe e_1 and e_2 as follows. Expand the smallest singularity-free rectangle containing e vertically as far as possible — the four singularities on the boundary of the resulting rectangle R define the tetrahedron T whose bottom edge is e . The edges e_1 and e_2 are the two edges of this tetrahedron that are neither the top nor the bottom edges such that the interiors of the rectangles

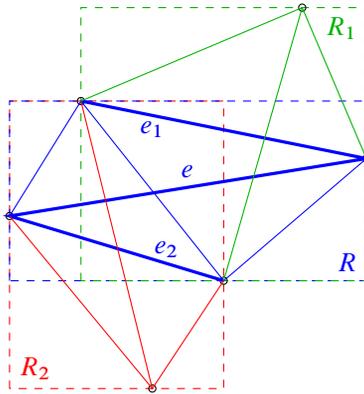


Figure 2: The edges e_1 and e_2 representing the vertices of Δ such that there is an edge from those vertices to the vertex represented by e .

R_1 and R_2 obtained by expanding the smallest singularity-free rectangle containing e_1 and e_2 vertically cover the interior of R . This is because e_1 and e_2 are the bottom edges of tetrahedra T_1 and T_2 determined by R_1 and R_2 , respectively, and both T_1 and T_2 have a bottom triangle that is a top triangle of T .

We also label each edge of Δ by an element of G , called the *drift* of the edge. To do this, choose a lift \tilde{t} of the triangle t corresponding to an edge of Δ in the veering triangulation of \tilde{M} . If the bottom edges of the tetrahedra right above and right below \tilde{t} are $g\tilde{e}$ and $g'\tilde{e}'$, respectively, then the drift of the edge of Δ corresponding to t is $g^{-1}g'$. The drift measures how much the coefficient of the edge $g\tilde{e}$ changes as we proceed to the other edge $g'\tilde{e}'$. Note that this definition is independent of the choice of the lift \tilde{t} .

Recall that a digraph is strongly connected if there is a path from any vertex to any other vertex. We will need the following lemma later:

Lemma 3.6 *The graph Δ is strongly connected.*

Proof Since pseudo-Anosov homeomorphisms of surfaces have dense orbits [14], there is a dense flowline in M . A dense flowline visits every tetrahedron of the veering triangulation infinitely often. Associated to this flowline is a bi-infinite path in Δ visiting every vertex infinitely often. This shows that Δ is strongly connected. \square

3.5 The extended graph Δ^*

We also define a graph Δ^* , obtained by adding some more labeled edges to Δ , one for each tetrahedron of the veering triangulation of M . For each tetrahedron T , we create

an edge from the top edge to the bottom edge of T . To define the label of this edge, choose a lift of T in the veering triangulation of \tilde{M} . If the bottom and top edges of the lift are $g_1\tilde{e}_1$ and $g_2\tilde{e}_2$, then our edge points from e_2 to e_1 and has drift $g_2^{-1}g_1$. This definition is also independent of the choice of the lift.

To distinguish between the edges of Δ and the edges of Δ^* that are not in Δ , we will call the two types of edges *triangle edges* and *tetrahedron edges*, respectively, as a reminder that they correspond to triangles and tetrahedra of the veering triangulation.

3.6 Computing the stashing sets

In this section we explain how the stashing sets $\text{Stash}(e, e')$ can be computed from the digraph Δ^* . We begin with a few definitions.

A *path* in the digraph Δ^* is a sequence of edges $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ of Δ^* with $n \geq 1$ such that the endpoint of ε_i is the same as the starting point of ε_{i+1} for all $i = 1, \dots, n - 1$. We caution the reader that we cannot refer to edges of Δ^* simply by their endpoints, since there might be multiple edges between vertices; see [Figure 11](#), for example.

The *drift* of a path is the product of the drifts of the edges of the path. Formally, the drift of the path $\varepsilon_1\varepsilon_2 \dots \varepsilon_n$ is

$$\prod_{i=1}^n \text{drift}(\varepsilon_i) \in G,$$

where $\text{drift}(\varepsilon_i) \in G$ denotes the drift of the edge ε_i .

A *good path* is a path whose first edge is a tetrahedron edge and whose remaining edges are triangle edges.

Proposition 3.7 (stashing sets via good paths) *We have*

$$\text{Stash}(e, e') = \{\text{drift}(\gamma) : \gamma \text{ is a good path from } e \text{ to } e' \text{ in } \Delta^*\}.$$

Proof First we show that the right-hand side contains the left-hand side. Suppose $g' \in \text{Stash}(e, e')$, which means that $\tilde{e} > g'\tilde{e}'$. Let p be the intersection of the images of \tilde{e} and $g'\tilde{e}'$ in \tilde{S} under the projection $\tilde{M} \rightarrow \tilde{S}$ by collapsing the flowlines. The preimage of p is a flowline that intersects \tilde{e} and $g'\tilde{e}'$. Consider the subinterval I of this flowline between \tilde{e} and $g'\tilde{e}'$.

If I does not intersect an edge of the veering triangulation aside from its endpoints, then it passes through a sequence of tetrahedra in a way that it enters each subsequent

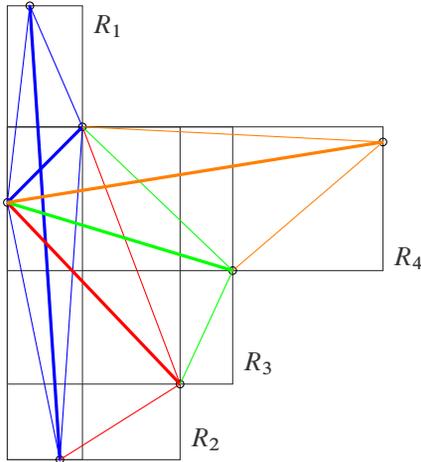


Figure 3: The edges of the veering triangulation of \tilde{M} corresponding to a good path in Δ^* . Every edge after the first one (the blue edge with the largest slope) is below the first edge.

tetrahedron through the interior of a face. Denote the bottom edges of these tetrahedra by $g_1\tilde{e}_1, \dots, g_k\tilde{e}_k$, where $g_1\tilde{e}_1$ is the bottom edge of the tetrahedron whose top edge is \tilde{e} and $g_k\tilde{e}_k = g'\tilde{e}'$. Then there is a tetrahedron edge in Δ^* from e to e_1 with drift g_1 and a triangle edge from e_i to e_{i+1} for all $i = 1, \dots, k - 1$ with drift $g_i^{-1}g_{i+1}$. Hence there is indeed a good path in Δ^* from e to $e_k = e'$ with drift $g_k = g'$.

If I does intersect an edge of the veering triangulation aside from its endpoints, then we can perturb I slightly so that it passes from one tetrahedron to the next through the interior of a face. From this sequence of tetrahedra, we obtain a good path with drift g just like in the previous case. The good path we get depends on how the perturbation is done, but any of them works for our purposes.

For the other direction, let $\varepsilon_1\varepsilon_2 \dots \varepsilon_n$ be a good path in Δ^* . Let e_k be the endpoint of ε_k for all $k = 1, \dots, n$ and let e_0 be the starting point of ε_1 . For all $k = 1, \dots, n$, let R_k be the maximal singularity-free rectangle obtained by expanding the smallest singularity-free rectangle containing $\text{drift}(\varepsilon_1 \dots \varepsilon_k)\tilde{e}_k$ vertically as far as possible. Note that \tilde{e}_0 and $\text{drift}(\varepsilon_1)\tilde{e}_1$ are the top and bottom edges of the tetrahedron corresponding to R_1 . In particular, \tilde{e}_0 intersects R_1 at its two horizontal sides. Each R_{k+1} is shorter and wider than R_k , so, by induction, we see that \tilde{e}_0 also intersects R_n at its two horizontal sides (Figure 3). Therefore, $\text{drift}(\varepsilon_1 \dots \varepsilon_n)\tilde{e}_n$, being the arc connecting the vertical sides of R_n , is indeed below \tilde{e}_0 . Hence, $\text{drift}(\varepsilon_1 \dots \varepsilon_n) \in \text{Stash}(e_0, e_n)$. \square

The drift of a path in Δ^* is independent of the order of edges of a path. So it will often be useful to think of a path in Δ^* as a nonnegative integer-valued function on the edges of Δ^* , where the value on each edge is the number of times that edge appears in the path. This viewpoint allows us to define the *sum* of two paths by taking the sum of the corresponding functions.

By a *cycle* in the digraph Δ^* , we mean a path $\varepsilon_1 \dots \varepsilon_n$ consisting of triangle edges such that the starting point of ε_1 coincides with the endpoint of ε_n . A *minimal cycle* is a cycle that traverses every vertex at most once. A *minimal good path* in Δ^* is a good path $\varepsilon_1 \dots \varepsilon_n$ such that the endpoints of ε_i are pairwise distinct for $i = 1, \dots, n$. It is allowed, however, that the starting point of ε_1 coincides with one of the other vertices traversed. There are finitely many minimal cycles and minimal good paths in Δ^* .

Let $\gamma = \varepsilon_1 \dots \varepsilon_n$ be a good path traversing the vertices e_0, \dots, e_n and let $\gamma_1, \dots, \gamma_k$ be cycles in Δ^* . We call the collection of paths $\gamma, \gamma_1, \dots, \gamma_k$ *connected* if

- (1) the triangle edges appearing in these paths (that is, every edge other than ε_1) form a connected subgraph of Δ^* when the orientations of the edges are ignored, and
- (2) at least one cycle γ_i traverses the vertex e_1 when $n = 1$.

We have the following decomposition lemma of good paths:

Lemma 3.8 (decompositions of good paths) *The sum of a connected collection of a minimal good path and a finite number of minimal cycles in Δ^* is a good path. Conversely, every good path in Δ^* can be written as such a sum.*

Proof To prove the first statement, we build up the sum step by step, adding one minimal cycle at a time. Denote by η_0 the minimal good path of the collection. There must be a minimal cycle from the collection that forms a connected union together with η_0 . Their sum η_1 is a good path. Then there must be another minimal good cycle from the collection that forms a connected union with η_1 . Their sum is a good path η_2 . Repeating this process until all cycles are added, we obtain the first statement.

For the second statement, let $\gamma = \varepsilon_1 \dots, \varepsilon_n$ be a good path in Δ^* . If it is minimal, we are done. If it is not minimal, then there are $1 \leq i < j \leq n$ such that the endpoints of ε_i and ε_j agree. Moreover, we can choose i and j so that $j - i$ is as small as possible. Then the subpath $\varepsilon_{i+1} \dots \varepsilon_j$ is a minimal cycle and γ can be written as a sum of this minimal cycle and a good path shorter than γ . We can repeat this process of removing

minimal cycles until the remaining good path is minimal. It is straightforward to verify that the collection of summands is connected. Hence, we obtain the second statement. \square

An immediate corollary of Proposition 3.7 and Lemma 3.8 is the following:

Corollary 3.9 *Denote by $\mathcal{P}_{e,e'} \subset G$ the set of drifts of minimal good paths from e to e' and by $\mathcal{B} \subset G$ the set of drifts of the minimal cycles of Δ^* . The set $\text{Stash}(e, e')$ is the set of products $pb_1^{\alpha_1} \dots b_k^{\alpha_k}$, where $p \in \mathcal{P}_{e,e'}$, $k \geq 0$ is an integer, $b_1, \dots, b_k \in \mathcal{B}$, $\alpha_1, \dots, \alpha_k$ are positive integers and p, b_1, \dots, b_k are drifts of a minimal good path and minimal cycles that form a connected collection.*

We will use Corollary 3.9 to compute exact values of the asymptotic translation length in the arc complex in Section 7. For the proof of Theorem 1.1, the following approximation of the stashing sets will be more convenient:

Corollary 3.10 *Denote by $\mathcal{P}_{e,e'} \subset G$ the set of drifts of minimal good paths from e to e' and by $\mathcal{B} \subset G$ the set of drifts of the minimal cycles of Δ^* . Furthermore, let $\mathcal{P}'_{e,e'} = \mathcal{P}_{e,e'} \prod_{b \in \mathcal{B}} b$. Then*

$$\mathcal{P}'_{e,e'} \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} \subset \text{Stash}(e, e') \subset \mathcal{P}_{e,e'} \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}},$$

where $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ denotes the monoid generated by \mathcal{B} .

By the product of two sets X and Y , we mean

$$XY = \{xy : x \in X, y \in Y\}.$$

Proof The second containment is a trivial consequence of Corollary 3.9. The first containment follows from Corollary 3.9 and the fact the union of any minimal good path from e to e' with all the minimal cycles is always a connected collection. This is because, by Lemma 3.6, the graph of triangle edges is strongly connected, so the union of all cycles or, equivalently, the union of all minimal cycles is a strongly connected graph containing all vertices. \square

4 Estimating the stashing sets

4.1 Monoids and cones

We begin this section by proving some general lemmas. We will use these lemmas to estimate the stashing sets at the end of the section.

For any $\mathcal{B} \subset \mathbb{R}^n$ and $\mathcal{E} \subset \mathbb{R}$, we introduce the notation

$$\langle \mathcal{B} \rangle_{\mathcal{E}} = \left\{ \sum_{i=1}^k \eta_i b_i : b_i \in \mathcal{B}, \eta_i \in \mathcal{E} \right\}$$

for the set generated by \mathcal{B} with coefficients in \mathcal{E} . (The empty sum is allowed in the definition and it is defined to be zero.) For example, $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ is the cone generated by \mathcal{B} and $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ is the monoid generated by \mathcal{B} .

Lemma 4.1 *Let $\mathcal{B} \subset \mathbb{Z}^n$ be a finite set and let $C = \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ be the cone generated by \mathcal{B} in \mathbb{R}^n . Then there exists some $x^* \in \mathbb{Z}^n$ such that*

$$\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} \cap (x^* + C) = \langle \mathcal{B} \rangle_{\mathbb{Z}} \cap (x^* + C).$$

In words, the lemma says that the sets $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ and $\langle \mathcal{B} \rangle_{\mathbb{Z}}$ are equal inside the translated cone $x^* + C$. From this viewpoint, it is clear that the lemma also holds for any element of the cone $x^* + C$ instead of x^* .

Proof The left-hand side is clearly contained in the right-hand side for any $x^* \in \mathbb{Z}^n$. We will find some $x^* \in \mathbb{Z}^n$ such that the reverse containment also holds. Let $\mathcal{B} = \{b_1, \dots, b_m\}$ and consider the compact subset $K = \{ \sum_{i=1}^m \kappa_i b_i : 0 \leq \kappa_i \leq 1 \}$ of C . Each element of $K \cap \langle \mathcal{B} \rangle_{\mathbb{Z}}$ can be represented in the form $\sum_{i=1}^m \eta_i b_i$ with $\eta_i \in \mathbb{Z}$. By choosing such an expression for each element, we may choose a positive integer η^* so that all η_i that appear in these finitely many representations satisfy $-\eta^* \leq \eta_i$.

We claim that the reverse containment in the lemma holds for $x^* = \eta^* \sum_{i=1}^m b_i$. To see this, let $x \in \langle \mathcal{B} \rangle_{\mathbb{Z}} \cap (x^* + C)$. Since $x \in x^* + C$, we have $x = \sum_{i=1}^m \alpha_i b_i$ for $\alpha_i \in \mathbb{R}$ and $\alpha_i \geq \eta^*$. We can rewrite this representation of x as

$$x = \sum_{i=1}^m [\alpha_i] b_i + \sum_{i=1}^m \{\alpha_i\} b_i,$$

where $[\alpha_i]$ and $\{\alpha_i\}$ denote the integer and fractional parts of α_i . The first of the two terms on the right is in $\langle \mathcal{B} \rangle_{\mathbb{Z}}$ and so is x , therefore the second term on the right is also in $\langle \mathcal{B} \rangle_{\mathbb{Z}}$. It is also in K ; therefore, we can replace it by $\sum_{i=1}^m \eta_i b_i$, where $\eta_i \in \mathbb{Z}$ and $\eta_i \geq -\eta^*$. Since $[\alpha_i] \geq \eta^*$, we obtain a representation of x as a sum of the b_i with nonnegative integer coefficients. Therefore, $x \in \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ and the right-hand side in the lemma is indeed contained in the left-hand side. □

Lemma 4.2 *Let $n \geq 2$ and let $\mathcal{B} \subset \mathbb{Z}^n$ be a finite set. Let $\mathcal{D} \subset \text{Hom}(\mathbb{R}^n, \mathbb{R})$ such that the cone $\langle \mathcal{D} \rangle_{\mathbb{R}_{\geq 0}}$ has nonempty interior in the n -dimensional vector space $\text{Hom}(\mathbb{R}^n, \mathbb{R})$. Assume that $\text{Frob}(\phi|_{\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}}) < \infty$ for every primitive integral point ϕ in $\langle \mathcal{D} \rangle_{\mathbb{R}_{\geq 0}}$. Then the following statements hold:*

- (i) $\langle \mathcal{B} \rangle_{\mathbb{Z}} = \mathbb{Z}^n$.
- (ii) The cone $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ has nonempty interior.
- (iii) There exists $x \in \mathbb{Z}^n$ such that $\mathbb{Z}^n \cap (x + \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}) \subset \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$.

Proof We begin by proving (i). If $\langle \mathcal{B} \rangle_{\mathbb{Z}}$ was a proper subgroup of \mathbb{Z}^n , then there would exist a surjective homomorphism $\phi: \mathbb{Z}^n \rightarrow \mathbb{Z}$ such that $\phi(\langle \mathcal{B} \rangle_{\mathbb{Z}})$ is not all of \mathbb{Z} . Since ϕ is surjective, it is a primitive integral point in $\text{Hom}(\mathbb{R}^n, \mathbb{R})$.

Let ϕ_0 be any primitive integral point in the interior of $\langle \mathcal{D} \rangle_{\mathbb{R}_{\geq 0}}$ that is not a scalar multiple of ϕ . For a large enough positive integer N , $N\phi_0 + \phi$ is in the interior of $\langle \mathcal{D} \rangle_{\mathbb{R}_{\geq 0}}$. We may choose a basis for the 2-dimensional lattice obtained as the intersection of $\text{Hom}(\mathbb{Z}^n, \mathbb{Z})$ with the 2-dimensional subspace spanned by ϕ_0 and ϕ in $\text{Hom}(\mathbb{R}^n, \mathbb{R})$ such that ϕ_0 has coordinates $(0, 1)$. Let (q, r) be the coordinates of ϕ . Note that $q \neq 0$ and q and r are relatively prime. From this, we see that, for any integer N , the point $\psi_N = Nq\phi_0 + \phi$ is primitive since it has coordinates $(q, r + Nq)$. Since $\phi(\langle \mathcal{B} \rangle_{\mathbb{Z}})$ is not all of \mathbb{Z} , there is some integer $d \geq 2$ that divides every element of this image. Choosing $N = ad$ for some large positive integer a , we can see that all elements of $\psi_{ad}(\langle \mathcal{B} \rangle_{\mathbb{Z}})$ are divisible by d . But this contradicts the fact that $\text{Frob}(\psi_{ad}|_{\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}}) < \infty$.

The statement (ii) is a straightforward corollary of (i).

By Lemma 4.1, there is some $x \in \mathbb{Z}^n$ such that

$$\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} \cap (x + \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}) = \langle \mathcal{B} \rangle_{\mathbb{Z}} \cap (x + \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}).$$

Using that $\langle \mathcal{B} \rangle_{\mathbb{Z}} = \mathbb{Z}^n$ from (i) and that the left-hand side is contained in $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$, we obtain (iii). □

Lemma 4.3 *For all $e \in E$, we have $\text{Stash}(e, e) \subset \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$.*

Proof Observe that, for any tetrahedron edge of Δ^* from e to e' , there is a path of triangle edges from e to e' with the same drift. To see this, choose a flowline close to e that does not intersect any edges of the veering triangulation. This flowline has a subarc that starts at the tetrahedron whose bottom edge is e and ends at the tetrahedron whose bottom edge is e' and whose top edge is e . Between the two tetrahedra, the

flowline intersects a sequence of tetrahedra. This sequence defines a path from e to e' with the required properties. As a consequence, for any good path in Δ^* starting and ending at the same vertex, there is a cycle in Δ with the same drift. So the statement follows by [Proposition 3.7](#). \square

We are now equipped with the tools to prove the following statement, describing the structure of the stashing sets:

Proposition 4.4 *Let $\mathcal{B} \subset G$ be the set of drifts of the minimal cycles of Δ^* . Then the cone $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ has nonempty interior and there exists $g \in G$ such that*

$$G \cap g \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \subset \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}.$$

Moreover, for every $e, e' \in E$, there exist $g_1, g_2 \in G$ such that

$$G \cap g_1 \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \subset \text{Stash}(e, e') \subset G \cap g_2 \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}.$$

Proof By [Corollary 3.4](#),

$$(4-1) \quad \text{Frob}(\phi|_{\text{Stash}(e,e)}) < \infty$$

for every primitive integral point ϕ in $-\mathbb{R}_+\mathcal{F}$. If we replace $\text{Stash}(e, e)$ with the set $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ (which is larger by [Lemma 4.3](#)) in (4-1), the statement remains true. Therefore, we may use [Lemma 4.2](#) with $G \cong \mathbb{Z}^n$ and any generator set \mathcal{D} for the cone $-\mathbb{R}_+\mathcal{F}$ to obtain that $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ has nonempty interior and that there exists $g \in G$ such that

$$(4-2) \quad G \cap g \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \subset \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}.$$

Moreover, we obtain that there are $g_1, g_2 \in G$ such that

$$G \cap g_1 \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \subset \mathcal{P}'_{e,e'} \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} \subset \text{Stash}(e, e') \subset \mathcal{P}_{e,e'} \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} \subset G \cap g_2 \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}},$$

where the first containment follows from (4-2), the second and third containments were shown in [Corollary 3.10](#), and the last containment follows from the simple observation that g_2 can be chosen so that $g_2^{-1} \mathcal{P}_{e,e'} \subset \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ since $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ has nonempty interior. \square

4.2 Duality of cones

The goal of this section is to prove [Proposition 4.5](#) below, which states that the cone over the fibered face consists of precisely those cohomology classes that take nonpositive values on the cone $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$.

Proposition 4.5 *Let $\mathcal{B} \subset G$ be the set of drifts of the minimal cycles of Δ^* . Then the interior of the cone $\mathbb{R}_+\mathcal{F} \subset H^1(M; \mathbb{R})$ can be described as*

$$\{\phi \in H^1(M; \mathbb{R}) : \phi(x) < 0 \text{ for all } x \in \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}\}.$$

We remark that an analogous statement was proven by Fried [6, Theorem D]. In that paper, Fried defined the set of *homology directions* of a flow on an n -dimensional closed manifold and showed that the integral cohomology classes that correspond to a fibration of the manifold over the circle are exactly the ones that take positive values on the set of homology directions. Unfortunately, Fried’s proof assumes that the manifold is closed, so the theorem cannot be directly applied in our case. In the end of the introduction, Fried mentions that under appropriate hypotheses the results carry over also to compact manifolds by doubling the manifold along the boundary, but details are not given. In order to make the proof of Proposition 4.5 as transparent as possible, instead of extending Fried’s theorem to the nonclosed case and relating the set of homology directions to our cone $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$, we give a direct proof of Proposition 4.5, following Fried’s strategy but in the combinatorial spirit of this paper.

Before giving the proof of Proposition 4.5, we prove a few brief lemmas.

Given any nonzero $\phi \in H^1(M; \mathbb{Z})$, not necessarily in the fibered cone, consider the infinite cyclic covering $M_\phi \rightarrow M$ corresponding to ϕ . This covering induces an infinite cyclic covering $\Delta_\phi \rightarrow \Delta$ of the graph Δ modeling the veering triangulation of M . We define the drift of each edge of Δ_ϕ as the drift of its projection in Δ .

Define an integer-valued function on the vertices of Δ_ϕ as follows. By associating to each tetrahedron of the veering triangulation its bottom edge, each tetrahedron in the veering triangulation of \tilde{M} can be referred to as $g\tilde{e}$ for some $g \in G$ and $e \in E$. Since M_ϕ is a quotient of \tilde{M} , where two edges $g_1\tilde{e}$ and $g_2\tilde{e}$ have the same image if and only if $\phi(g_1) = \phi(g_2)$, the integer $\phi(g)$ is a well-defined invariant of the image of any edge $g\tilde{e}$ in M_ϕ . This way we obtain an integer associated to each vertex v of Δ_ϕ , which we will denote by $\phi(v)$.

Lemma 4.6 *For any nonzero $\phi \in H^1(M; \mathbb{Z})$, there exists some $Q > 0$ such that, if γ is a path in Δ_ϕ starting at v and ending at v' , then*

$$\phi(v') - \phi(v) = q + \sum_{i=1}^k \phi(b_i)$$

for some $q \in \mathbb{Z}$ with $|q| \leq Q$ and some $k \geq 0$ and $b_i \in \mathcal{B}$.

Proof It is straightforward to verify from the definition of $\phi(v)$ and $\phi(v')$ that

$$\phi(v') - \phi(v) = \phi(\text{drift}(\gamma)) = \phi(\text{drift}(\pi(\gamma))),$$

where $\pi(\gamma)$ is the projection of γ in the graph Δ with finitely many vertices. As in [Lemma 3.8](#), we can decompose $\pi(\gamma)$ as a sum of minimal cycles and a path that does not contain a cycle. By setting

$$Q = \max\{|\phi(\text{drift}(\delta))| : \delta \text{ is a path in } \Delta \text{ containing no cycles}\},$$

we obtain the statement of the lemma. □

Lemma 4.7 *Let $\phi \in H^1(M; \mathbb{Z})$ such that $\phi(b) < 0$ for all $b \in \mathcal{B}$. If $\dots v_{-1}v_0v_1\dots$ is a bi-infinite path in Δ_ϕ , then $\lim_{n \rightarrow \infty} \phi(v_n) = -\infty$ and $\lim_{n \rightarrow -\infty} \phi(v_n) = \infty$.*

Proof Using [Lemma 4.6](#) and its notation, we have

$$\lim_{n \rightarrow \infty} \phi(v_n) - \phi(v_0) = \lim_{n \rightarrow \infty} q_n + \sum_{i=1}^{k_n} \phi(b_{i,n}),$$

where $q_n \in \mathbb{Z}$ with $|q_n| \leq Q$ and $b_{i,n} \in \mathcal{B}$. Since the $\phi(b_{i,n})$ are negative integers and $\lim_{n \rightarrow \infty} k_n = \infty$, we have $\lim_{n \rightarrow \infty} \phi(v_n) = -\infty$. The proof of the limit as $n \rightarrow -\infty$ is analogous. □

Lemma 4.8 *Let $\phi \in H^1(M; \mathbb{Z})$ be such that $\phi(b) < 0$ for all $b \in \mathcal{B}$. Let v_0 be a vertex of Δ_ϕ and let V_+ be the set of vertices (including v_0) that are endpoints of a path starting at v_0 . Then there exist $N_1, N_2 \in \mathbb{Z}$ such that*

$$\{v \in \Delta_\phi : \phi(v) \leq N_1\} \subset V_+ \subset \{v \in \Delta_\phi : \phi(v) \leq N_2\}.$$

Proof If $v \in V_+$, then, by [Lemma 4.6](#), we have

$$\phi(v') = \phi(v_0) + q + \sum_{i=1}^k \phi(b_i),$$

where $q \in \mathbb{Z}$ with $|q| \leq Q$ and $b_i \in \mathcal{B}$. Hence, the second containment in the lemma holds with $N_2 = \phi(v_0) + Q$. For the first containment, observe that there is some $N_1 < 0$ such that every integer less than N_1 can be written in the form $\sum_{i=1}^k \phi(b_i)$. This follows from the fact that the cone $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ has nonempty interior and that the monoid $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ contains every integral point in some translate of $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ ([Proposition 4.4](#)). □

Proof of Proposition 4.5 First we will show that, if ϕ is a primitive integral point in the interior of $\mathbb{R}_+\mathcal{F}$, then $\phi(g) < 0$ for all $g \in \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$. It suffices to show this for all $g \in \mathcal{B}$. Let γ be a cycle in Δ with drift g . Corresponding to the cycle γ is a sequence of tetrahedra T_0, \dots, T_m in the veering triangulation of \widetilde{M} such that, for each $i = 1, \dots, m$, the tetrahedra T_{i-1} and T_i share a face and T_i is below T_{i-1} . Moreover, $T_m = gT_0$. Therefore, multiplication by g translates T_0 to a tetrahedron below it.

By convention (see Section 2.1), the cohomology class ϕ evaluates to positive integers on loops of M whose lift “goes up” (the endpoint of the lift is higher than the starting point) in the infinite cyclic cover $S \times \mathbb{R} \rightarrow M$ corresponding to ϕ . As we see from the tetrahedron sequence, loops representing g lift to paths that “go down” in $S \times \mathbb{R}$. Therefore, $\phi(g) < 0$ indeed.

Consider the (open) cone

$$D = \{ \phi \in H^1(M; \mathbb{R}) : \phi(g) < 0 \text{ for all } g \in \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \} \subset H^1(M; \mathbb{R}).$$

What we have just proved implies that the interior of $\mathbb{R}_+\mathcal{F}$ is contained in D . To prove the proposition, we need to prove that, conversely, D is contained in the interior of $\mathbb{R}_+\mathcal{F}$. If this was not true, then D would contain a primitive integral class on the boundary of $\mathbb{R}_+\mathcal{F}$. (The boundary faces of the cone $\mathbb{R}_+\mathcal{F}$ are defined by rational equations, so primitive integral points are projectively dense on the boundary of $\mathbb{R}_+\mathcal{F}$.) Since primitive integral classes on the boundary of $\mathbb{R}_+\mathcal{F}$ are known not to correspond to fibrations, it suffices to show that if $\phi \in D$ is a primitive integral class, then ϕ is dual to a fibration.

A *cut* of Δ_ϕ is a way of dividing the vertices of Δ_ϕ into two disjoint nonempty sets $V_{-\infty}$ and V_∞ that are closed under “going forward” and “going backward”, respectively. More precisely, if there is an edge from v_1 to v_2 in Δ_ϕ , then $v_1 \in V_{-\infty}$ implies $v_2 \in V_{-\infty}$ and $v_2 \in V_\infty$ implies $v_1 \in V_\infty$.

To see that cuts exists, let v be a vertex of Δ_ϕ and let $V_{-\infty}$ be the set of vertices (including v) that are endpoints of a path starting at v and let V_∞ be the set of the remaining vertices. It is clear that $V_{-\infty}$ and V_∞ are closed under going forward and going backward, respectively. It follows from Lemmas 4.7 and 4.8 that both $V_{-\infty}$ and V_∞ are nonempty.

Next, we associate an embedded surface in M_ϕ to each cut. Given a cut $V_{-\infty} \cup V_\infty$, let Σ be the union of triangles of the veering triangulation corresponding to the edges starting at a point of V_∞ and ending at a point of $V_{-\infty}$. To show that Σ is a surface,



Figure 4: The immersed subgraph Γ_e in Δ whose vertices correspond to the tetrahedra adjacent to e .

we need to prove that there are two triangles meeting at every edge. (We are gluing together ideal triangles — their vertices are not part of the 3–manifold — therefore, we do not need to check that the links of the vertices are circles.) Let e be an edge of the veering triangulation of M_ϕ and let T and T' be the tetrahedra whose bottom and top edges are e , respectively. The tetrahedra adjacent to e define an immersed subgraph Γ_e of Δ_ϕ with the structure shown in Figure 4.

Observe that either

- (1) all vertices of Γ_e are in $V_{-\infty}$,
- (2) all vertices of Γ_e are in V_∞ , or
- (3) $T \in V_\infty$, $T' \in V_{-\infty}$ and exactly two edges of Γ_e start in V_∞ and end in $V_{-\infty}$, with one edge on each of the two paths from T to T' in Γ_e .

Hence there are indeed either zero or two triangles meeting at every edge and Σ is an embedded surface.

Next, observe that each flowline in M_ϕ intersects Σ at exactly one point. This is because the tetrahedra intersected by the flowline give rise to a bi-infinite path $\dots v_{-1}v_0v_1\dots$ in Δ_ϕ . By Lemmas 4.7 and 4.8, there exists some $i_0 \in \mathbb{Z}$ such that $v_i \in V_{-\infty}$ if $i \geq i_0$ and $v_i \in V_\infty$ otherwise. Therefore, the flowline intersects exactly one triangle of Σ : the one corresponding to the edge from v_{i_0-1} to v_{i_0} . When the flowline intersects some edges of the veering triangulation, the corresponding bi-infinite path is not unique, but it is straightforward to verify that such flowlines also intersect Σ in one point.

As a corollary, we obtain a homeomorphism $\Sigma \times \mathbb{R} \rightarrow M_\phi$ defined by the formula $(x, t) \mapsto g_t(x)$, where g_t denotes the flow on M_ϕ .

Let $h: M_\phi \rightarrow M_\phi$ be the generator of the deck group of the covering $M_\phi \rightarrow M$ such that $\phi(h(v)) = \phi(v) - 1$ for every vertex v of Δ_ϕ . Our final step is to replace Σ with a homotopic surface Σ' in M_ϕ such that $h(\Sigma')$ is disjoint from and homotopic to Σ' . This will show that the covering $M_\phi \rightarrow M$ comes from a fibration.

Every surface Σ' in M_ϕ intersecting every flowline once can be represented by a continuous function $u: \Sigma \rightarrow \mathbb{R}$ such that $\Sigma' = \{g_{u(x)}(x) : x \in \Sigma\}$. For example, the function corresponding to Σ is the constant zero function.

Let n be a positive integer and consider the surfaces $\Sigma, h(\Sigma), \dots, h^n(\Sigma)$, which all correspond to cuts of Δ_ϕ , and therefore intersect every flowline once. Let $u_0, \dots, u_n: \Sigma \rightarrow \mathbb{R}$ be the corresponding functions. Let $u' = (u_0 + \dots + u_{n-1})/n$ and let Σ' be the corresponding surface. The function corresponding to $h(\Sigma')$ is $(u_1 + \dots + u_n)/n$, which is strictly larger than u' at every point of Σ if n is large enough, since $u_n > u_0$ if n is large enough. Therefore, Σ' is an embedded surface intersecting every flowline exactly once such that $h(\Sigma')$ is homotopic to and disjoint from Σ' . Hence, $M_\phi \rightarrow M$ comes from a fibration, and that is what we wanted to show. \square

5 Lemmas on cones, lattices and volumes

This section contains various lemmas on cones, lattices and volumes in Euclidean spaces that will be used to prove the main theorems. All results in this section are self-contained and independent of 3-manifold theory.

5.1 Occupancy coefficients

Let V be an n -dimensional real vector space, let $K \subset V$ be a compact set with nonempty interior and let $\Lambda \subset V$ be a lattice. The *occupancy coefficient* $\text{occ}(\Lambda, K)$ of K with respect to the lattice Λ is the ratio

$$(5-1) \quad \text{occ}(\Lambda, K) = \frac{\text{vol}(K')}{\text{vol}(V/\Lambda)},$$

where $K' = a + bK$ with $a, b \in \mathbb{R}$, is a set of maximal volume that is obtained from K by dilatation and translation and does not contain any point of Λ in its interior.

It seems difficult to compute occupancy coefficients in general, but some basic facts can easily be deduced.

Lemma 5.1 *The occupancy coefficient of a connected set in a 1-dimensional vector space equals 1.*

Proof Identifying V with \mathbb{R} , we have $\Lambda = a\mathbb{Z}$ for some $a > 0$. Our connected set K is an interval. The longest interval K' that does not contain a point of Λ in its interior has length a . By (5-1), we have $\text{occ}(\Lambda, K) = a/a = 1$. \square

Lemma 5.2 *The occupancy coefficient is always at least 1.*

Proof If a set $K \subset V$ has volume less than $\text{vol}(V/\Lambda)$, then its image in the n -torus V/Λ is not everything; therefore, there is a translate of K that is disjoint from Λ . So $\text{vol}(K') \geq \text{vol}(V/\Lambda)$ for the set K' with maximal volume and therefore the occupancy coefficient is at least 1. \square

5.2 Occupancy coefficients as the lattice is varied

For any compact set $K \subset V$ with nonempty interior, introduce the notation

$$\min \text{occ}(K) = \inf_{\Lambda \subset V} \text{occ}(\Lambda, K),$$

where Λ ranges over the lattices in V . By [Lemma 5.2](#), $\min \text{occ}(K) \geq 1$ holds for all K .

Lemma 5.3 *Let $K \subset V$ be a compact connected set with nonempty interior in an n -dimensional vector space V . Consider the set*

$$(5-2) \quad \text{occs}(K) = \{\text{occ}(\Lambda, K) : \Lambda \subset V \text{ is a lattice}\}.$$

If $n = 1$, then $\text{occs}(K) = \{1\}$. If $n \geq 2$, then $\text{occs}(K)$ is a half-infinite interval whose left endpoint is $\min \text{occ}(K)$.

Proof The $n = 1$ case follows from [Lemma 5.1](#).

The space of lattices is connected and the occupancy coefficient is a continuous function, so $\text{occs}(K)$ is an interval. It is clear that the left endpoint of this interval is $\min \text{occ}(K)$. It remains to show that Λ can be chosen so that $\text{occ}(\Lambda, K)$ is arbitrarily large when $n \geq 2$.

Let e_1, \dots, e_n be a basis for V and consider the sequence of lattices generated by $ce_1, ce_2, \dots, ce_{n-1}, e_n$ as $c \rightarrow 0$. The covolumes of these lattices go to zero. However, a dilated and translated copy of K that lies between the hyperplanes Σ and $e_n + \Sigma$, where Σ is the hyperplane spanned by e_1, \dots, e_{n-1} , is disjoint from all these lattices; hence, $\text{vol}(K')$ in (5-1) is bounded from below as $c \rightarrow \infty$. So, indeed, the occupancy coefficient can be arbitrarily large. \square

5.3 The main technical lemma on Frobenius numbers

The following technical lemma is at the heart of the proof of [Theorem 1.1](#):

Lemma 5.4 *Let Λ be a lattice in an n -dimensional real vector space V . Let $C = \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ be a cone with nonempty interior, generated by a finite set $\mathcal{B} \subset V$. Let e_0 be a point in the interior of C and let $x_1, x_2 \in V$ be arbitrary. Then there is a constant $K = K(\Lambda, C, e_0, x_1, x_2) > 0$ such that the following holds.*

Let $\Lambda' \subset \Lambda$ be such that

$$(5-3) \quad \Lambda \cap (x_1 + C) \subset \Lambda' \subset \Lambda \cap (x_2 + C).$$

Let $\beta : V \rightarrow \mathbb{R}$ be a linear function with $\beta(e_0) = 1$ that takes rational values on Λ and positive values on $C - \{0\}$. Let $\bar{\beta}$ be the unique positive scalar multiple of the β such that $\bar{\beta}(\Lambda) = \mathbb{Z}$. Let P be a polytope that is the intersection of the hyperplane $\beta^{-1}(0)$ and $y - C$ for some $y \in V$ with $\beta(y) > 0$. Then

$$\left| \text{Frob}(\bar{\beta}|_{\Lambda'}) - n^{-1} \sqrt{\frac{\text{occ}(\Lambda \cap \beta^{-1}(0), P) \cdot \text{vol}(\mathbb{R}^n / \Lambda)}{n \text{vol}(C \cap \beta^{-1}([0, 1]))}} \bar{\beta}(e_0)^{1+1/(n-1)} \right| \leq K \bar{\beta}(e_0).$$

Proof Let P_0 be a polytope in the hyperplane $\beta^{-1}(0)$, obtained from P by a dilatation and translation whose interior does not contain any point of Λ and whose $(n-1)$ -dimensional volume is maximal with respect to this property. Let $y_0 \in V$ be such that $P_0 = \beta^{-1}(0) \cap (y_0 - C)$. Let $\Lambda_0 = \beta^{-1}(0) \cap \Lambda$ be the lattice in the hyperplane $\beta^{-1}(0)$.

Step 1 (upper bound on the Frobenius number) The following inequalities hold:

$$(5-4) \quad \begin{aligned} \text{Frob}(\bar{\beta}|_{\Lambda'}) &\leq \text{Frob}(\bar{\beta}|_{\Lambda \cap (x_1 + C)}) \\ &= \text{Frob}(\bar{\beta}|_{\Lambda \cap (x_1 + \Lambda_0 + C)}) \\ &\leq \text{Frob}(\bar{\beta}|_{\Lambda \cap \{z \in \mathbb{R}^n : \beta(z) \geq \beta(x_1 + y_0)\}}) \\ &< \bar{\beta}(x_1 + y_0). \end{aligned}$$

The first inequality follows from the containment $\Lambda \cap (x_1 + C) \subset \Lambda'$. The equality holds because $\bar{\beta}(\Lambda_0) = 0$. For the second inequality, note that, if z satisfies $\beta(z) > \beta(y_0)$, then $z \in C + \Lambda_0$. This is because the polytope $\beta^{-1}(0) \cap (z - C)$ is a scaled-up copy of P_0 , so it contains some $x^* \in \Lambda_0$ in its interior and therefore $z \in x^* + C \subset \Lambda_0 + C$. So, if $\beta(z) > \beta(y_0 + x_1)$, then $z \in x_1 + \Lambda_0 + C$. Finally, the last inequality simply follows from the definition of the Frobenius number.

Step 2 (lower bound on the Frobenius number) Let Q be the polytope that is the intersection of C and $e_0 - C$ (see Figure 5). Let $m = m(C, e_0, \Lambda) > 0$ be a number

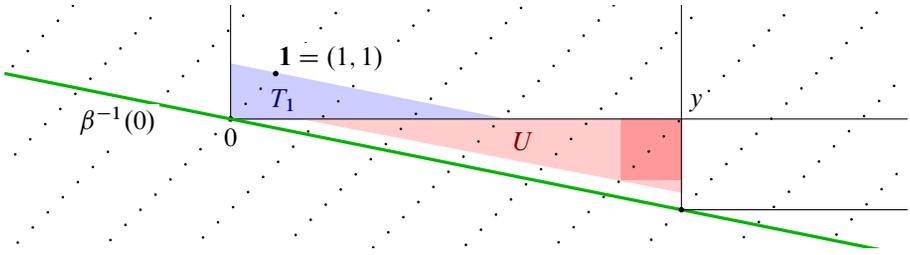


Figure 5

such that any translate of mQ contains some point of the lattice Λ in its interior. We claim that

$$(5-5) \quad \text{Frob}(\bar{\beta}|_{\Lambda'}) \geq \text{Frob}(\bar{\beta}|_{\Lambda \cap (x_2 + C)}) > \bar{\beta}(y_0 + x_2) - m\bar{\beta}(e_0).$$

The first inequality is a consequence of the containment $\Lambda' \subset \Lambda \cap (x_2 + C)$. For the second inequality, consider the polytope $U = \beta^{-1}([\beta(y_0) - m, \beta(y_0)]) \cap (y_0 - C)$ that contains the polytope $y_0 - mQ$.

The translate $x_2 + U$ of U contains a point $p \in \Lambda$ in its interior. The side of $x_2 + U$ opposite to $x_2 + y_0$ is contained in the level set $\beta^{-1}(\beta(x_2 + y_0) - m)$. Therefore, $\beta(p) > \beta(y_0 + x_2) - m$ and

$$\bar{\beta}(p) > \bar{\beta}(y_0 + x_2) - m\bar{\beta}(e_0).$$

The second inequality in (5-5) now follows from putting this together with the inequality $\text{Frob}(\bar{\beta}|_{\Lambda \cap (C + x_2)}) \geq \bar{\beta}(p)$, which holds because the elements of Λ on which $\bar{\beta}$ takes the value $\bar{\beta}(p)$ are exactly the points of $p + \Lambda_0$, none of which are contained in $C + x_2$, since the set $U + \Lambda_0$ is disjoint from $C + x_2$. (Figure 5 shows the case $x_2 = 0$.)

Let Y_{y_0} be the pyramid $(y_0 - C) \cap \beta^{-1}([0, \infty))$.

Step 3 (expressing the volume of the pyramid Y_{y_0} , first way) We claim that

$$(5-6) \quad \text{vol}(Y_{y_0}) = \frac{1}{n} \bar{\beta}(y_0) \text{occ}(\Lambda_0, P) \text{vol}(\mathbb{R}^n / \Lambda).$$

One can see this by comparing the pyramid Y_{y_0} with a pyramid Y whose base is a parallelepiped spanned by a basis of the lattice Λ_0 in the hyperplane $\beta^{-1}(0)$ and whose tip is some $v \in \Lambda$ with $\bar{\beta}(v) = 1$. Note that a parallelepiped Z that contains the base of Y as a face and e_0 as a vertex is a fundamental domain for Λ , since it is spanned by n linearly independent elements of Λ and it only contains elements of Λ at its vertices, since $\bar{\beta}$ is a primitive integral class with $\bar{\beta}(e_0) = 1$. So $\text{vol}(Z) = \text{vol}(\mathbb{R}^n / \Lambda)$.

Recall that the volume formulas for a pyramid and a parallelepiped are $(1/n)bh$ and bh , respectively, where b is the $(n-1)$ -dimensional area of the base and h is the height. Since Y and Z have the same base and same height, but Y is a pyramid and Z is a parallelepiped, we have

$$(5-7) \quad \text{vol}(Y) = \frac{1}{n} \text{vol}(Z) = \frac{1}{n} \text{vol}(\mathbb{R}^n / \Lambda).$$

Both Y_{y_0} and Y are pyramids with a base on the hyperplane $\beta^{-1}(0)$. To compare the volumes, we need to compare their heights and the areas of their bases. The base of Y_{y_0} is P_0 and the base of Y is a fundamental domain for Λ_0 . Therefore, the ratio of the areas of the bases is the occupancy coefficient $\text{vol}(P_0) / \text{vol}(\beta^{-1}(0) / \Lambda_0) = \text{occ}(\Lambda_0, P_0) = \text{occ}(\Lambda_0, P)$. The tips of Y_{y_0} and Y are at the level sets $\bar{\beta}^{-1}(\bar{\beta}(y_0))$ and $\bar{\beta}^{-1}(1)$, respectively; therefore, the height of Y_{y_0} is $\bar{\beta}(y_0)$ times the height of Y . The formula (5-6) follows from (5-7) and the comparisons between the bases and the heights.

Step 4 (expressing the volume of the pyramid Y_{y_0} , second way) Let Y_{e_0} be the pyramid $(e_0 - C) \cap \beta^{-1}([0, \infty))$. Using the similarity between the polytopes Y_{y_0} and Y_{e_0} , we have

$$(5-8) \quad \text{vol}(Y_{y_0}) = \frac{\bar{\beta}(y_0)^n}{\bar{\beta}(e_0)^n} \text{vol}(Y_{e_0}).$$

Step 5 (expressing $\bar{\beta}(y_0)$) Using the equality between the right-hand sides of (5-6) and (5-8) and solving for $\bar{\beta}(y_0)$, we obtain

$$(5-9) \quad \bar{\beta}(y_0) = \sqrt[n-1]{\frac{\text{occ}(\Lambda_0, P) \text{vol}(\mathbb{R}^n / \Lambda)}{n \text{vol}(Y_{e_0})}} \bar{\beta}(e_0)^{1+1/(n-1)}.$$

Step 6 (conclusion) The pyramid Y_{e_0} is isometric to the pyramid $C \cap \beta^{-1}([0, 1])$. The statement of the lemma now follows from (5-9), the upper and lower bounds (5-4)–(5-5) on the Frobenius number and the fact that all the error terms $\bar{\beta}(x_1)$, $\bar{\beta}(x_2)$ and $m\bar{\beta}(e_0)$ are constant multiples of $\bar{\beta}(e_0)$, where the constant depends only on x_1 , x_2 , C , Λ and e_0 . □

Remark 5.5 For a given C , e_0 and Λ as in Lemma 5.4 and any $K > 0$, there are only finitely many linear functions $\bar{\beta}: V \rightarrow \mathbb{R}$ taking positive values on $C - \{0\}$ and integer values on Λ such that $\bar{\beta}(e_0) < K$. To see this, let $v_1, \dots, v_n \in C$ form a basis for Λ such that e_0 is in the interior of the cone generated by the v_i . That is, $e_0 = c_1v_1 + \dots + c_nv_n$ with $c_i > 0$ for $i = 1, \dots, n$. Then $\bar{\beta}(e_0) = c_1\bar{\beta}(v_1) + \dots + c_n\bar{\beta}(v_n)$, where the $\bar{\beta}(v_i)$

are positive integers. From this, we see that there are only finitely many choices for the $\bar{\beta}(v_i)$ that make this sum less than K . Since the $\bar{\beta}(v_i)$ determine $\bar{\beta}$, we obtain that there are indeed finitely many possibilities for $\bar{\beta}$.

From Lemma 5.4 and Remark 5.5, we obtain the following:

Corollary 5.6 *Let V, C, e_0, Λ and Λ' be as in Lemma 5.4. Consider a sequence of pairwise distinct linear functions $\{\beta_k\}_{k \in \mathbb{N}}$ and associated polytopes $\{P_k\}_{k \in \mathbb{N}}$ as in Lemma 5.4. Then*

$$\lim_{k \rightarrow \infty} \frac{\text{Frob}(\bar{\beta}_k |_{\Lambda'})}{\bar{\beta}_k(e_0)^{1+1/(n-1)}} - \sqrt[n-1]{\frac{\text{occ}(\Lambda \cap \beta_k^{-1}(0), P_k) \cdot \text{vol}(V/\Lambda)}{n \text{vol}(C \cap \beta_k^{-1}([0, 1]))}} = 0.$$

5.4 Cones with a tetrahedron base

In the following lemmas, $\pi_i : \mathbb{R}^n \rightarrow \mathbb{R}$ denotes the projection to the i^{th} coordinate.

Lemma 5.7 *Let $n \geq 1$ be an integer and let $\alpha = (\alpha_1, \dots, \alpha_n)$ be such that $\sum_{i=1}^n \alpha_i = 1$ and $\alpha_i > 0$ for each i . Let $\beta_\alpha = \sum_{i=1}^n \alpha_i \pi_i$ and denote by T_α the tetrahedron $\mathbb{R}_{\geq 0}^n \cap \beta_\alpha^{-1}([0, 1])$. Then*

$$\text{vol}(T_\alpha) = \frac{1}{n!} \prod_{i=1}^n \frac{1}{\alpha_i}.$$

Proof One vertex of T_α is the origin; the other n vertices are the intersections of the hyperplane $\beta_\alpha^{-1}(1)$ with the coordinate axes. For example, the intersection with the first axis is the point $(x_1, 0, \dots, 0)$ that satisfies

$$1 = \beta_\alpha(x_1, 0, \dots, 0) = \alpha_1 x_1,$$

which yields $x_1 = 1/\alpha_1$. Similarly, we obtain that the only nonzero coordinates of the other intersection points are $1/\alpha_i$. The parallelepiped spanned by the intersection points has volume $\prod_{i=1}^n 1/\alpha_i$ and the tetrahedron spanned by them has volume $1/n!$ times that. □

5.5 Projective convergence of lattices

The occupancy coefficient term in Corollary 5.6 is not very well behaved, since the lattices $\Lambda \cap \beta_k^{-1}(0)$ may vary wildly even when the linear functions β_k converge. It

will be useful to single out subsequences where the lattices are stable in a sense. This section introduces some lemmas and terminology for this.

Lemma 5.8 *If $a_1, \dots, a_n \in \mathbb{Z}$ and $v_1, \dots, v_n \in \mathbb{R}^{n+1}$ are the columns of the $(n+1) \times n$ matrix*

$$(5-10) \quad \begin{pmatrix} a_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \ddots & a_{n-1} & 1 \\ 0 & 0 & 0 & \cdots & 0 & a_n \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

then $\langle v_1, \dots, v_n \rangle_{\mathbb{Z}} = \langle v_1, \dots, v_n \rangle_{\mathbb{R}} \cap \mathbb{Z}^{n+1}$.

Proof The group $\langle v_1, \dots, v_n \rangle_{\mathbb{Z}}$ is a finite-index subgroup of $\langle v_1, \dots, v_n \rangle_{\mathbb{R}} \cap \mathbb{Z}^{n+1}$. To show that they are equal, it suffices to show that there is some $v_{n+1} \in \mathbb{Z}^{n+1}$ such that v_1, \dots, v_{n+1} form a basis for \mathbb{Z}^{n+1} . The vector $v_{n+1} = (0, \dots, 0, 1, 0)^T$ has this property, since, by adding this vector as the last column of the matrix above, we obtain a matrix whose determinant is ± 1 . □

If Λ and $\{\Lambda_k\}_{k \in \mathbb{N}}$ are discrete subgroups of rank r in a vector space $V \cong \mathbb{R}^n$, then we say that $\Lambda_k \rightarrow \Lambda$ *projectively* if there is a basis $\{v^1, \dots, v^r\}$ of Λ , a positive constant c_k and a basis $\{v_k^1, \dots, v_k^r\}$ of Λ_k for every $k \in \mathbb{N}$ such that $\lim_{k \rightarrow \infty} c_k v_k^i = v^i$ for every $i = 1, \dots, r$.

Lemma 5.9 *Let Λ be a lattice in an n -dimensional vector space V . Let Σ be a hyperplane in V and let Λ_0 be any lattice in Σ . Then there exists a sequence $\{\Sigma_k\}_{k \in \mathbb{N}}$ of hyperplanes in V such that $\Sigma_k \cap \Lambda$ is a lattice in Σ_k for all k and $\Sigma_k \cap \Lambda \rightarrow \Lambda_0$ projectively.*

Proof First we prove the statement in a special case and then we use this special case to prove the general case.

In the special case, we assume that $\Sigma \cap \Lambda$ is a lattice in Σ . In this case, we can choose an isomorphism $\iota: V \rightarrow \mathbb{R}^n$ that identifies Λ with \mathbb{Z}^n and the hyperplane Σ with the orthogonal complement of $(0, \dots, 0, 1)^T$. In this special case, we further assume that Λ_0 is *rational*; that is, there is a positive constant c such that $c\Lambda_0 \subset \Lambda$. Then we may

choose the isomorphism ι so that the columns of the $n \times (n - 1)$ matrix

$$\begin{pmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{n-1} \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

form a basis for $\iota(c\Lambda_0)$ for some $b_1, \dots, b_{n-1} \in \mathbb{Z}$. Applying Lemma 5.8 with $a_i^{(k)} = kb_i$ for $i = 1, \dots, n - 1$ and all $k \in \mathbb{N}$ yields a sequence of hyperplanes (spanned by the columns of the matrix (5-10)) whose pullbacks by ι satisfy the required properties.

To prove the general case, we take a sequence of hyperplanes Σ_m converging to Σ_0 such that $\Sigma_m \cap \Lambda$ is a lattice in Σ_m for all m and, for each m , we take a rational lattice $\Lambda_0^{(m)}$ in Σ_m so that the lattices $\Lambda_0^{(m)}$ converge to Λ_0 projectively. As we have already shown, we can construct sequences of lattices of the required properties converging projectively to each $\Lambda_0^{(m)}$. From this, the statement of the lemma follows also for Λ_0 . \square

5.6 Projection of cones

The following lemma will be used for projections of the cone dual to the cone $\mathbb{R}_+\mathcal{F}$. Such projections naturally correspond to slices of the fibered face \mathcal{F} .

Lemma 5.10 *Let $C = \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \subset \mathbb{R}^n$ be the cone generated by some subset $\mathcal{B} \subset \mathbb{R}^n$. Suppose C has nonempty interior and let $p: \mathbb{R}^n \rightarrow \mathbb{R}^k$ be a linear map such that $p(\mathbb{Z}^n)$ is a lattice in \mathbb{R}^k . Then there exists $x \in \mathbb{R}^k$ such that $p(\mathbb{Z}^n) \cap (x + p(C)) \subset p(\mathbb{Z}^n \cap C)$.*

Proof Since $p(\mathbb{Z}^n)$ is a lattice in \mathbb{R}^k , the subset $A = p^{-1}(0) \cap \mathbb{Z}^n$ is a lattice in the $(n - k)$ -dimensional kernel $p^{-1}(0)$. Let $K > 0$ be large enough that $B(y, K)$, the ball of radius K centered at y , contains some $a \in A$ for all $y \in p^{-1}(0)$. Since the cone C has nonempty interior, there is some $c_0 \in C$ such that $B(c_0, K) \subset C$. But then $B(c, K) \subset C$ for all $c \in c_0 + C$.

We claim that any point $b \in p(\mathbb{Z}^n)$ that lies in $p(c_0 + C) = p(c_0) + p(C)$ is the image of some point of \mathbb{Z}^n that lies in C . This will prove the lemma with $x = p(c_0)$. To see this, let $c \in c_0 + C$ be such that $p(c) = b$. We know that $B(c, K) \subset C$, and the ball $B(c, K)$ is centered around a point of the translate $p^{-1}(0) + c$ of the subspace $p^{-1}(0)$. Moreover, this translated subspace $p^{-1}(0) + c$, being a preimage of b , contains some point in \mathbb{Z}^n and hence contains a translate of A . As a consequence, $(p^{-1}(0) + c) \cap B(c, K) \cap \mathbb{Z}^n$ is nonempty and any element of it is a point of \mathbb{Z}^n that lies in C and maps to b . \square

6 Proof of the main theorem

We are now ready to prove our main theorem.

In the statement of the theorem, we use the following notions of duality. Let V be a vector space, let $C = \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$ be a cone with nonempty interior, generated by some subset $\mathcal{B} \subset V$, and let $\Lambda \subset V$ be a lattice. Then the dual of the triple (V, C, Λ) is the triple (V^*, C^*, Λ^*) , where V^* is the dual vector space of V , the cone $C^* \subset V^*$ is the set of linear functions $V \rightarrow \mathbb{R}$ that take nonnegative values on C , and $\Lambda^* \subset V^*$ is the lattice consisting of linear functions that take integer values on Λ . Note that, if $\phi: V \rightarrow \mathbb{R}$ is a linear function that takes positive values on C , then $\phi \in \text{int}(C^*)$.

Theorem 1.1 is a direct corollary of the following theorem, which additionally describes the bounding function g :

Theorem 6.1 *Let M be a connected 3-manifold that admits a complete finite-volume hyperbolic metric. Let \mathcal{F} be a fully punctured fibered face of the unit ball of the Thurston norm on $H^1(M; \mathbb{R})$. Let $1 \leq d \leq \dim(H^1(M; \mathbb{R})) - 1$, let Ω be a rational d -dimensional slice of \mathcal{F} cut out by the $(d + 1)$ -dimensional subspace Σ , let C be the cone $\langle \Omega \rangle_{\mathbb{R}_{\geq 0}}$ in Σ and consider the lattice $\Lambda = \Sigma \cap H^1(M; \mathbb{Z})$ in Σ . Consider the dual triple $(\Sigma^*, C^*, \Lambda^*)$ of the triple (Σ, C, Λ) .*

Let $\text{Graph}(\mu_d|_{\Omega}) \subset \Omega \times \mathbb{R}$ be the graph of the normalized asymptotic translation length function μ_d , restricted to Ω . Let $g: \text{int}(\Omega) \rightarrow \mathbb{R}_+$ be the function defined by the formula

$$(6-1) \quad g(\phi) = \sqrt[d]{\frac{(d + 1) \text{vol}_{\Lambda^*}(C^* \cap \beta_{\phi}^{-1}([0, 1]))}{\min \text{occ}(C^* \cap \beta_{\phi}^{-1}(1))}},$$

where vol_{Λ^*} is the translation-invariant volume form on Σ^* with respect to which Λ^* has covolume 1 and β_{ϕ} denotes the linear function $\Sigma^* \rightarrow \mathbb{R}$ corresponding to the element $\phi \in \Sigma$ in the dual space of Σ^* .

Then the set of accumulation points of the graph $\text{Graph}(\mu_d|_{\Omega})$ is

$$\{(\omega, g(\omega)) : \omega \in \text{int}(\Omega)\}$$

if $d = 1$ and

$$\{(\omega, r) : \omega \in \text{int}(\Omega), 0 \leq r \leq g(\omega)\} \cup (\partial\Omega \times [0, \infty))$$

if $d \geq 2$. Moreover, the function g is continuous and $g(\phi) \rightarrow \infty$ as $\phi \rightarrow \partial\Omega$.

Proof Once again, we break the proof into several steps.

Step 1 (asymptotic behavior of the weighted graphs $W(\phi)$) Every element of $H_1(M; \mathbb{R})$ defines a linear function $H^1(M; \mathbb{R}) \rightarrow \mathbb{R}$ and this linear function restricts to a linear function $\Sigma \rightarrow \mathbb{R}$. This way we obtain a natural linear map $p: H_1(M; \mathbb{R}) \rightarrow \Sigma^*$. It is easy to see that $p(G) = \Lambda^*$.

By [Proposition 4.5](#), the elements of the cone C take nonpositive values on the cone $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$; therefore, $p(\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}) \subset -C^*$. Conversely, every element of $-C^*$ is a linear function $\Sigma \rightarrow \mathbb{R}$ that takes nonpositive values on C and every such linear function is a restriction of a linear function $H^1(M; \mathbb{R}) \rightarrow \mathbb{R}$ that takes nonpositive values on the cone $\mathbb{R}^+ \mathcal{F}$ over the fibered face. By the other direction of [Proposition 4.5](#), this last linear function corresponds to an element of $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}$. Hence, we have $p(\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}) = -C^*$.

Fix some $e, e' \in E$. By [Proposition 4.4](#), there are $g_1, g_2 \in G$ such that

$$G \cap g_1 \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} \subset \text{Stash}(e, e') \subset G \cap g_2 \langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}}.$$

We claim that it follows that there are some $v_1, v_2 \in \Sigma^*$ such that

$$\Lambda^* \cap (v_1 + C^*) \subset -p(\text{Stash}(e, e')) \subset \Lambda^* \cap (v_2 + C^*),$$

using additive notation in the vector space Σ^* . The existence of v_1 follows from [Lemma 5.10](#). The existence of v_2 is simply a consequence of the identity $p(A \cap B) \subset p(A) \cap p(B)$.

Let $e_0: \Sigma \rightarrow \mathbb{R}$ be the linear function that takes the value 1 on Ω . Note that e_0 is in the interior of C^* .

We now wish to apply [Corollary 5.6](#) with $V = \Sigma^*$, $C = C^*$, $e_0, \Lambda = \Lambda^*$ and $\Lambda' = -p(\text{Stash}(e, e'))$ to conclude that

$$(6-2) \quad \lim_{\phi \rightarrow \phi_0} \frac{\text{Frob}(\bar{\beta}_\phi |_{\Lambda'})}{\bar{\beta}_\phi(e_0)^{1+1/d}} - a \sqrt{\frac{\text{occ}(\Lambda^* \cap \beta_\phi^{-1}(0), P_\phi) \cdot \text{vol}(\Sigma^*/\Lambda^*)}{(d+1) \text{vol}(C^* \cap \beta_\phi^{-1}([0, 1]))}} = 0,$$

where $P_\phi = \beta_\phi^{-1}(0) \cap (y - C^*)$ for some $y \in \text{int}(C^*)$ and $\text{vol}(\cdot)$ is any translation-invariant volume form on Σ^* . The cohomology classes ϕ are rational points of the interior of Ω and ϕ_0 is an arbitrary element of Ω .

It is straightforward to check that the hypotheses of [Corollary 5.6](#) are satisfied. For example, the classes β_ϕ take positive values on C^* , since ϕ is assumed to be in the

interior of Ω and hence in the interior of C . The equality $\beta_\phi(e_0) = 1$ holds because e_0 takes the value 1 on Ω and ϕ is in Ω . The rest of the hypotheses are also satisfied; therefore, Corollary 5.6 applies and (6-2) holds.

Denote by $w_\phi(ee')$ the weight of the edge ee' in the weighted graph $W(\phi)$. Using the definition (3-2), we have

$$w_\phi(ee') = \text{Frob}(\bar{\phi}|_{-\text{Stash}(e,e')}) = \text{Frob}(\bar{\beta}_\phi|_{\Lambda'}).$$

To see that the second equality holds, first note that $\beta_\phi(p(x)) = \phi(x)$ for all $x \in H_1(M; \mathbb{R})$. So $\phi(G) \subset \mathbb{Q}$ is the same discrete subgroup as $\beta_\phi(\Lambda^*) = \beta_\phi(p(G)) \subset \mathbb{Q}$, so $\bar{\phi} = c\phi$ and $\bar{\beta}_\phi = c\beta_\phi$ hold with the same constant $c > 0$. Therefore, $\bar{\phi}(x) = \bar{\beta}_\phi(p(x))$ for all $x \in H_1(M; \mathbb{R})$.

Finally, we have $\beta_\phi(e_0) = 1 = \|\phi\|$ whenever $\phi \in \Omega$. After multiplying by c , we obtain $\bar{\beta}_\phi(e_0) = \|\bar{\phi}\|$. Applying these substitutions to (6-2) yields

$$(6-3) \quad \lim_{\phi \rightarrow \phi_0} \frac{w_\phi(ee')}{\|\bar{\phi}\|^{1+1/d}} - d \sqrt{\frac{\text{occ}(\Lambda^* \cap \beta_\phi^{-1}(0), P_\phi) \cdot \text{vol}(\Sigma^*/\Lambda^*)}{(d+1) \text{vol}(C^* \cap \beta_\phi^{-1}([0, 1]))}} = 0.$$

Step 2 (accumulation points in $\text{int}(\Omega) \times \mathbb{R}$) Assume that ϕ_0 is in the interior of Ω . Let us apply Lemma 5.9 with $V = \Sigma^*$, $\Lambda = \Lambda^*$, $\Sigma = \beta_{\phi_0}^{-1}(0)$ and for a lattice Λ_0 in Σ such that $\text{occ}(\Lambda_0, P_{\phi_0}) = \alpha$ for some $\alpha \in \text{occs}(P_{\phi_0})$. Such a lattice Λ_0 exists by the definition of the set $\text{occs}(P_{\phi_0})$ in (5-2). Lemma 5.9 guarantees that there exists a sequence $\phi_k \rightarrow \phi_0$ such that

$$\lim_{k \rightarrow \infty} \text{occ}(\Lambda^* \cap \beta_{\phi_k}^{-1}(0), P_{\phi_k}) = \text{occ}(\Lambda_0, P_{\phi_0}) = \alpha$$

and therefore

$$(6-4) \quad \lim_{k \rightarrow \infty} \frac{w_{\phi_k}(ee')}{\|\bar{\phi}_k\|^{1+1/d}} = d \sqrt{\frac{\alpha \text{vol}(\Sigma^*/\Lambda^*)}{(d+1) \text{vol}(C^* \cap \beta_{\phi_0}^{-1}([0, 1]))}}$$

for each edge ee' . Since ϕ_0 is in the interior of Ω , the set $C^* \cap \beta_{\phi_0}^{-1}([0, 1])$ is a pyramid of finite volume.

Now recall from (1-2) that $\mu_d(\phi) = \|\bar{\phi}\|^{1+1/d} \ell_{\mathcal{A}}(\bar{\phi})$. Moreover, by Proposition 3.5, $\ell_{\mathcal{A}}(\bar{\phi})$ equals the reciprocal of the maximal average cycle weight in the graph $W(\phi)$. By (6-4), the weight of each edge has the same asymptotics for the sequence ϕ_k ; therefore, the average weight of every cycle also has the same asymptotics. So we can

replace $w_{\phi_k}(ee')$ by $1/\ell_A(\bar{\phi}_k)$ in (6-4) and the limit still holds. Taking the reciprocal of both sides, we obtain

$$\begin{aligned} \lim_{k \rightarrow \infty} \mu_d(\phi_k) &= \sqrt[d]{\frac{(d+1) \text{vol}(C^* \cap \beta_{\phi_0}^{-1}([0, 1]))}{\alpha \text{vol}(\Sigma^*/\Lambda^*)}} \\ &= \sqrt[d]{\frac{(d+1) \text{vol}_{\Lambda^*}(C^* \cap \beta_{\phi_0}^{-1}([0, 1]))}{\alpha}}. \end{aligned}$$

Such a sequence $\phi_k \rightarrow \phi$ exists for every $\alpha \in \text{occs}(P_{\phi_0})$. Using Lemma 5.3 and the fact that the polytope P_{ϕ_0} in the hyperplane $\beta_{\phi_0}^{-1}(0)$ and the polytope $C^* \cap \beta_{\phi_0}^{-1}(1)$ in the hyperplane $\beta_{\phi_0}^{-1}(1)$ are homothetic, we obtain that the accumulation points of $\text{Graph}(\mu_d|_{\Omega})$ in $\text{int}(\Omega) \times \mathbb{R}$ are as specified in the theorem.

Step 3 (accumulation points in $\partial\Omega \times \mathbb{R}$) If $\phi \rightarrow \phi_0$, then the pyramid $C^* \cap \beta_{\phi}^{-1}([0, 1])$ converges to the set $C^* \cap \beta_{\phi_0}^{-1}([0, 1])$. If $\phi_0 \in \partial\Omega$ and hence $\phi_0 \in \partial C$, then this limit set is unbounded and has infinite volume. By Lemma 5.1, the occupancy coefficient is 1 if $d = 1$; therefore, the expression under the root in (6-3) goes to 0. Hence, $\lim_{\phi \rightarrow \phi_0} w_{\phi}(ee')/\|\bar{\phi}\|^{1+1/d} = 0$ and $\lim_{\phi \rightarrow \phi_0} \mu_d(\phi) = \infty$ when $\phi_0 \in \partial\Omega$ and $d = 1$. Therefore, $\text{Graph}(\mu_d|_{\Omega})$ does not have any accumulation point in $\partial\Omega \times \mathbb{R}$ when $d = 1$.

Similarly, the limit of the pyramids $C^* \cap \beta_{\phi}^{-1}([0, 1])$ in the definition of g is an unbounded set when $\phi \rightarrow \partial\Omega$. This shows that $g(\phi) \rightarrow \infty$ whenever $\phi \rightarrow \partial\Omega$. It is clear that g is continuous.

It is now automatic that the set of accumulation points of $\text{Graph}(\mu_d|_{\Omega})$ in $\partial\Omega \times \mathbb{R}$ is $\partial\Omega \times [0, \infty)$ when $d \geq 2$, since g goes to infinity at $\partial\Omega$ and the set of accumulation points is closed. □

Next, we prove Theorem 1.4.

Proof of Theorem 1.4 Choose $\omega_1, \dots, \omega_{d+1}$, the vertices of the simplex Ω , as the basis for Σ . This choice of basis naturally defines coordinates on Σ and the dual space Σ^* . With these coordinates, we have $C \cong \mathbb{R}_{\geq 0}^{d+1} \cong C^*$.

By Theorem 6.1, the function g takes the form

$$g(\phi) = \sqrt[d]{\frac{(d+1) \text{vol}_{\Lambda^*}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_{\phi}^{-1}([0, 1]))}{\min \text{occ}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_{\phi}^{-1}(1))}}.$$

Using the definition of g^* in the theorem and denoting β_ϕ for $\phi = \sum_{i=1}^{d+1} \alpha_i \omega_i$ by β_α , where $\alpha = (\alpha_1, \dots, \alpha_{d+1})$, we can rewrite this as

$$(6-5) \quad g^*(\alpha) = \sqrt[d]{\frac{(d+1) \operatorname{vol}_{\Lambda^*}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_\alpha^{-1}([0, 1]))}{\min \operatorname{occ}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_\alpha^{-1}(1))}}.$$

Note that $\mathbb{R}_{\geq 0}^{d+1} \cap \beta_\alpha^{-1}(1)$ is a d -dimensional simplex for all α . All d -dimensional simplices have the same minimal occupancy coefficient, since they differ only by a linear transformation. Therefore, the denominator is a constant O_d , depending only on d . (It is straightforward to check that this definition of O_d is equivalent to the definition provided in the introduction after [Theorem 1.4](#).)

The volume in the numerator can be written as

$$(6-6) \quad \operatorname{vol}_{\Lambda^*}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_\alpha^{-1}([0, 1])) = \frac{\operatorname{vol}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_\alpha^{-1}([0, 1]))}{\operatorname{vol}(\mathbb{R}^{d+1}/\Lambda^*)},$$

where $\operatorname{vol}(\cdot)$ denotes the standard volume form on \mathbb{R}^{d+1} . Note that $\beta_\alpha = \sum_{i=1}^{d+1} \alpha_i \pi_i$, where $\pi_i : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ is the projection to the i th coordinate. So we can apply [Lemma 5.7](#) to obtain that

$$(6-7) \quad \operatorname{vol}(\mathbb{R}_{\geq 0}^{d+1} \cap \beta_\alpha^{-1}([0, 1])) = \frac{1}{(d+1)!} \prod_{i=1}^{d+1} \frac{1}{\alpha_i}.$$

Finally, the covolume of Λ^* equals the reciprocal of the covolume of Λ . By our choice of basis, the volume form on $\Sigma \cong \mathbb{R}^{d+1}$ is the one with respect to which the lattice $\Gamma = \langle \omega_1, \dots, \omega_{d+1} \rangle_{\mathbb{Z}}$ has covolume 1. So

$$(6-8) \quad \operatorname{vol}(\mathbb{R}^{d+1}/\Lambda^*) = \frac{1}{\operatorname{vol}(\mathbb{R}^{d+1}/\Lambda)} = \frac{1}{\operatorname{vol}_\Gamma(\Sigma/\Lambda)} = \operatorname{vol}_\Lambda(\Sigma/\Gamma).$$

Putting together (6-5), (6-6), (6-7) and (6-8), we obtain that

$$g^*(\alpha) = \sqrt[d]{\frac{(1/d!) \prod_{i=1}^{d+1} 1/\alpha_i}{O_d \operatorname{vol}_\Lambda(\Sigma/\Gamma)}}.$$

This is what we wanted to prove. The fact that $O_d = 1$ follows from [Lemma 5.1](#). \square

Using [Theorem 6.1](#), we can also prove [Theorem 1.8](#).

Proof of Theorem 1.8 The isomorphism $i : \Sigma_1 \rightarrow \Sigma_2$ induces a dual isomorphism $i^* : \Sigma_2^* \rightarrow \Sigma_1^*$ of the dual spaces. By indexing the objects in the statement of [Theorem 6.1](#)

by 1 and 2, corresponding to the manifolds M_1 and M_2 , respectively, the isomorphism i^* identifies C_2^* with C_1^* . From

$$\theta = \frac{\text{vol}(\Sigma_2/\Lambda_2)}{\text{vol}(\Sigma_2/i(\Lambda_1))},$$

we obtain that

$$\theta = \frac{\text{vol}(\Sigma_1/\Lambda_1^*)}{\text{vol}(\Sigma_1/i^*(\Lambda_2))}$$

and therefore $\text{vol}_{i^*(\Lambda_2)} = \theta \text{vol}_{\Lambda_1^*}$. The functions β_ϕ are also identified in the sense that, for $\phi_1 \in \Omega_1$, we have $\beta_{i(\phi_1)}(x) = \beta_{\phi_1}(i^*(x))$ for all $x \in \Sigma_2$. So

$$\begin{aligned} g_2(i(\phi_1)) &= \sqrt[d]{\frac{(d+1) \text{vol}_{\Lambda_2^*}(C_2^* \cap \beta_{i(\phi_1)}^{-1}([0, 1]))}{\min \text{occ}(C_2^* \cap \beta_{i(\phi_1)}^{-1}(1))}} \\ &= \sqrt[d]{\frac{(d+1) \text{vol}_{i^*(\Lambda_2^*)}(C_1^* \cap \beta_{\phi_1}^{-1}([0, 1]))}{\min \text{occ}(C_1^* \cap \beta_{\phi_1}^{-1}(1))}} \\ &= \sqrt[d]{\frac{(d+1)\theta \text{vol}_{\Lambda_1^*}(C_1^* \cap \beta_{\phi_1}^{-1}([0, 1]))}{\min \text{occ}(C_1^* \cap \beta_{\phi_1}^{-1}(1))}} = \theta^{1/d} g_1(\phi_1) \end{aligned}$$

by [Theorem 6.1](#). □

Finally, we show that the bounding function g in [Theorem 1.4](#) is convex.

Lemma 6.2 *For any $\delta > 0$ and integer $n \geq 1$, the function*

$$f(\alpha_1, \dots, \alpha_n) = \prod_{i=1}^n \alpha_i^{-\delta}$$

is convex on its natural domain $\{(\alpha_1, \dots, \alpha_n) : \alpha_i > 0 \text{ for } i = 1, \dots, n\}$.

Proof The Hessian of $\log f$ is a diagonal matrix with diagonal entries δ/α_i^2 . This matrix is positive definite; therefore, f is logarithmically convex. Every logarithmically convex function is also convex, since a composition of a convex function with the increasing convex function e^x is also convex. Hence, f is indeed convex. □

7 An example

In this section, we consider the simplest pseudo-Anosov braid on three strands, describe the veering triangulation of its mapping torus, and compute the asymptotic translation length in the arc complex for infinitely many fibrations of this 3-manifold. The purpose

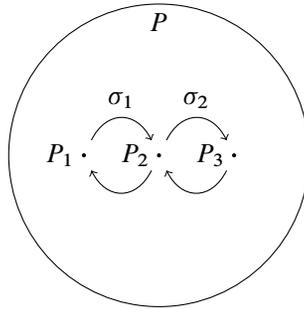


Figure 6: The half-twists σ_1 and σ_2 .

of this computation is two-fold: to illustrate the methods of Section 3 on a concrete example and to show that it seems very difficult to find an explicit formula for the normalized asymptotic translation length functions μ_d defined in (1-2).

A good reference for the pseudo-Anosov theory appearing in this section (invariant train tracks, measured foliations and translation surfaces) is [4, Chapters 14 and 15].

Let S be the sphere punctured at four points P_1, P_2, P_3 and P . Let σ_1 and σ_2 be the half-twists illustrated in Figure 6.

Invariant train tracks

The train track τ on the left of Figure 7 is invariant under $f = \sigma_1\sigma_2^{-1}$ (read from left to right) and the train track τ^{-1} on the right is invariant under f^{-1} . On each train track, measures are parametrized by measures on two of the branches. The action of f and f^{-1} on the measures are $(x_1, x_2) \mapsto (x_2 + 2x_1, x_1 + x_2)$ and $(y_1, y_2) \mapsto (y_2 + 2y_1, y_1 + y_2)$. In other words, both maps are described by the matrix $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, whose eigenvalues are φ^2 and φ^{-2} , where φ is the golden ratio, the largest root of $x^2 - x - 1$. The eigenvector corresponding to φ^2 is $(\varphi, 1)$. Therefore the unstable foliation \mathcal{F}^u is represented by the measure $(x_1, x_2) = (\varphi, 1)$ on τ and the stable foliation \mathcal{F}^s is represented by $(y_1, y_2) = (1, \varphi^{-1})$ on τ^{-1} . (The invariant measured foliations are well defined only up to scaling. We choose the scaling in a way that will be convenient later on.)



Figure 7: The invariant train tracks τ and τ^{-1} .

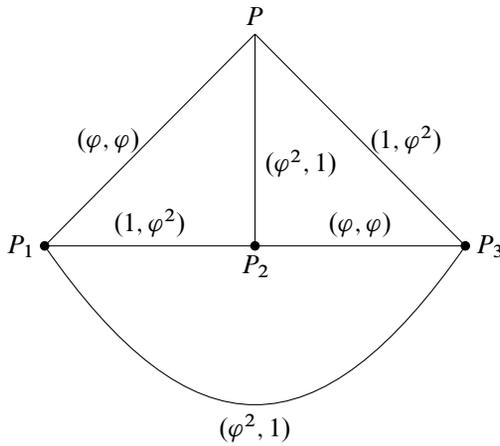


Figure 8: The measures of the edges with respect to \mathcal{F}^u (second coordinate) and \mathcal{F}^s (first coordinate).

The half-translation surface

Our next goal is to draw a picture of the half-translation surface whose horizontal foliation is \mathcal{F}^u and whose vertical foliation is \mathcal{F}^s . For this, consider the ideal triangulation of S consisting of four triangles, shown in Figure 8. The measures of the edges with respect to \mathcal{F}^u and \mathcal{F}^s can be obtained from the measured train tracks. These measures on the edges provide the widths and heights of the edges in the half-translation surface. Using these coordinates for the edges, we obtain the upper left picture in Figure 9 showing the half-translation surface defined by \mathcal{F}^u and \mathcal{F}^s .

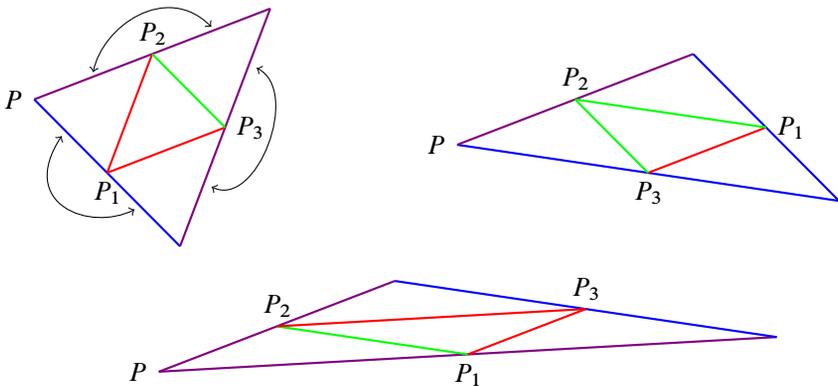


Figure 9: The half-translation surface defined by \mathcal{F}^u and \mathcal{F}^s . Pairs of boundary edges are identified by 180° rotations.

The veering triangulation

We can now use Guéritaud’s construction to find the veering triangulation of the mapping torus M . Flipping the edges P_1P_2 and PP_3 yields the upper right triangulation in Figure 9. Then flipping P_2P_3 and PP_1 yields the third triangulation in Figure 9. This triangulation is the image of the initial triangulation under f (stretched horizontally by φ^2 and compressed vertically by φ^{-2}). So the veering triangulation τ is obtained by gluing two tetrahedra below the initial triangulation, then two tetrahedra under that, and finally mapping the top (initial) triangulation to the bottom (final) triangulation by f .

Therefore, τ consists of four tetrahedra, four edges and eight faces. The four edges are colored by blue, red, purple and green. Note that the top and bottom edges of the tetrahedra and either blue and red or green and purple. For each tetrahedra, the other four edges are colored by four different colors.

The infinite cyclic cover of S

The homology of S is generated by the loops c_1, c_2 and c_3 around the punctures P_1, P_2 and P_3 , respectively. We have $f(c_1) = c_3, f(c_2) = c_1$ and $f(c_3) = c_2$. Therefore the f -invariant cohomology is $H^1(S; \mathbb{Z})^f = \langle \alpha \rangle$, where $\alpha(c_1) = \alpha(c_2) = \alpha(c_3) = 1$. Let t be a generator for $H = \text{Hom}(H^1(S; \mathbb{Z})^f, \mathbb{Z}) \cong \mathbb{Z}$. By evaluating elements of $H^1(S; \mathbb{Z})^f$ on loops, we obtain a surjective homomorphism $\pi_1(S) \rightarrow H$. Corresponding to this homomorphism is an infinite cyclic covering $\tilde{S} \rightarrow S$. For more details about the theory, see [11, Section 3].

To construct \tilde{S} explicitly, cut the upper left surface in Figure 9 along the edges PP_1, PP_2 and PP_3 , take infinitely many copies of this cut-up surface, and reglue the edges according to the labeling in the left column of Figure 10 to obtain a surface \tilde{S} . (Ignore the meaning of the labels for now; we will elaborate on that later.) The action of the deck transformation t is translating each triangle to the triangle below it. To check that this is the right infinite cyclic covering, all we need to check is that the loops c_1, c_2 and c_3 , oriented clockwise, all lift to paths in \tilde{S} connecting some point x to tx ; therefore, c_1, c_2 and c_3 all map to t under the homomorphism $\pi_1(S) \rightarrow H$.

The maximal abelian cover of M

To construct the maximal abelian cover \tilde{M} of M and its veering triangulation, we start with the triangulated surface in the left column of Figure 10 and we build down by

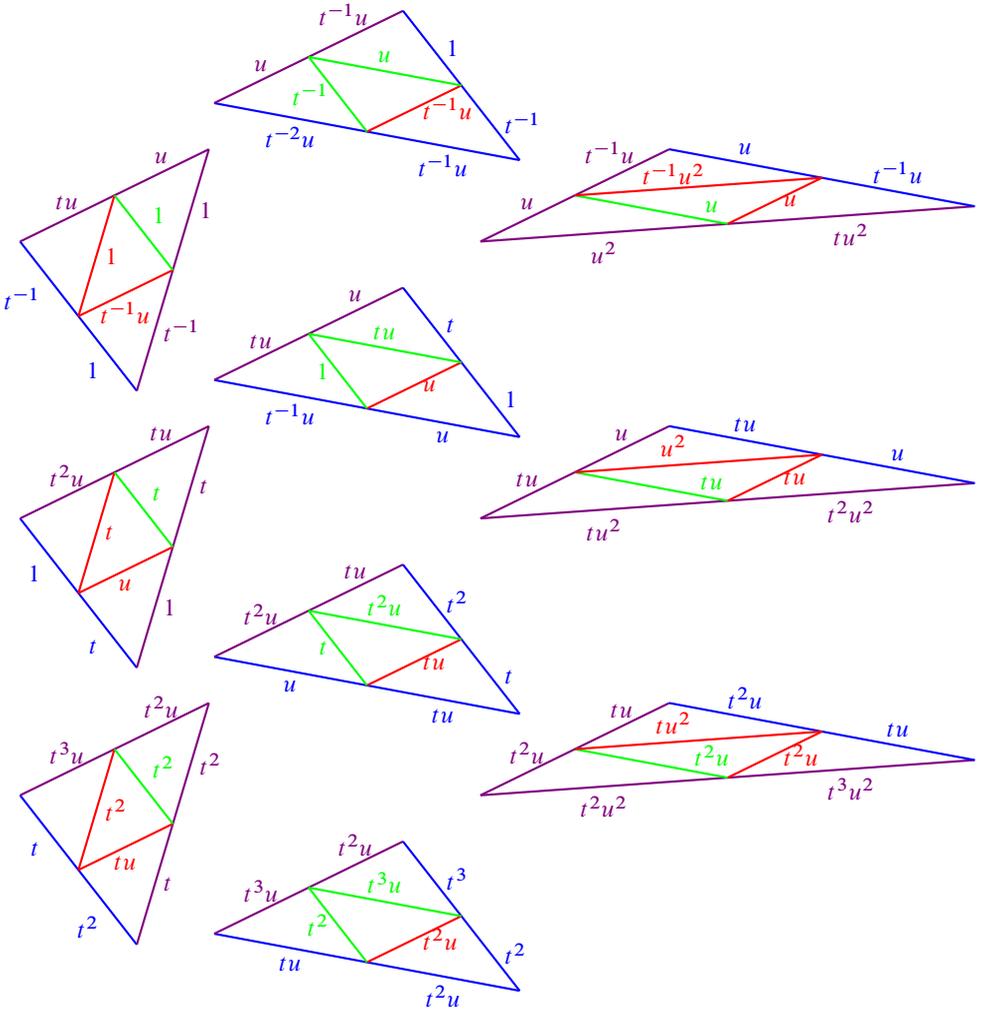


Figure 10: Part of the 2-skeleton of the veering triangulation on the maximal abelian cover \tilde{M} of M . The picture continues in all four directions indefinitely. The deck transformation t acts by translating down by one triangle. The deck transformation u acts by translating to the right by two columns, rotating each triangle by 180° , and stretching horizontally by φ^2 and vertically by $1/\varphi^2$.

gluing tetrahedra below it in the same way as we did for the construction of the veering triangulation of M .

First, we glue tetrahedra to all quadrilaterals whose diagonals are lifts of the edges P_1P_2 and PP_3 . The bottom of the resulting cell complex is triangulated as shown in the second column (after some rearranging of the triangles to make the second column

look similar to the first column). Then we glue another round of tetrahedra to the bottom again obtain a cell complex whose bottom is triangulated as shown in the third column. Take infinitely many copies of this cell complex, indexed by \mathbb{Z} . Choose a lift \tilde{f} of f identifying the triangulated surface in the first column with the triangulated surface in the third column, and use this to glue together the top of copy i with the bottom of copy $i + 1$. There is an isomorphism u of the resulting cell complex that maps copy $i + 1$ to copy i . (One should think of copy $i + 1$ to be above copy i in the flow. Therefore, u shifts downward.) Our 3-manifold M is the quotient of this cell complex by the group generated by t and u .

Labeling the edges by G

The labeling of the edges in [Figure 10](#) can be found as follows. First, for each color, label exactly one edge in the left column by 1. These edges are the chosen lifts of the four edges of the veering triangulation of M . For simplicity, we have chosen all four lifts in the upper left triangle.

Using the t -action, the translates of the four edges in the middle left and bottom left triangle should get the labels t and t^2 , respectively. Using the identification of the boundary edges, we can label all edges in the left column except one red and two purple edges in each triangle.

Now, using the u -action, we can label all edges in the right column except one red and two purple edges in each triangle.

The second column is obtained from the first column by flipping two edges for each triangle, and the third column is obtained from the second column analogously. This yields identifications between certain edges in the first and second columns and also in the second and third columns. In fact, the one purple and two red edges in each triangle in the first column are present in the third column, where they are already labeled. Copying this labeling to the first column yields a complete labeling of edges there.

Now, using the u -action, we obtain a complete labeling of the third column as well. Finally, using the identifications between the first and second and the second and third columns, respectively, it is possible to fully label the second column as well.

7.1 The graph Δ^*

Using [Figure 10](#), it is straightforward to construct the graph Δ^* defined in [Section 3.5](#). See [Figure 11](#).

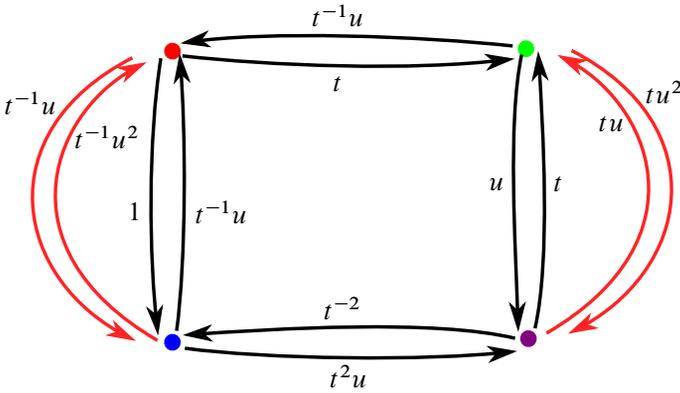


Figure 11: The graph Δ^* corresponding to the fibered face containing the pseudo-Anosov braid $f = \sigma_1\sigma_2^{-1}$. The black edges are the triangle edges and the red edges are the tetrahedron edges. Therefore the graph Δ is the subgraph consisting of black edges.

7.2 Minimal cycles and minimal good paths

The minimal cycles are listed in Table 1 with their drift. Denoting the set of drifts of minimal cycles by \mathcal{B} as in Section 3.6, we have

$$\mathcal{B} = \{t^{-1}u, u, tu, t^{-2}u^2, t^2u^2\}$$

and

$$\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} = \{t^a u^b : |a| \leq b\}.$$

The minimal good paths are listed in Table 2.

7.3 Determining the stashing sets $\text{Stash}(e, e')$

For a minimal good path γ , denote by $\text{Stash}(\gamma)$ the set of drifts of good paths that decompose as the union of γ with minimal cycles (see Lemma 3.8).

Proposition 7.1 *If γ is a minimal good path in Δ^* , then $\text{Stash}(\gamma) = \text{drift}(\gamma) \cdot D$, where*

$$D = \begin{cases} \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} - \{tu\} & \text{if } \gamma = RBR, BR, RBR, BRB, \\ \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} - \{t^{-1}u\} & \text{if } \gamma = GP, PG, GPG, PGP, \\ \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} & \text{otherwise.} \end{cases}$$

cycle	RBR	RGR	BPB	GPG	RBPGR	RGPBR
drift	$t^{-1}u$	u	u	tu	t^2u^2	$t^{-2}u^2$

Table 1: The drifts of minimal cycles.

path drift	<i>RB</i> $t^{-1}u$	<i>RBR</i> $t^{-2}u^2$	<i>RBP</i> tu^2	<i>RBRG</i> $t^{-1}u^2$	<i>RBPG</i> t^2u^2	<i>RBRGP</i> $t^{-1}u^3$	<i>RBPGR</i> tu^3
path drift	<i>BR</i> $t^{-1}u^2$	<i>BRB</i> $t^{-1}u^2$	<i>BRG</i> u^2	<i>BRBP</i> tu^3	<i>BRGP</i> u^3	<i>BRBPG</i> t^2u^3	<i>BRGPB</i> $t^{-2}u^3$
path drift	<i>GP</i> tu^2	<i>GPG</i> t^2u^2	<i>GPB</i> $t^{-1}u^2$	<i>GPGR</i> tu^3	<i>GPBR</i> $t^{-2}u^3$	<i>GPGRB</i> tu^3	<i>GPBRG</i> $t^{-1}u^3$
path drift	<i>PG</i> tu	<i>PGP</i> tu^2	<i>PGR</i> u^2	<i>PGPB</i> $t^{-1}u^2$	<i>PGRB</i> u^2	<i>PGPBR</i> $t^{-2}u^3$	<i>PGRBP</i> t^2u^3

Table 2: The drifts of minimal good paths.

Proof If γ consists of three or four edges, then it forms a connected collection with all minimal cycles. So, in these cases, we have $D = \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$.

If $\gamma = RBP$, then the only minimal cycle γ does not form a connected collection with RGR . But the drift of the cycle BPB is the same as the drift of the cycle RGR , so we still have $D = \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$. We obtain $D = \langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ similarly for $\gamma = BRG, GPB$ and PGB .

For the remaining four possibilities for γ starting with R or B (namely RB, BR, RBR and BRB), the cycles $RBR, RBPGR$ and $RGPBR$ form a connected collection with γ , so $t^{-1}u, t^2u^2, t^{-2}u^2 \in D$. Also, at least one of RGR and BPB forms a connected collection with γ , so $u \in D$. However, the cycle GPG does not form a connected collection with γ , so $tu \notin D$.

It remains to show that all elements of $\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} - \{tu\}$ that are not nonnegative integral linear combinations of $u, t^{-1}u$ and t^2u^2 are in D . This follows from the fact that, once we add to γ a minimal cycle with drift u or t^2u^2 , we can now add the cycle GPG to obtain a connected collection. So the translated cones $u\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ and $t^2u^2\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$ are contained in D . This completes the proof in the case that γ starts with R or B .

The proof is analogous if γ starts with G or P . □

Proposition 7.1 allows us to determine the stashing set $\text{Stash}(e, e')$ for every pair $e, e' \in \{B, R, G, P\}$ using the formula

$$\text{Stash}(e, e') = \bigcup \{ \text{Stash}(\gamma) : \gamma \text{ is a minimal good path from } e \text{ to } e' \}.$$

For example,

$$\begin{aligned} \text{Stash}(R, R) &= \text{Stash}(RBR) \cup \text{Stash}(RBPGR) \\ &= t^{-2}u^2(\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} - \{tu\}) \cup tu^3\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}, \end{aligned}$$

since $\text{drift}(RBR) = t^{-2}u^2$ and $\text{drift}(RBPGR) = tu^3$. We can visualize that computation as follows.

Draw the cone $\langle t^{-1}u, u, tu \rangle_{\mathbb{Z}_{\geq 0}}$ as in Figure 12. The horizontal axis is the t -axis and the vertical axis is the u -axis. Mark the point $t^{-2}u^2$ with a “left tick” and the point tu^3 with a cross, indicating the coloring of the sets $t^{-2}u^2(\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}} - \{tu\})$ and $tu^3\langle \mathcal{B} \rangle_{\mathbb{Z}_{\geq 0}}$. The union of these two sets forms $\text{Stash}(R, R)$. Figure 12 illustrates the computation for all $\text{Stash}(e, e')$.

7.4 The fibered face

Denote by $\phi_{a,b}$ the element of $H^1(M)$ such that $\phi_{a,b}(t) = a$ and $\phi_{a,b}(u) = b$.

Lemma 7.2 *The monodromy $f = \sigma_1\sigma_2^{-1}$ corresponds to $\phi_{0,-1}$.*

Proof The homology class t can be represented by loops in M that are in the fiber dual to f , and loops representing u wind around the fibration dual to f once in the opposite direction of the flow. □

Let \mathcal{F} be the fibered face such that the cone $\mathbb{R}_+\mathcal{F}$ contains $\phi_{0,-1}$.

Lemma 7.3 $\mathbb{R}_+\mathcal{F} = \{\phi_{a,b} : |a| \leq -b, b < 0\}$.

Proof By Proposition 4.5, the cone $\mathbb{R}_+\mathcal{F}$ contains precisely those cohomology classes that take nonpositive values on $\langle \mathcal{B} \rangle_{\mathbb{R}_{\geq 0}} = \langle t^{-1}u, tu \rangle_{\mathbb{R}_{\geq 0}}$. □

Lemma 7.4 $\|\phi_{0,-1}\| = 2$.

Proof The fiber dual to f is a four-punctured sphere and has Euler characteristic -2 . □

Lemma 7.5 *We have $\|\phi_{a,b}\| = -2b$ whenever $\phi_{a,b} \in \mathbb{R}_+\mathcal{F}$.*

Proof This follows from Lemma 7.4 and the fact that the Thurston norm on the cone $\mathbb{R}_+\mathcal{F}$ has the symmetry $\|\phi_{a,b}\| = \|\phi_{-a,b}\|$. This symmetry can be seen from the following symmetry δ of the veering triangulation of M . The symmetry δ maps the upper left triangulation in Figure 9 to the upper right triangulation by vertical reflection and a horizontal stretch. This map extends to a symmetry of the whole

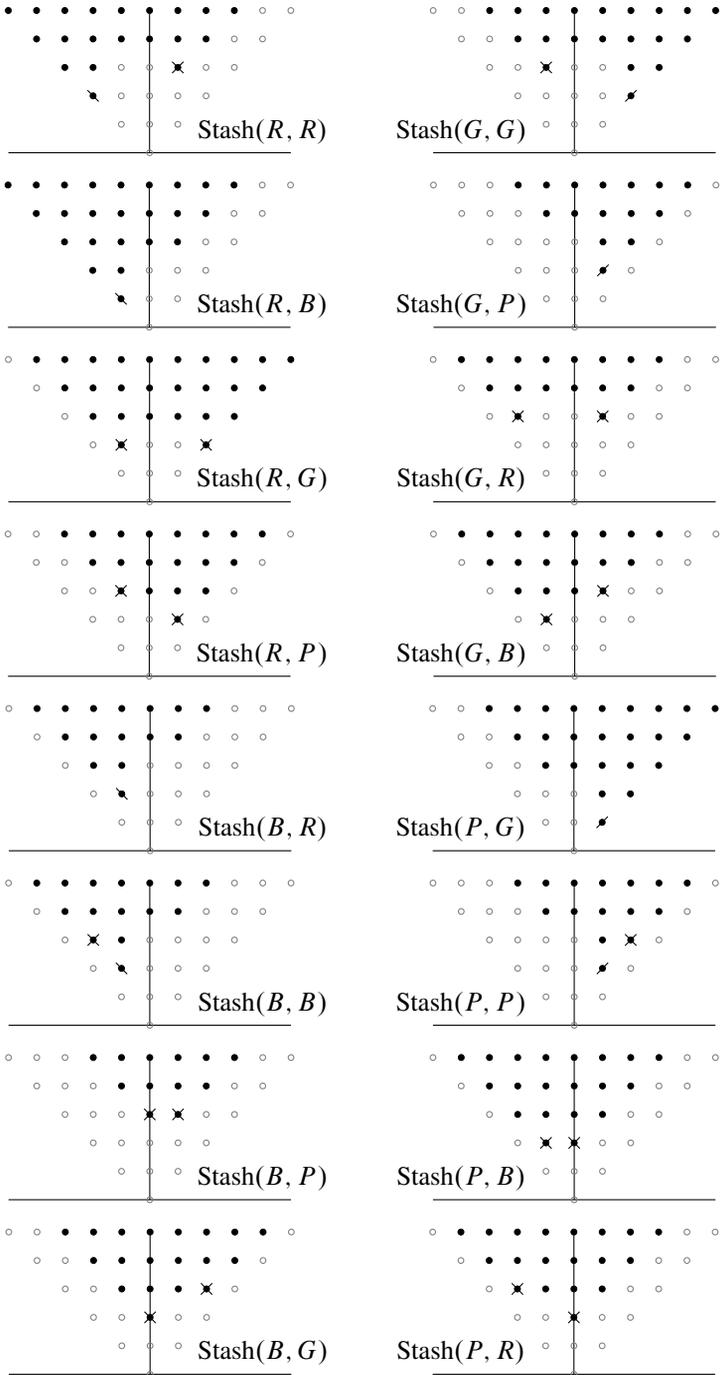


Figure 12: The sets $\text{Stash}(e, e')$ for all pairs $e, e' \in \{R, B, G, P\}$.

veering triangulation and exchanges the red edge with the green edge and the blue edge with the purple edge. The symmetry $\delta: M \rightarrow M$ maps the fiber S to a homotopic fiber by an orientation-reversing map, but preserves the orientation of the flow. Therefore, the action on homology is $\delta_*(u) = u$ and $\delta_*(t) = -t$. So, in the cone $\mathbb{R}_+\mathcal{F}$, two integral points $\phi_{-a,b}$ and $\phi_{a,b}$ are dual to homeomorphic fibers and hence their Thurston norms are equal. \square

Corollary 7.6 $\mathcal{F} = \{\phi_{a,-1/2} : |a| \leq \frac{1}{2}\}.$

7.5 Accumulation points of the graph of μ_1

We will apply [Theorem 1.4](#) to the fibered face \mathcal{F} to find the accumulation points of the graph of the function μ_1 . According to the theorem, the set of accumulation points is the graph of a continuous function $g: \text{int}(\mathcal{F}) \rightarrow \mathbb{R}_+.$

Proposition 7.7 $g(\phi_{a,-1/2}) = \frac{2}{(\frac{1}{2}-a)(\frac{1}{2}+a)}.$

Proof By [Theorem 1.4](#), we have

$$g(\alpha\omega_1 + (1-\alpha)\omega_2) = g^*(\alpha, 1-\alpha) = \frac{1}{\text{vol}_\Lambda(\Sigma/\langle\omega_1, \omega_2\rangle_{\mathbb{Z}}) \cdot \alpha(1-\alpha)},$$

where $\omega_1 = \phi_{-1/2,-1/2}$ and $\omega_2 = \phi_{1/2,-1/2}$, $\Sigma = H^1(M; \mathbb{R})$ and $\Lambda = H^1(M; \mathbb{Z}).$ Since

$$\det \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} = \frac{1}{2},$$

the covolume in the denominator is $\frac{1}{2}.$ So

$$g(\phi_{-\alpha/2+(1-\alpha)/2,-1/2}) = \frac{2}{\alpha(1-\alpha)}$$

and, by substituting $a = \frac{1}{2} - \alpha,$ we obtain the desired formula. \square

7.6 Exact values of the function μ_1

Proposition 7.8 *For the primitive integral cohomology classes ϕ listed in [Table 3](#), the asymptotic translation length $\ell_{\mathcal{A}}(\phi)$ of the monodromy corresponding to ϕ is as shown in the table.*

For each fibration, the table also shows the cycles of Δ^ with maximal average weight.*

ϕ	$\ell_{\mathcal{A}}(\phi)$	maximal cycles
$\phi_{0,-1}$	$\frac{2}{3}$	<i>BPB, RGR</i>
$\phi_{1,-2}$	$\frac{1}{6}$	<i>BB</i>
$\phi_{1,-3}$	$\frac{1}{9}$	<i>BB, RR</i>
$\phi_{1,-4}$	$\frac{1}{13}$	<i>RR</i>
$\phi_{1,-k}$ ($k \geq 5$ odd)	$2/(k+1)^2$	<i>BB</i>
$\phi_{1,-k}$ ($k \geq 6$ even)	$2/(k^2+2k-1)$	<i>BPB</i>

Table 3

We remark that these cycles correspond to bi-infinite geodesics in the arc complex of the fiber that are invariant under some power of the monodromy.

Proof Using the stashing sets $\text{Stash}(e, e')$ shown in Figure 12, we can determine the weighted graphs $W(\phi)$ using the definition (3-2).

Case 1 ($\phi_{0,-1}, \phi_{1,-2}, \phi_{1,-3}$ and $\phi_{1,-4}$) In these cases, we find the graphs $W(\phi)$ shown in Figure 13 by a case-by-case inspection of each set in Figure 12. The maximum averages are $\frac{3}{2}, 6, 9$ and 13 ; therefore, the asymptotic translation lengths are $\frac{2}{3}, \frac{1}{6}, \frac{1}{9}$ and $\frac{1}{13}$, respectively, by Proposition 3.5.

Case 2 ($k \geq 5$ is odd) Note that $t^{(k-1)/2}u^{(k+3)/2}$ and $t^{-(k+1)/2}u^{(k+1)/2}$ are not in $\text{Stash}(B, B)$, and both evaluate to $-\frac{1}{2}(k+1)^2$ by $\phi_{1,-k}$. Hence, $\frac{1}{2}(k+1)^2 \notin -\phi_{1,-k}(\text{Stash}(B, B))$. One easily verifies that $\frac{1}{2}(k+1)^2$ is in fact the largest integer that is not contained in $-\phi_{1,-k}(\text{Stash}(B, B))$. Since $\text{Stash}(B, R) = \text{Stash}(B, B)$, it is also the largest integer not contained in $-\phi_{1,-k}(\text{Stash}(B, R))$.

The set

$$Z_k = \{t^a u^b : -\frac{1}{2}(k-1) \leq a \leq \frac{1}{2}(k-1), b \geq \frac{1}{2}(k+3)\}$$

is contained in $\text{Stash}(R, R), \text{Stash}(R, B), \text{Stash}(R, G), \text{Stash}(R, P), \text{Stash}(B, G), \text{Stash}(G, G), \text{Stash}(G, R), \text{Stash}(G, B), \text{Stash}(P, G), \text{Stash}(P, B)$ and $\text{Stash}(P, R)$, so the weights of the corresponding edges are less than

$$k \cdot \frac{1}{2}(k+3) - \frac{1}{2}(k-1) = \frac{1}{2}(k^2+2k+1) = \frac{1}{2}(k+1)^2.$$

The set tZ_k is contained in $\text{Stash}(G, P) = \text{Stash}(P, P)$, so the weights of the corresponding edges are strictly less than $\frac{1}{2}(k+1)^2 - 1$. For the only remaining pair, (B, P) , one can check that $w(B, P) = \frac{1}{2}(k+1)^2 - 1$.

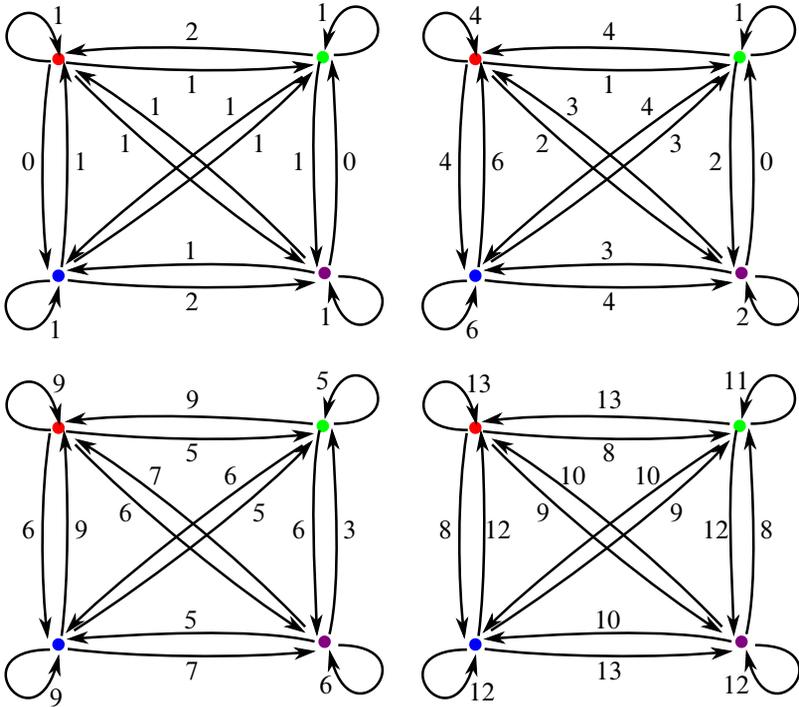


Figure 13: The graphs $W(\phi_{0,-1})$ (top left), $W(\phi_{1,-2})$ (top right), $W(\phi_{1,-3})$ (bottom left) and $W(\phi_{1,-4})$ (bottom right).

Therefore, $w(B, B) = w(B, R) = \frac{1}{2}(k + 1)^2$ and the weights of the other edges are strictly smaller. Therefore, the largest average cycle weight is $\frac{1}{2}(k + 1)^2$, realized only by the loop on the blue vertex.

Case 3 ($k \geq 6$ is even) The set

$$W_k = \{t^a u^b : -\frac{1}{2}k \leq a \leq \frac{1}{2}(k - 2), b \geq \frac{1}{2}(k + 2)\} - \{t^{(k-2)/2} u^{(k+2)/2}\}$$

is contained in $\text{Stash}(R, R)$, $\text{Stash}(R, B)$, $\text{Stash}(R, G)$, $\text{Stash}(B, R)$, $\text{Stash}(B, B)$, $\text{Stash}(G, R)$ and $\text{Stash}(G, B)$; therefore, the weights of the corresponding edges are at most

$$X_k = k \cdot \frac{1}{2}(k + 2) - \frac{1}{2}(k - 2) = \frac{1}{2}(k^2 + k + 2).$$

The set tW_k is contained in $\text{Stash}(R, P)$, $\text{Stash}(B, G)$, $\text{Stash}(G, G)$, $\text{Stash}(P, G)$, $\text{Stash}(P, B)$ and $\text{Stash}(P, R)$; therefore, the weights of the corresponding edges are at most $X_k - 1$.

The set t^2W_k is contained in $\text{Stash}(G, P)$, $\text{Stash}(P, P)$, therefore the weights of the corresponding edges are at most $X_k - 2$.

There is one remaining edge: BP . Neither $t^{-(k-2)/2}u^{(k+2)/2}$ nor $t^{(k+2)/2}u^{(k+4)/2}$ is in $\text{Stash}(B, P)$ and both expressions evaluate to $\frac{1}{2}(k^2 + 3k - 2)$, so one can verify that $w(B, P) = \frac{1}{2}(k^2 + 3k - 2)$. One can also check that $w(P, B)$ is in fact exactly $X_k - 1$; therefore, the cycle BPB has average weight $\frac{1}{2}(k^2 + 2k - 1)$. Since this weight is larger than X_k , no other cycle can have the same or larger average weight. \square

Finally, we give the proof of [Theorem 1.6](#).

Proof of Theorem 1.6 Parametrize the fibered face \mathcal{F} with the interval $[-1, 1]$, using the map $\phi_{a,b} \mapsto a/b$.

Using [Lemma 7.5](#), we have $\|\phi_{0,-1}\|^2 = 4$ and $\|\phi_{\pm 1,-k}\|^2 = 4k^2$ for every integer $k \geq 2$. Together with [Proposition 7.8](#), we obtain the values of $\mu_1(t)$ in [Theorem 1.6](#) for $t \leq 0$. By the symmetry discussed in the proof of [Lemma 7.5](#), we have $\mu_1(t) = \mu_1(-t)$ for all $t \in (-1, 1)$, which yields the claimed values of $\mu_1(t)$ when $t > 0$.

Using the substitution $t = a/(-\frac{1}{2})$, whence $a = -\frac{1}{2}t$, in [Proposition 7.7](#), we obtain that the set of accumulation points of the graph of $\mu_1(t)$ is the graph of

$$\frac{2}{(\frac{1}{2} + \frac{1}{2}t)(\frac{1}{2} - \frac{1}{2}t)} = \frac{8}{1 - t^2},$$

as claimed. Finally, it is straightforward to check that $\mu_1(t) < 8/(1 - t^2)$ for all values of t for which we have determined $\mu_1(t)$. \square

References

- [1] **I Agol**, *Ideal triangulations of pseudo-Anosov mapping tori*, from “Topology and geometry in dimension three” (W Li, L Bartolini, J Johnson, F Luo, R Myers, JH Rubinstein, editors), Contemp. Math. 560, Amer. Math. Soc., Providence, RI (2011) 1–17 [MR](#) [Zbl](#)
- [2] **H Baik, E Kin, H Shin, C Wu**, *Asymptotic translation lengths and normal generation for pseudo-Anosov monodromies of fibered 3-manifolds*, *Algebr. Geom. Topol.* 23 (2023) 1363–1398 [MR](#) [Zbl](#)
- [3] **H Baik, H Shin, C Wu**, *An upper bound on the asymptotic translation lengths on the curve graph and fibered faces*, *Indiana Univ. Math. J.* 70 (2021) 1625–1637 [MR](#) [Zbl](#)
- [4] **B Farb, D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) [MR](#) [Zbl](#)
- [5] **D Fried**, *Flow equivalence, hyperbolic systems and a new zeta function for flows*, *Comment. Math. Helv.* 57 (1982) 237–259 [MR](#) [Zbl](#)

- [6] **D Fried**, *The geometry of cross sections to flows*, Topology 21 (1982) 353–371 [MR](#) [Zbl](#)
- [7] **D Futer**, **S Schleimer**, *Cusp geometry of fibered 3–manifolds*, Amer. J. Math. 136 (2014) 309–356 [MR](#) [Zbl](#)
- [8] **V Gadre**, **C-Y Tsai**, *Minimal pseudo-Anosov translation lengths on the complex of curves*, Geom. Topol. 15 (2011) 1297–1312 [MR](#) [Zbl](#)
- [9] **F Guéritaud**, *Veering triangulations and Cannon–Thurston maps*, J. Topol. 9 (2016) 957–983 [MR](#) [Zbl](#)
- [10] **E Kin**, **H Shin**, *Small asymptotic translation lengths of pseudo-Anosov maps on the curve complex*, Groups Geom. Dyn. 13 (2019) 883–907 [MR](#) [Zbl](#)
- [11] **CT McMullen**, *Polynomial invariants for fibered 3–manifolds and Teichmüller geodesics for foliations*, Ann. Sci. École Norm. Sup. 33 (2000) 519–560 [MR](#) [Zbl](#)
- [12] **YN Minsky**, **SJ Taylor**, *Fibered faces, veering triangulations, and the arc complex*, Geom. Funct. Anal. 27 (2017) 1450–1496 [MR](#) [Zbl](#)
- [13] **J-P Otal**, *Le théorème d’hyperbolisation pour les variétés fibrées de dimension 3*, Astérisque 235, Soc. Math. France, Paris (1996) [MR](#) [Zbl](#)
- [14] **V Poénaru**, *Classification des difféomorphismes des surfaces*, from “Travaux de Thurston sur les surfaces” (A Fathi, F Laudenbach, V Poénaru, editors), Astérisque 66–67, Soc. Math. France, Paris (1979) Exposé 9 [MR](#) [Zbl](#)
- [15] **JL Ramírez Alfonsín**, *The Diophantine Frobenius problem*, Oxford Lecture Series in Mathematics and its Applications 30, Oxford Univ. Press (2005) [MR](#) [Zbl](#)
- [16] **WP Thurston**, *A norm for the homology of 3–manifolds*, Mem. Amer. Math. Soc. 339, Amer. Math. Soc., Providence, RI (1986) 99–130 [MR](#) [Zbl](#)
- [17] **AD Valdivia**, *Asymptotic translation length in the curve complex*, New York J. Math. 20 (2014) 989–999 [MR](#) [Zbl](#)

Department of Mathematics, Georgia Institute of Technology
Atlanta, GA, United States

strennerb@gmail.com

Received: 14 March 2021

Revised: 30 January 2022

A uniformizable spherical CR structure on a two-cusped hyperbolic 3–manifold

YUEPING JIANG

JIEYAN WANG

BAOHUA XIE

Let $\langle I_1, I_2, I_3 \rangle$ be the complex hyperbolic $(4, 4, \infty)$ triangle group. We prove Schwartz’s conjecture that $\langle I_1, I_2, I_3 \rangle$ is discrete and faithful if and only if $I_1 I_3 I_2 I_3$ is nonelliptic. If $I_1 I_3 I_2 I_3$ is parabolic, we show that the even subgroup $\langle I_2 I_3, I_2 I_1 \rangle$ is the holonomy representation of a uniformizable spherical CR structure on the two-cusped hyperbolic 3–manifold $s782$ in SnapPy notation.

20H10, 22E40, 51M10, 57M50

1 Introduction

Let $\mathbb{H}_{\mathbb{C}}^2$ be the complex hyperbolic plane and $\mathrm{PU}(2, 1)$ be its holomorphic isometry group; see [Section 2](#) for more details. It is well known that $\mathbb{H}_{\mathbb{C}}^2$ is one of the rank-one symmetric spaces and $\mathrm{PU}(2, 1)$ is a semisimple Lie group. $\mathbb{H}_{\mathbb{C}}^2$ can be viewed as the unit ball in \mathbb{C}^2 equipped with the Bergman metric. Its ideal boundary $\partial\mathbb{H}_{\mathbb{C}}^2$ is the 3–sphere S^3 . We study the geometry of discrete subgroups of $\mathrm{PU}(2, 1)$.

Let M be a 3–manifold. A *spherical CR structure* on M is a system of coordinate charts into S^3 such that the transition functions are restrictions of elements of $\mathrm{PU}(2, 1)$. Any spherical CR structure on M determines a pair (ρ, d) , where $\rho: \pi_1(M) \rightarrow \mathrm{PU}(2, 1)$ is the holonomy and $d: \tilde{M} \rightarrow S^3$ is the developing map. There is a special spherical CR structure. A *uniformizable spherical CR structure* on M is a homeomorphism between M and a quotient space Ω/Γ , where Γ is a discrete subgroup of $\mathrm{PU}(2, 1)$ and $\Omega \subset \partial\mathbb{H}_{\mathbb{C}}^2$ is the discontinuity region of Γ . An interesting problem in complex hyperbolic geometry is to find (uniformizable) spherical CR structures on hyperbolic 3–manifolds.

Geometric structures modeled on the boundary of complex hyperbolic space are rather difficult to construct. The first example of a spherical CR structure existing on a cusped

hyperbolic 3–manifold was discovered by Schwartz. In [23], Schwartz constructed a uniformizable spherical CR structure on the Whitehead link complement. He also constructed a closed hyperbolic 3–manifold that admits a uniformizable spherical CR structure in [26] at almost the same time.

Let M_8 be the complement of the figure eight knot. In [9], Falbel constructed two different representations ρ_1 and ρ_2 of $\pi_1(M_8)$ in $\text{PU}(2, 1)$, and proved that ρ_1 is the holonomy of a spherical CR structure on M_8 . In [11], Falbel and Wang proved that ρ_2 is also the holonomy of a spherical CR structure on M_8 . In [7], Deraux and Falbel constructed a uniformizable spherical CR structure on M_8 whose holonomy is ρ_2 . In [6], Deraux proved that there is a 1–parameter family of spherical CR uniformizations of the figure eight knot complement. This family is in fact a deformation of the uniformization constructed in [7].

Let us return to the Whitehead link complement. It admits a uniformizable spherical CR structure which is different from Schwartz's. In the recent work [22], Parker and Will also constructed a spherical CR uniformization of the Whitehead link complement. By applying spherical CR Dehn surgery theorems to the uniformizations of the Whitehead link complement, one can get infinitely many manifolds which admit uniformizable spherical CR structures. In [28], Schwartz proved a spherical CR Dehn surgery theorem, and applied it to the spherical CR uniformization of the Whitehead link complement constructed in [23] to obtain infinitely many closed hyperbolic 3–manifolds which admit uniformizable spherical CR structures. In [2], Acosta applied the spherical CR Dehn surgery theorem he proved in [1] to the spherical CR uniformization of the Whitehead link complement constructed by Parker and Will in [22] to obtain infinitely many one-cusped hyperbolic 3–manifolds which admit uniformizable spherical CR structures. In particular, the spherical CR uniformization of the complement of the figure eight knot constructed by Deraux and Falbel [7] is contained in this family.

There are some hyperbolic 3–manifolds described in the SnapPy census (see [4]) which admit spherical CR structures. In [5], Deraux proved that the cusped hyperbolic 3–manifold $m009$ admits a uniformizable spherical CR structure whose holonomy representation was constructed by Falbel, Koseleff and Rouillier in [10]. In [16; 18], Ma and Xie proved that the cusped hyperbolic 3–manifolds $m038$, $s090$, $m295$ and 6_1^3 admit spherical CR uniformizations. They also gave the second explicit example of a closed hyperbolic 3–orbifold which admits a uniformizable spherical CR structure in [17].

We show that the two-cusped hyperbolic 3–manifold $s782$ admits a uniformizable spherical CR structure. By studying the action of the even subgroup of a discrete

complex hyperbolic triangle group on $\mathbb{H}_{\mathbb{C}}^2$, we prove that the quotient space of its discontinuity region is homeomorphic to $s782$. That means the holonomy representation of the spherical CR uniformization of $s782$ is a triangle group.

Now let us talk about the complex hyperbolic triangle groups. Let $\Delta_{p,q,r}$ be the abstract (p, q, r) reflection triangle group with the presentation

$$\langle \sigma_1, \sigma_2, \sigma_3 \mid \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = (\sigma_2\sigma_3)^p = (\sigma_3\sigma_1)^q = (\sigma_1\sigma_2)^r = \text{id} \rangle,$$

where p, q and r are positive integers, or ∞ in which case the corresponding relation disappears. A *complex hyperbolic (p, q, r) triangle group* is a representation of $\Delta_{p,q,r}$ in $\text{PU}(2, 1)$, which maps the generators to complex involutions fixing complex lines in $\mathbb{H}_{\mathbb{C}}^2$. The study of complex hyperbolic triangle groups was begun by Goldman and Parker, and in [13] they studied the complex hyperbolic (∞, ∞, ∞) triangle groups. They conjectured that a representation of $\Delta_{\infty,\infty,\infty}$ into $\text{PU}(2, 1)$ is discrete and faithful if and only if the image of $\sigma_1\sigma_2\sigma_3$ is nonelliptic. The Goldman–Parker conjecture was proved by Schwartz in [24] (and with a better proof in [27]). In particular, the representation with the image of $\sigma_1\sigma_2\sigma_3$ being parabolic is closely related with the holonomy of the spherical CR uniformization of the Whitehead link complement constructed in [23]. In the survey [25], a series of conjectures on complex hyperbolic triangle groups are put forward.

Conjecture 1.1 (Schwartz [25]) *Suppose that $p \leq q \leq r$. Let $\langle I_1, I_2, I_3 \rangle$ be a complex hyperbolic (p, q, r) triangle group. Then $\langle I_1, I_2, I_3 \rangle$ is a discrete and faithful representation of $\Delta_{p,q,r}$ if and only if $I_1I_3I_2I_3$ and $I_1I_2I_3$ are nonelliptic. Moreover:*

- *If $3 \leq p < 10$, then $\langle I_1, I_2, I_3 \rangle$ is discrete and faithful if and only if $I_1I_3I_2I_3$ is nonelliptic.*
- *If $p > 13$, then $\langle I_1, I_2, I_3 \rangle$ is discrete and faithful if and only if $I_1I_2I_3$ is nonelliptic.*

In a recent work [22], Parker and Will proved [Conjecture 1.1](#) for complex hyperbolic $(3, 3, \infty)$ groups. They also showed that, when $I_1I_3I_2I_3$ is parabolic, the quotient of $\mathbb{H}_{\mathbb{C}}^2$ by the group $\langle I_2I_3, I_2I_1 \rangle$ is a complex hyperbolic orbifold whose boundary is a spherical CR uniformization of the Whitehead link complement. In [21], Parker, Wang and Xie proved [Conjecture 1.1](#) for complex hyperbolic $(3, 3, n)$ groups with $n \geq 4$. Furthermore, Acosta [2] showed that when $I_1I_3I_2I_3$ is parabolic the group $\langle I_2I_3, I_2I_1 \rangle$ is the holonomy representation of a uniformizable spherical CR structure on the Dehn surgery of the Whitehead link complement on one cusp of type $(1, n - 3)$.

We give a proof of [Conjecture 1.1](#) for the complex hyperbolic $(4, 4, \infty)$ triangle groups and further analyze the group when $I_1 I_3 I_2 I_3$ is parabolic. Our result is as follows:

Theorem 1.2 *Let $\langle I_1, I_2, I_3 \rangle$ be a complex hyperbolic $(4, 4, \infty)$ triangle group. Then $\langle I_1, I_2, I_3 \rangle$ is a discrete and faithful representation of $\Delta_{4,4,\infty}$ if and only if $I_1 I_3 I_2 I_3$ is nonelliptic. Moreover, when $I_1 I_3 I_2 I_3$ is parabolic, the quotient of $\mathbb{H}_{\mathbb{C}}^2$ by the group $\langle I_2 I_3, I_2 I_1 \rangle$ is a complex hyperbolic orbifold whose boundary is a spherical CR uniformization of the two-cusped hyperbolic 3-manifold $s782$ in the SnapPy census.*

In [\[29\]](#), Wyss-Gallifent studied the complex hyperbolic $(4, 4, \infty)$ triangle groups. He discovered several discrete groups with $I_1 I_3 I_2 I_3$ being regular elliptic of finite order and conjectured that there should be countably infinitely many. It would be very interesting to know what the manifold at infinity is for the group with $I_1 I_3 I_2 I_3$ being regular elliptic of finite order.

Our method is to construct Ford domains for the triangle groups acting on $\mathbb{H}_{\mathbb{C}}^2$. The space of complex hyperbolic $(4, 4, \infty)$ triangle groups $\langle I_1, I_2, I_3 \rangle$ is parametrized by the angle $\theta \in [0, \frac{\pi}{2})$; see [Section 3](#). Let $S = I_2 I_3$, $T = I_2 I_1$ and $\Gamma = \langle S, T \rangle$. Here S is regular elliptic of order 4, and T is parabolic fixing the point at infinity. For each group in the parameter space, the Ford domain D is the intersection of the closures of the exteriors of the isometric spheres for the elements $S, S^{-1}, S^2, (S^{-1}T)^2$ and their conjugations by the powers of T . The combination of D is the same except when $I_1 I_3 I_2 I_3$ is parabolic, in which case there are additional parabolic fixed points. D is preserved by the subgroup $\langle T \rangle$ and is a fundamental domain for the cosets of $\langle T \rangle$ in Γ . Its ideal boundary $\partial_{\infty} D$ is the complement of a tubular neighborhood of the T -invariant \mathbb{R} -circle (or horotube defined in [\[28\]](#)). By intersecting $\partial_{\infty} D$ with a fundamental domain for $\langle T \rangle$ acting on $\partial \mathbb{H}_{\mathbb{C}}^2$, we obtain a fundamental domain for Γ acting on its discontinuity region; see [Section 4](#).

When $I_1 I_3 I_2 I_3$ is parabolic, that is $\theta = \frac{\pi}{3}$, there are four additional parabolic fixed points fixed by $T^{-1} S^2, S^2 T^{-1}, S T^{-1} S$ and $T^{-1} S T^{-1} S T$, except the point at infinity which is the fixed point of T ; see [Section 5](#). By studying the combinatorial properties of the fundamental domain for Γ acting on its discontinuity region $\Omega(\Gamma)$, we prove that the quotient $\Omega(\Gamma)/\Gamma$ is homeomorphic to the two-cusped hyperbolic 3-manifold $s782$. Motivated by the work of Acosta [\[2\]](#), we guess that there are similar structures on its surgeries.

Acknowledgments We thank Jiming Ma for his help in the proof of [Theorem 5.22](#). Xie also would like to thank Jiming Ma for numerous helpful discussions on complex

hyperbolic geometry during his visit to Fudan University. We would like to thank the referee for comments which improved a previous version of this paper. Jiang was supported by NSFC 12271148. Wang was supported by NSFC 11701165. Xie was supported by NSFC 11871202 and Hunan Provincial Natural Science Foundation of China 2018JJ3024.

2 Background

The purpose of this section is to briefly introduce complex hyperbolic geometry. One can refer to Goldman’s book [12] for more details.

2.1 Complex hyperbolic plane

Let $\langle z, w \rangle = w^* H z$ be the Hermitian form on \mathbb{C}^3 associated to H , where H is the Hermitian matrix

$$H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Then \mathbb{C}^3 is the union of the negative cone V_- , null cone V_0 and positive cone V_+ , where

$$V_- = \{z \in \mathbb{C}^3 - \{0\} : \langle z, z \rangle < 0\}, \quad V_0 = \{z \in \mathbb{C}^3 - \{0\} : \langle z, z \rangle = 0\},$$

and

$$V_+ = \{z \in \mathbb{C}^3 - \{0\} : \langle z, z \rangle > 0\}.$$

Definition 2.1 Let $P: \mathbb{C}^3 - \{0\} \rightarrow \mathbb{C}P^2$ be the projectivization map. Then the *complex hyperbolic plane* $\mathbb{H}_{\mathbb{C}}^2$ is defined to be $P(V_-)$, and its boundary $\partial\mathbb{H}_{\mathbb{C}}^2$ is defined to be $P(V_0)$. This is the *Siegel domain model* of $\mathbb{H}_{\mathbb{C}}^2$.

There is another model of $\mathbb{H}_{\mathbb{C}}^2$.

Definition 2.2 The *ball model* of $\mathbb{H}_{\mathbb{C}}^2$ is the unit ball in \mathbb{C}^2 , which is given by the Hermitian matrix $J = \text{diag}(1, 1, -1)$. In this model, $\partial\mathbb{H}_{\mathbb{C}}^2$ is then the 3-dimensional sphere $S^3 \subset \mathbb{C}^2$. The *Cayley transform* C is given by

$$C = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

It satisfies $C^* H C = J$ and interchanges the Siegel domain model and the ball model of $\mathbb{H}_{\mathbb{C}}^2$.

Let $\mathcal{N} = \mathbb{C} \times \mathbb{R}$ be the Heisenberg group with product

$$[z, t] \cdot [\zeta, v] = [z + \zeta, t + v - 2 \operatorname{Im}(\bar{z}\zeta)].$$

Then, in the Siegel domain model of $\mathbb{H}_{\mathbb{C}}^2$, the boundary of the complex hyperbolic plane $\partial\mathbb{H}_{\mathbb{C}}^2$ can be identified to the union $\mathcal{N} \cup \{q_{\infty}\}$, where q_{∞} is the point at infinity. The *standard lift* of q_{∞} and $q = [z, t] \in \mathcal{N}$ in \mathbb{C}^3 are

$$(2-1) \quad \mathbf{q}_{\infty} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} \frac{1}{2}(-|z|^2 + it) \\ z \\ 1 \end{bmatrix}.$$

The closure of the complex hyperbolic plane $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$ can be identified to the union of $\mathcal{N} \times \mathbb{R}_{\geq 0}$ and $\{q_{\infty}\}$. Any point $q = (z, t, u) \in \mathcal{N} \times \mathbb{R}_{\geq 0}$ has the standard lift

$$\mathbf{q} = \begin{bmatrix} \frac{1}{2}(-|z|^2 - u + it) \\ z \\ 1 \end{bmatrix}.$$

Here (z, t, u) is called the *horospherical coordinates* of $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$. Let $d(u, v)$ be the distance between two points $u, v \in \mathbb{H}_{\mathbb{C}}^2$. Then the *Bergman metric* on the complex hyperbolic plane is given by the distance formula

$$\cosh^2\left(\frac{1}{2}d(u, v)\right) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle},$$

where $\mathbf{u}, \mathbf{v} \in \mathbb{C}^3$ are lifts of u and v .

Definition 2.3 The *Cygan metric* d_{Cyg} on $\partial\mathbb{H}_{\mathbb{C}}^2 - \{q_{\infty}\}$ is defined to be

$$(2-2) \quad d_{\text{Cyg}}(p, q) = |2\langle \mathbf{p}, \mathbf{q} \rangle|^{1/2} = \left| |z - w|^2 - i(t - s + 2 \operatorname{Im}(z\bar{w})) \right|^{1/2},$$

where $p = [z, t]$ and $q = [w, s]$.

The Cygan metric satisfies the properties of a distance. The *extended Cygan metric* on $\mathbb{H}_{\mathbb{C}}^2$ is given by the formula

$$(2-3) \quad d_{\text{Cyg}}(p, q) = \left| |z - w|^2 + |u - v| - i(t - s + 2 \operatorname{Im}(z\bar{w})) \right|^{1/2},$$

where $p = (z, t, u)$ and $q = (w, s, v)$.

The formula $d_{\text{Cyg}}(p, q) = |2\langle \mathbf{p}, \mathbf{q} \rangle|^{1/2}$ remains valid even if one of p or q lies on $\partial\mathbb{H}_{\mathbb{C}}^2$. A *Cygan sphere* is a sphere for the extended Cygan distance.

There are two kinds of 2-dimensional totally real totally geodesic subspaces of $\mathbb{H}_{\mathbb{C}}^2$: complex lines and Lagrangian planes.

Definition 2.4 Let v^\perp be the orthogonal space of $v \in V_+$ with respect to the Hermitian form. The intersection of the projective line $P(v^\perp)$ with $\mathbb{H}_\mathbb{C}^2$ is called a *complex line*. The vector v is its *polar vector*.

The ideal boundary of a complex line on $\partial\mathbb{H}_\mathbb{C}^2$ is called a \mathbb{C} -circle. In the Heisenberg group, \mathbb{C} -circles are either vertical lines or ellipses whose projections on the z -plane are circles.

Let $\mathbb{H}_\mathbb{R}^2 = \{(x_1, x_2) \in \mathbb{H}_\mathbb{C}^2 : x_1, x_2 \in \mathbb{R}\}$ be the set of real points. $\mathbb{H}_\mathbb{R}^2$ is a Lagrangian plane. All the Lagrangian planes are the images of $\mathbb{H}_\mathbb{R}^2$ by isometries of $\mathbb{H}_\mathbb{C}^2$. The ideal boundary of a Lagrangian plane is called an \mathbb{R} -circle. In the Heisenberg group, \mathbb{R} -circles are either straight lines or lemniscate curves whose projections on the z -plane are the figure eight.

2.2 Isometries

Let $SU(2, 1)$ be the special unitary matrix preserving the Hermitian form. Then the projective unitary group $PU(2, 1) = SU(2, 1)/\{I, \omega I, \omega^2 I\}$ is the holomorphic isometry group of $\mathbb{H}_\mathbb{C}^2$, where $\omega = \frac{1}{2}(-1 + i\sqrt{3})$ is a primitive cubic root of unity. Note that complex conjugation also preserves the Bergman distance, and the full isometry group of $\mathbb{H}_\mathbb{C}^2$ is generated by $PU(2, 1)$ and complex conjugation; see Section 3.4 of [20].

Definition 2.5 Any isometry $g \in PU(2, 1)$ is *loxodromic* if it has exactly two fixed points on $\partial\mathbb{H}_\mathbb{C}^2$, *parabolic* if it has exactly one fixed point on $\partial\mathbb{H}_\mathbb{C}^2$, and *elliptic* if it has at least one fixed point in $\mathbb{H}_\mathbb{C}^2$.

The types of isometries can be determined by the traces of their matrix realizations; see Theorem 6.2.4 of Goldman [12]. Now suppose that $A \in SU(2, 1)$ has real trace. Then A is elliptic if $-1 \leq \text{tr}(A) < 3$. Moreover, A is unipotent if A is not the identity and $\text{tr}(A) = 3$. In particular, if $\text{tr}(A) = -1, 0, 1$, A is elliptic of order 2, 3 or 4, respectively.

There is a special class of elliptic elements of order two:

Definition 2.6 The *complex involution* on complex line C with polar vector n is

$$(2-4) \quad I_C(z) = -z + \frac{2\langle z, n \rangle}{\langle n, n \rangle} n.$$

It is obvious that I_C is a holomorphic isometry fixing the complex line C .

There is a special class of unipotent elements in $PU(2, 1)$:

Definition 2.7 A left Heisenberg translation associated to $[z, t] \in \mathcal{N}$ is given by

$$(2-5) \quad T_{[z,t]} = \begin{bmatrix} 1 & -\bar{z} & \frac{1}{2}(-|z|^2 + it) \\ 0 & 1 & z \\ 0 & 0 & 1 \end{bmatrix}.$$

It is obvious that $T_{[z,t]}$ fixes q_∞ and maps $[0, 0] \in \mathcal{N}$ to $[z, t]$.

2.3 Isometric spheres and Ford polyhedron

Suppose that $g = (g_{ij})_{i,j=1}^3 \in PU(2, 1)$ does not fix q_∞ . Then it is obvious that $g_{31} \neq 0$. We first recall the definition of isometric spheres and relevant properties; see for instance [20].

Definition 2.8 The isometric sphere of g , denoted by $\mathcal{I}(g)$, is the set

$$(2-6) \quad \mathcal{I}(g) = \{p \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : |\langle p, q_\infty \rangle| = |\langle p, g^{-1}(q_\infty) \rangle|\}.$$

The isometric sphere $\mathcal{I}(g)$ is the Cygan sphere with center

$$g^{-1}(q_\infty) = [\bar{g}_{32}/\bar{g}_{31}, 2 \operatorname{Im}(\bar{g}_{33}/\bar{g}_{31})]$$

and radius $r_g = \sqrt{2/|g_{31}|}$.

The interior of $\mathcal{I}(g)$ is the set

$$(2-7) \quad \{p \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : |\langle p, q_\infty \rangle| > |\langle p, g^{-1}(q_\infty) \rangle|\}.$$

The exterior of $\mathcal{I}(g)$ is the set

$$(2-8) \quad \{p \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : |\langle p, q_\infty \rangle| < |\langle p, g^{-1}(q_\infty) \rangle|\}.$$

The isometric spheres are paired as follows:

Lemma 2.9 [12, Section 5.4.5] Let g be an element in $PU(2, 1)$ which does not fix q_∞ . Then g maps $\mathcal{I}(g)$ to $\mathcal{I}(g^{-1})$ and the exterior of $\mathcal{I}(g)$ to the interior of $\mathcal{I}(g^{-1})$. Also, for any unipotent transformation $h \in PU(2, 1)$ fixing q_∞ , we have $\mathcal{I}(g) = \mathcal{I}(hg)$.

Since isometric spheres are Cygan spheres, we now recall some facts about Cygan spheres. Let $S_{[0,0]}(r)$ be the Cygan sphere with center $[0, 0]$ and radius $r > 0$. Then

$$(2-9) \quad S_{[0,0]}(r) = \{(z, t, u) \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : (|z|^2 + u)^2 + t^2 = r^4\}.$$

The geographical coordinates on the Cygan sphere will play an important role in our calculation; see Section 2.5 of [22].

Definition 2.10 The point $q = q(\alpha, \beta, w) \in S_{[0,0]}(r)$ with geographical coordinates (α, β, w) is the point whose lift to \mathbb{C}^3 is

$$(2-10) \quad \mathbf{q} = \mathbf{q}(\alpha, \beta, w) = \begin{bmatrix} -\frac{1}{2}r^2 e^{-i\alpha} \\ rwe^{i(-\alpha/2+\beta)} \\ 1 \end{bmatrix},$$

where $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\beta \in [0, \pi)$ and $w \in [-\sqrt{\cos(\alpha)}, \sqrt{\cos(\alpha)}]$. The ideal boundary of $S_{[0,0]}(r)$ on $\partial\mathbb{H}_{\mathbb{C}}^2$ consists of the points with $w = \pm\sqrt{\cos(\alpha)}$.

We are interested in the intersection of Cygan spheres.

Proposition 2.11 [22, Proposition 2.10] *The intersection of two Cygan spheres is connected.*

Remark 2.12 This intersection is often called a *Giraud disk*.

The following property should be useful to describe the intersection of Cygan spheres; see Proposition 2.12 of [22] or Example 5.1.8 of [12].

Proposition 2.13 *Let $S_{[0,0]}(r)$ be a Cygan sphere with geographical coordinates (α, β, w) .*

- (1) *The level sets of α are complex lines, called slices of $S_{[0,0]}(r)$.*
- (2) *The level sets of β are Lagrangian planes, called meridians of $S_{[0,0]}(r)$.*
- (3) *The set of points with $w = 0$ is the spine of $S_{[0,0]}(r)$. It is a geodesic contained in every meridian.*

A central part of this paper is constructing a polyhedron for a finitely generated subgroup of $\text{PU}(2, 1)$.

Definition 2.14 Let G be a discrete subgroup of $\text{PU}(2, 1)$. The *Ford polyhedron* D_G for G is the set

$$D_G = \{p \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : |\langle p, q_{\infty} \rangle| \leq |\langle p, g^{-1}q_{\infty} \rangle| \text{ for all } g \in G \text{ with } g(q_{\infty}) \neq q_{\infty}\}.$$

That is to say D_G is the intersection of closures of the exteriors of all the isometric spheres for elements of G which do not fix q_{∞} . In fact, the Ford polyhedron is the limit of Dirichlet polyhedra as the center point goes to q_{∞} .

3 The parameter space of complex hyperbolic $(4, 4, \infty)$ triangle groups

In this section, we give a parameter space of the complex hyperbolic $(4, 4, \infty)$ triangle groups.

Let $\theta \in [0, \frac{\pi}{2})$. Let I_1, I_2 and I_3 be the complex involutions on the complex lines C_1, C_2 and C_3 in complex hyperbolic space $\mathbb{H}_{\mathbb{C}}^2$ with polar vectors $\mathbf{n}_1, \mathbf{n}_2$ and \mathbf{n}_3 , respectively. By conjugating elements in $\text{PU}(2, 1)$, one can normalize so that $\partial C_3 = \{[z, 0] \in \mathcal{N} : |z| = \sqrt{2}\}$, $\partial C_1 = \{[z_1, t] \in \mathcal{N} : t \in \mathbb{R}\}$ and $\partial C_2 = \{[z_2, t] \in \mathcal{N} : t \in \mathbb{R}\}$. That is, ∂C_3 is the circle in the z -plane of the Heisenberg group with center the origin and radius $\sqrt{2}$, and ∂C_1 (resp. ∂C_2) is the vertical line whose projection on the z -plane of the Heisenberg group is the point z_1 (resp. z_2). Thus the polar vectors of the complex lines can be written as

$$\mathbf{n}_1 = \begin{bmatrix} z_1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{n}_2 = \begin{bmatrix} z_2 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{n}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Since $\text{tr}(I_1 I_3) = \text{tr}(I_2 I_3) = 1$, we have $|z_1| = |z_2| = 1$. Then, up to rotation about the t -axis of the Heisenberg group, the \mathbb{C} -circles ∂C_1 and ∂C_2 can be normalized to be the sets $\partial C_1 = \{[-e^{-i\theta}, t] \in \mathcal{N} : t \in \mathbb{R}\}$ and $\partial C_2 = \{[e^{i\theta}, t] \in \mathcal{N} : t \in \mathbb{R}\}$.

Note that the \mathbb{C} -circles ∂C_1 and ∂C_2 coincide with each other if $\theta = \frac{\pi}{2}$. According to (2-4), the complex involutions I_1, I_2 and I_3 on the complex lines are given as

$$I_1 = \begin{bmatrix} -1 & 2e^{i\theta} & 2 \\ 0 & 1 & 2e^{-i\theta} \\ 0 & 0 & -1 \end{bmatrix}, \quad I_2 = \begin{bmatrix} -1 & -2e^{-i\theta} & 2 \\ 0 & 1 & -2e^{i\theta} \\ 0 & 0 & -1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Proposition 3.1 *Let $\theta \in [0, \frac{\pi}{2})$, and I_1, I_2 and I_3 be defined as above. Then $\langle I_1, I_2, I_3 \rangle$ is a complex hyperbolic $(4, 4, \infty)$ triangle group. Furthermore, the element $I_1 I_3 I_2 I_3$ is nonelliptic if and only if $0 \leq \theta \leq \frac{\pi}{3}$.*

Proof By computing the products of two involutions

$$I_2 I_3 = \begin{bmatrix} 2 & 2e^{-i\theta} & -1 \\ -2e^{i\theta} & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad I_3 I_1 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & -2e^{-i\theta} \\ -1 & e^{i\theta} & 2 \end{bmatrix},$$

and

$$I_2 I_1 = \begin{bmatrix} 1 & -4 \cos(\theta) & -4(1 + e^{-2i\theta}) \\ 0 & 1 & 4 \cos(\theta) \\ 0 & 0 & 1 \end{bmatrix}.$$

It is easy to verify that $I_2 I_3$ and $I_3 I_1$ are elliptic of order 4 and $I_2 I_1$ is unipotent. Thus $\langle I_1, I_2, I_3 \rangle$ is a complex hyperbolic $(4, 4, \infty)$ triangle group.

Since the trace of $I_1 I_3 I_2 I_3$ is $\text{tr}(I_1 I_3 I_2 I_3) = 7 + 8 \cos(2\theta)$, the element $I_1 I_3 I_2 I_3$ is elliptic if and only if

$$-1 \leq \text{tr}(I_1 I_3 I_2 I_3) = 7 + 8 \cos(2\theta) < 32,$$

that is, $\frac{\pi}{3} < \theta \leq \frac{\pi}{2}$. Thus $I_1 I_3 I_2 I_3$ is nonelliptic if and only if $0 \leq \theta \leq \frac{\pi}{3}$. Moreover, when $\theta = \frac{\pi}{3}$, the element $I_1 I_3 I_2 I_3$ is parabolic. □

If $\theta = 0$, all entries of I_1, I_2 and I_3 are in the ring of integers \mathbb{Z} , and if $\theta = \frac{\pi}{3}$ all entries of I_1, I_2 and I_3 are in the ring of Eisenstein integers $\mathbb{Z}[-1 + i\frac{\sqrt{3}}{2}]$. In both cases, the group $\langle I_1, I_2, I_3 \rangle$ is arithmetic. Thus, we have the following proposition:

Proposition 3.2 (1) *If $\theta = 0$, then the group $\langle I_1, I_2, I_3 \rangle$ is discrete and preserves a Lagrangian plane.*

(2) *If $\theta = \frac{\pi}{3}$, then the group $\langle I_1, I_2, I_3 \rangle$ is discrete.*

4 The Ford domain

For $\theta \in [0, \frac{\pi}{3}]$, let $S = I_2 I_3$ and $T = I_2 I_1$. Then $\Gamma = \langle S, T \rangle$ is a subgroup of $\langle I_1, I_2, I_3 \rangle$ of index two. In this section, we will mainly prove that Γ is discrete. Our method is to construct a candidate Ford domain D (see Definition 4.12), then apply the Poincaré polyhedron theorem to show that D is a fundamental domain for the cosets of $\langle T \rangle$ in Γ , and also that Γ is discrete.

4.1 The isometric spheres

Definition 4.1 For $k \in \mathbb{Z}$, let

- \mathcal{I}_k^+ be the isometric sphere $\mathcal{I}(T^k S T^{-k}) = T^k \mathcal{I}(S)$ and c_k^+ be its center,
- \mathcal{I}_k^- be the isometric sphere $\mathcal{I}(T^k S^{-1} T^{-k}) = T^k \mathcal{I}(S^{-1})$ and c_k^- be its center,
- \mathcal{I}_k^* be the isometric sphere $\mathcal{I}(T^k S^2 T^{-k}) = T^k \mathcal{I}(S^2)$ and c_k^* be its center,

- \mathcal{I}_k^\diamond be the isometric sphere $\mathcal{I}(T^k(S^{-1}T)^2T^{-k}) = T^k\mathcal{I}((S^{-1}T)^2)$ and c_k^\diamond be its center.

Note that S and $S^{-1}T$ both have order 4, so $S^2 = S^{-2}$ and $(S^{-1}T)^2 = (S^{-1}T)^{-2}$. The centers and radii of the isometric spheres \mathcal{I}_k^+ , \mathcal{I}_k^- , \mathcal{I}_k^\star and \mathcal{I}_k^\diamond are listed in the following table:

isometric sphere	center	radius
\mathcal{I}_k^+	$c_k^+ = [4k \cos(\theta), 8k \sin(2\theta)]$	$\sqrt{2}$
\mathcal{I}_k^-	$c_k^- = [4k \cos(\theta) + 2e^{i\theta}, 0]$	$\sqrt{2}$
\mathcal{I}_k^\star	$c_k^\star = [4k \cos(\theta) + e^{i\theta}, 4k \sin(2\theta)]$	1
\mathcal{I}_k^\diamond	$c_k^\diamond = [4k \cos(\theta) - e^{-i\theta}, 4k \sin(2\theta)]$	1

Proposition 4.2 Let $k \in \mathbb{Z}$.

- (1) There is an antiholomorphic involution τ such that $\tau(\mathcal{I}_k^+) = \mathcal{I}_{-k}^+$, $\tau(\mathcal{I}_k^-) = \mathcal{I}_{-k-1}^-$ and $\tau(\mathcal{I}_k^\star) = \mathcal{I}_{-k}^\diamond$.
- (2) The complex involution I_2 interchanges \mathcal{I}_k^\star and \mathcal{I}_{-k}^\star , interchanges \mathcal{I}_k^+ and \mathcal{I}_{-k}^- , and interchanges \mathcal{I}_k^\diamond and $\mathcal{I}_{-k+1}^\diamond$.

Proof (1) Let $\tau: \mathbb{C}^3 \rightarrow \mathbb{C}^3$ be given by

$$\tau: \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \mapsto \begin{bmatrix} \bar{z}_1 \\ -\bar{z}_2 \\ \bar{z}_3 \end{bmatrix}.$$

Then τ^2 is the identity. It is easy to see that τ fixes the polar vector \mathbf{n}_3 , and interchanges the polar vectors \mathbf{n}_1 and \mathbf{n}_2 . Thus τ conjugates I_3 to itself, I_1 to I_2 and vice versa. So τ conjugates T to T^{-1} , S to $T^{-1}S$, S^{-1} to $S^{-1}T$ and S^2 to $(T^{-1}S)^2 = (S^{-1}T)^2$. This implies that $\tau(\mathcal{I}_k^+) = \mathcal{I}_{-k}^+$, $\tau(\mathcal{I}_k^-) = \mathcal{I}_{-k-1}^-$ and $\tau(\mathcal{I}_k^\star) = \mathcal{I}_{-k}^\diamond$.

(2) The statement is easily obtained by the facts $I_2SI_2 = S^{-1}$ and $I_2TI_2 = T^{-1}$. \square

Before we consider the intersections of two isometric spheres, we would like to give a useful technical lemma. Suppose that $q \in \mathcal{I}_0^+$. Then by (2-10) the lift of $q = q(\alpha, \beta, w)$ is given by

$$(4-1) \quad \mathbf{q} = \mathbf{q}(\alpha, \beta, w) = \begin{bmatrix} -e^{-i\alpha} \\ \sqrt{2}we^{i(-\alpha/2+\beta)} \\ 1 \end{bmatrix},$$

where $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\beta \in [0, \pi)$ and $w \in [-\sqrt{\cos(\alpha)}, \sqrt{\cos(\alpha)}]$.

Definition 4.3 Let (α, β, w) be the geographical coordinates of \mathcal{I}_0^+ . We define

$$\begin{aligned} f_0^*(\alpha, \beta, w) &= 2w^2 + 1 + \cos(\alpha) - \sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta - \theta) - 2\sqrt{2}w \cos(\frac{1}{2}\alpha + \beta - \theta), \\ f_0^-(\alpha, \beta, w) &= 2w^2 + 1 + \cos(\alpha) - \sqrt{2}w \cos(\frac{1}{2}\alpha + \beta - \theta) - 2\sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta - \theta), \\ f_{-1}^-(\alpha, \beta, w) &= 2w^2 + 1 + \cos(\alpha) + \sqrt{2}w \cos(\frac{1}{2}\alpha + \beta + \theta) + 2\sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta + \theta). \end{aligned}$$

Lemma 4.4 Suppose that $\theta \in [0, \frac{\pi}{3}]$. Let $f_0^*(\alpha, \beta, w)$, $f_0^-(\alpha, \beta, w)$ and $f_{-1}^-(\alpha, \beta, w)$ be the functions in Definition 4.3. Suppose that $q \in \mathcal{I}_0^+$. Then

- (1) q lies on \mathcal{I}_0^* (resp. in its interior or exterior) if and only if $f_0^*(\alpha, \beta, w) = 0$ (resp. is negative or positive),
- (2) q lies on \mathcal{I}_0^- (resp. in its interior or exterior) if and only if $f_0^-(\alpha, \beta, w) = 0$ (resp. is negative or positive),
- (3) q lies on \mathcal{I}_{-1}^- (resp. in its interior or exterior) if and only if $f_{-1}^-(\alpha, \beta, w) = 0$ (resp. is negative or positive).

Proof (1) Any point $q \in \mathcal{I}_0^+$ lies on \mathcal{I}_0^* (resp. in its interior or exterior) if and only if the Cygan distance between q and the center of \mathcal{I}_0^* is 1 (resp. less than 1 or greater than 1). Using (4-1), the difference between the Cygan distance from q to the center of \mathcal{I}_0^* and 1 is

$$\begin{aligned} d_{\text{Cyg}}(q, c_0^*)^4 - 1 &= 4|-e^{-i\alpha} + \sqrt{2}we^{i(-\alpha/2+\beta-\theta)} - \frac{1}{2}|^2 - 1 \\ &= 4(2w^2 + 1 + \cos(\alpha) - \sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta - \theta) - 2\sqrt{2}w \cos(\frac{1}{2}\alpha + \beta - \theta)) \\ &= 4f_0^*(\alpha, \beta, w). \end{aligned}$$

Hence, q lies on \mathcal{I}_0^* (resp. in its interior or exterior) if and only if $f_0^*(\alpha, \beta, w) = 0$ (resp. negative or positive).

The rest of the proof runs as before. □

4.2 The intersection of isometric spheres

Proposition 4.5 Suppose that $\theta \in [0, \frac{\pi}{3}]$. Then each pair of isometric spheres in $\{\mathcal{I}_k^+ : k \in \mathbb{Z}\}$ is disjoint in $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$.

Proof It suffices to show that \mathcal{I}_0^+ and \mathcal{I}_k^+ are disjoint for $|k| \geq 1$. Observe that T is a Heisenberg translation associated with $[4 \cos(\theta), 8 \sin(2\theta)]$. According to the Cygan metric given in (2-2), the Cygan distance between the centers of \mathcal{I}_0^+ and \mathcal{I}_k^+ is

$$4\sqrt{|k| \cos(\theta)} \cdot |k \cos(\theta) - i \sin(\theta)|^{1/2} \geq 4\sqrt{\cos(\theta)} \geq 2\sqrt{2}.$$

Thus the Cygan distance between the centers of \mathcal{I}_0^+ and \mathcal{I}_k^+ is bigger than the sum of the radii, except when $k = \pm 1$ and $\theta = \frac{\pi}{3}$. This implies that \mathcal{I}_0^+ and \mathcal{I}_k^+ are disjoint for all $|k| \geq 2$.

When $k = \pm 1$ and $\theta = \frac{\pi}{3}$, although the Cygan distance between the centers of \mathcal{I}_0^+ and $\mathcal{I}_{\pm 1}^+$ is the sum of the radii, we claim that they are still disjoint. Using the symmetry τ in Proposition 4.2, we only need to show that $\mathcal{I}_0^+ \cap \mathcal{I}_1^+ = \emptyset$. Suppose that $q \in \mathcal{I}_0^+$. Using the geographical coordinates of $q = q(\alpha, \beta, w)$ given in (4-1), we can compute the difference between the Cygan distance of q and the center of \mathcal{I}_1^+ with its radius. That is,

$$\begin{aligned} d_{\text{Cyg}}(q, [2, 4\sqrt{3}])^4 - 4 &= 4|-e^{-i\alpha} + \sqrt{2}we^{i(-\alpha/2+\beta-\theta)} - 4e^{i\pi/3}|^2 - 4 \\ &= 32\left(w^2 + \frac{\sqrt{2}}{2}w(\cos(\frac{1}{2}\alpha + \beta) + 2\cos(\frac{1}{2}\alpha - \beta + \frac{\pi}{3})) + \cos(\alpha + \frac{\pi}{3}) + 2\right) \\ &= 32f(\alpha, \beta, w). \end{aligned}$$

Here $f(\alpha, \beta, w)$ can be seen as a quadratic function of w . Let

$$B = \frac{\sqrt{2}}{2}(\cos(\frac{1}{2}\alpha + \beta) + 2\cos(\frac{1}{2}\alpha - \beta + \frac{\pi}{3}))$$

and $C = \cos(\alpha + \frac{\pi}{3}) + 2$. If $B^2 - 4C < 0$, then it is obvious that $f(\alpha, \beta, w) > 0$. If $B^2 - 4C \geq 0$, then $B \leq -2\sqrt{C}$ (which is impossible by numerical computation) or $B \geq 2\sqrt{C}$. In this case we have $B - 2\sqrt{\cos(\alpha)} \geq B - 2\sqrt{C} \geq 0$ since $\cos(\alpha) \leq C$. This means that the symmetry axes of f lie on the left side of $w = -\sqrt{\cos(\alpha)}$. Besides, one can compute numerically that $f(\alpha, \beta, -\sqrt{\cos(\alpha)}) > 0$ on the range of α and β . So, we have $f(\alpha, \beta, w) > 0$. This means that every point on \mathcal{I}_0^+ lies outside of \mathcal{I}_1^+ . Hence $\mathcal{I}_0^+ \cap \mathcal{I}_1^+ = \emptyset$. □

By a similar argument, we have the following propositions:

Proposition 4.6 Suppose that $\theta \in [0, \frac{\pi}{3}]$. Then

- (1) \mathcal{I}_0^+ and \mathcal{I}_k^- are disjoint in $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$, except possibly when $k = -1, 0$,
- (2) \mathcal{I}_0^+ and \mathcal{I}_k^* are disjoint in $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$, except possibly when $k = -1, 0$,
- (3) \mathcal{I}_0^+ and \mathcal{I}_k^\diamond are disjoint in $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$, except possibly when $k = 0, 1$.

Proposition 4.7 Suppose that $\theta \in [0, \frac{\pi}{3}]$. Then:

- (1) \mathcal{I}_0^\star and \mathcal{I}_k^\star are disjoint in $\mathbb{H}_\mathbb{C}^2$. Furthermore, when $\theta = \frac{\pi}{3}$ the closures of \mathcal{I}_0^\star and \mathcal{I}_{-1}^\star (respectively, \mathcal{I}_1^\star) are tangent at the parabolic fixed point of $T^{-1}S^2$ (respectively, $T(T^{-1}S^2)T^{-1}$) on $\partial\mathbb{H}_\mathbb{C}^2$.
- (2) \mathcal{I}_0^\star and \mathcal{I}_k^\diamond are disjoint in $\mathbb{H}_\mathbb{C}^2 \cup \partial\mathbb{H}_\mathbb{C}^2$, except possibly when $k = 0, 1$.

Lemma 4.8 Suppose that $\theta \in [0, \frac{\pi}{3}]$. Then $\mathcal{I}_0^+ \cap \mathcal{I}_0^\star \cap \mathcal{I}_{-1}^- = \emptyset$ except when $\theta = \frac{\pi}{3}$, in which case the triple intersection is the point $[e^{i2\pi/3}, -\sqrt{3}] \in \partial\mathbb{H}_\mathbb{C}^2$. Moreover, this point is the parabolic fixed point of $T^{-1}S^2$.

Proof Suppose that $q \in \mathcal{I}_0^+$. Using Lemma 4.4, the geographical coordinates (α, β, w) of $q \in \mathcal{I}_0^+ \cap \mathcal{I}_0^\star \cap \mathcal{I}_{-1}^-$ should satisfy

$$(4-2) \quad 2w^2 + 1 + \cos(\alpha) - \sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta - \theta) - 2\sqrt{2}w \cos(\frac{1}{2}\alpha + \beta - \theta) = 0,$$

$$(4-3) \quad 2w^2 + 1 + \cos(\alpha) + \sqrt{2}w \cos(\frac{1}{2}\alpha + \beta + \theta) + 2\sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta + \theta) = 0.$$

Subtracting (4-2) and (4-3),

$$2\sqrt{2}w \cos(\beta)(\cos(\frac{1}{2}\alpha + \theta) + 2\cos(\frac{1}{2}\alpha - \theta)) = 0.$$

This implies that either $w = 0$ or $\beta = \frac{\pi}{2}$, since $(\cos(\frac{1}{2}\alpha + \theta) + 2\cos(\frac{1}{2}\alpha - \theta)) \neq 0$ for $\theta \in [0, \frac{\pi}{3}]$. We know that the points with $w = 0$ lie in the meridian with $\beta = \frac{\pi}{2}$. Therefore, a necessary condition for $q \in \mathcal{I}_0^+ \cap \mathcal{I}_0^\star \cap \mathcal{I}_{-1}^-$ is that $\beta = \frac{\pi}{2}$.

Substituting $\beta = \frac{\pi}{2}$ into (4-2) and simplifying,

$$(4-4) \quad 2w^2 + 2\cos^2(\frac{1}{2}\alpha) + \sqrt{2}w(\sin(\frac{1}{2}\alpha)\cos(\theta) - 3\cos(\frac{1}{2}\alpha)\sin(\theta)) = 0.$$

Let $b(\alpha, \theta) = \sin(\frac{1}{2}\alpha)\cos(\theta) - 3\cos(\frac{1}{2}\alpha)\sin(\theta)$. It is easy to see that, for every α , the function $\theta \mapsto b(\alpha, \theta)$ is decreasing on $[0, \frac{\pi}{3}]$.

The left-hand side of (4-4) can be seen as a quadratic function of w with positive leading coefficient. Thus (4-4) has at least one solution only if $b^2 - 8\cos^2(\frac{1}{2}\alpha) \geq 0$, that is $b \geq 2\sqrt{2}\cos(\frac{1}{2}\alpha)$, which is impossible since $b \leq b(\alpha, 0) = \sin(\frac{1}{2}\alpha)$, or $b \leq -2\sqrt{2}\cos(\frac{1}{2}\alpha)$. Since $\sqrt{\cos(\alpha)} \leq \cos(\frac{1}{2}\alpha)$, we have $b + 2\sqrt{2}\sqrt{\cos(\alpha)} \leq b + 2\sqrt{2}\cos(\frac{1}{2}\alpha) \leq 0$. This means that the symmetry axes of the quadratic function lie on the right-hand side of $w = \sqrt{\cos(\alpha)}$.

Besides, one can compute that

$$\begin{aligned} & b\sqrt{\cos(\alpha)} + \sqrt{2}(\cos(\alpha) + \cos^2(\frac{1}{2}\alpha)) \\ & \geq (\sin(\frac{1}{2}\alpha)\cos(\frac{\pi}{3}) - 3\cos(\frac{1}{2}\alpha)\sin(\frac{\pi}{3}))\sqrt{\cos(\alpha)} + \sqrt{2}(\cos(\alpha) + \cos^2(\frac{1}{2}\alpha)) \\ & = \frac{\sqrt{2}}{2}(\frac{1}{2}\sqrt{\cos(\alpha)} + \frac{\sqrt{2}}{2}\sin(\frac{1}{2}\alpha))^2 + \frac{3\sqrt{2}}{2}(\frac{\sqrt{3}}{2}\sqrt{\cos(\alpha)} - \frac{\sqrt{2}}{2}\cos(\frac{1}{2}\alpha))^2. \end{aligned}$$

Then $b\sqrt{\cos(\alpha)} + \sqrt{2}(\cos(\alpha) + \cos^2(\frac{1}{2}\alpha)) \geq 0$. If it is 0, then $\alpha = -\frac{\pi}{3}$ and $\theta = \frac{\pi}{3}$. This means that for $w \in [-\sqrt{\cos(\alpha)}, \sqrt{\cos(\alpha)}]$ equation (4-4) has no solution except when $\theta = \frac{\pi}{3}$ and $\alpha = -\frac{\pi}{3}$, in which case $w = \sqrt{\cos(\alpha)} = \frac{\sqrt{2}}{2}$.

Hence $q \in \mathcal{I}_0^+ \cap \mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ if and only if $\theta = \frac{\pi}{3}$, $\alpha = -\frac{\pi}{3}$ and $w = \sqrt{\cos(\alpha)} = \frac{\sqrt{2}}{2}$. When $\theta = \frac{\pi}{3}$, $T^{-1}S^2$ is unipotent and its fixed point is the eigenvector with eigenvalue 1. One can compute that its fixed point is $[e^{i2\pi/3}, -\sqrt{3}] \in \partial\mathbb{H}_{\mathbb{C}}^2$, which equals the point $q(-\frac{\pi}{3}, \frac{\pi}{2}, \frac{\sqrt{2}}{2})$. □

Proposition 4.9 *Suppose that $\theta \in [0, \frac{\pi}{3}]$. Then:*

- (1) *The intersection $\mathcal{I}_0^* \cap \mathcal{I}_0^\diamond$ lies in the interior of \mathcal{I}_0^+ .*
- (2) *The intersection $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ either is empty or lies in the interior of \mathcal{I}_0^+ . Further, when $\theta = \frac{\pi}{3}$, there is a unique point on the ideal boundary of $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ on $\partial\mathbb{H}_{\mathbb{C}}^2$, which is fixed by $T^{-1}S^2$, lying on the ideal boundary of \mathcal{I}_0^+ .*

Proof (1) Let $p = (z, t, u) \in \mathcal{I}_0^* \cap \mathcal{I}_0^\diamond$, then p satisfies the equations

$$||z - e^{i\theta}|^2 + u - i(t + 2 \operatorname{Im}(ze^{-i\theta}))| = 1, \quad ||z + e^{-i\theta}|^2 + u - i(t + 2 \operatorname{Im}(-ze^{i\theta}))| = 1.$$

Set $z = |z|e^{i\phi}$. By simplifying, we have

$$\begin{aligned} &||z|^2 + 1 + u - 2|z| \cos(\phi - \theta) - i(t + 2|z| \sin(\phi - \theta))| = 1, \\ &||z|^2 + 1 + u + 2|z| \cos(\phi + \theta) - i(t - 2|z| \sin(\phi + \theta))| = 1. \end{aligned}$$

Now set

$$(4-5) \quad |z|^2 + 1 + u - 2|z| \cos(\phi - \theta) - i(t + 2|z| \sin(\phi - \theta)) = e^{i\alpha},$$

$$(4-6) \quad |z|^2 + 1 + u + 2|z| \cos(\phi + \theta) - i(t - 2|z| \sin(\phi + \theta)) = e^{i\beta}.$$

Since

$$\cos \alpha = |z|^2 + 1 + u - 2|z| \cos(\phi - \theta) = (|z| - \cos(\phi - \theta))^2 + \sin^2(\phi - \theta) + u \geq 0$$

and

$$\cos \beta = |z|^2 + 1 + u + 2|z| \cos(\phi + \theta) = (|z| + \cos(\phi + \theta))^2 + \sin^2(\phi + \theta) + u \geq 0,$$

we have $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$ and $-\frac{\pi}{2} \leq \beta \leq \frac{\pi}{2}$. Thus $\cos(\frac{1}{2}\beta - \frac{1}{2}\alpha) \geq 0$. By computing the difference of (4-5) and (4-6), we have

$$(4-7) \quad z = \frac{e^{i\beta} - e^{i\alpha}}{4 \cos(\theta)} = \pm \frac{\sin(\frac{1}{2}\beta - \frac{1}{2}\alpha)}{2 \cos(\theta)} e^{i(\pm\pi/2 + \beta/2 + \alpha/2)}.$$

Thus $\phi = \pm \frac{\pi}{2} + \frac{1}{2}\beta + \frac{1}{2}\alpha$. Therefore,

$$\begin{aligned} & (|z|^2 + u)^2 + t^2 \\ &= (\cos(\alpha) - 1 + 2|z| \cos(\phi - \theta))^2 + (\sin(\alpha) + 2|z| \sin(\phi - \theta))^2 \\ &= 2 + 4|z|^2 - 2 \cos(\alpha) - 4|z| \cos(\phi - \theta) + 4|z| \cos(\phi - \theta - \alpha) \\ &\leq 2 + 4|z|^2 - 2(|z|^2 + 1 - 2|z| \cos(\phi - \theta)) - 4|z| \cos(\phi - \theta) + 4|z| \cos(\phi - \theta - \alpha) \\ &= 2|z|^2 + 4|z| \cos(\phi - \theta - \alpha) = 2|z|^2 + 4|z| \cos\left(\pm \frac{\pi}{2} - \theta + \frac{1}{2}\beta - \frac{1}{2}\alpha\right) \\ &= 2|z|^2 + 4|z|(\pm \sin(\theta) \cos\left(\frac{1}{2}\beta - \frac{1}{2}\alpha\right) \mp \cos(\theta) \sin\left(\frac{1}{2}\beta - \frac{1}{2}\alpha\right)) \\ &= \frac{\sin^2\left(\frac{1}{2}\beta - \frac{1}{2}\alpha\right)}{2 \cos^2(\theta)} + \tan(\theta) \sin(\beta - \alpha) - 2 \sin^2\left(\frac{1}{2}\beta - \frac{1}{2}\alpha\right). \end{aligned}$$

Since $\theta \in [0, \frac{\pi}{3}]$, we have that $\sin^2(\frac{1}{2}\beta - \frac{1}{2}\alpha)/(2 \cos^2(\theta)) \leq 2 \sin^2(\frac{1}{2}\beta - \frac{1}{2}\alpha)$ and $\tan(\theta) \sin(\beta - \alpha) \leq \sqrt{3}$. This implies that $(|z|^2 + u)^2 + t^2 < 4$, so the intersection $\mathcal{I}_0^* \cap \mathcal{I}_0^\diamond$ lies in the interior of \mathcal{I}_0^+ .

(2) Suppose that $p = (z, t, u) \in \mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$. Then p satisfies the equations

$$\begin{aligned} 1 &= \left| |z - e^{i\theta}|^2 + u - i(t + 2 \operatorname{Im}(ze^{-i\theta})) \right| = \left| |z|^2 + u + 1 - it - 2ze^{-i\theta} \right|, \\ 2 &= \left| |z + 2e^{-i\theta}|^2 + u - i(t + 2 \operatorname{Im}(-2ze^{i\theta})) \right| = \left| |z|^2 + u + 4 - it + 4ze^{i\theta} \right|. \end{aligned}$$

Now set

$$(4-8) \quad |z|^2 + u + 1 - it - 2ze^{-i\theta} = e^{i\beta},$$

$$(4-9) \quad |z|^2 + u + 4 - it + 4ze^{i\theta} = 2e^{i\alpha}.$$

By computing the difference of (4-8) and (4-9), we have

$$(4-10) \quad z = \frac{2e^{i\alpha} - e^{i\beta} - 3}{4e^{i\theta} + 2e^{-i\theta}}.$$

According to (4-8),

$$(4-11) \quad u = \cos(\beta) - |ze^{-i\theta} - 1|^2,$$

$$(4-12) \quad t = -\sin(\beta) - 2 \operatorname{Im}(ze^{-i\theta}).$$

Since

$$\cos \beta = u + |ze^{-i\theta} - 1|^2 \geq 0 \quad \text{and} \quad 2 \cos \alpha = u + |ze^{i\theta} + 2|^2 \geq 0,$$

we have $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$ and $-\frac{\pi}{2} \leq \beta \leq \frac{\pi}{2}$.

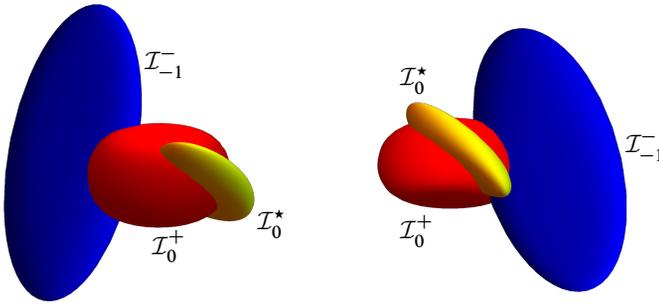


Figure 1: The ideal boundaries of the three spheres \mathcal{I}_0^+ , \mathcal{I}_{-1}^- and \mathcal{I}_0^* on $\partial\mathbb{H}_{\mathbb{C}}^2$. (On the left is the case when $\theta = 0$ and on the right is $\theta = \frac{\pi}{3}$.)

Now let us consider the case when $\theta = \frac{\pi}{3}$. Substituting $\alpha = \frac{\pi}{6}$ and $\beta = 0$ into (4-10), (4-11) and (4-12), we obtain the point

$$p_0 = \left(\frac{1}{3}(-3 + \sqrt{3}) + i\frac{1}{6}(3 + \sqrt{3}), \frac{1}{6}(3 - 7\sqrt{3}), \frac{1}{6}(-13 + 8\sqrt{3})\right) \in \mathcal{I}_0^* \cap \mathcal{I}_{-1}^-.$$

One can compute that p_0 lies in the interior of \mathcal{I}_0^+ , since

$$||z|^2 + u - it|^2 = |e^{i\beta} - 1 + 2ze^{-i\theta}|^2 = 4|z|^2 = \frac{20}{3} - 2\sqrt{3} < 4.$$

We know that $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ is connected from Proposition 2.11. Thus, according to Lemma 4.8, $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ lies in the interior of \mathcal{I}_0^+ except the point $[e^{i2\pi/3}, -\sqrt{3}]$, which lies on the ideal boundary of \mathcal{I}_0^+ ; see Figure 1.

Observe that coordinates of the centers of \mathcal{I}_0^* and \mathcal{I}_{-1}^- are continuous on θ . Thus the geometric positions of the spheres \mathcal{I}_0^* and \mathcal{I}_{-1}^- depend continuously on θ . When $\theta = 0$, since the Cygan distance between the centers of \mathcal{I}_0^* and \mathcal{I}_{-1}^- is bigger than the sum of their radii, one can see that $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^- = \emptyset$. When $\theta = \frac{\pi}{3}$, we have shown that $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ lies in the interior of \mathcal{I}_0^+ . We also have $\mathcal{I}_0^+ \cap \mathcal{I}_0^* \cap \mathcal{I}_{-1}^- = \emptyset$ for $\theta \in [0, \frac{\pi}{3})$ by Lemma 4.8. Hence, the intersection $\mathcal{I}_0^* \cap \mathcal{I}_{-1}^-$ is either empty or contained in the interior of \mathcal{I}_0^+ . □

Proposition 4.10 Suppose that $\theta \in [0, \frac{\pi}{3}]$. For $k \in \mathbb{Z}$, the three isometric spheres \mathcal{I}_k^+ , \mathcal{I}_k^- and \mathcal{I}_k^* (resp. \mathcal{I}_k^+ , \mathcal{I}_{k-1}^- and \mathcal{I}_k^\diamond) have the following properties:

- The intersections $\mathcal{I}_k^+ \cap \mathcal{I}_k^-$, $\mathcal{I}_k^- \cap \mathcal{I}_k^*$ and $\mathcal{I}_k^* \cap \mathcal{I}_k^+$ (resp. $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^-$, $\mathcal{I}_{k-1}^- \cap \mathcal{I}_k^\diamond$ and $\mathcal{I}_k^\diamond \cap \mathcal{I}_k^+$) are discs.
- The intersection $\mathcal{I}_k^+ \cap \mathcal{I}_k^- \cap \mathcal{I}_k^*$ (resp. $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^- \cap \mathcal{I}_k^\diamond$) is a union of two geodesics which are crossed at the fixed point of $T^k S T^{-k}$ (resp. $T^k (S^{-1} T) T^{-k}$) in $\mathbb{H}_{\mathbb{C}}^2$

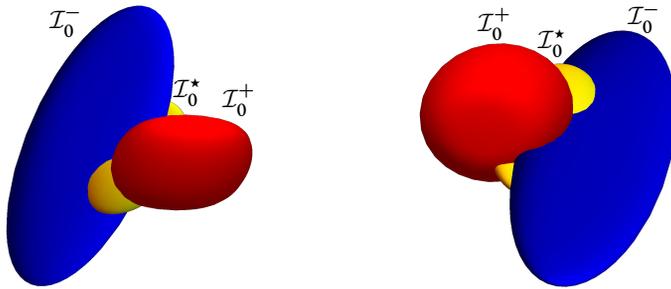


Figure 2: The ideal boundaries of the three spheres \mathcal{I}_0^+ , \mathcal{I}_0^- and \mathcal{I}_0^* on $\partial\mathbb{H}_{\mathbb{C}}^2$. (On the left is the case when $\theta = 0$ and on the right is $\theta = \frac{\pi}{3}$.)

and whose four endpoints are on $\partial\mathbb{H}_{\mathbb{C}}^2$. Moreover, the four rays from the fixed point to the four endpoints are cyclically permuted by $T^k ST^{-k}$ (resp. $T^k(S^{-1}T)T^{-k}$).

Proof According to Definition 4.1 and Proposition 4.2, it suffices to consider the isometric spheres \mathcal{I}_0^+ , \mathcal{I}_0^- and \mathcal{I}_0^* ; see Figure 2.

Let $q \in \mathcal{I}_0^+$. Consider the geographical coordinates (α, β, w) of q in (4-1). By Lemma 4.4, if q lies on $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$, then α, β and w should satisfy the equation

$$(4-13) \quad 2w^2 + 1 + \cos(\alpha) - \sqrt{2}w \cos(\frac{1}{2}\alpha + \beta - \theta) - 2\sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta - \theta) = 0.$$

Similarly, if q lies on $\mathcal{I}_0^+ \cap \mathcal{I}_0^*$, then α, β and w should satisfy the equation

$$(4-14) \quad 2w^2 + 1 + \cos(\alpha) - \sqrt{2}w \cos(-\frac{1}{2}\alpha + \beta - \theta) - 2\sqrt{2}w \cos(\frac{1}{2}\alpha + \beta - \theta) = 0.$$

Thus $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$ is the set of solutions of (4-13) and $\mathcal{I}_0^+ \cap \mathcal{I}_0^*$ is the set of solutions of (4-14). One can easily verify that the geographical coordinates of the point $q(0, \theta, \frac{\sqrt{2}}{2}) \in \mathbb{H}_{\mathbb{C}}^2$ satisfy (4-13) and (4-14), so the intersections $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$ and $\mathcal{I}_0^+ \cap \mathcal{I}_0^*$ are topological discs from Proposition 2.11.

The intersection of these two sets gives the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$. Now let us solve the system of equations (4-13) and (4-14). Let $t = \beta - \theta$. Subtracting (4-13) and (4-14) and simplifying, we obtain

$$2w \sin(\frac{1}{2}\alpha) \sin(t) = 0.$$

Thus $w = 0$ (this is impossible), $\alpha = 0$, or $t = 0$. If $t = 0$, then setting $\beta = \theta$ in (4-13), we get

$$2w^2 - 3\sqrt{2} \cos(\frac{1}{2}\alpha)w + 1 + \cos(\alpha) = 2(w - \frac{1}{2}\sqrt{2} \cos(\frac{1}{2}\alpha))(w - \sqrt{2} \cos(\frac{1}{2}\alpha)) = 0.$$

Note that the solutions of the above equation for w should satisfy $w^2 \leq \cos(\alpha)$. Thus

$$w = \frac{1}{2}\sqrt{2}\cos\left(\frac{1}{2}\alpha\right) \quad \text{with } \cos(\alpha) \geq \frac{1}{3}.$$

If $\alpha = 0$, then (4-13) becomes

$$(4-15) \quad 2w^2 - 3\sqrt{2}\cos(t)w + 2 = 0.$$

Note that the solutions of (4-15) for w should satisfy $w^2 \leq \cos(\alpha) = 1$. Thus the solutions of (4-15) are

$$w = \frac{3\cos(t) - \sqrt{9\cos^2(t) - 8}}{2\sqrt{2}} \quad \text{with } \frac{2\sqrt{2}}{3} \leq \cos(t) \leq 1,$$

and

$$w = \frac{3\cos(t) + \sqrt{9\cos^2(t) - 8}}{2\sqrt{2}} \quad \text{with } -1 \leq \cos(t) \leq -\frac{2\sqrt{2}}{3}.$$

So the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$ is the union $\mathcal{L}_1 \cup \mathcal{C}_1 \cup \mathcal{C}_2$, where

$$\mathcal{L}_1 = \left\{ q(\alpha, t + \theta, w) \in \mathcal{I}_0^+ : \cos(\alpha) \geq \frac{1}{3}, t = 0 \text{ and } w = \frac{1}{2}\sqrt{2}\cos\left(\frac{1}{2}\alpha\right) \right\},$$

$$\mathcal{C}_1 = \left\{ q(0, t + \theta, w) \in \mathcal{I}_0^+ : \frac{2\sqrt{2}}{3} \leq \cos(t) \leq 1 \text{ and } w = \frac{3\cos(t) - \sqrt{9\cos^2(t) - 8}}{2\sqrt{2}} \right\},$$

and

$$\mathcal{C}_2 = \left\{ q(0, t + \theta, w) \in \mathcal{I}_0^+ : -1 \leq \cos(t) \leq -\frac{2\sqrt{2}}{3}, w = \frac{3\cos(t) + \sqrt{9\cos^2(t) - 8}}{2\sqrt{2}} \right\}.$$

Note that \mathcal{L}_1 lies in a Lagrangian plane of \mathcal{I}_0^+ , and $\mathcal{C}_1 \cup \mathcal{C}_2$ lies in a complex line of \mathcal{I}_0^+ . It is obvious that \mathcal{C}_1 is an arc. One of its endpoints is $q(0, \theta, \frac{\sqrt{2}}{2}) \in \mathbb{H}_{\mathbb{C}}^2$, which is the fixed point of S . The other one is $q(0, \arccos(\frac{2\sqrt{2}}{3}) + \theta, 1) \in \partial\mathbb{H}_{\mathbb{C}}^2$. Similarly, \mathcal{C}_2 is an arc with endpoints $q(0, \theta, \frac{\sqrt{2}}{2})$ and $q(0, \arccos(-\frac{2\sqrt{2}}{3}) + \theta, -1) \in \partial\mathbb{H}_{\mathbb{C}}^2$. Thus $\mathcal{C}_1 \cup \mathcal{C}_2$ is connected. The endpoints of \mathcal{L}_1 are $q(\arccos(\frac{1}{3}), \theta, \frac{\sqrt{3}}{3})$ and $q(-\arccos(\frac{1}{3}), \theta, \frac{\sqrt{3}}{3})$, which are on $\partial\mathbb{H}_{\mathbb{C}}^2$. It is easy to see that \mathcal{L}_1 intersects with $\mathcal{C}_1 \cup \mathcal{C}_2$ at the point $q(0, \theta, \frac{\sqrt{2}}{2}) \in \mathbb{H}_{\mathbb{C}}^2$.

Moreover, $\mathcal{C}_1 \cup \mathcal{C}_2$ is a geodesic. In fact, the complex line containing $\mathcal{C}_1 \cup \mathcal{C}_2$ is $\mathcal{C} = \{(-1, z) \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : |z| \leq \sqrt{2}\}$. This is a disc bounded by the circle with center the origin and radius $\sqrt{2}$, while $\mathcal{C}_1 \cup \mathcal{C}_2$ lies in the circle with center $\frac{3}{2}e^{i\theta}$ and radius $\frac{1}{2}$, which is orthogonal to the boundary of the complex line. The Cayley transform given in Definition 2.2 maps \mathcal{C} to the vertical axis $\{(0, z) \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : |z| \leq 1\}$ in the ball

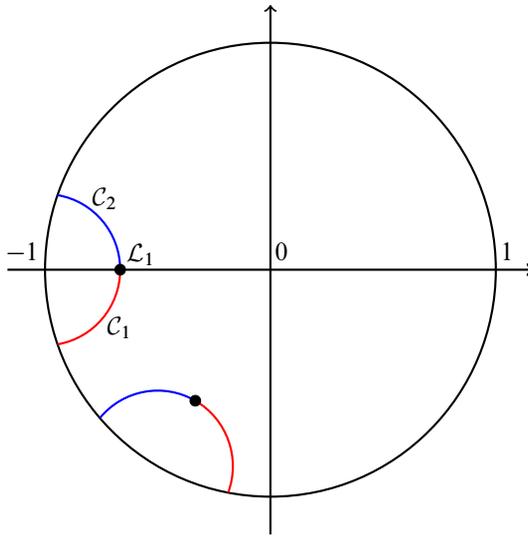


Figure 3: The triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$ viewed on the vertical axis in the ball model of $\mathbb{H}_{\mathbb{C}}^2$. The blue curve is \mathcal{C}_1 and the red one is \mathcal{C}_2 . The black point on the curve is the projection of \mathcal{L}_1 on the complex line. The left curve is the case when $\theta = 0$ and the one on the lower left is the case when $\theta = \frac{\pi}{3}$.

model of $\mathbb{H}_{\mathbb{C}}^2$. Thus \mathcal{C} is isometric to the Poincaré disc. Then $\mathcal{C}_1 \cup \mathcal{C}_2$ is mapped by the Cayley transform to an arc contained in the circle with center $-3e^{i\theta}/(2\sqrt{2})$ and radius $1/(2\sqrt{2})$, which is orthogonal to the unit circle. Hence $\mathcal{C}_1 \cup \mathcal{C}_2$ is a geodesic; see Figure 3.

The Cayley transform maps \mathcal{L}_1 to $\{(-\tan(\frac{1}{2}\alpha)i, -e^{i\theta}/\sqrt{2}) \in \mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2 : \cos(\alpha) \geq \frac{1}{3}\}$. Thus \mathcal{L}_1 and $\mathcal{C}_1 \cup \mathcal{C}_2$ are crossed at the point $(0, -e^{i\theta}/\sqrt{2}) \in \mathbb{H}_{\mathbb{C}}^2$, which is the image of $q(0, \theta, \frac{\sqrt{2}}{2})$ under the Cayley transform.

It is easy to check that $S(q(0, \theta, \frac{\sqrt{2}}{2})) = q(0, \theta, \frac{\sqrt{2}}{2})$, and that the other four points are cyclically permuted by S :

$$\begin{array}{ccc}
 q(0, \arccos(\frac{2\sqrt{2}}{3}) + \theta, 1) & \xrightarrow{S} & q(\arccos(\frac{1}{3}), \theta, \frac{\sqrt{3}}{3}) \\
 \uparrow s & & \downarrow s \\
 q(-\arccos(\frac{1}{3}), \theta, \frac{\sqrt{3}}{3}) & \xleftarrow{S} & q(0, \arccos(-\frac{2\sqrt{2}}{3}) + \theta, -1)
 \end{array}$$

Moreover, it is easy to verify that $\mathcal{C}_1 \cup \mathcal{C}_2 = S(\mathcal{L}_1)$, $S^2(\mathcal{L}_1) = \mathcal{L}_1$ and $S^2(\mathcal{C}_1) = \mathcal{C}_2$. Thus, the four rays from the fixed point to the four endpoints are cyclically permuted by S . \square

By applying powers of T and the symmetries in Proposition 4.2 to Propositions 4.6, 4.7 and 4.9, all pairwise intersections of the isometric spheres can be summarized:

Corollary 4.11 *Suppose that $\theta \in [0, \frac{\pi}{3}]$. Let $S = \{\mathcal{I}_k^\pm, \mathcal{I}_k^*, \mathcal{I}_k^\diamond : k \in \mathbb{Z}\}$ be the set of all the isometric spheres. Then for all $k \in \mathbb{Z}$:*

- (1) \mathcal{I}_k^+ is contained in the exterior of all the isometric spheres in S except $\mathcal{I}_k^-, \mathcal{I}_{k-1}^-, \mathcal{I}_k^*, \mathcal{I}_{k-1}^*, \mathcal{I}_k^\diamond$ and $\mathcal{I}_{k+1}^\diamond$. Moreover, $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^*$ (resp. $\mathcal{I}_k^+ \cap \mathcal{I}_{k+1}^\diamond$) is either empty or contained in the interior of \mathcal{I}_{k-1}^- (resp. \mathcal{I}_k^-). When $\theta = \frac{\pi}{3}$, $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^*$ (resp. $\mathcal{I}_k^+ \cap \mathcal{I}_{k+1}^\diamond$) will be tangent with \mathcal{I}_{k-1}^- (resp. \mathcal{I}_k^-) on $\partial\mathbb{H}_\mathbb{C}^2$ at the parabolic fixed point of $T^k(S^2T)T^{-k}$ (resp. $T^k(S^{-1}TS^{-1})T^{-k}$).
- (2) \mathcal{I}_k^- is contained in the exterior of all the isometric spheres in S except $\mathcal{I}_k^+, \mathcal{I}_{k+1}^+, \mathcal{I}_k^*, \mathcal{I}_{k+1}^*, \mathcal{I}_k^\diamond$ and $\mathcal{I}_{k+1}^\diamond$. Moreover, $\mathcal{I}_k^- \cap \mathcal{I}_k^\diamond$ (resp. $\mathcal{I}_k^- \cap \mathcal{I}_{k+1}^*$) is either empty or contained in the interior of \mathcal{I}_k^+ (resp. \mathcal{I}_{k+1}^+). When $\theta = \frac{\pi}{3}$, $\mathcal{I}_k^- \cap \mathcal{I}_k^\diamond$ (resp. $\mathcal{I}_k^- \cap \mathcal{I}_{k+1}^*$) will be tangent with \mathcal{I}_k^+ (resp. \mathcal{I}_{k+1}^+) on $\partial\mathbb{H}_\mathbb{C}^2$ at the parabolic fixed point of $T^k(ST^{-1}S)T^{-k}$ (resp. $T^k(S^2T^{-1})T^{-k}$).
- (3) \mathcal{I}_k^* is contained in the exterior of all the isometric spheres in S except $\mathcal{I}_k^\pm, \mathcal{I}_{k+1}^+, \mathcal{I}_{k-1}^-, \mathcal{I}_k^\diamond$ and $\mathcal{I}_{k+1}^\diamond$. Moreover, $\mathcal{I}_k^* \cap \mathcal{I}_k^\diamond$ (resp. $\mathcal{I}_k^* \cap \mathcal{I}_{k+1}^\diamond$) is contained in the interior of \mathcal{I}_k^+ (resp. \mathcal{I}_k^-). $\mathcal{I}_k^* \cap \mathcal{I}_{k+1}^+$ is described in (1), and $\mathcal{I}_k^* \cap \mathcal{I}_{k-1}^-$ is described in (2). When $\theta = \frac{\pi}{3}$, \mathcal{I}_k^* will be tangent with \mathcal{I}_{k+1}^* (resp. \mathcal{I}_{k-1}^-) on $\partial\mathbb{H}_\mathbb{C}^2$ at the parabolic fixed point of $T^k(S^2T^{-1})T^{-k}$ (resp. $T^k(T^{-1}S^2)T^{-k}$).
- (4) \mathcal{I}_k^\diamond is contained in the exterior of all the isometric spheres in S except $\mathcal{I}_k^\pm, \mathcal{I}_{k-1}^+, \mathcal{I}_k^*$ and \mathcal{I}_{k-1}^* . Moreover, $\mathcal{I}_k^\diamond \cap \mathcal{I}_k^*$ and $\mathcal{I}_k^\diamond \cap \mathcal{I}_{k-1}^*$ are described in (3). $\mathcal{I}_k^\diamond \cap \mathcal{I}_k^-$ is described in (2) and $\mathcal{I}_k^\diamond \cap \mathcal{I}_{k-1}^+$ is described in (1). When $\theta = \frac{\pi}{3}$, \mathcal{I}_k^\diamond will be tangent with $\mathcal{I}_{k+1}^\diamond$ (resp. \mathcal{I}_{k-1}^*) on $\partial\mathbb{H}_\mathbb{C}^2$ at the parabolic fixed point of $T^k(T^{-1}S^2)T^{-k}$ (resp. $T^k(S^2T^{-1})T^{-k}$).

4.3 The Ford domain

Definition 4.12 Let D be the intersection of the closures of the exteriors of the isometric spheres $\mathcal{I}_k^+, \mathcal{I}_k^-, \mathcal{I}_k^*$ and \mathcal{I}_k^\diamond , for $k \in \mathbb{Z}$.

Definition 4.13 For $k \in \mathbb{Z}$, let s_k^+ (resp. s_k^-, s_k^* and s_k^\diamond) be the side of D contained in the isometric sphere \mathcal{I}_k^+ (resp. $\mathcal{I}_k^-, \mathcal{I}_k^*$ and \mathcal{I}_k^\diamond).

Definition 4.14 A ridge is defined to be the 2-dimensional connected intersections of two sides.

By Corollary 4.11, the ridges are $s_k^+ \cap s_k^-, s_k^+ \cap s_k^*, s_k^+ \cap s_{k-1}^-, s_k^+ \cap s_k^\diamond, s_k^- \cap s_k^*$ and $s_{k-1}^- \cap s_k^\diamond$ for $k \in \mathbb{Z}$, and the sides and ridges are related as follows:

- The side s_k^+ is bounded by the ridges $s_k^+ \cap s_k^-, s_k^+ \cap s_k^*, s_k^+ \cap s_{k-1}^-$ and $s_k^+ \cap s_k^\diamond$.
- The side s_k^- is bounded by the ridges $s_k^+ \cap s_k^-, s_k^- \cap s_k^*, s_k^- \cap s_{k+1}^+$ and $s_0^- \cap s_{k+1}^\diamond$.
- The side s_k^* is bounded by the ridges $s_k^+ \cap s_k^*$ and $s_k^- \cap s_k^*$.
- The side s_k^\diamond is bounded by the ridges $s_k^\diamond \cap s_{k-1}^-$ and $s_k^+ \cap s_k^\diamond$.

Proposition 4.15 *The ridges $s_k^+ \cap s_k^-, s_k^+ \cap s_k^*, s_k^+ \cap s_{k-1}^-, s_k^+ \cap s_k^\diamond, s_k^- \cap s_k^*$ and $s_{k-1}^- \cap s_k^\diamond$ for $k \in \mathbb{Z}$ are all topologically the union of two sectors.*

Proof The ridge $s_k^+ \cap s_k^-$ is contained in $\mathcal{I}_k^+ \cap \mathcal{I}_k^-$. According to Proposition 4.10, $\mathcal{I}_k^+ \cap \mathcal{I}_k^-$ is topologically a disc and $\mathcal{I}_k^+ \cap \mathcal{I}_k^- \cap \mathcal{I}_k^*$ is the union of two crossed geodesics. The two crossed geodesics divide the disc into four sectors, one opposite pair of which will lie in the interior of the isometric sphere \mathcal{I}_k^* . Thus $s_k^+ \cap s_k^-$ is the other opposite pair of the four sectors in the disc. More precisely, up to the powers of T , let us consider $s_0^+ \cap s_0^-$. Let Δ be the disc $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$ described in (4-13) and the two crossed geodesics $\mathcal{L}_1 \cup \mathcal{C}_1 \cup \mathcal{C}_2$ be as described in Proposition 4.10. By Proposition 4.2, the complex involution I_2 preserves Δ and $\mathcal{L}_1 \cup \mathcal{C}_1 \cup \mathcal{C}_2$. Recall that I_2 fixes the complex line \mathcal{C}_2 with polar vector n_2 described in Section 3. One can compute that the intersection $\mathcal{C}_2 \cap \Delta$ is the curve

$$(4-16) \quad \mathcal{C}_2 \cap \Delta = \{q(\alpha, \frac{1}{2}\alpha + \theta, \frac{\sqrt{2}}{2}) \in \mathcal{I}_0^+ : \cos(\alpha) \geq \frac{1}{3}\}.$$

Of course $\mathcal{C}_2 \cap \Delta$ intersects with $\mathcal{L}_1 \cup \mathcal{C}_1 \cup \mathcal{C}_2$ at the fixed point of S , and divides Δ into two parts. I_2 fixes $\mathcal{C}_2 \cap \Delta$ and interchanges \mathcal{L}_1 and $\mathcal{C}_1 \cup \mathcal{C}_2$. Thus $\mathcal{C}_2 \cap \Delta$ is contained in the union of two opposite sectors. By Lemma 4.4, $\mathcal{C}_2 \cap \Delta$ lies on the closure of the exterior of \mathcal{I}_0^* , since $f_0^*(\alpha, \frac{1}{2}\alpha + \theta, \frac{\sqrt{2}}{2}) = 1 - \cos(\alpha) \geq 0$. Therefore, the union of two opposite sectors containing $\mathcal{C}_2 \cap \Delta$ is the ridge $s_0^+ \cap s_0^-$. Moreover, this ridge is preserved by I_2 . By using the parametrization of the Giraud disk in [8], we can draw the Giraud disk $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$ and the intersection with the isometric spheres $\mathcal{I}_{-1}^-, \mathcal{I}_0^*, \mathcal{I}_{-1}^*, \mathcal{I}_0^\diamond$ and \mathcal{I}_1^\diamond ; see Figure 4.

The other ridges can be described by a similar argument. □

Proposition 4.16 (1) *The side s_k^+ (resp. s_k^-) is a topological solid cylinder in $\mathbb{H}_{\mathbb{C}}^2 \cup \partial\mathbb{H}_{\mathbb{C}}^2$. The intersection of ∂s_k^+ (resp. ∂s_k^-) with $\mathbb{H}_{\mathbb{C}}^2$ is the disjoint union of two topological discs.*

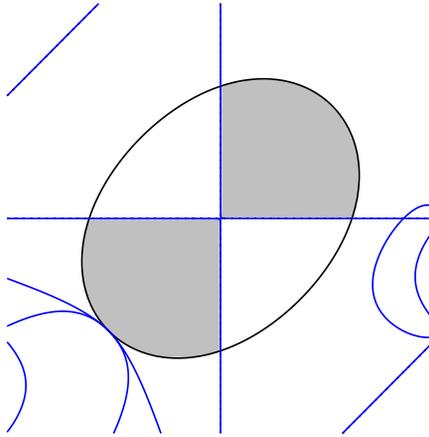


Figure 4: The ridge $s_0^+ \cap s_0^-$ (the shaded region in the intersection of $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$) in the plane with spinal coordinates introduced in [8]. The triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$ is the two mutually perpendicular lines. Compare to [8, Figure 16].

- (2) The side s_k^* (resp. s_k^\diamond) is a topological solid light cone in $\mathbb{H}_\mathbb{C}^2 \cup \partial\mathbb{H}_\mathbb{C}^2$. The intersection of ∂s_k^* (resp. ∂s_k^\diamond) with $\mathbb{H}_\mathbb{C}^2$ is the light cone.

Proof (1) The side s_k^+ is contained in the isometric sphere \mathcal{I}_k^+ . By Corollary 4.11, s_k^+ might intersect with the sides contained in the isometric spheres $\mathcal{I}_k^-, \mathcal{I}_{k-1}^-, \mathcal{I}_k^*, \mathcal{I}_{k-1}^*, \mathcal{I}_k^\diamond$ and $\mathcal{I}_{k+1}^\diamond$.

Let Δ_1 be the union of the ridges $s_k^+ \cap s_k^-$ and $s_k^+ \cap s_k^*$, and Δ_2 be the union of the ridges $s_k^+ \cap s_{k-1}^-$ and $s_k^+ \cap s_k^\diamond$. By Proposition 4.10, Δ_1 contains the cross $\mathcal{I}_k^+ \cap \mathcal{I}_k^- \cap \mathcal{I}_k^*$. By Proposition 4.15, Δ_1 is a union of four sectors which are patched together along the cross. Hence, Δ_1 is topologically either a disc or a light cone. By a straightforward computation, the ideal boundary of Δ_1 on $\mathbb{H}_\mathbb{C}^2$ is a simple closed curve on the ideal boundary of \mathcal{I}_k^+ ; see Figure 2. Thus Δ_1 is a topological disc. By a similar argument, Δ_2 is a topological disc.

If $\theta \neq \frac{\pi}{3}$ then $\mathcal{I}_k^+ \cap \mathcal{I}_k^* \cap \mathcal{I}_{k-1}^- = \emptyset$, so Δ_1 and Δ_2 are disjoint, and if $\theta = \frac{\pi}{3}$ they intersect at two points on $\partial\mathbb{H}_\mathbb{C}^2$; see Figures 9 and 10. Note that isometric spheres are topological balls and their pairwise intersections are connected. So, s_k^+ is a topological solid cylinder; see Figure 5. A similar argument describes s_k^- .

- (2) The side s_k^* is contained in the isometric sphere \mathcal{I}_k^* . According to Corollary 4.11, s_k^* only intersects with s_k^+ and s_k^- . Let Δ_3 be the union of $s_k^+ \cap s_k^*$ and $s_k^- \cap s_k^*$. By

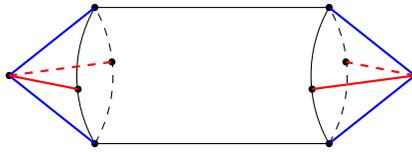


Figure 5: A schematic view of the side $s_k^+ (s_k^-)$.

Propositions 4.10 and 4.15, Δ_3 is a union of four sectors which are patched together along the cross $\mathcal{I}_k^+ \cap \mathcal{I}_k^- \cap \mathcal{I}_k^*$. By a computation for the case $k = 0$, one can see that the ideal boundary of Δ_3 is a union of two disjoint simple closed curves in the ideal boundary of \mathcal{I}_k^* ; see Figure 2. Thus Δ_3 is a light cone. Hence, s_k^* is topologically a solid light cone; see Figure 6. A similar argument describes s_k^\diamond . \square

By applying a Poincaré polyhedron theorem in $\mathbb{H}_{\mathbb{C}}^2$ as stated for example in [22], [8] or [19] (see [3] for a version in the hyperbolic plane), we have our main result:

Theorem 4.17 Suppose that $\theta \in [0, \frac{\pi}{3}]$. Let D be as in Definition 4.12. Then D is a fundamental domain for the cosets of $\langle T \rangle$ in Γ . Moreover, Γ is discrete and has the presentation

$$\Gamma = \langle S, T \mid S^4 = (T^{-1}S)^4 = \text{id} \rangle.$$

Proof The sides of D are $s_k^+ s_k^-$, s_k^* and s_k^\diamond . The ridges of D are $s_k^+ \cap s_k^-$, $s_k^+ \cap s_k^*$, $s_k^+ \cap s_{k-1}^-$, $s_k^+ \cap s_k^\diamond$, $s_k^- \cap s_k^*$ and $s_{k-1}^- \cap s_k^\diamond$. To obtain the side-pairing maps and ridge cycles, by applying powers of T , it suffices to consider the case where $k = 0$.

The side-pairing maps The side s_0^+ is contained in the isometric sphere $\mathcal{I}(S)$ and s_0^- in the isometric sphere $\mathcal{I}(S^{-1})$. The ridge $s_0^+ \cap s_0^-$ is contained in the disc $\mathcal{I}(S) \cap \mathcal{I}(S^{-1})$, which is defined by the triple equality

$$|\langle z, \mathbf{q}_\infty \rangle| = |\langle z, S^{-1}(\mathbf{q}_\infty) \rangle| = |\langle z, S(\mathbf{q}_\infty) \rangle|$$

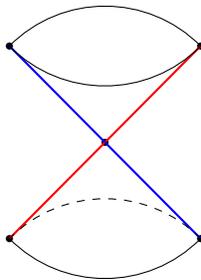


Figure 6: A schematic view of the side $s_k^* (s_k^\diamond)$.

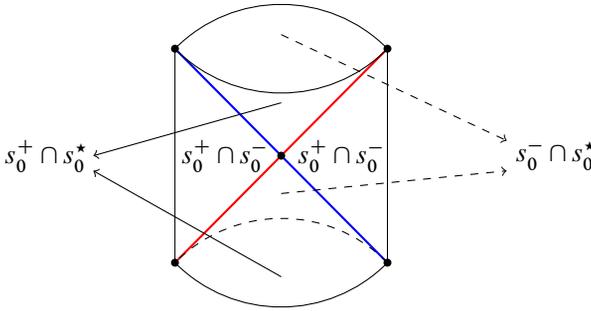


Figure 7: A schematic view of the ridges $s_0^+ \cap s_0^-$, $s_0^+ \cap s_0^*$ and $s_0^- \cap s_0^*$. The thick cross is the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$, that is the union $\mathcal{L}_1 \cup \mathcal{C}_1 \cup \mathcal{C}_2$ described in Proposition 4.10.

and the ridge $s_0^- \cap s_0^*$ is contained in the disc $\mathcal{I}(S^{-1}) \cap \mathcal{I}(S^2)$, which is defined by the triple equality

$$|\langle z, \mathbf{q}_\infty \rangle| = |\langle z, S(\mathbf{q}_\infty) \rangle| = |\langle z, S^{-2}(\mathbf{q}_\infty) \rangle|.$$

Since S maps q_∞ to $S(q_\infty)$, $S^{-1}(q_\infty)$ to q_∞ and $S(q_\infty)$ to $S^2(q_\infty) = S^{-2}(q_\infty)$, S maps the disc $\mathcal{I}(S) \cap \mathcal{I}(S^{-1})$ to the disc $\mathcal{I}(S^{-1}) \cap \mathcal{I}(S^2)$. Note that

$$\mathcal{I}(S) \cap \mathcal{I}(S^{-1}) \cap \mathcal{I}(S^2)$$

is the union of two crossed geodesics (see Proposition 4.10) whose four rays are cyclically permuted by S . Since the ridge $s_0^+ \cap s_0^-$ lies in the closure of the exterior of the isometric sphere $\mathcal{I}(S^2)$, according to (4-16), the point $q(\frac{\pi}{3}, \frac{\pi}{6} + \theta, \frac{\sqrt{2}}{2})$ is contained in $s_0^+ \cap s_0^-$. One can easily verify that $S(q(\frac{\pi}{3}, \frac{\pi}{6} + \theta, \frac{\sqrt{2}}{2}))$ lies in the exterior of $\mathcal{I}(S)$. Thus $S(q(\frac{\pi}{3}, \frac{\pi}{6} + \theta, \frac{\sqrt{2}}{2}))$ is contained in $s_0^- \cap s_0^*$, which lies in the closure of the exterior of the isometric sphere $\mathcal{I}(S)$.

Hence S maps the ridge $s_0^+ \cap s_0^-$ to the ridge $s_0^- \cap s_0^*$. Similarly, S maps $s_0^+ \cap s_0^*$ to $s_0^+ \cap s_0^-$; see Figure 7. Since S maps $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^\diamond$ to $\mathcal{I}_0^- \cap \mathcal{I}_1^+ \cap \mathcal{I}_1^\diamond$, a similar argument shows that S maps $s_0^+ \cap s_{-1}^-$ to $s_0^- \cap s_1^\diamond$ and $s_0^+ \cap s_0^\diamond$ to $s_0^- \cap s_1^+$.

By a similar argument, s_0^* (resp. s_0^\diamond) is mapped to itself by the elliptic element of order two S^2 (resp. $(T^{-1}S)^2 = (S^{-1}T)^2$), which sends $s_0^+ \cap s_0^*$ to $s_0^- \cap s_0^*$ (resp. $s_0^\diamond \cap s_{-1}^-$ to $s_0^+ \cap s_0^\diamond$) and vice versa.

Hence, the side-pairing maps are

$$T^k S T^{-k} : s_k^+ \rightarrow s_k^-, \quad T^k S^2 T^{-k} : s_k^* \rightarrow s_k^* \quad \text{and} \quad T^k (T^{-1} S)^2 T^{-k} : s_k^\diamond \rightarrow s_k^\diamond.$$

The cycle transformations According to the side-pairing maps, the ridge cycles are

$$(s_k^+ \cap s_k^-, s_k^+, s_k^-) \xrightarrow{T^k S T^{-k}} (s_k^* \cap s_k^-, s_k^*, s_k^-) \xrightarrow{T^k S^2 T^{-k}} (s_k^+ \cap s_k^*, s_k^+, s_k^*) \xrightarrow{T^k S T^{-k}} (s_k^+ \cap s_k^-, s_k^+, s_k^-)$$

and

$$(s_k^+ \cap s_{k-1}^-, s_k^+, s_{k-1}^-) \xrightarrow{T^k (T^{-1} S) T^{-k}} (s_k^\diamond \cap s_{k-1}^-, s_k^\diamond, s_{k-1}^-) \xrightarrow{T^k (T^{-1} S)^2 T^{-k}} (s_k^+ \cap s_k^\diamond, s_k^+, s_k^\diamond) \xrightarrow{T^k (T^{-1} S) T^{-k}} (s_k^+ \cap s_{k-1}^-, s_k^+, s_{k-1}^-).$$

Thus the cycle transformations are

$$T^k S T^{-k} \cdot T^k S^2 T^{-k} \cdot T^k S T^{-k} = T^k S^4 T^{-k}$$

and

$$T^k (T^{-1} S) T^{-k} \cdot T^k (T^{-1} S)^2 T^{-k} \cdot T^k (T^{-1} S) T^{-k} = T^k (T^{-1} S)^4 T^{-k},$$

which are equal to the identity map, since $S^4 = \text{id}$ and $(T^{-1} S)^4 = \text{id}$.

The local tessellation There are exactly two copies of D along each side, since the sides are contained in isometric spheres and the side-pairing maps send the exteriors to the interiors. Thus there is nothing to verify for the points in the interior of every side.

According to the ridge cycles and cycle transformations, there are exactly three copies of D along each ridge.

$s_k^+ \cap s_k^-$, $s_k^* \cap s_k^-$ and $s_k^+ \cap s_k^*$ These three ridges are in one cycle. Thus, we only need to consider the ridge $s_k^+ \cap s_k^-$. Since the cycle transformation of $s_k^+ \cap s_k^-$ is

$$T^k S T^{-k} \cdot T^k S^2 T^{-k} \cdot T^k S T^{-k} = \text{id},$$

the three copies of D along $s_k^+ \cap s_k^-$ are D , $T^k S^{-1} T^{-k}(D)$ and $T^k S T^{-k}(D)$. We know that $s_k^+ \cap s_k^-$ is contained in $\mathcal{I}(T^k S T^{-k}) \cap \mathcal{I}(T^k S^{-1} T^{-k})$, which is defined by the triple equality

$$|\langle z, \mathbf{q}_\infty \rangle| = |\langle z, T^k S^{-1} T^{-k}(\mathbf{q}_\infty) \rangle| = |\langle z, T^k S T^{-k}(\mathbf{q}_\infty) \rangle|.$$

For z in the neighborhoods of $s_k^+ \cap s_k^-$ in D , $|\langle z, \mathbf{q}_\infty \rangle|$ is the smallest of the three quantities in the above triple equality.

For z in the neighborhoods of $s_k^* \cap s_k^-$ in D , $|\langle z, \mathbf{q}_\infty \rangle|$ is at most $|\langle z, T^k S T^{-k}(\mathbf{q}_\infty) \rangle|$ or $|\langle z, T^k S^{-2} T^{-k}(\mathbf{q}_\infty) \rangle|$. Applying $T^k S^{-1} T^{-k}$ gives a neighborhood of $s_k^+ \cap s_k^-$ in

$T^k S^{-1} T^{-k}(D)$, where $|\langle z, T^k S^{-1} T^{-k}(\mathbf{q}_\infty) \rangle|$ is the smallest of the three quantities in the above triple equality.

For z in the neighborhoods of $s_k^+ \cap s_k^*$ in D , $|\langle z, \mathbf{q}_\infty \rangle|$ is at most $|\langle z, T^k S^{-1} T^{-k}(\mathbf{q}_\infty) \rangle|$ or $|\langle z, T^k S^{-2} T^{-k}(\mathbf{q}_\infty) \rangle|$. Applying $T^k S T^{-k}$ gives a neighborhood of $s_k^+ \cap s_k^-$ in $T^k S T^{-k}(D)$, where $|\langle z, T^k S T^{-k}(\mathbf{q}_\infty) \rangle|$ is the smallest of the three quantities in the above triple equality.

Thus the union of D , $T^k S^{-1} T^{-k}(D)$ and $T^k S T^{-k}(D)$ forms a regular neighborhood of each point in $s_k^+ \cap s_k^-$.

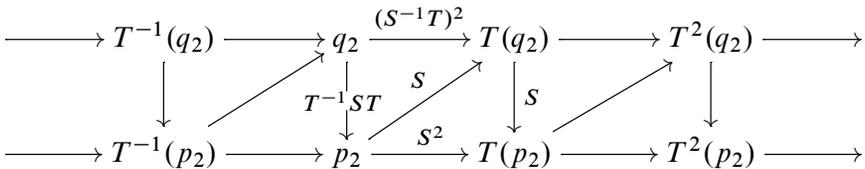
$s_k^+ \cap s_{k-1}^-$, $s_k^\diamond \cap s_{k-1}^-$ and $s_k^+ \cap s_k^\diamond$ We need only consider the ridge $s_k^+ \cap s_{k-1}^-$. Since the cycle transformation of $s_k^+ \cap s_{k-1}^-$ is

$$T^k(T^{-1}S)T^{-k} \cdot T^k(T^{-1}S)^2T^{-k} \cdot T^k(T^{-1}S)T^{-k} = \text{id},$$

the union of D , $T^k(T^{-1}S)^{-1}T^{-k}(D)$ and $T^k(T^{-1}S)T^{-k}(D)$ forms a regular neighborhood of each point in $s_k^+ \cap s_{k-1}^-$ by a similar argument as in the previous case.

Consistent system of horoballs When $\theta = \frac{\pi}{3}$, there are accidental ideal vertices on D . The sides s_k^* and s_{k+1}^* will be asymptotic on $\partial\mathbb{H}_\mathbb{C}^2$ at the fixed point of the parabolic element $T^k(S^2T^{-1})T^{-k}$, and the sides s_k^\diamond and s_{k+1}^\diamond will be asymptotic on $\partial\mathbb{H}_\mathbb{C}^2$ at the fixed point of the parabolic element $T^k(ST^{-1}S)T^{-k}$. To show that there is a consistent system of horoballs it suffices to show that all the cycle transformations fixing a given cusp are nonloxodromic.

Let p_2 be the fixed point of $T^{-1}S^2$ and q_2 be the fixed point of $(T^{-1}S)^2T$ (the coordinates of p_2 and q_2 are given in Definition 5.1, or see Figure 9). Then all the accidental ideal vertices $\{T^k(p_2)\}$ and $\{T^k(q_2)\}$ are related by the side-pairing maps as follows:



Thus, up to powers of T , all the cycle transformations are $S \cdot S \cdot S^{-2} = \text{id}$ and

$$T^{-1}ST \cdot (S^{-1}T)^{-2} \cdot S = (T^{-1}S^2)^2 = (I_1 I_3 I_2 I_3)^2,$$

which is parabolic. This means that p_2 is fixed by the parabolic element $(T^{-1}S^2)^2$.

Therefore, D is a fundamental domain for the cosets of $\langle T \rangle$ in Γ . The side-pairing maps and T will generate the group Γ . The reflection relations are

$$(T^k S^2 T^{-k})^2 = \text{id} \quad \text{and} \quad (T^k (T^{-1} S)^2 T^{-k})^2 = \text{id}.$$

The cycle relations are

$$T^k S^4 T^{-k} = \text{id} \quad \text{and} \quad T^k (T^{-1} S)^4 T^{-k} = \text{id}.$$

Thus Γ is discrete and has the presentation

$$\Gamma = \langle S, T \mid S^4 = (T^{-1} S)^4 = \text{id} \rangle. \quad \square$$

Since Γ is a subgroup of $\langle I_1, I_2, I_3 \rangle$ of index 2:

Corollary 4.18 *Let $\langle I_1, I_2, I_3 \rangle$ be a complex hyperbolic $(4, 4, \infty)$ triangle group as in Proposition 3.1. Then $\langle I_1, I_2, I_3 \rangle$ is discrete and faithful if and only if $I_1 I_3 I_2 I_3$ is nonelliptic.*

This answers Conjecture 1.1 on the complex hyperbolic $(4, 4, \infty)$ triangle group.

5 The manifold at infinity

In this section, we study the group Γ in the case when $\theta = \frac{\pi}{3}$. That is, the group $\Gamma = \langle S, T \rangle = \langle I_2 I_3, I_2 I_1 \rangle$ with $T^{-1} S^2 = I_1 I_3 I_2 I_3$ being parabolic.

In this case, the Ford domain D has additional ideal vertices on $\partial \mathbb{H}_{\mathbb{C}}^2$, which are parabolic fixed points corresponding to the conjugators of $T^{-1} S^2$. By intersecting a fundamental domain for $\langle T \rangle$ acting on $\partial \mathbb{H}_{\mathbb{C}}^2$ with the ideal boundary of D , we obtain a fundamental domain for Γ acting on its discontinuity region $\Omega(\Gamma)$.

Topologically, this fundamental domain is the product of an unknotted cylinder and a ray; see Proposition 5.13. By cutting and gluing we obtain two polyhedra \mathcal{P}_+ and \mathcal{P}_- ; see Proposition 5.18. Gluing \mathcal{P}_- to \mathcal{P}_+ by S^{-1} , we obtain a polyhedron \mathcal{P} . By studying the combinatorial properties of \mathcal{P} , we show that the quotient $\Omega(\Gamma)/\Gamma$ is homeomorphic to the two-cusped hyperbolic 3-manifold $s782$.

Let U be the ideal boundary of D on $\partial \mathbb{H}_{\mathbb{C}}^2$. Let \tilde{s}_k^+ (resp. \tilde{s}_k^- , \tilde{s}_k^* , and \tilde{s}_k^\diamond) be the ideal boundary of the side of D contained in the ideal boundary of the isometric sphere \mathcal{I}_k^+ (resp. \mathcal{I}_k^- , \mathcal{I}_k^* and \mathcal{I}_k^\diamond). Then the union of all the sides $\{\tilde{s}_k^+\}$, $\{\tilde{s}_k^-\}$, $\{\tilde{s}_k^*\}$ and $\{\tilde{s}_k^\diamond\}$ for $k \in \mathbb{Z}$ form the boundary of U .

5.1 The vertices of U

Definition 5.1 In the Heisenberg coordinates we define the points

$$\begin{aligned}
 q_2 &= \left[\frac{1}{2}(-3+i\sqrt{3}), -\sqrt{3} \right], & p_6 &= \left[\frac{1}{6}(2-\sqrt{6}+i(2\sqrt{3}+\sqrt{2})), -\frac{4}{3}\sqrt{2} \right], \\
 q_3 &= \left[\frac{1}{2}(1+i\sqrt{3}), \sqrt{3} \right], & p_7 &= \left[\frac{1}{6}(2+\sqrt{6}+i(2\sqrt{3}-\sqrt{2})), \frac{4}{3}\sqrt{2} \right], \\
 p_2 &= \left[\frac{1}{2}(-1+i\sqrt{3}), -\sqrt{3} \right], & p_8 &= \left[\frac{1}{6}(-2+\sqrt{6}+i(2\sqrt{3}+\sqrt{2})), \frac{4}{3}\sqrt{2} \right], \\
 p_3 &= \left[\frac{1}{2}(3+i\sqrt{3}), \sqrt{3} \right], & p_9 &= \left[\frac{1}{6}(-2-\sqrt{6}+i(2\sqrt{3}-\sqrt{2})), -\frac{4}{3}\sqrt{2} \right], \\
 p_4 &= \left[\frac{1}{6}(4+\sqrt{6}+i(4\sqrt{3}-\sqrt{2})), 0 \right], & p_{10} &= \left[\frac{1}{6}(-4+\sqrt{6}+i(4\sqrt{3}+\sqrt{2})), 0 \right], \\
 p_5 &= \left[\frac{1}{6}(4-\sqrt{6}+i(4\sqrt{3}+\sqrt{2})), 0 \right], & p_{11} &= \left[\frac{1}{6}(-4-\sqrt{6}+i(4\sqrt{3}-\sqrt{2})), 0 \right],
 \end{aligned}$$

and

$$p_{12} = T(p_9), \quad p_{13} = T(p_8), \quad p_{14} = T(p_{11}), \quad p_{15} = T(p_{10}).$$

By [Proposition 4.10](#) and [Corollary 4.11](#), we have the following:

Proposition 5.2 *The points in [Definition 5.1](#) satisfy:*

- p_4, p_5, p_6 and p_7 are the four points on the ideal boundary of $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$, which is described in [Proposition 4.10](#).
- p_8, p_9, p_{10} and p_{11} are the four points on the ideal boundary of $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^\diamond$.
- p_{12}, p_{13}, p_{14} and p_{15} are the four points on the ideal boundary of $\mathcal{I}_1^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_1^\diamond$.
- p_2 (resp. p_3) is the parabolic fixed point of $T^{-1}S^2$ (resp. S^2T^{-1}), which is the intersection $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^* \cap \mathcal{I}_{-1}^*$ (resp. $\mathcal{I}_1^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^* \cap \mathcal{I}_1^*$).
- q_3 (resp. q_2) is the parabolic fixed point of $ST^{-1}S$ (resp. $T^{-1}ST^{-1}ST$), which is the intersection of the four isometric spheres $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^\diamond \cap \mathcal{I}_1^\diamond$ (resp. $\mathcal{I}_{-1}^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^\diamond \cap \mathcal{I}_{-1}^\diamond$).

Proof As described in [Proposition 4.10](#), all of the triple intersections $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^*$, $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^\diamond$ and $\mathcal{I}_1^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_1^\diamond$ have exactly four points lying on $\partial\mathbb{H}_\mathbb{C}^2$. When writing the standard lifts of p_4, p_5, p_6 and p_7 , one can see that they are the four points in the proof of [Proposition 4.10](#). Thus the first item is proved.

By [Proposition 4.2](#), the four points of $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^\diamond$ are the images of p_4, p_5, p_6 and p_7 under the antiholomorphic involution τ , which are p_{11}, p_{10}, p_8 and p_9 .

The second item and the fact that $\mathcal{I}_1^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_1^\diamond = T(\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^\diamond)$ imply the third item.

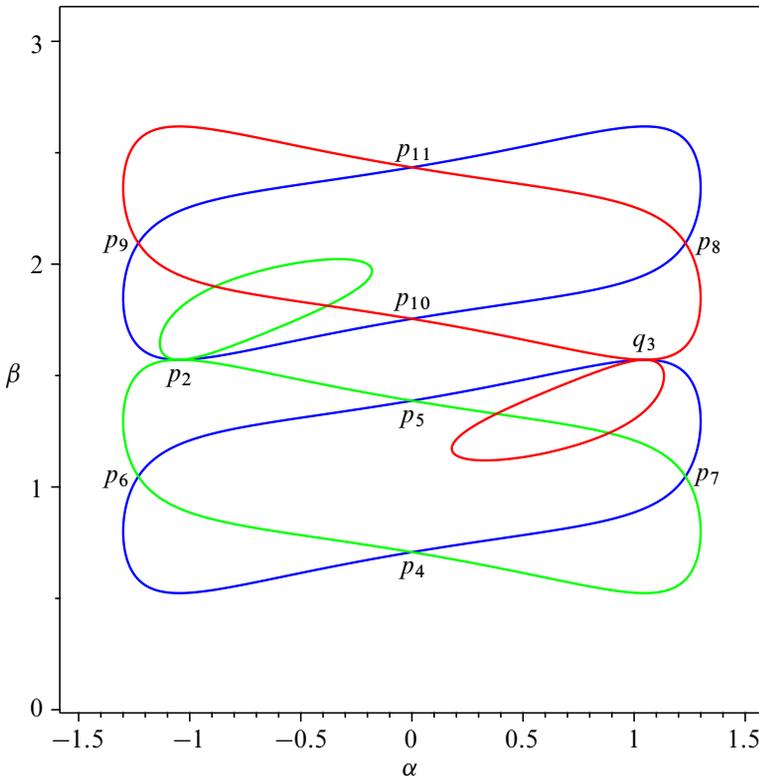


Figure 8: Intersections of the isometric spheres $\mathcal{I}_0^-, \mathcal{I}_{-1}^-, \mathcal{I}_0^*, \mathcal{I}_{-1}^*, \mathcal{I}_0^\diamond$ and \mathcal{I}_1^\diamond with \mathcal{I}_0^+ in $\partial\mathbb{H}_\mathbb{C}^2$, viewed in geographical coordinates. Here $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ in the vertical coordinate and $\beta \in [0, \pi]$ in the horizontal one. The region exterior to the six Jordan closed curves has two connect components. One of them is the topological octagon with vertices $p_2, p_6, p_4, p_7, q_3, p_8, p_{11}$ and p_9 . The other one is a topological quadrilateral with vertices p_2, p_5, q_3 and p_{10} .

We will only prove the statement for p_2 in the last two items; the others follow by similar arguments. By Lemma 4.8, p_2 is the parabolic fixed point of $T^{-1}S^2$ and is the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^*$. By Corollary 4.11(3), \mathcal{I}_0^* is tangent with \mathcal{I}_{-1}^* at p_2 . □

5.2 The sides of U

Now we study the combinatorial properties of the sides; see Figures 8 and 9.

Proposition 5.3 *The interior of the side \tilde{s}_0^+ has connected components*

- an octagon, denoted by \mathcal{O}_0^+ , with vertices $p_2, p_6, p_4, p_7, q_3, p_8, p_{11}$ and p_9 ,
- a quadrilateral, denoted by \mathcal{Q}_0^+ , with vertices p_2, p_5, q_3 and p_{10} .

Proof By Proposition 4.16, when $\theta < \frac{\pi}{3}$, the side \tilde{s}_0^+ is topologically an annulus bounded by two disjoint simple closed curves which are the union of the ideal boundaries of the ridges $s_0^+ \cap s_{-1}^-$ and $s_0^+ \cap s_0^\diamond$, respectively $s_0^+ \cap s_0^-$ and $s_0^+ \cap s_0^*$. When $\theta = \frac{\pi}{3}$, these two curves intersect at two points, which divide \tilde{s}_0^+ into two parts. That is to say the interior of the side \tilde{s}_0^+ has two connected components.

By Proposition 5.2, the ideal boundary of the ridge $s_0^+ \cap s_{-1}^-$ (resp. $s_0^+ \cap s_0^\diamond$) is a union of two disjoint Jordan arcs $[p_9, p_{10}]$ and $[p_8, p_{11}]$ (resp. $[p_{10}, p_8]$ and $[p_{11}, p_9]$), and the ideal boundary of the ridge $s_0^+ \cap s_0^-$ (resp. $s_0^+ \cap s_0^*$) is a union of two disjoint Jordan arcs $[p_5, p_7]$ and $[p_4, p_6]$ (resp. $[p_7, p_4]$ and $[p_6, p_5]$). Since p_2 is the intersection of four isometric spheres $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^* \cap \mathcal{I}_{-1}^*$, it lies on $[p_9, p_{10}]$ and $[p_6, p_5]$. Similarly, q_3 lies on $[p_5, p_7]$ and $[p_{10}, p_8]$.

By Proposition 4.2, the antiholomorphic involution τ preserves \tilde{s}_0^+ and interchanges its boundaries. It is easy to check that τ interchanges p_2 and q_3 , p_5 and p_{10} , p_4 and p_{11} , p_6 and p_8 , and p_7 and p_9 . Thus one part of \tilde{s}_0^+ is a quadrilateral with vertices p_2, p_5, q_3 and p_{10} , denoted by \mathcal{O}_0^+ . The other is an octagon with vertices $p_2, p_6, p_4, p_7, q_3, p_8, p_{11}$ and p_9 , denoted by \mathcal{Q}_0^+ . Both of them are preserved by τ . □

According to the symmetry I_2 in Proposition 4.2, which interchanges \mathcal{I}_0^+ and \mathcal{I}_0^- :

Proposition 5.4 *The interior of the side \tilde{s}_0^- has connected components*

- an octagon, denoted by \mathcal{O}_0^- , with vertices $p_5, p_6, p_4, p_3, p_{15}, p_{13}, p_{14}$ and q_3 ,
- a quadrilateral, denoted by \mathcal{Q}_0^- , with vertices p_3, p_7, q_3 and p_{12} .

Proof Note that side s_0^- is bounded by the ridges $s_0^- \cap s_1^+, s_0^- \cap s_1^\diamond, s_0^- \cap s_0^+$ and $s_0^- \cap s_0^*$. By Proposition 4.2, the side s_0^- is isometric to s_0^+ under the complex involution I_2 . Thus its ideal boundary \tilde{s}_0^- will be also isometric to \tilde{s}_0^+ . This implies that \tilde{s}_0^- has the same combinatorial properties as \tilde{s}_0^+ . One can check that

$$I_2: (q_3, p_5, p_2, p_{10}, p_8, p_{11}, p_9, p_6) \leftrightarrow (q_3, p_7, p_3, p_{12}, p_{14}, p_{13}, p_{15}, p_4).$$

Thus one part of \tilde{s}_0^- is an octagon, denoted by \mathcal{O}_0^- , whose vertices are $p_5, p_6, p_4, p_3, p_{15}, p_{13}, p_{14}$ and q_3 . The other is a quadrilateral, denoted by \mathcal{Q}_0^- , whose vertices are p_3, p_7, q_3 and p_{12} ; see Figure 9. □

Remark 5.5 The vertex q_3 lies on the \mathbb{C} -circle associated to I_2 , that is, the ideal boundary of the complex line fixed by I_2 . One can also observe that p_2 is fixed by I_1 .

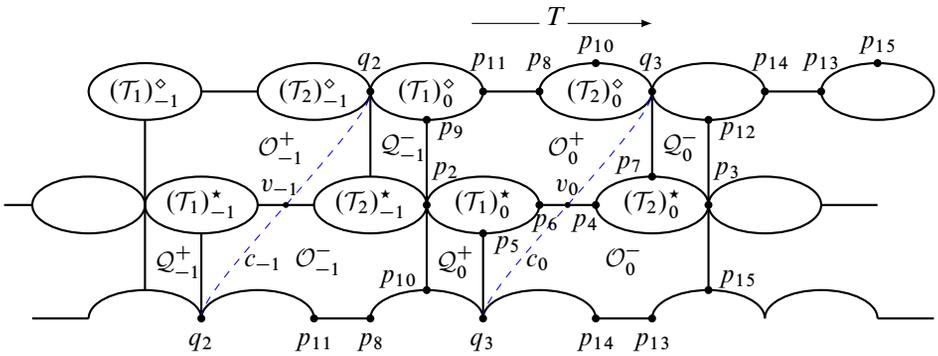


Figure 9: A combinatorial picture of ∂U . The top and bottom curves are identified. \mathcal{O}_0^{\pm} (resp. \mathcal{O}_{-1}^{\pm}) is divided by c_0 (resp. c_{-1}) into a quadrilateral \mathcal{Q}_0^{\pm} (resp. \mathcal{Q}_{-1}^{\pm}) and a heptagon \mathcal{H}_0^{\pm} (resp. \mathcal{H}_{-1}^{\pm}), v_0 is the intersection of c_0 with the arc $[p_4, p_6]$ and v_{-1} is the intersection of c_{-1} with the arc $[T^{-1}(p_4), T^{-1}(p_6)]$.

Proposition 5.6 *The interior of side \tilde{s}_0^{\star} is a union of two disjoint triangles, denoted by $(\mathcal{T}_1)_0^{\star}$ and $(\mathcal{T}_2)_0^{\star}$, whose vertices are p_2, p_5 and p_6 and p_3, p_4 and p_7 , respectively.*

Proof By Proposition 4.16, the side \tilde{s}_0^{\star} is the union of two disjoint discs, which are bounded by the ideal boundary of the ridges $s_0^+ \cap s_0^{\star}$ and $s_0^- \cap s_0^{\star}$.

As stated in Proposition 5.2, the ideal boundary of $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_0^{\star}$ contains the four points p_4, p_5, p_6 and p_7 . Thus \tilde{s}_0^{\star} is the union of two disjoint bigons, one with vertices p_5 and p_6 , and the other with p_4 and p_7 . Proposition 5.2 also tells us that p_2 and p_3 lie on different components of the boundaries of the two bigons.

Therefore, both of the components are triangles, denoted by $(\mathcal{T}_1)_0^{\star}$ and $(\mathcal{T}_2)_0^{\star}$, whose vertices are p_2, p_5 and p_6 and p_3, p_4 and p_7 , respectively. □

According to the symmetry τ in Proposition 4.2, the side \tilde{s}_0^{\diamond} has the same topological properties as the side \tilde{s}_0^{\star} . Thus by a similar argument:

Proposition 5.7 *The interior of side \tilde{s}_0^{\diamond} is a union of two disjoint triangles, denoted by $(\mathcal{T}_1)_0^{\diamond}$ and $(\mathcal{T}_2)_0^{\diamond}$, whose vertices are q_2, p_9 and p_{11} , and q_3, p_8 and p_{10} , respectively.*

5.3 A fundamental domain for the subgroup $\langle T \rangle$

Proposition 5.8 *Let $L = \{[x + i\frac{\sqrt{3}}{2}, \sqrt{3}x] \in \mathcal{N} : x \in \mathbb{R}\}$. Then L is a T -invariant \mathbb{R} -circle. Furthermore, L is contained in the complement of D .*

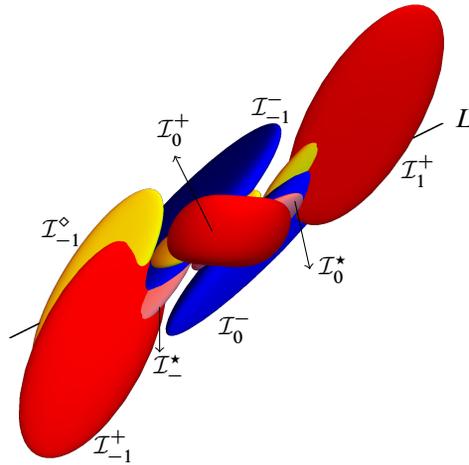


Figure 10: A realistic picture of the ideal boundaries of the isometric spheres. $\mathcal{I}_0^+, \mathcal{I}_1^+$ and \mathcal{I}_{-1}^+ (red), \mathcal{I}_0^- and \mathcal{I}_{-1}^- (blue), \mathcal{I}_0^* and \mathcal{I}_{-1}^* (pink), $\mathcal{I}_0^\circ, \mathcal{I}_{-1}^\circ$ and \mathcal{I}_1° (yellow). The line L is the T -invariant \mathbb{R} -circle.

Proof It is obvious that L is an \mathbb{R} -circle, since it is the image of the x -axis of \mathcal{N} by a Heisenberg translation along the y -axis. For any point $[x + i\frac{\sqrt{3}}{2}, \sqrt{3}x] \in L$, we have $T([x + i\frac{\sqrt{3}}{2}, \sqrt{3}x]) = [(x + 2) + i\frac{\sqrt{3}}{2}, \sqrt{3}(x + 2)]$, which lies in L . Thus L is a T -invariant \mathbb{R} -circle; see Figure 10.

Note that T acts on L as a translation through 2. To show L is contained in the complement of D , it suffices to show that a segment with length 2 is contained in the interior of some isometric spheres. By considering the Cygan distance between a point in L and the center of an isometric sphere, one can compute that the segments $\{[x + i\frac{\sqrt{3}}{2}, \sqrt{3}x] : -\frac{1}{2} \leq x \leq \frac{1}{2}\}$ and $\{[x + i\frac{\sqrt{3}}{2}, \sqrt{3}x] : \frac{1}{2} \leq x \leq \frac{3}{2}\}$ lie in the interiors of \mathcal{I}_0^+ and \mathcal{I}_0^- , respectively. □

Definition 5.9 Let

$$\Sigma_{-1} = \{[-\frac{3}{2} + iy, t] \in \mathcal{N} : y, t \in \mathbb{R}\} \quad \text{and} \quad \Sigma_0 = \{[\frac{1}{2} + iy, t] \in \mathcal{N} : y, t \in \mathbb{R}\}$$

be two planes in the Heisenberg group.

In fact, the vertical planes Σ_{-1} and Σ_0 are boundaries of fans in the sense of [14]. Let D_T be the region between Σ_{-1} and Σ_0 , that is

$$D_T = \{[x + iy, t] \in \mathcal{N} : -\frac{3}{2} \leq x \leq \frac{1}{2}\}.$$

It is clear that $\Sigma_0 = T(\Sigma_{-1})$. Thus D_T is a fundamental domain for $\langle T \rangle$ acting on $\partial\mathbb{H}_{\mathbb{C}}^2$.

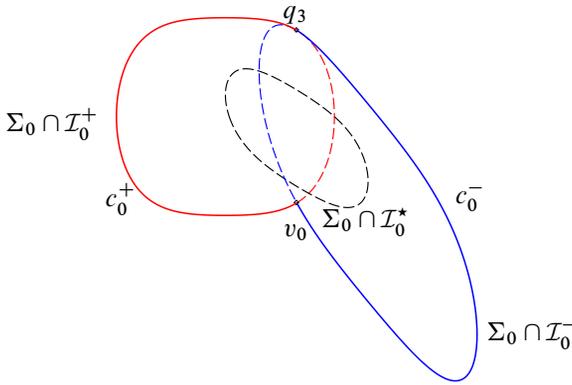


Figure 11: The intersection of Σ_0 with \mathcal{I}_0^+ , \mathcal{I}_0^- and \mathcal{I}_0^* viewed in Σ_0 . Here $c_0 = c_0^+ \cup c_0^-$ is a simple closed curve, where c_0^+ and c_0^- are the solid line parts of $\Sigma_0 \cap \mathcal{I}_0^+$ and $\Sigma_0 \cap \mathcal{I}_0^-$, respectively.

Proposition 5.10 *The intersections of Σ_0 and Σ_{-1} with the isometric spheres \mathcal{I}_k^\pm , \mathcal{I}_k^* and \mathcal{I}_k^\diamond are empty, except:*

- $\Sigma_0 \cap \mathcal{I}_0^\pm$ and $\Sigma_0 \cap \mathcal{I}_0^*$ are circles, and $\Sigma_0 \cap \mathcal{I}_0^\diamond = \Sigma_0 \cap \mathcal{I}_1^\diamond = \{q_3\}$.
- $\Sigma_{-1} \cap \mathcal{I}_{-1}^\pm$ and $\Sigma_{-1} \cap \mathcal{I}_{-1}^*$ are circles, and $\Sigma_{-1} \cap \mathcal{I}_{-1}^\diamond = \Sigma_{-1} \cap \mathcal{I}_0^\diamond = \{q_2\}$.

Proof Since the isometric spheres are strictly convex, their intersections with a plane are either a topological circle, a point or empty. Note that $\Sigma_0 = T(\Sigma_{-1})$. Thus it suffices to consider the intersections of Σ_0 with the isometric spheres. By a strait computation, each one of $\Sigma_0 \cap \mathcal{I}_0^\pm$ and $\Sigma_0 \cap \mathcal{I}_0^*$ is a circle; see Figure 11. □

Lemma 5.11 *The plane Σ_0 (resp. Σ_{-1}) is preserved by I_2 (resp. $T^{-1}I_2T$). The intersection $\Sigma_0 \cap \partial U$ (resp. $\Sigma_{-1} \cap \partial U$) is a simple closed curve c_0 (resp. c_{-1}) in the union $\tilde{s}_0^+ \cup \tilde{s}_0^-$ (resp. $\tilde{s}_{-1}^+ \cup \tilde{s}_{-1}^-$), which contains the points q_3 and $v_0 = [\frac{1}{2} + i\frac{\sqrt{3}}{2}, -\sqrt{3}]$ (resp. q_2 and $v_{-1} = T^{-1}(v_0)$).*

Proof It suffices to consider Σ_0 . The \mathbb{C} -circle associated to I_2 , that is, the ideal boundary of the complex line fixed by I_2 , is $\{[\frac{1}{2} + i\frac{\sqrt{3}}{2}, t] \in \mathcal{N} : t \in \mathbb{R}\}$, which is contained in Σ_0 . Thus Σ_0 is preserved by I_2 .

It is obvious that Σ_0 contains q_3 , which is the tangent point of \mathcal{I}_0^\diamond and \mathcal{I}_1^\diamond . The intersections $\Sigma_0 \cap \mathcal{I}_0^+$, $\Sigma_0 \cap \mathcal{I}_0^-$ and $\Sigma_0 \cap \mathcal{I}_0^*$ are circles by Proposition 5.10. One can compute that the intersection $\Sigma_0 \cap \mathcal{I}_0^+ \cap \mathcal{I}_0^-$ contains q_3 , and $v_0 = [\frac{1}{2} + i\frac{\sqrt{3}}{2}, -\sqrt{3}]$; see Figure 11. These two points divide the circles on \mathcal{I}_0^+ and \mathcal{I}_0^- into two arcs. Let c_0^+

be the arc with endpoints q_3 and v_0 on \mathcal{I}_0^+ lying in the exterior of \mathcal{I}_0^- and c_0^- be the one on \mathcal{I}_0^- lying in the exterior of \mathcal{I}_0^+ . Then $c_0 = c_0^+ \cup c_0^-$ is a simple closed curve. Observe that $\Sigma_0 \cap \mathcal{I}_0^*$ lies in the union of the interiors of \mathcal{I}_0^+ and \mathcal{I}_0^- . Thus Σ_0 does not intersect \tilde{s}_0^* . By Proposition 5.10, c_0^+ lies on \tilde{s}_0^+ and c_0^- lies on \tilde{s}_0^- . Therefore, the intersection $\Sigma_0 \cap \partial U$ is c_0 , which is a simple closed curve containing q_3 and v_0 . \square

5.4 The topology of U

Proposition 5.12 *Let U^c be the closure of the complement of U in \mathcal{N} . Then the closure of the intersection $U^c \cap D_T$ is a solid tube homeomorphic to a 3–ball.*

Proof It suffices to show that the boundary of $U^c \cap D_T$ is a 2–sphere. Now let us consider the cell structure of $U^c \cap D_T$; see Figure 9. According to Lemma 5.11, the intersection of U^c with Σ_0 (resp. Σ_{-1}) is a topological disc with two vertices, q_3 and v_0 (resp. q_2 and v_{-1}), and two edges, c_0^\pm (resp. c_{-1}^\pm). Also, c_0 (resp. c_{-1}) divides \mathcal{O}_0^\pm (resp. \mathcal{O}_{-1}^\pm) into a quadrilateral \mathcal{Q}'_0^\pm (resp. \mathcal{Q}'_{-1}^\pm) and a heptagon \mathcal{H}_0^\pm (resp. \mathcal{H}_{-1}^\pm).

Since p_2, p_5 and $T^{-1}(p_4)$ are contained in D_T , one can see that D_T contains $\mathcal{Q}'_0^-, \mathcal{Q}'_{-1}^+, \mathcal{H}_0^+$ and \mathcal{H}_{-1}^- . Besides, D_T contains $\mathcal{Q}_0^+, \mathcal{Q}_{-1}^-, (\mathcal{T}_1)_0^\diamond, (\mathcal{T}_2)_0^\diamond, (\mathcal{T}_1)_0^*$ and $(\mathcal{T}_2)_{-1}^*$. Thus the boundary of $U^c \cap D_T$ consists of 12 faces, 23 edges and 13 vertices; see the region between c_0 and c_{-1} in Figure 9. Therefore the Euler characteristic of the boundary of $U^c \cap D_T$ is 2. So the boundary of $U^c \cap D_T$ is a 2–sphere. \square

Propositions 5.8 and 5.12 imply the following result:

Proposition 5.13 *$U \cap D_T$ is the product of an unknotted cylinder with a ray, which is homeomorphic to $S^1 \times [0, 1] \times \mathbb{R}_{\geq 0}$.*

Proof As stated in Proposition 5.8, U^c contains the line L . Thus $U^c \cap D_T$ is a tubular neighborhood of $L \cap D_T$. It cannot be knotted. Therefore $\partial U \cap D_T$ is an unknotted cylinder homeomorphic to $S^1 \times [0, 1]$. One can see that $U \cap \Sigma_0$ is the product of c_0 with a ray, and $U \cap \Sigma_{-1}$ is the product of c_{-1} with a ray. Both of them are homeomorphic to $S^1 \times \mathbb{R}_{\geq 0}$. Hence $U \cap D_T$ is the product of an unknotted cylinder with a ray, and is homeomorphic to $S^1 \times [0, 1] \times \mathbb{R}_{\geq 0}$. \square

Applying powers of T , Proposition 5.13 immediately implies the following corollary:

Corollary 5.14 *U is the product of an unknotted cylinder with a ray homeomorphic to $S^1 \times \mathbb{R} \times \mathbb{R}_{\geq 0}$.*

Remark 5.15 U is the complement of a tubular neighborhood of the T -invariant \mathbb{R} -circle L , that is, a horotube for T . (See [28] for the definition of a horotube.)

5.5 The manifold

Definition 5.16 Suppose that the cylinder $S^1 \times [0, 1]$ has a combinatorial cell structure with finite faces $\{F_i\}$. A *canonical subdivision* on $S^1 \times [0, 1] \times \mathbb{R}_{\geq 0}$ is a finite union of 3-dimensional pieces $\{\hat{F}_i\}$ where $\hat{F}_i = F_i \times \mathbb{R}_{\geq 0}$.

Proposition 5.17 *There is a canonical subdivision on $U \cap D_T$.*

Proof As described in the proof of Proposition 5.12, the combinatorial cell structure of $\partial U \cap D_T$ has 10 faces, $\mathcal{Q}'_0, \mathcal{Q}'_{-1}, \mathcal{H}_0^+, \mathcal{H}_{-1}^+, \mathcal{Q}_0^+, \mathcal{Q}_{-1}^+, (\mathcal{T}_1)_0^\diamond, (\mathcal{T}_2)_0^\diamond, (\mathcal{T}_1)_0^\star$ and $(\mathcal{T}_2)_{-1}^\star$. By Proposition 5.13, $U \cap D_T$ is the union of 3-dimensional pieces $\hat{\mathcal{Q}}_0^-, \hat{\mathcal{Q}}_{-1}^+, \hat{\mathcal{H}}_0^+, \hat{\mathcal{H}}_{-1}^+, \hat{\mathcal{Q}}_0^+, \hat{\mathcal{Q}}_{-1}^+, (\hat{\mathcal{T}}_1)_0^\diamond, (\hat{\mathcal{T}}_2)_0^\diamond, (\hat{\mathcal{T}}_1)_0^\star$ and $(\hat{\mathcal{T}}_2)_{-1}^\star$. Combinatorially, these 3-dimensional pieces are the cone from q_∞ to the faces of $\partial U \cap D_T$. □

Let $\Omega(\Gamma)$ be the discontinuity region of Γ acting on $\partial\mathbb{H}_\mathbb{C}^2$. Then $U \cap D_T$ is obviously a fundamental domain for Γ . By cutting and gluing, we can obtain the following fundamental domain for Γ acting on $\Omega(\Gamma)$:

Proposition 5.18 *Let \mathcal{P}_+ be the union of $\hat{\mathcal{H}}_0^+, (\hat{\mathcal{T}}_1)_0^\diamond, (\hat{\mathcal{T}}_2)_0^\diamond, T(\hat{\mathcal{Q}}_{-1}^+), T(\hat{\mathcal{Q}}_{-1}^-)$ and $T((\hat{\mathcal{T}}_2)_{-1}^\star)$. Let \mathcal{P}_- be the union of $\hat{\mathcal{Q}}_0^+, (\hat{\mathcal{T}}_1)_0^\star, \hat{\mathcal{Q}}_0^-$ and $T(\hat{\mathcal{H}}_{-1}^-)$. Then $\mathcal{P}_+ \cup \mathcal{P}_-$ is a fundamental domain for Γ acting on $\Omega(\Gamma)$. Moreover, \mathcal{P}_+ (resp. \mathcal{P}_-) is combinatorially an eleven-pyramid (resp. nine-pyramid) with cone vertex q_∞ and base*

$$\mathcal{O}_0^+ \cup \mathcal{Q}_0^- \cup (\mathcal{T}_1)_0^\diamond \cup (\mathcal{T}_2)_0^\diamond \cup (\mathcal{T}_2)_0^\star \quad (\text{resp. } \mathcal{O}_0^- \cup \mathcal{Q}_0^+ \cup (\mathcal{T}_1)_0^\star).$$

Proof Since $\Sigma_0 = T(\Sigma_{-1})$ and $c_0 = T(c_{-1})$, $U \cap D_T$ and $T(U \cap D_T)$ can be glued together along $c_0 \times \mathbb{R}_{\geq 0}$. Note that $U \cap D_T$ is a fundamental domain for Γ acting on $\Omega(\Gamma)$ and has a subdivision as described in Proposition 5.17. Therefore $\mathcal{P}_+ \cup \mathcal{P}_-$ is also a fundamental domain.

As described in Proposition 5.12, c_0 (resp. c_{-1}) divides \mathcal{O}_0^\pm (resp. \mathcal{O}_{-1}^\pm) into a quadrilateral \mathcal{Q}'_0^\pm (resp. \mathcal{Q}'_{-1}^\pm) and a heptagon \mathcal{H}_0^\pm (resp. \mathcal{H}_{-1}^\pm). Note that $\mathcal{O}_0^\pm = T(\mathcal{O}_{-1}^\pm)$ and $(\mathcal{T}_2)_0^\star = T((\mathcal{T}_2)_{-1}^\star)$. Thus the base of \mathcal{P}_+ (resp. \mathcal{P}_-) is $\mathcal{O}_0^+ \cup \mathcal{Q}_0^- \cup (\mathcal{T}_1)_0^\diamond \cup (\mathcal{T}_2)_0^\diamond \cup (\mathcal{T}_2)_0^\star$ (resp. $\mathcal{O}_0^- \cup \mathcal{Q}_0^+ \cup (\mathcal{T}_1)_0^\star$), which is combinatorially a hendecagon (resp. an enneagon); see Figures 9 and 12. □

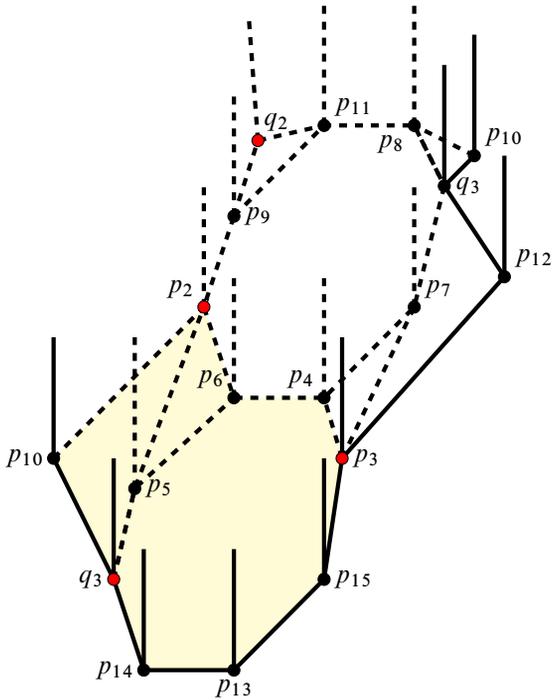


Figure 12: A schematic view of the fundamental domain of Γ on $\Omega(\Gamma)$. The red vertices are the parabolic fixed points. The yellow polygon is $\mathcal{O}_0^- \cup \mathcal{Q}_0^+ \cup (\mathcal{T}_1)_0^*$.

Definition 5.19 Let $p'_2 = S^{-1}(p_2)$, $p_1 = S^{-1}(q_\infty) = [0, 0]$ and $p'_{10} = S^{-1}(p_{10})$.

Lemma 5.20 Let \mathcal{P} be the union $\mathcal{P}_+ \cup S^{-1}(\mathcal{P}_{-1})$. Then \mathcal{P} is combinatorially a polyhedron with 8 triangular faces, 4 square faces, 2 pentagonal faces and 2 hexagonal faces. The faces of \mathcal{P} are paired as

$$\begin{aligned}
 T &: (q_\infty, p_2, p_9, q_2) \mapsto (q_\infty, p_3, p_{12}, q_3), \\
 S^{-1}T &: (q_\infty, q_2, p_{11}, p_8, p_{10}) \mapsto (p_1, p_2, p_9, p_{11}, p_8), \\
 (S^{-1}T)^2 &: (q_2, p_9, p_{11}) \mapsto (q_3, p_8, p_{10}), \\
 S^{-1} &: (q_\infty, p_{10}, q_3) \mapsto (p_1, p'_{10}, p_2), \\
 S^{-1}T^{-1}S &: (p_1, p_8, q_3) \mapsto (p_1, p'_{10}, p'_2), \\
 S^{-2} &: (q_3, p_{12}, p_3, p_7) \mapsto (p'_2, p'_{10}, p_2, p_6), \\
 S^{-1} &: (q_\infty, p_2, p_6, p'_2, p_4, p_3) \mapsto (p_1, p'_2, p_4, p_3, p_7, q_3), \\
 S^{-1} &: (p_6, p'_2, p_4) \mapsto (p_4, p_3, p_7).
 \end{aligned}$$

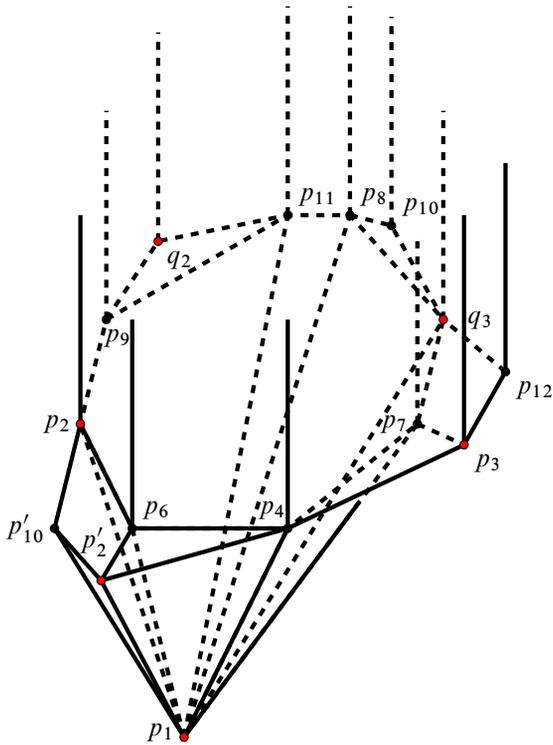


Figure 13: The combinatorial picture of \mathcal{P} . The red vertices of \mathcal{P} are the parabolic fixed points.

Proof The bases of \mathcal{P}_+ and \mathcal{P}_{-1} are paired as

$$\begin{aligned} S: \mathcal{Q}_0^+ &\rightarrow \mathcal{Q}_0^-, & (p_2, p_5, q_3, p_{10}) &\mapsto (q_3, p_7, p_3, p_{12}), \\ S^2: (\mathcal{T}_1)_0^* &\rightarrow (\mathcal{T}_2)_0^*, & (p_2, p_5, p_6) &\mapsto (p_3, p_4, p_7), \\ (S^{-1}T)^2: (\mathcal{T}_1)_0^\diamond &\rightarrow (\mathcal{T}_2)_0^\diamond, & (q_2, p_9, p_{11}) &\mapsto (q_3, p_8, p_{10}), \end{aligned}$$

and

$$S^{-1}: \mathcal{O}_0^- \rightarrow \mathcal{O}_0^+, \quad (p_3, p_4, p_6, p_5, q_3, q_{14}, p_{13}, p_{15}) \mapsto (q_3, p_7, p_4, p_6, p_2, p_9, p_{11}, p_8).$$

Thus $S^{-1}(\mathcal{P}_{-1})$ and \mathcal{P}_+ are glued along \mathcal{O}_0^+ . According to Lemma 2.9, $S^{-1}(\mathcal{P}_{-1})$ lies in the interior of \mathcal{I}_0^+ , since \mathcal{P}_{-1} lies in the exterior of \mathcal{I}_0^+ . Moreover, $p_1 = S^{-1}(q_\infty) = [0, 0]$ is the center of the isometric sphere \mathcal{I}_0^+ ; see Figure 13. \square

Proposition 5.21 Let Ω be the discontinuity region of Γ acting on $\mathbb{H}_{\mathbb{C}}^2$. Then the fundamental group of Ω/Γ has a presentation

$$\langle u, v, w \mid w^{-1}vuw^{-1}v^{-1}wu = v^2wuw^{-3}u = \text{id} \rangle.$$

Proof Let x_i for $i = 1, 2, 3, 4, 5, 6, 7, 8$ be the corresponding gluing maps of \mathcal{P} given in Lemma 5.20. These are the generators of the fundamental group of Ω/Γ .

By considering the edge cycles of \mathcal{P} under the gluing maps, we have the relations

$$\begin{aligned} x_7^{-1} \cdot x_5 \cdot x_7 \cdot x_1 &= \text{id}, & x_2^{-1} \cdot x_4 \cdot x_1 &= \text{id}, & x_2 \cdot x_3^{-1} \cdot x_4^{-1} \cdot x_6 \cdot x_1 &= \text{id}, \\ x_3^{-1} \cdot x_5^{-1} \cdot x_6 \cdot x_1 &= \text{id}, & x_2 \cdot x_3 \cdot x_2 &= \text{id}, & x_4^{-1} \cdot x_5 \cdot x_2 &= \text{id}, \\ x_7 \cdot x_8 \cdot x_6 &= \text{id}, & x_8 \cdot x_7 \cdot x_6 &= \text{id}, & x_8^{-1} \cdot x_7 &= \text{id}. \end{aligned}$$

For example, the edge cycle of $[q_\infty, p_2]$ is

$$[q_\infty, p_2] \xrightarrow{x_1} [q_\infty, p_3] \xrightarrow{x_7} [p_1, q_3] \xrightarrow{x_5} [p_1, p'_2] \xrightarrow{x_7^{-1}} [q_\infty, p_2].$$

Thus

$$x_7^{-1} \cdot x_5 \cdot x_7 \cdot x_1 = \text{id}.$$

This is the first relation. The others can be given by a similar argument.

Simplifying the relations and setting $u = x_1$, $v = x_2$ and $w = x_7$, we obtain the presentation of the fundamental group of Ω/Γ . □

Now we are ready to show the following theorem:

Theorem 5.22 *Let Ω be the discontinuity region of Γ acting on $\mathbb{H}^2_{\mathbb{C}}$. Then the quotient space Ω/Γ is homeomorphic to the two-cusped hyperbolic 3–manifold $s782$ in the SnapPy census.*

Proof Let $M = \Omega/\Gamma$. According to Proposition 5.21, the fundamental group of M has a presentation

$$\pi_1(M) = \langle u, v, w \mid w^{-1}vu^{-1}v^{-1}wu = v^2wuw^{-3}u = \text{id} \rangle.$$

The manifold $s782$ is a two-cusped hyperbolic 3–manifold with finite volume. Its fundamental group, provided by SnapPy, has a presentation

$$\pi_1(s782) = \langle a, b, c \mid a^2cb^4c = abca^{-1}b^{-1}c^{-1} = \text{id} \rangle.$$

Using Magma, we get an isomorphism $\Psi: \pi_1(M) \rightarrow \pi_1(s782)$ given by

$$\Psi(u) = c^{-1}b^{-1}, \quad \Psi(v) = b^{-1} \quad \text{and} \quad \Psi(w) = a.$$

Therefore M will be the connected sum of $s782$ and a closed 3–manifold with trivial fundamental group by the prime decompositions of 3–manifolds [15]. The solution of the Poincaré conjecture implies that the closed 3–manifold is the 3–sphere. Thus M is homeomorphic to $s782$. □

References

- [1] **M Acosta**, *Spherical CR Dehn surgeries*, Pacific J. Math. 284 (2016) 257–282 [MR](#) [Zbl](#)
- [2] **M Acosta**, *Spherical CR uniformization of Dehn surgeries of the Whitehead link complement*, Geom. Topol. 23 (2019) 2593–2664 [MR](#) [Zbl](#)
- [3] **A F Beardon**, *The geometry of discrete groups*, Graduate Texts in Math. 91, Springer (1983) [MR](#) [Zbl](#)
- [4] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy, a computer program for studying the geometry and topology of 3-manifolds* (2016) Available at <http://snappy.computop.org>
- [5] **M Deraux**, *On spherical CR uniformization of 3-manifolds*, Exp. Math. 24 (2015) 355–370 [MR](#) [Zbl](#)
- [6] **M Deraux**, *A 1-parameter family of spherical CR uniformizations of the figure eight knot complement*, Geom. Topol. 20 (2016) 3571–3621 [MR](#) [Zbl](#)
- [7] **M Deraux**, **E Falbel**, *Complex hyperbolic geometry of the figure-eight knot*, Geom. Topol. 19 (2015) 237–293 [MR](#) [Zbl](#)
- [8] **M Deraux**, **J R Parker**, **J Paupert**, *New non-arithmetic complex hyperbolic lattices*, Invent. Math. 203 (2016) 681–771 [MR](#) [Zbl](#)
- [9] **E Falbel**, *A spherical CR structure on the complement of the figure eight knot with discrete holonomy*, J. Differential Geom. 79 (2008) 69–110 [MR](#) [Zbl](#)
- [10] **E Falbel**, **P-V Koseleff**, **F Rouillier**, *Representations of fundamental groups of 3-manifolds into $\mathrm{PGL}(3, \mathbb{C})$: exact computations in low complexity*, Geom. Dedicata 177 (2015) 229–255 [MR](#) [Zbl](#)
- [11] **E Falbel**, **J Wang**, *Branched spherical CR structures on the complement of the figure-eight knot*, Michigan Math. J. 63 (2014) 635–667 [MR](#) [Zbl](#)
- [12] **W M Goldman**, *Complex hyperbolic geometry*, Clarendon, New York (1999) [MR](#) [Zbl](#)
- [13] **W M Goldman**, **J R Parker**, *Complex hyperbolic ideal triangle groups*, J. Reine Angew. Math. 425 (1992) 71–86 [MR](#) [Zbl](#)
- [14] **W M Goldman**, **J R Parker**, *Dirichlet polyhedra for dihedral groups acting on complex hyperbolic space*, J. Geom. Anal. 2 (1992) 517–554 [MR](#) [Zbl](#)
- [15] **J Hempel**, *3-Manifolds*, Ann. of Math. Stud. 86, Princeton Univ. Press (1976) [MR](#) [Zbl](#)
- [16] **J Ma**, **B Xie**, *Spherical CR uniformization of the magic 3-manifold*, preprint (2021) [arXiv 2106.06668](https://arxiv.org/abs/2106.06668)
- [17] **J Ma**, **B Xie**, *Menger curve and spherical CR uniformization of a closed hyperbolic 3-orbifold*, preprint (2022) [arXiv 2201.04765](https://arxiv.org/abs/2201.04765)

- [18] **J Ma, B Xie**, *Three-manifolds at infinity of complex hyperbolic orbifolds*, preprint (2022) [arXiv 2205.11167](https://arxiv.org/abs/2205.11167)
- [19] **G D Mostow**, *On a remarkable class of polyhedra in complex hyperbolic space*, Pacific J. Math. 86 (1980) 171–276 [MR](#) [Zbl](#)
- [20] **J R Parker**, *Notes on complex hyperbolic geometry*, University of Durham (2003) Available at <https://maths.dur.ac.uk/users/j.r.parker/img/NCHG.pdf>
- [21] **J R Parker, J Wang, B Xie**, *Complex hyperbolic $(3, 3, n)$ triangle groups*, Pacific J. Math. 280 (2016) 433–453 [MR](#) [Zbl](#)
- [22] **J R Parker, P Will**, *A complex hyperbolic Riley slice*, Geom. Topol. 21 (2017) 3391–3451 [MR](#) [Zbl](#)
- [23] **RE Schwartz**, *Degenerating the complex hyperbolic ideal triangle groups*, Acta Math. 186 (2001) 105–154 [MR](#) [Zbl](#)
- [24] **RE Schwartz**, *Ideal triangle groups, dented tori, and numerical analysis*, Ann. of Math. 153 (2001) 533–598 [MR](#) [Zbl](#)
- [25] **RE Schwartz**, *Complex hyperbolic triangle groups*, from “Proceedings of the International Congress of Mathematicians, II” (T Li, editor), Higher Ed. (2002) 339–349 [MR](#) [Zbl](#)
- [26] **RE Schwartz**, *Real hyperbolic on the outside, complex hyperbolic on the inside*, Invent. Math. 151 (2003) 221–295 [MR](#) [Zbl](#)
- [27] **RE Schwartz**, *A better proof of the Goldman–Parker conjecture*, Geom. Topol. 9 (2005) 1539–1601 [MR](#) [Zbl](#)
- [28] **RE Schwartz**, *Spherical CR geometry and Dehn surgery*, Ann. of Math. Stud. 165, Princeton Univ. Press (2007) [MR](#) [Zbl](#)
- [29] **JO Wyss-Gallifent**, *Complex hyperbolic triangle groups*, PhD thesis, University of Maryland (2000) [MR](#) Available at <https://www.proquest.com/docview/304626210>

*School of Mathematics, Hunan University
Changsha, China*

*School of Mathematics, Hunan University
Changsha, China*

*School of Mathematics, Hunan University
Changsha, China*

ypjiang@hnu.edu.cn, jywang@hnu.edu.cn, xiexbh@hnu.edu.cn

Received: 19 April 2021 Revised: 27 February 2022

A connection between cut locus, Thom space and Morse–Bott functions

SOMNATH BASU
SACHCHIDANAND PRASAD

Associated to every closed, embedded submanifold N in a connected Riemannian manifold M , there is the distance function d_N which measures the distance of a point in M from N . We analyze the square of this function and show that it is Morse–Bott on the complement of the cut locus $\text{Cu}(N)$ of N provided M is complete. Moreover, the gradient flow lines provide a deformation retraction of $M - \text{Cu}(N)$ to N . If M is a closed manifold, then we prove that the Thom space of the normal bundle of N is homeomorphic to $M/\text{Cu}(N)$. We also discuss several interesting results which are either applications of these or related observations regarding the theory of cut locus. These results include, but are not limited to, a computation of the local homology of singular matrices, a classification of the homotopy type of the cut locus of a homology sphere inside a sphere, a deformation of the indefinite unitary group $U(p, q)$ to $U(p) \times U(q)$ and a geometric deformation of $\text{GL}(n, \mathbb{R})$ to $O(n, \mathbb{R})$ which is different from the Gram–Schmidt retraction.

[53B21](#), [53C22](#), [55P10](#); [32B20](#), [57R19](#), [58C05](#)

1. Introduction	4186
2. Preliminaries	4189
3. Main results	4198
4. Applications to Lie groups	4218
Appendix A. The continuity of the map (3-2)	4227
Appendix B. Derivative of the square root map	4229
References	4232

1 Introduction

On a Riemannian manifold M , the distance function $d_N(\cdot) := d(N, \cdot)$ from a closed subset N is fundamental in the study of variational problems. For instance, the viscosity solution of the Hamilton–Jacobi equation is given by the flow of the gradient vector of the distance function d_N when N is the smooth boundary of a relatively compact domain in manifolds; see Li and Nirenberg [17] and Mantegazza and Mennucci [18]. Although the distance function d_N is not differentiable at N , squaring the function removes this issue. Associated to N and the distance function d_N is a set $\text{Cu}(N)$, the cut locus of N in M . The cut locus of a point, a notion initiated by Poincaré [23], has been extensively studied (see Kobayashi [16] for a survey as well as Buchner [4], Myers [20], Sakai [26] and Wolter [29]). There has been work on the structure of the cut locus of submanifolds. One may refer to the works of Hebda [9; 10], Sabau and Tanaka [25] and Singh [28]. Suitable simple examples indicate that $M - \text{Cu}(N)$ topologically deforms to N . One of our main results is the following:

Theorem A (Theorem 3.32) *Let N be a closed embedded submanifold of a complete Riemannian manifold M and $d_N: M \rightarrow \mathbb{R}$ denote the distance function with respect to N . If $f = d_N^2$, then its restriction to $M - \text{Cu}(N)$ is a Morse–Bott function, with N as the critical submanifold. Moreover, $M - \text{Cu}(N)$ deforms to N via the gradient flow of f .*

It is observed that this deformation takes infinite time. To obtain a strong deformation retract, one reparametrizes the flow lines to be defined over $[0, 1]$. It can be shown (Lemma 3.18) that the cut locus $\text{Cu}(N)$ is a strong deformation retract of $M - N$. A primary motivation for Theorem A came from understanding the cut locus of $N = O(n, \mathbb{R})$ inside $M = M(n, \mathbb{R})$, equipped with the Euclidean metric. We show in Section 2.2 that the cut locus is the set Sing of singular matrices and the deformation of its complement is not the Gram–Schmidt deformation but rather the deformation obtained from the polar decomposition, ie $A \in \text{GL}(n, \mathbb{R})$ deforms to $A\sqrt{A^T A}^{-1}$. Combining this with a result of Hebda [9, Theorem 1.4], we are able to compute the local homology of Sing (see Lemma 2.15 and Corollary 2.16).

Theorem B *For $A \in M(n, \mathbb{R})$,*

$$H_{n^2-1-i}(\text{Sing}, \text{Sing} - A) \cong \tilde{H}^i(O(n-k, \mathbb{R})),$$

where $A \in \text{Sing}$ has rank $k < n$.

When the cut locus is empty, we deduce that M is diffeomorphic to the normal bundle ν of N in M . In particular, M deforms to N . Among applications, we discuss two families of examples. We reprove the known fact that $GL(n, \mathbb{R})$ deforms to $O(n, \mathbb{R})$ for any choice of left-invariant metric on $GL(n, \mathbb{R})$ which is right- $O(n, \mathbb{R})$ -invariant. However, this deformation is not obtained topologically but by Morse–Bott flows. For a natural choice of such a metric, this deformation (4-2) is not the Gram–Schmidt deformation but one obtained from the polar decomposition. We also consider $U(p, q)$, the group preserving the indefinite form of signature (p, q) on \mathbb{C}^n . We show (Theorem 4.6) that $U(p, q)$ deforms to $U(p) \times U(q)$ for the left-invariant metric given by $\langle X, Y \rangle := \text{tr}(X^*Y)$. In particular, we show that the exponential map is surjective for $U(p, q)$ (Corollary 4.8). To our knowledge, this method is different from the standard proof.

For a Riemannian manifold we have the exponential map at $p \in M$, $\exp_p : T_p M \rightarrow M$. Let ν denote the normal bundle of N in M . We will modify the exponential map (see Section 3.2) to define the *rescaled exponential* $\widetilde{\exp} : D(\nu) \rightarrow M$, the domain of which is the unit disk bundle of ν . The main result (Theorem 3.16) here is the observation that there is a connection between the cut locus $\text{Cu}(N)$ and Thom space $\text{Th}(\nu) := D(\nu)/S(\nu)$ of ν .

Theorem C *Let N be an embedded submanifold inside a closed, connected Riemannian manifold M . If ν denotes the normal bundle of N in M , then there is a homeomorphism*

$$\widetilde{\exp} : D(\nu)/S(\nu) \xrightarrow{\cong} M/\text{Cu}(N).$$

This immediately leads to a long exact sequence in homology (see (3-6))

$$\dots \rightarrow H_j(\text{Cu}(N)) \xrightarrow{i_*} H_j(M) \xrightarrow{q} \widetilde{H}_j(\text{Th}(\nu)) \xrightarrow{\partial} H_{j-1}(\text{Cu}(N)) \rightarrow \dots .$$

This is a useful tool in characterizing the homotopy type of the cut locus. We list a few applications and related results.

Theorem D *Let N be a homology k -sphere embedded in a Riemannian manifold M^d homeomorphic to S^d .*

- (1) *If $d \geq k + 3$, then $\text{Cu}(N)$ is homotopy equivalent to S^{d-k-1} . Moreover, if M and N are real analytic and the embedding is real analytic, then $\text{Cu}(N)$ is a simplicial complex of dimension at most $d - 1$.*
- (2) *If $d = k + 2$, then $\text{Cu}(N)$ has the homology of S^1 . There exist homology 3-spheres in S^5 for which $\text{Cu}(N) \simeq S^1$. However, for nontrivial knots K in S^3 , the cut locus is not homotopy equivalent to S^1 .*

The above results are a combination of Theorems 3.24 and 3.9 and Example 3.29. In general, the structure of the cut locus may be wild (see Gluck and Singer [7], Itoh and Sabau [13] and Itoh and Vîlcu [15]). Myers [20] had shown that, if M is a real analytic sphere, then $\text{Cu}(p)$ is a finite tree each of whose edges is an analytic curve with finite length. Buchner [4] later generalized this result to the cut locus of a point in higher-dimensional manifolds. Theorem 3.9, which states that the cut locus of an analytic submanifold (in an analytic manifold) is a simplicial complex, is a natural generalization of Buchner's result (and its proof). We attribute it to Buchner although it is not present in the original paper. This analyticity assumption also helps us to compute the homotopy type of the cut locus of a finite set of points in any closed, orientable, real analytic surface of genus g (Theorem 3.27). In Example 3.29 we make some observations about the cut locus of embedded homology spheres of codimension 2. This includes the case of real analytic knots in the round sphere \mathbb{S}^3 .

We apply our study of gradient of distance-squared function to two families of Lie groups: $\text{GL}(n, \mathbb{R})$ and $U(p, q)$. With a particular choice of left-invariant Riemannian metric which is right-invariant with respect to a maximally compact subgroup K , we analyze the geodesics and the cut locus of K . In both cases, we obtain that G deforms to K via Morse–Bott flow (Lemma 4.1 and Theorem 4.6). Although these results are deducible from classical results of Cartan and Iwasawa, our method is geometric and specific to suitable choices of Riemannian metrics. It also makes very little use of structure theory of Lie algebras.

Organization of the paper In Section 2 we first recall basic definitions of Morse–Bott functions and cut locus of a subset (see Section 2.1). In Section 2.2 we analyze the distance function from $O(n, \mathbb{R})$ in $M(n, \mathbb{R})$. This highlights and motivates Theorem A as well as allows for computation of local homology of singular matrices (Theorem B). In Section 3 we first recall some relevant basic definitions from geometry (see Section 3.1). We make some observations about the differentiability of the distance function (following Wolter [29]) and show that the cut locus is a simplicial complex for an analytic pair (following Buchner [4]). In Section 3.2 we prove Theorem C and discuss some applications, including Theorem D. In Section 3.3 we prove Theorem A. In Section 4 we discuss two specific examples: we analyze the cut locus of $O(n, \mathbb{R})$ inside $\text{GL}(n, \mathbb{R})$ in Section 4.1 and the cut locus of $U(p) \times U(q)$ inside $U(p, q)$ in Section 4.2. In Appendix A we prove Proposition 3.14, the continuity of the map s (see (3-2)). This result is crucial for Section 3.2. In Appendix B we compute the

derivative of the square root map for positive-definite matrices (Lemma B.1). We also analyze the differentiability of the map $A \mapsto \text{tr}(\sqrt{A^T A})$ in Lemma B.2.

Acknowledgements Basu acknowledges the support of the SERB MATRICS grant MTR/2017/000807. Prasad was supported by a UGC (NET)-JRF fellowship.

2 Preliminaries

We recall the notion of Morse function and Morse–Bott function in Section 2.1, keeping in mind the square of the distance function from a submanifold being a potential Morse–Bott function, which we will analyze in Section 3.3. We also recall the definition of cut locus of a subset in a Riemannian manifold. In Example 2.7 we observe that the join of spheres being a sphere can be observed geometrically via cut locus. In Section 2.2 we analyze the cut locus of orthogonal matrices and compute the relative homology of the cut locus (2-8). Along the way, we note that the geometric deformation of $\text{GL}(n, \mathbb{R})$ to $O(n, \mathbb{R})$, obtained via the distance-squared function, is *not* the Gram–Schmidt deformation.

2.1 Background

Given a smooth n -dimensional manifold M , we say that a point $p \in M$ is a *critical point* of a smooth function $f: M \rightarrow \mathbb{R}$ if

$$df_p: T_p M \rightarrow T_{f(p)} \mathbb{R}$$

vanishes. In a coordinate neighborhood $(\phi = (x_1, x_2, \dots, x_n), U)$ around p , for all $j = 1, 2, \dots, n$ we have

$$\frac{\partial(f \circ \phi^{-1})}{\partial x_j}(\phi(p)) = 0.$$

A critical point p is called *nondegenerate* if the determinant of the Hessian matrix

$$\text{Hess}_p(f) := \left(\frac{\partial^2(f \circ \phi^{-1})}{\partial x_i \partial x_j}(\phi(p)) \right)$$

is nonzero. Let us denote the set of all critical points of f by $\text{Cr}(f)$. If all the critical points are nondegenerate, then f is said to be a *Morse function*. Morse–Bott functions are generalizations of Morse functions, where we are allowed to have nondegenerate critical submanifolds.

Definition 2.1 (Morse–Bott functions) Let M be a Riemannian manifold. A smooth submanifold $N \subset M$ is said to be a *nondegenerate critical submanifold* of f if $N \subseteq \text{Cr}(f)$ and, for any $p \in N$, $\text{Hess}_p(f)$ is nondegenerate in the direction normal to N at p . The function f is said to be *Morse–Bott* if the connected components of $\text{Cr}(f)$ are nondegenerate critical submanifolds.

Note that “ $\text{Hess}_p(f)$ is nondegenerate in the direction normal to N at p ” means for any $V \in (T_p N)^\perp$ there exists $W \in (T_p N)^\perp$ such that $\text{Hess}_p(f)(V, W) \neq 0$.

Example 2.2 Let $M = \mathbb{R}^{n+1}$ equipped with the Euclidean metric d . If $N = \mathbb{S}^n$ is the unit sphere, then the distance between a point $p \in \mathbb{R}^{n+1}$ and N is given by

$$d(N, p) := \inf_{q \in N} d(q, p).$$

We shall denote by d^2 the square of the distance. Now consider the function

$$f: M \rightarrow \mathbb{R}, \quad x \mapsto d^2(N, x) = (\|x\| - 1)^2.$$

The function $f: M - \{0\}$ is a Morse–Bott function with $N = \mathbb{S}^n$ as the critical submanifold.

The trace function on $\text{SO}(n, \mathbb{R})$, $U(n, \mathbb{C})$ and $\text{Sp}(n, \mathbb{C})$ is a Morse–Bott function (see Banyaga and Hurtubise [1, Exercise 22, page 90]). We refer the interested reader to [1] for basic results on Morse–Bott theory.

We shall now define the cut locus for a point. The notion of cut locus was first introduced for convex surfaces by Poincaré [23] in 1905 under the name *la ligne de partage*, meaning *the dividing line*.

Definition 2.3 (cut locus) Let M be a complete Riemannian manifold and $p \in M$. If $\text{Cu}(p)$ denotes the *cut locus* of p , then a point q is in $\text{Cu}(p)$ if there exists a minimal geodesic joining p to q any extension of which beyond q is not minimal.

Recall that a minimal geodesic joining p and q is a geodesic that realizes the distance between p and q . The existence of minimal geodesics joining two given points is implied by completeness of the Riemannian manifold. Therefore, in almost all of the examples, the manifolds under consideration will be complete Riemannian manifolds. When $M = \mathbb{S}^n$, ie an n -sphere with the round metric induced from \mathbb{R}^{n+1} , for any $p \in \mathbb{S}^n$, the cut locus $\text{Cu}(p)$ will be the corresponding antipodal point. Later, in

Definition 3.11, we give a slightly different but equivalent definition of cut locus, following Sakai’s book [26, Section 4.1].

In order to have a definition of the cut locus for a submanifold (or a subset), we need to generalize the notion of a minimal geodesic.

Definition 2.4 A geodesic γ is called a *distance-minimal geodesic* joining N to p if there exists $q \in N$ such that γ is a minimal geodesic joining q to p and $\ell(\gamma) = d(N, p) =: d_N(p)$. We will refer to such geodesics as *N -geodesics*.

If N is an embedded submanifold, then an N -geodesic is necessarily orthogonal to N . This follows from the first variational principle. We are ready to define the cut locus for $N \subset M$.

Definition 2.5 (cut locus) Let M be a Riemannian manifold and N be any nonempty subset of M . If $\text{Cu}(N)$ denotes the *cut locus of N* , then we say that $q \in \text{Cu}(N)$ if and only if there exists a distance-minimal geodesic joining N to q such that any extension of it beyond q is not a distance-minimal geodesic.

The cut locus of a sphere (see [Example 2.2](#)) is its center. The set $\text{Cu}(p)$ is closed; see Postnikov [24, Exercise 28.4, page 363]. In general, the cut locus of a subset need not be closed, as the following example, due to Sabau and Tanaka [25], illustrates.

Example 2.6 (Sabau and Tanaka 2016) Consider \mathbb{R}^2 with the Euclidean inner product. Let $\{\theta_n\}$, with $\theta_1 \in (0, \pi)$, be a decreasing sequence converging to 0. Let $\overline{B(\mathbf{0}, 1)}$ be the closed unit ball centered at $(0, 0)$. Let $B_n := B(q_n, 1)$ be the open ball with radius 1 and centered at q_n . We have chosen q_n so that it does not belong to $\overline{B(\mathbf{0}, 1)}$ and denotes the center of the circle passing through $p_n = (\cos \theta_n, \sin \theta_n)$ and $p_{n+1} = (\cos \theta_{n+1}, \sin \theta_{n+1})$. Define $N \subset \mathbb{R}^2$ by

$$N := \overline{B(\mathbf{0}, 1)} \setminus \bigcup_{n=1}^{\infty} B(q_n, 1).$$

See [Figure 1](#). Note that N is a closed set and the sequence $\{q_n\}$ of cut points of N converges to the point $(2, 0)$. However, $(2, 0)$ is not a cut point of N .

In [Theorem 3.30](#) we will prove that, for a submanifold N , the set $\text{Cu}(N)$ is closed, by showing that it is the closure of the set of all points in M which have at least two minimal geodesics joining N to $p \in M$.

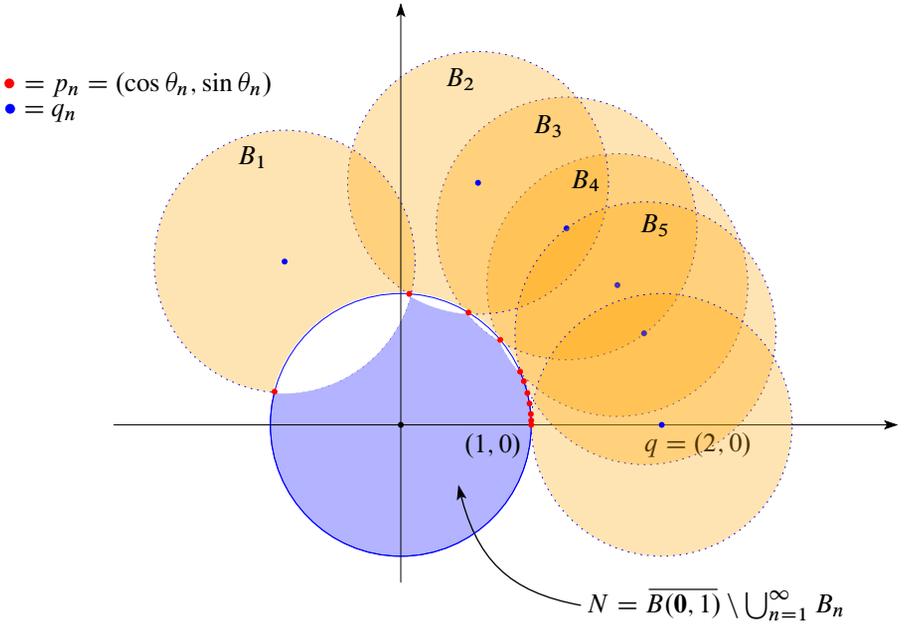


Figure 1: The cut locus need not be closed.

Example 2.7 (join induced by cut locus) Let $S_i^k \hookrightarrow S^n$ denote the embedding of the k -sphere in the first $k + 1$ coordinates and S_l^{n-k-1} denote the embedding of the $(n-k-1)$ -sphere in the last $n - k$ coordinates. It can be seen that $\text{Cu}(S_i^k) = S_l^{n-k-1}$. In fact, starting at a point $p \in S_i^k$ and traveling along a unit-speed geodesic in a direction normal to $T_p S_i^k$, we obtain a cut point at a distance $\frac{\pi}{2}$ from S_i^k . Moreover, in this case, $\text{Cu}(S_l^{n-k-1}) = S_i^k$ and the n -sphere S^n can be expressed as the union of geodesic

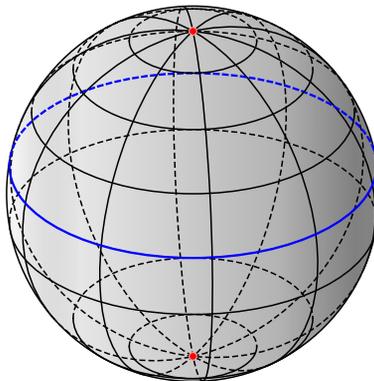


Figure 2: The cut locus of the equator in S^2 .

segments joining S_i^k to S_l^{n-k-1} . This is a geometric variant of the fact that the n -sphere is the (topological) join of S^k and S^{n-k-1} . We also observe that $S^n - S_l^{n-k-1}$ deforms to S_i^k while $S^n - S_i^k$ deforms to S_l^{n-k-1} .

In our example, let ν_i^{n-k} and ν_l^{k+1} denote the normal bundles of S_i^k and S_l^{n-k-1} , respectively. We may express S^n as the union of normal disk bundles $D(\nu_i)$ and $D(\nu_l)$. These disk bundles are trivial and are glued along their common boundary $S_i^k \times S_l^{n-k-1}$ to produce S^n . Moreover, S_i^k is an analytic submanifold of the real analytic Riemannian manifold S^n with the round metric. There is a generalization of this phenomenon due to Omori [21, Lemmas 1.3–1.5 and Theorem 3.1].

Theorem 2.8 (Omori 1968) *Let M be a compact, connected, real analytic Riemannian manifold which has an analytic submanifold N such that the cut point of N with respect to every geodesic which starts from N and whose initial direction is orthogonal to N has a constant distance π from N . Then $N' = \text{Cu}(N)$ is an analytic submanifold and M has a decomposition $M = DN \cup_\varphi DN'$, where DN and DN' are the normal disk bundles of N and N' , respectively, and φ is the gluing map.*

2.2 An illuminating example

Let $M = M(n, \mathbb{R})$, the set of $n \times n$ matrices, and $N = O(n, \mathbb{R})$, the set of all orthogonal $n \times n$ matrices. Let $A, B \in M(n, \mathbb{R})$. We fix the standard flat Euclidean metric on $M(n, \mathbb{R})$ by identifying it with \mathbb{R}^{n^2} . This induces a distance function given by

$$d(A, B) := \sqrt{\text{tr}((A - B)^T(A - B))}.$$

Consider the distance-squared function

$$f : \text{GL}(n, \mathbb{R}) \rightarrow \mathbb{R}, \quad A \mapsto d_{O(n, \mathbb{R})}^2(A).$$

Lemma 2.9 *The function f can be explicitly expressed as*

$$(2-1) \quad f(A) = n + \text{tr}(A^T A) - 2 \text{tr}(\sqrt{A^T A}).$$

Proof Let $A \in \text{GL}(n, \mathbb{R})$ be any invertible matrix. Then

$$(2-2) \quad \begin{aligned} f(A) &= \inf_{B \in O(n, \mathbb{R})} \text{tr}((A - B)^T(A - B)) \\ &= \inf_{B \in O(n, \mathbb{R})} [\text{tr}(A^T A) - \text{tr}(A^T B) - \text{tr}(B^T A) + \text{tr}(B^T B)] \\ &= \text{tr}(A^T A) + \inf_{B \in O(n, \mathbb{R})} [-2 \text{tr}(A^T B)] + n \\ &= \text{tr}(A^T A) + n - 2 \sup_{B \in O(n, \mathbb{R})} \text{tr}(A^T B). \end{aligned}$$

In order to maximize the function

$$h_A: O(n, \mathbb{R}) \rightarrow \mathbb{R}, \quad B \mapsto \text{tr}(A^T B),$$

for any invertible matrix A , we may first assume that A is a diagonal matrix with positive entries. Then

$$|h_A(B)| = |\text{tr}(A^T B)| = \left| \sum_{i=1}^n a_{ii} b_{ii} \right| \leq \sum_{i=1}^n |a_{ii} b_{ii}| \leq \sum_{i=1}^n a_{ii} = \text{tr}(A^T) = h_A(I).$$

Thus, one of the maximizers is $B = I$. For a general nonsingular matrix A , we will use the *singular value decomposition* (SVD). Write $A = UDV^T$, where U and V are $n \times n$ orthogonal matrices and D is a diagonal matrix with positive entries. For any $B \in O(n, \mathbb{R})$, using the cyclic property of trace, we can see that

$$\text{tr}(A^T B) = \text{tr}(D(U^T B V)).$$

Since $U^T B V$ is an orthogonal matrix, maximizing over B reduces to the earlier observation that B will be a maximizer if $U^T B V = I$, which implies $B = UV^T$.

Since A is invertible, by the polar decomposition, there exists an orthogonal matrix Q and a symmetric positive-definite matrix $S = \sqrt{A^T A}$ such that $A = QS$. Since S is a symmetric matrix, we can diagonalize it, ie $S = P\tilde{D}P^T$, where $P \in O(n, \mathbb{R})$ and \tilde{D} is a diagonal matrix with the eigenvalues of S as its diagonal entries. Thus,

$$A = QS = QP\tilde{D}P^T.$$

Set $U = QP$ and $V = P$ to obtain the SVD of A . In particular, the minimizer is given by

$$B = Q = A\sqrt{A^T A}^{-1}.$$

Therefore,

$$f(A) = n + \text{tr}(A^T A) - 2 \text{tr}(\sqrt{A^T A})$$

for invertible matrices.

In order to compute f for a noninvertible matrix A , we note that $\text{GL}(n, \mathbb{R})$ is dense in $M(n, \mathbb{R})$ and that $\sqrt{A^T A}$ is well defined for $A \in M(n, \mathbb{R})$. The continuity of the map $A \mapsto \sqrt{A^T A}$ on $M(n, \mathbb{R})$ implies that the same formula (2-1) for f applies to A as well. □

In order to understand the differentiability of f , it suffices to analyze the function $A \mapsto \text{tr}(\sqrt{A^T A})$.

Lemma 2.10 *The map $g: M(n, \mathbb{R}) \rightarrow \mathbb{R}, A \mapsto \text{tr}(\sqrt{A^T A})$, is differentiable if and only if A is invertible.*

The proof of this postponed to the appendix (see [Lemma B.2](#)).

Let us define the function

$$\phi: M(n, \mathbb{R}) \rightarrow \mathbb{R}, \quad A \mapsto \text{tr}(\sqrt{A^T A}).$$

We claim that

$$(2-3) \quad D\phi_A(H) = \text{tr}\left(\int_0^\infty e^{-t\sqrt{A^T A}}(A^T H + H^T A)e^{-t\sqrt{A^T A}} dt\right).$$

The following lemma (see [Lemma B.1](#) for a proof) along with chain rule will prove our claim:

Lemma 2.11 *Let A be a positive-definite matrix and $\psi: A \mapsto \sqrt{A}$. Then*

$$D\psi_A(H) = \int_0^\infty e^{-t\sqrt{A}} H e^{-t\sqrt{A}} dt$$

for any symmetric matrix H .

We may drastically simplify, using basic analysis and linear algebra, the derivative of ϕ given by (2-3) to obtain

$$D\phi_A(H) = \langle A\sqrt{A^T A}^{-1}, H \rangle.$$

For any $A \in \text{GL}(n, \mathbb{R})$,

$$Df_A = 2A - 2A\sqrt{A^T A}^{-1} = -2A(\sqrt{A^T A}^{-1} - I).$$

Hence, the negative gradient of the function f , restricted to $\text{GL}(n, \mathbb{R})$, is given by

$$-\nabla f|_A = 2A(\sqrt{A^T A}^{-1} - I).$$

The critical points are orthogonal matrices. If $\gamma(t)$ is an integral curve of $-\nabla f$ initialized at A , then $\gamma(0) = A$ and

$$(2-4) \quad \frac{d\gamma}{dt} = -2\gamma(t) + 2\gamma(t)\sqrt{\gamma(t)^T \gamma(t)}^{-1} = -2\gamma(t) + 2(\gamma(t)^T)^{-1}\sqrt{\gamma(t)^T \gamma(t)}.$$

Take the test solution of (2-4) given by

$$(2-5) \quad \gamma(t) = Ae^{-2t} + (1 - e^{-2t})(A^T)^{-1}\sqrt{A^T A} = Ae^{-2t} + (1 - e^{-2t})A\sqrt{A^T A}^{-1}.$$

In order to show that $\gamma(t)$ satisfies (2-4), we may verify the simplifications

$$\begin{aligned} \gamma(t)^T \gamma(t) &= (\sqrt{A^T A} e^{-2t} + (1 - e^{-2t}) I)^2, \\ \sqrt{\gamma(t)^T \gamma(t)}^T &= (\sqrt{A^T A} A^{-1} \gamma(t))^T = \gamma(t)^T (A^T)^{-1} \sqrt{A^T A}. \end{aligned}$$

This implies that

$$(\gamma(t)^T)^{-1} \sqrt{\gamma(t)^T \gamma(t)} = (A^T)^{-1} \sqrt{A^T A}.$$

The right-hand side of (2-4), with the test solution, can be simplified to

$$-2Ae^{-2t} + 2e^{-2t} (A^T)^{-1} \sqrt{A^T A},$$

which is the derivative of γ . Thus, $\gamma(t)$, as defined in (2-5), is the required flow line which deforms $GL(n, \mathbb{R})$ to $O(n, \mathbb{R})$. In particular, $GL^+(n, \mathbb{R})$ deforms to $SO(n, \mathbb{R})$ and the other component of $GL(n, \mathbb{R})$ deforms to $O(n, \mathbb{R}) \setminus SO(n, \mathbb{R})$. We note, however, that this deformation takes infinite time to perform the retraction.

Remark 2.12 A modified curve

$$(2-6) \quad \eta(t) = A(1 - t) + tA\sqrt{A^T A}^{-1},$$

with the same image as γ , defines an actual deformation retraction of $GL(n, \mathbb{R})$ to $O(n, \mathbb{R})$. Apart from its origin via the distance function, this is a geometric deformation in the following sense. Given $A \in GL(n, \mathbb{R})$, consider its columns as an ordered basis. This deformation deforms the ordered basis according to the length of the basis vectors and mutual angles between pairs of basis vectors in a geometrically uniform manner. This is in sharp contrast with Gram–Schmidt orthogonalization, also a deformation of $GL(n, \mathbb{R})$ to $O(n, \mathbb{R})$, which is asymmetric as it never changes the direction of the first column, the modified second column only depends on the first two columns, and so on.

We now show that f is Morse–Bott. The tangent space $T_I O(n, \mathbb{R})$ consists of skew-symmetric matrices while the normal vectors at I_n are the symmetric matrices. As left translation by an orthogonal matrix is an isometry of $M(n, \mathbb{R})$, normal vectors at $A \in O(n, \mathbb{R})$ are of the form AW for symmetric matrices W . Since

$$Df_A(H) = 2\langle A, H \rangle - 2\langle A\sqrt{A^T A}^{-1}, H \rangle,$$

the relevant Hessian is

$$\text{Hess}(f)_A(H, H') = \lim_{t \rightarrow 0} \frac{Df_{A+tH'}(H) - Df_A(H)}{t}$$

with $H = AW$ and $H' = AW'$ for symmetric matrices W and W' . A standard computation leads to

$$\text{Hess}(f)_A(H, H') = 2 \text{tr}(H^T H') = 2\langle H, H' \rangle.$$

Therefore, the Hessian matrix restricted to $(T_A O(n, \mathbb{R}))^\perp$ is $2I_{n(n+1)/2}$. This is a recurring feature of distance-squared functions associated to embedded submanifolds (see Proposition 3.5).

There is a relationship between the local homology of cut loci and the reduced Čech cohomology of the link of a point in the cut locus. This is due to Theorem 1.4 of Hebda [9] and the remark following it.

Definition 2.13 Let N be an embedded submanifold of a complete smooth Riemannian manifold M . For each $q \in \text{Cu}(N)$, consider the set $\Lambda(q, N)$ of unit tangent vectors at q such that the associated geodesics realize the distance between q and N . This set is called the *link* of q with respect to N .

The set of points in N obtained by the endpoints of the geodesics associated to $\Lambda(q, N)$ will be called the *equidistant set*, denoted by $\text{Eq}(q, N)$, of q with respect to N .

Since the equidistant set $\text{Eq}(q, N)$, consisting of points which realize the distance $d_N(q)$, is obtained by exponentiating the points in $\Lambda(q, N)$, there is a natural surjection map from $\Lambda(q, N)$ to $\text{Eq}(q, N)$.

Theorem 2.14 (Hebda 1983) *Let N be a properly embedded submanifold of a complete Riemannian manifold M of dimension n . If $q \in \text{Cu}(N)$ and v is an element of $\Lambda := \Lambda(q, N)$, then there is an isomorphism*

$$(2-7) \quad \check{H}^i(\Lambda, v) \cong H_{n-1-i}(\text{Cu}(N), \text{Cu}(N) - q).$$

We are interested in computing $\Lambda(A, O(n, \mathbb{R}))$ for singular matrices A . Note that geodesics in $M(n, \mathbb{R})$, initialized at A , are straight lines and any two such geodesics can never meet other than at A . Therefore, there is a natural identification between the link and the equidistant set of A .

Lemma 2.15 *If $A \in M(n, \mathbb{R})$ is singular of rank k , then $\text{Eq}(A, O(n, \mathbb{R}))$ is homeomorphic to $O(n - k, \mathbb{R})$.*

Proof Using the singular value decomposition, we write $A = UDV^T$, where $U, V \in O(n, \mathbb{R})$ and D is a diagonal matrix with entries the eigenvalues of $\sqrt{A^T A}$. If we specify that the diagonal entries of D are arranged in decreasing order, then D is unique. Moreover, as A has rank $k < n$, the first k diagonal entries of D are positive while the last $n - k$ diagonal entries are zero. In order to find the matrices in $O(n, \mathbb{R})$ which realize the distance $d(A, O(n, \mathbb{R}))$, by (2-2), it suffices to find $B \in O(n, \mathbb{R})$ such that

$$\sup_{B \in O(n, \mathbb{R})} \operatorname{tr}(A^T B) = \sup_{B \in O(n, \mathbb{R})} \operatorname{tr}(VDU^T B) = \sup_{B \in O(n, \mathbb{R})} \operatorname{tr}(DU^T B V)$$

is maximized. However, $U^T B V \in O(n, \mathbb{R})$ has orthonormal rows and the specific form of D implies that the maximum happens if and only if $U^T B V$ has e_1, \dots, e_k as the first k rows, in order. Therefore, $U^T B V$ is a block orthogonal matrix, with blocks of I_k and $C \in O(n - k, \mathbb{R})$, ie $B \in U(I_k \times O(n - k, \mathbb{R}))V^T$. \square

Corollary 2.16 *Let Sing denote the space of singular matrices in $M(n, \mathbb{R})$. If $A \in \text{Sing}$ is of rank $k < n$, then there is an isomorphism*

$$(2-8) \quad \tilde{H}^i(O(n - k, \mathbb{R})) \cong H_{n^2 - 1 - i}(\text{Sing}, \text{Sing} - A).$$

Proof It follows from Lemma 2.15 that $\Lambda(A, O(n, \mathbb{R})) \cong O(n - k, \mathbb{R})$ if A has rank k . Since $O(n - k, \mathbb{R})$ is a manifold, the Čech and singular cohomology groups are isomorphic. The space Sing is a star-convex set, whence all homotopy and homology groups are that of a point. Applying (2-7) in our case, we obtain an isomorphism

$$\tilde{H}^i(O(n - k, \mathbb{R})) \cong H_{n^2 - 1 - i}(\text{Sing}, \text{Sing} - A)$$

between the reduced cohomology and local homology groups. In particular, the local homology of the cut locus at A detects the rank of A . \square

Similar computations hold for $U(n, \mathbb{C})$ and singular $n \times n$ complex matrices.

3 Main results

We recall some results about exponential maps and Fermi coordinates in Section 3.1. A result of Wolter [29] may be generalized to prove (Lemma 3.7) that the distance-squared function from a submanifold is not differentiable on the separating set. This result may be well known to experts, but the proof, following Wolter, is elementary. Buchner’s result [4] may be generalized to prove (Theorem 3.9) that the cut locus is a simplicial

complex for real analytic pairs. In [Section 3.2](#) we recall the notion of Thom space and apply it to the normal bundle of an embedded submanifold in a closed, connected Riemannian manifold. Our first main result, [Theorem 3.16](#), states that the quotient of the ambient manifold by the cut locus of the submanifold results in the Thom space of the normal bundle. As a consequence we obtain [Theorem 3.24](#), which says that a homology k -sphere inside a manifold homeomorphic to S^d has cut locus weakly homotopy equivalent to S^{d-k-1} provided $d - k \geq 3, k > 0$ and $H_{d-1}(\text{Cu}(N))$ is torsion-free. [Theorem 3.27](#) is another consequence about analytic surfaces. In [Section 3.3](#) we prove (see [Theorem 3.30](#)) that the cut locus of a submanifold is closed, essentially following Wolter’s arguments [[29](#)]. This leads us to the other main result, [Theorem 3.32](#), which proves that the complement of the cut locus $\text{Cu}(N)$ deforms to N .

3.1 Basic results

For understanding the geometry in the neighborhood of a submanifold, it is convenient to use Fermi coordinates, a generalization of normal coordinates. We shall briefly introduce Fermi coordinates and state some of their relevant properties. Let N be an embedded submanifold of a Riemannian manifold M . Let ν be the normal bundle of $N \subseteq M$, ie

$$\nu := \{(p, v) : p \in N, v \in (T_p N)^\perp\}.$$

In fact, ν is a subbundle of the restriction of TM to N . We define the *exponential map of the normal bundle* as

$$(3-1) \quad \exp_\nu : \nu \rightarrow M, \quad \exp_\nu(p, v) := \exp_p(v) \quad \text{for } (p, v) \in \nu.$$

We may write $\exp_\nu(v)$ in short and call this the *normal exponential map*.

Now we will list some lemmas; for proofs we refer to Gray [[8](#), Sections 2.1 and 2.3].

Lemma 3.1 *Let N be a topologically embedded submanifold of a Riemannian manifold M . Then the normal exponential map $\exp_\nu : \nu \rightarrow M$ maps a neighborhood of N in ν diffeomorphically onto a neighborhood of N in M .*

Let \mathcal{O}_N denote the largest neighborhood of the zero section of ν for which \exp_ν is a diffeomorphism. We shall later be able to describe this neighborhood in terms of a function s ; see (3-2). To define a system of Fermi coordinates, we need an arbitrary system of coordinates (y_1, \dots, y_k) defined in a neighborhood $\mathcal{U} \subset N$ of $p \in N$ together with orthogonal sections E_{k+1}, \dots, E_n of the restriction of ν to \mathcal{U} .

Definition 3.2 (Fermi coordinates) The Fermi coordinates (x_1, \dots, x_n) of $N \subset M$ centered at p (relative to a given coordinate system (y_1, \dots, y_k) on N and orthogonal sections E_{k+1}, \dots, E_n of ν) are defined by

$$x_l \left(\exp_\nu \left(\sum_{j=k+1}^n t_j E_j(p') \right) \right) = y_l(p') \quad \text{for } l = 1, \dots, k,$$

$$x_i \left(\exp_\nu \left(\sum_{j=k+1}^n t_j E_j(p') \right) \right) = t_i \quad \text{for } i = k + 1, \dots, n,$$

for $p' \in \mathcal{U}$ provided the numbers t_{k+1}, \dots, t_n are small enough that

$$t_{k+1} E_{k+1}(p') + \dots + t_n E_n(p') \in \mathcal{O}_N.$$

Since \exp_ν is a diffeomorphism on \mathcal{O}_N , $(x_1, \dots, x_k, x_{k+1}, \dots, x_n)$ defines a coordinate system near p . In fact, the restrictions to N of the coordinate vector fields $\partial/\partial x_{k+1}, \dots, \partial/\partial x_n$ are orthonormal.

Lemma 3.3 Let γ be a unit-speed geodesic normal to N with $\gamma(0) = p \in N$. If $u = \gamma'(0)$, then there is a system of Fermi coordinates (x_1, \dots, x_n) such that, for small enough t , ie for $(p, tu) \in \mathcal{O}_N$, we have

$$\frac{\partial}{\partial x_{k+1}} \Big|_{\gamma(t)} = \gamma'(t), \quad \frac{\partial}{\partial x_l} \Big|_p \in T_p N, \quad \frac{\partial}{\partial x_i} \Big|_p \in (T_p N)^\perp$$

for $1 \leq l \leq k$ and $k + 1 \leq i \leq n$. Furthermore, for $1 \leq j \leq n$,

$$(x_j \circ \gamma)(t) = t \delta_{j(k+1)}.$$

Definition 3.4 Let (x_1, \dots, x_n) be a system of Fermi coordinates for $N \subset M$. Define $\sigma(x_1, \dots, x_n)$ to be the nonnegative number satisfying

$$\sigma^2 = \sum_{i=k+1}^n x_i^2.$$

It is known that σ does not depend on the choice of Fermi coordinates.

Proposition 3.5 Let U be a neighborhood of N such that each point in U admits a unique unit-speed N -geodesic. If $p \in U$, then

$$\sigma(p) = d_N(p).$$

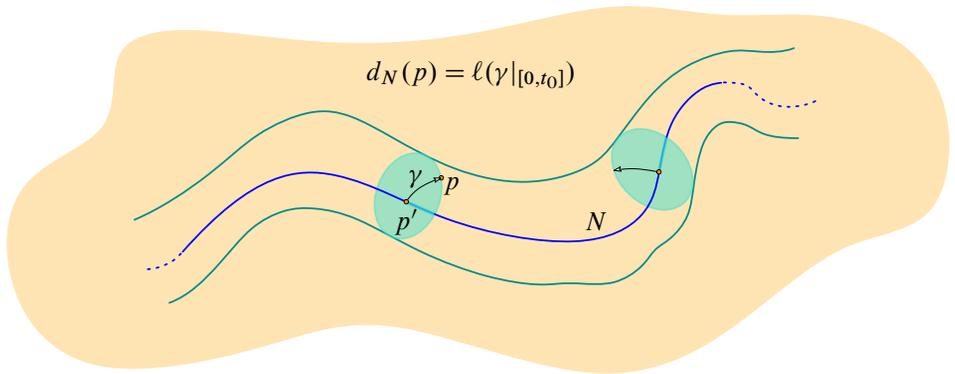


Figure 3: Distance via Fermi coordinates.

Proof Since the expression of σ is independent of the choice of the Fermi coordinates, we will make a special choice of the Fermi coordinates (x_1, \dots, x_n) . For $p \in U$, choose the unique unit-speed N -geodesic γ joining p to N . This geodesic meets N orthogonally at $\gamma(0) = p'$. Choose t_0 so that $\gamma(t_0) = p$; see Figure 3. According to Lemma 3.3, there is a system of Fermi coordinates (x_1, \dots, x_n) centered at p' such that $x_i(\gamma(t)) = t \delta_{i(k+1)}$. The sequence of equalities

$$\sigma(p) = x_{k+1}(\gamma(t_0)) = t_0 = d_N(p)$$

completes the proof. □

Corollary 3.6 Consider the distance-squared function with respect to a submanifold N in M . The Hessian of the distance-squared function at the critical submanifold N is nondegenerate in the normal direction.

Towards the regularity of the distance-squared function, the following observation will be useful. It is a routine generalization of [29, Lemma 1].

Lemma 3.7 Let M be a connected, complete Riemannian manifold and N be an embedded submanifold of M . Suppose two N -geodesics exist joining N to $q \in M$. Then $d_N^2 : M \rightarrow \mathbb{R}$ has no directional derivative at q for vectors in direction of those two N -geodesics.

Proof Let us assume that all the geodesics are parametrized by arc length. Let $\gamma_i : [0, \hat{t}] \rightarrow M$ for $i = 1, 2$ be two distinct geodesics with $\gamma_1(0), \gamma_2(0) \in N$ and $\gamma_1(l) = q = \gamma_2(l)$, where $l = d_N(q)$ and $0 < l < \hat{t}$. Let us suppose that the two

geodesics start at p_1 and p_2 and so $d(p_1, q) = l = d(p_2, q)$. Note that the directional derivative of d^2 at q in the direction of $\gamma'_i(q)$ from the left is given by

$$(d^2)'_-(q) := \lim_{\varepsilon \rightarrow 0^+} \frac{d_N^2(\gamma_i(l)) - d_N^2(\gamma_i(l - \varepsilon))}{\varepsilon} = \lim_{\varepsilon \rightarrow 0^+} \frac{l^2 - (l - \varepsilon)^2}{\varepsilon} = 2l.$$

Next, we claim that the derivative of the same function from the right is strictly bounded above by $2l$. Let $\omega \in (0, \pi]$ be the angle between the two geodesics γ_1 and γ_2 at q . Define the function

$$u(\tau) := d_N(\gamma_1(l - \varepsilon)) + d(\gamma_1(l - \varepsilon), \gamma_2(\tau + l)).$$

By the triangle inequality, we observe that

$$f(\tau) := (u(\tau))^2 \geq d^2(p_1, \gamma_2(\tau + l)) \geq d_N^2(\gamma_2(\tau + l)),$$

and equality holds at $\tau = 0$ and $(u(0))^2 = d_N^2(q) = l^2$. Thus, in order to prove the claim, it suffices to show that the derivative of f from the right, at $\tau = 0$, is bounded below by $2l$. We need to invoke a version of the cosine law for small geodesic triangles. Although this may be well known to experts, we will use the version that appears in Sharafutdinov’s work [27] (see also Daniilidis et al [6, Lemma 2.4] for a detailed proof). In our case, this means that

$$d^2(\gamma_1(l - \varepsilon), \gamma_2(\tau + l)) = \varepsilon^2 + \tau^2 + 2\varepsilon\tau \cos \omega + K(\tau)\varepsilon^2\tau^2,$$

where $|K(\tau)|$ is bounded and the side lengths are sufficiently small. Note that we are considering geodesic triangles with two vertices constant and the varying vertex being $\gamma_2(l + \tau)$. It follows from taking a square root and then expanding in powers of τ that

$$d(\gamma_1(l - \varepsilon), \gamma_2(\tau + l)) = \sqrt{\varepsilon^2 + \tau^2 + 2\varepsilon\tau \cos \omega} (1 + O(\tau^2)).$$

It follows that

$$u(\tau) = l - \varepsilon + \sqrt{\varepsilon^2 + \tau^2 + 2\varepsilon\tau \cos \omega} (1 + O(\tau^2)).$$

Therefore, $u'_+(\tau) = \cos \omega = d'_+(\gamma_1(l - \varepsilon), \gamma_2(l))$. Observe that

$$\begin{aligned} f'_+(\tau)|_{\tau=0} &= 2d_N(\gamma_1(l - \varepsilon))d'_+(\gamma_1(l - \varepsilon), \gamma_2(l)) + 2d(\gamma_1(l - \varepsilon), \gamma_2(l))d'_+(\gamma_1(l - \varepsilon), \gamma_2(l)) \\ &= 2d_N(\gamma_1(l - \varepsilon)) \cos \omega + 2d(\gamma_1(l - \varepsilon), \gamma_2(l)) \cos \omega \\ &= 2d_N(\gamma_1(l)) \cos \omega < 2l. \end{aligned}$$

Thus, we have proved the claim and subsequently the result. □

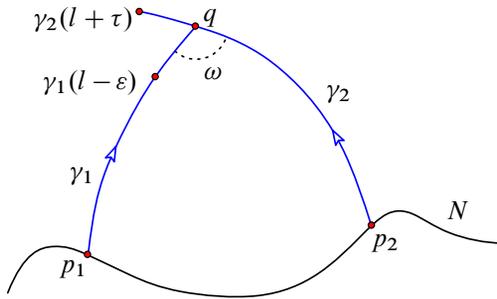


Figure 4: When two N -geodesics meet.

The above lemma prompts us to define the following set, the notation being consistent with Wolter’s paper [29]:

Definition 3.8 Let N be a subset of a Riemannian manifold M . The set $\text{Se}(N)$, called the *separating set*,¹ consists of all points $q \in M$ such that at least two distance-minimal geodesics from N to q exist.

If $q \in \text{Se}(N)$ but $q \notin \text{Cu}(N)$, then we have Figure 4, ie γ_1 is an N -geodesic beyond q while γ_2 is another N -geodesic for q . The triangle inequality applied to $\gamma_1(0)$, $q = \gamma_1(l)$ and $\gamma_2(l + \tau)$ implies that

$$d_N(\gamma_2(l + \tau)) < l + \tau,$$

while, for τ small enough, $d_N(\gamma_2(l + \tau)) = l + \tau$ as γ_2 is an N -geodesic beyond q . This contradiction establishes the well-known fact $\text{Se}(N) \subseteq \text{Cu}(N)$. In quite a few examples, these two sets are equal. In the case of $M = \mathbb{S}^n$ with $N = \{p\}$, the set $\text{Se}(N)$ consists of $-p$. There is an infinite family of minimal geodesics joining p to $-p$. An appropriate choice of a pair of such minimal geodesics would create a loop, which is permissible in the definition of $\text{Se}(N)$.

Regarding the question of cut loci being triangulable, we recall the result of Buchner [4] that the cut locus (of a point) of a real analytic Riemannian manifold (of dimension d) is a simplicial complex of dimension at most $d - 1$. It follows, without many changes, that the result holds for cut loci of submanifolds as well. Hence, we attribute the following result to Buchner:

¹We could not find any name for this set in the literature. This terminology is our own although this nomenclature is rarely used in the paper.

Theorem 3.9 (Buchner 1977) *Let N be an analytic submanifold of a real analytic manifold M . If M is of dimension d , then the cut locus $\text{Cu}(N)$ is a simplicial complex of dimension at most $d - 1$.*

The obvious modifications to the proof by Buchner are the following:

- (i) Choose ε to be such that there is a unique geodesic from p to q if $(p, q) < \varepsilon$ and, if $d_N(q) < \varepsilon$, then there is a unique N -geodesic to q .
- (ii) Consider the set $\Omega_N(t_0, t_1, \dots, t_k)$, the space of piecewise broken geodesics starting at N , and define $\Omega_N(t_0, t_1, \dots, t_k)^S$ analogously.
- (iii) The map

$$\Omega_N(t_0, t_1, \dots, t_k)^S \rightarrow N \times M \times \dots \times M, \quad \omega \mapsto (\omega(t_0), \omega(t_1), \dots, \omega(t_k)),$$

determines an analytic structure on $\Omega_N(t_0, t_1, \dots, t_k)^S$.

The remainder of the proof works essentially verbatim.

Remark 3.10 As we have seen in [Example 2.7](#), the dimension of the cut locus of a k -dimensional submanifold is $d - k - 1$. However, generically, we may not expect this to be true. In fact, for real analytic knots (except the unknot) in \mathbb{S}^3 , it is always the case that the cut locus cannot be homotopic to a (connected) 1-dimensional simplicial complex (see [Example 3.29](#)).

3.2 Thom space via cut locus

Let (M, g) be a complete Riemannian manifold with distance function d . The exponential map at p ,

$$\exp_p: T_p M \rightarrow M,$$

is defined on the tangent space. Moreover, there exists a minimal geodesic joining any two points in M . However, not all geodesics are distance-realizing. Given $v \in T_p M$ with $\|v\| = 1$, let γ_v be the geodesic initialized at p with velocity v . Let $S(TM)$ denote the unit tangent bundle and let $[0, \infty]$ be the one-point compactification of $[0, \infty)$. Define

$$s: S(TM) \rightarrow [0, \infty], \quad s(v) := \sup\{t \in [0, \infty) : \gamma_v|_{[0,t]} \text{ is minimal}\}.$$

Definition 3.11 (cut locus) Let M be a complete, connected Riemannian manifold. If $s(v) < \infty$ for some $v \in S(T_p M)$, then $\exp_p(s(v)v)$ is called a *cut point*. The collection of cut points is defined to be the cut locus of p .

As geodesics are locally distance-realizing, $s(v) > 0$ for any $v \in S(TM)$. The following result [26, Proposition 4.1] will be important for the underlying ideas in its proof:

Proposition 3.12 *The map $s : S(TM) \rightarrow [0, \infty]$, $u \mapsto s(u)$, is continuous.*

The proof relies on a characterization of $s(v)$ provided $s(v) < \infty$. A positive real number T is $s(v)$ if and only if $\gamma_v : [0, T] \rightarrow M$ is minimal and at least one of the following holds:

- (i) $\gamma_v(T)$ is the first conjugate point of p along γ_v .
- (ii) There exists $u \in S(T_p M)$ with $u \neq v$ and $\gamma_u(T) = \gamma_v(T)$.

Recall that, if $\gamma : [0, a] \rightarrow M$ is a geodesic, then $q = \gamma(t_0)$ is conjugate to $p = \gamma(0)$ along γ if \exp_p is singular at $t_0 \dot{\gamma}(0)$, ie $(D \exp_p)(t_0 \dot{\gamma}(0))$ is not of full rank.

Remark 3.13 If M is compact, then it has bounded diameter, which implies that $s(v) < \infty$ for any $v \in S(TM)$. The converse is also true: if M is complete and connected with $s(v) < \infty$ for any $v \in S(TM)$, then M has bounded diameter, whence it is compact.

We shall be concerned with closed Riemannian manifolds in what follows. Let N be an embedded submanifold inside a closed, ie compact without boundary, manifold M . Let ν denote the normal bundle of N in M with $D(\nu)$ denoting the unit disk bundle. In the context of $S(\nu)$, the unit normal bundle and the cut locus of N , distance-minimal geodesics or N -geodesics are relevant (see Definitions 2.4 and 2.5). We want to consider

$$(3-2) \quad s : S(\nu) \rightarrow [0, \infty), \quad s(v) := \sup\{t \in [0, \infty) : \gamma_v|_{[0,t]} \text{ is an } N\text{-geodesic}\}.$$

Notice that $0 < s(v) \leq s(v)$ for any $v \in S(\nu)$. In the special case when $N = \{p\}$, s is simply the restriction of s to $T_p M$. Analogous to Proposition 3.12, we have the following result:

Proposition 3.14 *The map $s : S(\nu) \rightarrow [0, \infty)$, as defined in (3-2), is continuous.*

As expected, the proof of Proposition 3.14 relies on a characterization of $s(v)$ similar to that of $s(v)$ (refer to Lemma A.2 and Bishop and Crittenden’s book [3, Exercise 23, page 241]).

Let us postpone the proofs (see Appendix A) and proceed with some immediate applications.

Definition 3.15 (rescaled exponential) The *rescaled exponential* or *s*-exponential map is defined to be

$$\widetilde{\exp}: D(\nu) \rightarrow M, \quad (p, v) \mapsto \begin{cases} \exp_p(s(\hat{v})v) & \text{if } v = \|v\|\hat{v} \neq 0, \\ p & \text{if } v = 0. \end{cases}$$

We are now ready to prove the main result of this section.

Theorem 3.16 Let N be an embedded submanifold inside a closed, connected Riemannian manifold M . If ν denotes the normal bundle of N in M , then there is a homeomorphism

$$\widetilde{\exp}: D(\nu)/S(\nu) \xrightarrow{\cong} M/\text{Cu}(N).$$

Proof It follows from Proposition 3.14 that the rescaled exponential is continuous. Moreover, $\widetilde{\exp}$ is surjective and $\widetilde{\exp}(S(\nu)) = \text{Cu}(N)$. If there exist $(p, v) \neq (q, w) \in D(\nu)$ such that

$$\widetilde{\exp}(p, v) = \widetilde{\exp}(q, w) = p',$$

then $d_N(p')$ can be computed in two ways to obtain

$$d_N(p') = s(\hat{v})\|v\| = s(\hat{w})\|w\|.$$

Thus, $T = d(p', N)$ is a number such that $\gamma_v: [0, T] \rightarrow M$ is an N -geodesic and $\gamma_v(T) = \gamma_w(T) = p'$. By Lemma A.2, we conclude that $T = s(\hat{v}) = s(\hat{w})$, whence $\|v\| = \|w\| = 1$. Therefore, $\widetilde{\exp}$ is injective on the interior of $D(\nu)$.

As $\text{Cu}(N)$ is closed and M is a compact metric space, the quotient space $M/\text{Cu}(N)$ is Hausdorff. As the quotient $D(\nu)/S(\nu)$ is compact, standard topological arguments imply the map induced by the rescaled exponential is a homeomorphism. \square

Recall that the *Thom space* $\text{Th}(E)$ of a real vector bundle $E \rightarrow B$ of rank k is $D(E)/S(E)$, where it is understood that we have chosen a Euclidean metric on E . If B is compact, then the Thom space $\text{Th}(E)$ is the one-point compactification of E . In general, we compactify the fibers and then collapse the section at infinity to a point to obtain $\text{Th}(E)$. Thus, Thom spaces obtained via two different metrics are homeomorphic. We will now revisit a basic property of Thom space via its connection to the cut locus. It can be seen that

$$(3-3) \quad \text{Cu}(N_1 \times N_2) = (\text{Cu}(N_1) \times M_2) \cup (M_1 \times \text{Cu}(N_2))$$

for an embedding $N_1 \times N_2$ inside $M_1 \times M_2$. If ν_j is the normal bundle of N_j inside M_j , then [Theorem 3.16](#) along with (3-3) implies that

$$\begin{aligned} \text{Th}(\nu_1 \oplus \nu_2) &\cong \frac{M_1 \times M_2}{(M_1 \times \text{Cu}(N_2)) \cup (\text{Cu}(N_1) \times M_2)} \cong \frac{M_1/\text{Cu}(N_1) \times M_2/\text{Cu}(N_2)}{M_1/\text{Cu}(N_1) \vee M_2/\text{Cu}(N_2)} \\ &\cong \text{Th}(\nu_1) \wedge \text{Th}(\nu_2). \end{aligned}$$

Let $N = N_1 \sqcup N_2$ be a disjoint union of connected manifolds of the same dimension. If $N \hookrightarrow M$, then let ν_j denote the normal bundle of N_j in M . If ν is the normal bundle of N in M , then

$$(3-4) \quad \text{Th}(\nu) \cong \text{Th}(\nu_1) \vee \text{Th}(\nu_2).$$

This implies that

$$M/\text{Cu}(N) \cong M/\text{Cu}(N_1) \vee M/\text{Cu}(N_2).$$

Example 3.17 Consider the two circles

$$N_1 = \{(\cos t, \sin t, 0, 0) \mid t \in \mathbb{R}\}, \quad N_2 = \{(0, 0, \cos t, \sin t) \mid t \in \mathbb{R}\}$$

in \mathbb{S}^3 . The link $N := N_1 \sqcup N_2$ has linking number 1. It can be checked that

$$\text{Cu}(N) = \left\{ \frac{1}{\sqrt{2}}(\cos s, \sin s, \cos t, \sin t) \mid s, t \in \mathbb{R} \right\}$$

is a torus. Note that $\text{Cu}(N_1) = N_2$ and vice versa as well as

$$\mathbb{S}^3/\text{Cu}(N_j) \cong (S^1 \times S^2)/(S^1 \times \infty),$$

where $S^1 \times S^2$ is the fiberwise compactification of the normal bundle of N_j . We conclude that

$$\mathbb{S}^3/\text{Cu}(N) \cong \left(\frac{S^1 \times S^2}{S^1 \times \infty} \right) \vee \left(\frac{S^1 \times S^2}{S^1 \times \infty} \right).$$

There are some topological similarities between $\text{Cu}(N)$ and $M - N$.

Lemma 3.18 *The cut locus $\text{Cu}(N)$ is a strong deformation retract of $M - N$. In particular, $(M, \text{Cu}(N))$ is a good pair and the number of path components of $\text{Cu}(N)$ equals that of $M - N$.*

Proof Consider the map $H: (M - N) \times [0, 1] \rightarrow M - N$ defined via the normal exponential map

$$H(q, t) = \begin{cases} \exp_\nu \left[\left\{ t \cdot s \left(\frac{\exp_\nu^{-1}(q)}{\|\exp_\nu^{-1}(q)\|} \right) + (1-t) \|\exp_\nu^{-1}(q)\| \right\} \frac{\exp_\nu^{-1}(q)}{\|\exp_\nu^{-1}(q)\|} \right] & \text{if } q \in M - (\text{Cu}(N) \cup N), \\ q & \text{if } q \in \text{Cu}(N). \end{cases}$$

If $q \in M - (\text{Cu}(N) \cup N)$, then let γ be the unique N -geodesic joining N to q . The path $H(q, t)$ is the image of this geodesic from q to the first cut point along γ . The continuity of s implies that H is continuous. It also satisfies $H(q, 0) = q$ and $H(q, 1) \in \text{Cu}(N)$. The claims about good pair and path components are clear. \square

Corollary 3.19 *If two embeddings $f, g: N \rightarrow M$ are ambient isotopic, then $\text{Cu}(f(N))$ and $\text{Cu}(g(N))$ are homotopy equivalent.*

Proof The hypothesis implies that there is a diffeomorphism $\varphi: M \rightarrow M$ such that $\varphi(f(N)) = g(N)$. Thus, $M - \text{Cu}(f(N))$ is homeomorphic to $M - \text{Cu}(g(N))$ and the claim follows from the lemma above. Note that, in the smooth category, the notion of isotopic and ambient isotopic are equivalent (refer to Section 8.1 of Hirsch’s book [12]). Thus, the same conclusion holds if we assume that the embeddings are isotopic. \square

Remark 3.20 Without the assumption of M being closed, the above result fails to be true. One may consider $M = S^1 \times \mathbb{R}$ with the natural product metric and $N = S^1$. In fact, the universal cover of M is $\mathbb{R} \times \mathbb{R}$ while that of N is \mathbb{R} . If we choose a periodic curve in \mathbb{R}^2 which is isotopic to the x -axis and has nonempty cut locus in \mathbb{R}^2 , then we may pass via the covering map to obtain an embedding g of N isotopic to the embedding f identifying N with $S^1 \times \{0\}$. For this pair, $\text{Cu}(f(N)) = \emptyset$ while $\text{Cu}(g(N)) \neq \emptyset$.

Several other identifications between topological invariants can be explored. For instance, if $\iota: N^k \hookrightarrow M^d$ is, as before, such that $M - N$ is path-connected, then

$$(3-5) \quad \iota_*: \pi_j(\text{Cu}(N)) \xrightarrow{\cong} \pi_j(M)$$

if $0 \leq j \leq d - k - 2$ while ι_* is a surjection for $j = d - k - 1$. The proof of this relies on a general position argument, ie being able to find a homotopy of the sphere that avoids N , followed by Lemma 3.18. Surjectivity of ι_* if $j \leq d - k - 1$ is imposed by the requirement that a sphere S^j in general position must not intersect N^k . Injectivity of ι for $j \leq d - k - 2$ is imposed by the condition that a homotopy $S^j \times [0, 1]$ in general position must not intersect N^k . This observation (3-5) generalizes a result of Sakai [26, Proposition 4.5(1)].

The inclusion $i: \text{Cu}(N) \hookrightarrow M$ induces a long exact sequence in homology

$$\cdots \rightarrow H_j(\text{Cu}(N)) \xrightarrow{i_*} H_j(M) \rightarrow H_j(M, \text{Cu}(N)) \xrightarrow{\partial} H_{j-1}(\text{Cu}(N)) \rightarrow \cdots .$$

As $(M, \text{Cu}(N))$ is a good pair (see Lemma 3.18), we replace the relative homology of $(M, \text{Cu}(N))$ with the reduced homology of $M/\text{Cu}(N) \cong \text{Th}(v)$. This results in the

long exact sequence

$$(3-6) \quad \cdots \rightarrow H_j(\text{Cu}(N)) \xrightarrow{i_*} H_j(M) \xrightarrow{q} \tilde{H}_j(\text{Th}(v)) \xrightarrow{\partial} H_{j-1}(\text{Cu}(N)) \rightarrow \cdots .$$

If $N = \{p\}$ is a point, then $\text{Th}(v) = S^d$ and (3-6) imply isomorphisms

$$i_* : H_j(\text{Cu}(p)) \xrightarrow{\cong} H_j(M), \quad i^* : H^j(M) \xrightarrow{\cong} H^j(\text{Cu}(p))$$

for $j \neq d, d - 1$ (see [26, Proposition 4.5(2)]).

Remark 3.21 The long exact sequence (3-6) can be interpreted as the dual to the long exact sequence in cohomology of the pair (M, N) . If $N = N_1 \sqcup \cdots \sqcup N_l$ is a disjoint union of submanifolds of dimension k_1, \dots, k_l , respectively, then the Thom isomorphism implies that

$$\begin{aligned} \tilde{H}_j(\text{Th}(v)) &\cong \tilde{H}_j(\text{Th}(v_1)) \oplus \cdots \oplus \tilde{H}_j(\text{Th}(v_l)) \\ &\cong H_{j-(d-k_1)}(N_1) \oplus \cdots \oplus H_{j-(d-k_l)}(N_l), \end{aligned}$$

where v_j is the normal bundle of N_j . Applying Poincaré duality to each N_j , we obtain isomorphisms

$$\tilde{H}_j(\text{Th}(v)) \cong \bigoplus_{i=1}^l H^{d-j}(N_i) = H^{d-j}(N).$$

Poincaré–Lefschetz duality applied to the pair (M, N) provides isomorphisms

$$(3-7) \quad \check{H}^j(M, N) \cong H_{d-j}(M - N).$$

As M and N are triangulable, Čech cohomology may be replaced by singular cohomology. Since $M - N$ deforms to $\text{Cu}(N)$ by Lemma 3.18, we have isomorphisms

$$(3-8) \quad H^j(M, N) \cong H_{d-j}(\text{Cu}(N)).$$

Combining all these isomorphisms, we obtain the long exact sequence in cohomology for (M, N) from (3-6).

Lemma 3.22 *Let N be a closed submanifold of M with l components. If M has dimension d , then $H_{d-1}(\text{Cu}(N))$ is free abelian of rank $l - 1$ and $H_{d-j}(\text{Cu}(N)) \cong H^j(M)$ if $j - 2 \geq k$, where k is the maximum of the dimensions of the components of N .*

Proof It follows from (3-7) that

$$H_{d-1}(\text{Cu}(N)) \cong H^1(M, N).$$

Consider the long exact sequence associated to the pair (M, N) ,

$$0 \rightarrow H^0(M, N) \rightarrow H^0(M) \xrightarrow{i^*} H^0(N) \rightarrow H^1(M, N) \rightarrow H^1(M) \rightarrow H^1(N) \rightarrow \dots$$

If N has l components, ie $N = N_1 \sqcup \dots \sqcup N_l$, where N_j has dimension k_j , then $H^1(M, N)$ is torsion-free. This follows from the fact that $i^*(1) = (1, \dots, 1)$ and $H^1(M)$ is free abelian. In particular, if $H^1(M) = 0$, then $H_{d-1}(\text{Cu}(N)) \cong \mathbb{Z}^{l-1}$.

The long exact sequence for the pair (M, N) implies that there are isomorphisms

$$(3-9) \quad H_{d-j}(\text{Cu}(N)) \cong H^j(M, N) \xrightarrow{\cong} H^j(M)$$

if $j \geq k + 2$, where $k = \max\{k_1, \dots, k_l\}$. □

Remark 3.23 The cut locus can be very hard to compute. For a general space, we have the notion of topological dimension. This notion coincides with the usual notion if the space is triangulable. However, Barratt and Milnor [2] proved that the singular homology of a space may be nonzero beyond its topological dimension. Čech (co)homology is better equipped to detect topological dimension and is the reason why one may prefer it over singular homology due to the generic fractal-like nature of cut loci (see the remarks following [Theorem C](#) in [Section 1](#)). Although the topological dimension of $\text{Cu}(N)$ is at most $d - 1$, it is not apparent that $H_{d-1}(\text{Cu}(N))$ is a free abelian group.

There are several applications of this discussion.

Theorem 3.24 *Let N be a smooth homology k -sphere embedded in a Riemannian manifold homeomorphic to S^d . If $d \geq k + 3$, then the cut locus $\text{Cu}(N)$ is homotopy equivalent to S^{d-k-1} .*

Proof As N has codimension at least 3, its complement is path-connected. It follows from (3-5) and [Lemma 3.18](#) that $M - N$ is $(d - k - 2)$ -connected. In particular, $M - N$ is simply connected and, by the Hurewicz isomorphism, $H_j(M - N) = 0$ if $j \leq d - k - 2$. Note that $H_d(M - N) = 0$ as $M - N$ is a noncompact manifold of dimension d .

If $k > 0$, then, by [Lemma 3.22](#), $H_{d-1}(M - N) = 0$. Moreover, by Poincaré–Lefschetz duality (3-7), the only nonzero higher homology of $M - N$ is $H_{d-k-1}(M - N) \cong \mathbb{Z}$. By the Hurewicz theorem, there is an isomorphism $\pi_{d-k-1}(M - N) \cong \mathbb{Z}$. Let

$$\alpha: S^{d-k-1} \rightarrow M - N$$

be a generator. The map α_* induces an isomorphism on all homology groups between two simply connected CW complexes. It follows from Whitehead’s theorem that α is a homotopy equivalence. Using Lemma 3.18, we obtain our homotopy equivalence $H_1 \circ \alpha: S^{d-k-1} \rightarrow \text{Cu}(N)$.

If $k = 0$, then, by Lemma 3.22, $H_{d-1}(M - N) \cong \mathbb{Z}$. Arguments similar to the $k > 0$ case now apply to obtain a homotopy equivalence with S^{d-1} . □

The above result was foreshadowed by Example 2.7, where we showed that the cut locus of $N = \mathbb{S}_i^k$ inside $M = \mathbb{S}^d$ is \mathbb{S}_i^{d-k-1} . It also differs from Poincaré–Lefschetz duality in that we are able to detect the exact homotopy type of the cut locus. In fact, when M and N are real analytic and the embedding is also real analytic, then, by Theorem 3.9, we infer that $\text{Cu}(N)$ is a simplicial complex of dimension at most $d - 1$. Towards this direction, Theorem 3.24 can be pushed further.

Proposition 3.25 *Let N be a real analytic homology k -sphere embedded in a real analytic homology d -sphere M . If $d \geq k + 3$, then the cut locus $\text{Cu}(N)$ is a simplicial complex of dimension at most $d - 1$, having the homology of the $(d - k - 1)$ -sphere with fundamental group isomorphic to that of M .*

The proof of this is a combination of ideas used in the proof of Theorem 3.24. The homotopy type cannot be deduced here due to the presence of a nontrivial fundamental group. An intriguing example can be obtained by combining Proposition 3.25 and the Poincaré homology sphere.

Example 3.26 (cut locus of 0-sphere in the Poincaré sphere) Let \tilde{I} be the binary icosahedral group. It is a double cover of I , the icosahedral group, and can be realized a subgroup of $\text{SU}(2)$. It is known that $H_1(\tilde{I}; \mathbb{Z}) = H_1(I; \mathbb{Z}) = 0$, ie it is perfect and the second homology of the classifying space $B\tilde{I}$ is zero. A presentation of \tilde{I} is given by

$$\tilde{I} = \langle s, t \mid (st)^2 = s^3 = t^5 \rangle.$$

In fact, if we construct a cell complex X of dimension 2 using the presentation above, then X has one 0-cell, two 1-cells and two 2-cells. The cellular chain complex, as computed from the presentation, is given by

$$0 \rightarrow \mathbb{Z}^2 \xrightarrow{\begin{pmatrix} -1 & 2 \\ 3 & -5 \end{pmatrix}} \mathbb{Z}^2 \xrightarrow{0} \mathbb{Z} \rightarrow 0.$$

Therefore, $H_1(X) = H_2(X) = 0$ while $\pi_1(X) = \tilde{I}$.

In contrast, consider the cut locus C of the 0–sphere in $SU(2)/\tilde{I}$, the Poincaré homology sphere. As $SU(2)$ is real analytic, so is the homology sphere. By [Proposition 3.25](#), C is a finite, connected simplicial complex of dimension 2 such that $\pi_1(C) \cong \tilde{I}$ and $H_*(C; \mathbb{Z}) \cong H_*(S^2; \mathbb{Z})$. The existence of this space is interesting for the following reason: although $X \vee S^2$ has the same topological invariants we are unable to determine whether $X \vee S^2$ is homotopy equivalent to C .

In the codimension two case, we have two results.

Theorem 3.27 *Let Σ be a closed, orientable, real analytic surface of genus g and N a nonempty finite subset. Then $\text{Cu}(N)$ is a connected graph, homotopy equivalent to a wedge product of $|N| + 2g - 1$ circles.*

Proof As $M - N$ is connected, [Lemma 3.18](#) implies that $\text{Cu}(N)$ is connected. It follows from [Theorem 3.9](#) that $\text{Cu}(N)$ is a finite 1–dimensional simplicial complex, ie a finite graph. In this case, $\text{Th}(v)$ is a wedge product of $|N|$ copies of S^2 (see (3-4)). We consider (3-6) with $j = 2$:

$$0 \xrightarrow{i_*} \mathbb{Z} \xrightarrow{q} \tilde{H}_2(\vee_{|N|} S^2) \xrightarrow{\partial} H_1(\text{Cu}(N)) \xrightarrow{i_*} H_1(\Sigma) \rightarrow 0.$$

Note that $H_{d-1}(M)$ is torsion-free, whence all the groups appearing in the long exact sequence are free abelian groups. This implies that

$$\dim_{\mathbb{Z}} H_1(\text{Cu}(N)) = 2g + |N| - 1.$$

As $\text{Cu}(N)$ is a connected finite graph, collapsing a maximal tree T results in a quotient space $\text{Cu}(N)/T$ which is homotopic to $\text{Cu}(N)$ as well as being a wedge product of $|N| + 2g - 1$ circles. □

Remark 3.28 Itoh and Vilcu [15] proved that every finite, connected graph can be realized as the cut locus (of a point) of some surface. There remains the question of orientability of the surface. As noted in the proof of [Theorem 3.27](#), if the surface is orientable and $|N| = 1$, then the graph has an even number of generating cycles. If Σ is nonorientable, then $\Sigma \cong (\mathbb{R}P^2)^{\#k}$ has nonorientable genus k and the oriented double cover of Σ has genus $g = k - 1$. Recall that $H_1(\Sigma) \cong \mathbb{Z}^{k-1} \oplus \mathbb{Z}_2$ and $H_2(\Sigma) = 0$. Looking at (3-6) with $j = 2$, we obtain

$$0 \rightarrow \mathbb{Z} \rightarrow H_1(\text{Cu}(p)) \rightarrow \mathbb{Z}^{k-1} \oplus \mathbb{Z}_2 \rightarrow 0.$$

Thus, $H_1(\text{Cu}(p)) \cong \mathbb{Z}^k$ as homology groups of graphs are free abelian. Let $B_\varepsilon(\text{Cu}(p))$ denote the ε -neighborhood of $\text{Cu}(p)$ in Σ . For ε sufficiently small, this is a surface such that $\overline{B_\varepsilon(\text{Cu}(p))}$ has one boundary component. The compact surface $B_\varepsilon(\text{Cu}(p))$ is reminiscent of ribbon graphs. The surface Σ can be obtained as the connect-sum of a disk centered at p and the closure of $B_\varepsilon(\text{Cu}(p))$. Therefore, nonorientability of Σ is equivalent to nonorientability of $B_\varepsilon(\text{Cu}(p))$. A similar observation appears in the unpublished work of Itoh and Vilcu [14, Theorem 3.7].

Example 3.29 (homology spheres of codimension two) In continuation of Theorem 3.24, let $N \hookrightarrow S^{k+2}$ be a homology sphere of dimension $k \geq 1$. Since N has codimension two, $S^{k+2} - N$ is path-connected and so is $\text{Cu}(N)$. We are not assuming that the metric on S^{k+2} is real analytic. Using (3-8) and the long exact sequence in cohomology of (M, N) , we infer that $H_1(\text{Cu}(N)) \cong \mathbb{Z}$ and all higher homology groups vanish. However, the Hurewicz theorem cannot be used here to establish that $\pi_1(\text{Cu}(N)) \cong \mathbb{Z}$.

In particular cases, we may conclude that $\text{Cu}(N)$ is homotopic to a circle. It was proved by Plotnick [22] that certain homology 3–spheres N , obtained by a Dehn surgery of type $1/2a$ on a knot, smoothly embed in S^5 with complement a homotopy circle. Since $M - N$ deforms to $\text{Cu}(N)$, it follows that there is a map $\alpha: S^1 \rightarrow \text{Cu}(N)$ inducing isomorphisms on homotopy and homology groups.

If $k = 1$, then a homology 1–sphere is just a knot K in S^3 . Since $S^3 - K$ deforms to $\text{Cu}(K)$, the fundamental group of the cut locus is the knot group. Moreover, in the case of real analytic knots in S^3 , the cut locus is a finite simplicial complex of dimension at most 2 (see Theorem 3.9). Except for the unknot, the knot group is never a free group, while the fundamental group of a connected, finite graph is free. This observation establishes that $\text{Cu}(K)$ is always a 2–dimensional simplicial complex whenever K is a nontrivial (real analytic) knot in S^3 .

3.3 Morse–Bott function associated to distance function

We first prove that the closure of $\text{Se}(N)$ is the cut locus, closely following the proof given in [29] for the case of a point.

Theorem 3.30 *Let $\text{Cu}(N)$ be the cut locus of a compact submanifold N of a complete Riemannian manifold M . The subset $\text{Se}(N)$ of $\text{Cu}(N)$ is dense in $\text{Cu}(N)$.*

Proof Let $q \in \text{Cu}(N)$ but not in $\text{Se}(N)$. Choose an N –geodesic γ , joining N to q , such that any extension of γ is not an N –geodesic. This geodesic γ is unique as $q \notin \text{Se}(N)$.

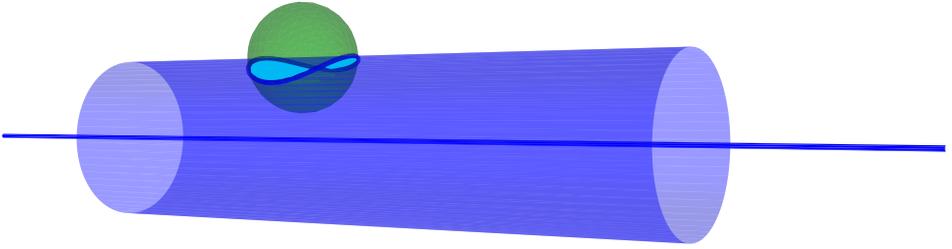


Figure 5: $\text{Co}(x_0, \delta)$.

We may write $\gamma(t) = \exp_v(tx)$, where $\gamma(0) = p \in N$ and $\gamma'(0) = x_0 \in S(v_p)$. It follows from the definition of s that $q = \exp_v(s(x_0)x_0)$. We need to show that every neighborhood of q in $\text{Cu}(N)$ must intersect $\text{Se}(N)$. Suppose it is false. Let $\delta > 0$ and consider $\overline{B}(x_0, \delta)$, the closed ball with center x_0 and radius δ . Define the cone

$$\text{Co}(x_0, \delta) := \{tx : 0 \leq t \leq 1, x \in \overline{B}(x_0, \delta) \cap S(v)\};$$

see Figure 5. Since $\overline{B}(x_0, \delta) \cap S(v)$ is homeomorphic to a closed $(n-1)$ -ball for sufficiently small δ , the cone will be homeomorphic to a closed Euclidean n -ball. Similarly, define another cone

$$\text{Co}^*(x_0, \delta) := \left\{ s\left(\frac{x}{\|x\|}\right)x \mid x \in \text{Co}(x_0, \delta), x \neq 0 \right\} \cup \{0\}.$$

Note that $s(x_0)$ is finite. As s is continuous, due to Proposition 3.14, for sufficiently small δ the term $s(x/\|x\|)$ is still finite, whence $\text{Co}^*(x_0, \delta)$ is well defined. We claim that $\text{Co}^*(x_0, \delta)$ is also homeomorphic to a closed Euclidean $(n-k)$ -ball. Indeed, a nonzero $x \in \text{Co}(x_0, \delta)$ implies $x = \lambda \hat{x}$ for some $\lambda \in (0, 1]$ and $\hat{x} \in \overline{B}(x_0, \delta) \cap S(v)$. Since $s(\hat{x})x = \lambda s(\hat{x})\hat{x}$, it follows that $\text{Co}^*(x_0, \delta)$ is the cone of the set

$$\{s(\hat{x})\hat{x} \mid \hat{x} \in \overline{B}(x_0, \delta) \cap S(v)\},$$

which is homeomorphic to $\overline{B}(x_0, \delta) \cap S(v)$. Now we have a dichotomy:

- (a) for a fixed small $\delta > 0$, the restriction of \exp_v to $\text{Co}^*(x_0, \delta)$ is a homeomorphism to its image because it is injective, or
- (b) for any $\delta > 0$, the restriction of \exp_v to $\text{Co}^*(x_0, \delta)$ is not injective.

If (b) holds, choose $v_n \neq w_n \in \text{Co}^*(x_0, 1/n)$ such that these map to q_n under \exp_v . Thus, $q_n \in \text{Se}(N)$ and compactness of $S(v)$ ensures that q_n converges to q . If (a) holds, then let $B(q, \varepsilon)$ denote the open ball in M centered at q with radius $\varepsilon > 0$. We claim that it intersects the complement of $\exp_v(\text{Co}^*(x_0, \delta))$ in M . But it is true as $s(x_0)x_0$ lies on the boundary of $\text{Co}^*(x_0, \delta)$ and hence it has a neighborhood in $\text{Co}^*(x_0, \delta)$

which is homeomorphic to a closed n -dimensional Euclidean half plane. Since \exp_ν , restricted to $\text{Co}^*(x_0, \delta)$ is a homeomorphism, the open ball $B(q, \varepsilon)$ must intersect the points outside the image of $\exp_\nu(\text{Co}^*(x_0, \delta))$.

Now take $\varepsilon = 1/n$. For each n , there exists $q_n \in B(q, 1/n)$ with $q_n \notin \exp_\nu(\text{Co}^*(x_0, \delta))$. Since M is complete, for each point q_n let γ_n be an N -geodesic joining $p_n \in N$ to q_n . We may invoke the following result from Busemann’s book [5, Theorem 5.16, page 24]. Let $\{\gamma_n\}$ be a sequence of rectifiable curves in a finitely compact set X such that the lengths $\ell(\gamma_n)$ are bounded. If the initial points p_n of γ_n form a bounded set, then $\{\gamma_n\}$ contains a subsequence γ_{n_k} which converges uniformly to a rectifiable curve $\tilde{\gamma}$ in X and

$$\ell(\tilde{\gamma}) \leq \liminf \ell(\gamma_{n_k}).$$

Since $\{p_n\}$ lie in the compact set N , we obtain a rectifiable curve $\tilde{\gamma}$ such that

$$\ell(\tilde{\gamma}) \leq \liminf \ell(\gamma_{n_k}) = \lim_k \ell(\gamma_{n_k}) = \lim_k d_N(q_{n_k}) = d_N(q).$$

Thus, $\tilde{\gamma}$ is actually an N -geodesic joining $p' = \lim_k p_{n_k}$ to q and the unit tangent vectors $x_{n_k} = \gamma'_{n_k}(0)$ at p_{n_k} converges to the unit tangent vector $\tilde{x} = \tilde{\gamma}'(0)$ at p' . Since x_0 is an interior point of the set $\overline{B(x_0, \delta)} \cap S(\nu)$, any sequence in $S(\nu)$ converging to x_0 must eventually lie in $\text{Co}(x_0, \delta)$. According to our choice, $q_{n_k} \notin \exp_\nu(\text{Co}^*(x_0, \delta))$ and the x_{n_k} all lie outside of $\text{Co}(x_0, \delta)$. Hence, $x_0 \neq \tilde{x}$ and $\gamma \neq \tilde{\gamma}$. Thus, there are two distinct N -geodesics γ and $\tilde{\gamma}$ joining N to q , a contradiction to $q \notin \text{Se}(N)$. \square

We have seen (in Lemma 3.7) that d_N^2 is smooth away from the cut locus. It follows from Theorem 3.30 that the cut locus is the closure of the singularity of d_N^2 . The following example suggests that d_N^2 can be differentiable at points in $\text{Cu}(N) - \text{Se}(N)$ but not twice differentiable:

Example 3.31 (cut locus of an ellipse) We discuss the regularity of the distance-squared function from an ellipse $x^2/a^2 + y^2/b^2 = 1$ (with $a > b > 0$) in \mathbb{R}^2 . For a discussion of the cut locus for ellipses inside \mathbb{S}^2 and ellipsoids, see Hebda [10, pages 90–91]. Let (x_0, y_0) be a point inside the ellipse lying in the first quadrant. The point closest to (x_0, y_0) and lying on the ellipse is given by

$$x = \frac{a^2 x_0}{t + a^2}, \quad y = \frac{b^2 y_0}{t + b^2},$$

where t is the unique root of the quartic

$$\left(\frac{ax_0}{t+a^2}\right)^2 + \left(\frac{by_0}{t+b^2}\right)^2 = 1$$

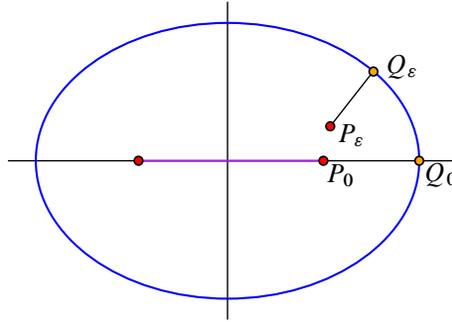


Figure 6: Cut locus of an ellipse.

in the interval $(-b^2, \infty)$. Given (α, β) with $\beta > 0$, we set

$$P_\varepsilon(\alpha, \beta) = \left(\frac{a^2 - b^2}{a} + \varepsilon\alpha, \varepsilon\beta \right);$$

this defines a straight line passing through $P_0(\alpha, \beta)$ in the direction of (α, β) . For $\varepsilon > 0$, $P_\varepsilon(\alpha, \beta)$ lies in the first quadrant and we denote by $t = t(\varepsilon)$ the unique relevant root of the quartic

$$\left(\frac{a((a^2 - b^2)/a + \varepsilon\alpha)}{t + a^2} \right)^2 + \left(\frac{b\varepsilon\beta}{t + b^2} \right)^2 = 1.$$

Simplifying this after dividing by ε and taking a limit $\varepsilon \rightarrow 0^+$, we obtain

$$\frac{2a\alpha}{a^2 - b^2} = \lim_{\varepsilon \rightarrow 0^+} \left(\left(\frac{2}{a^2 - b^2} \right) \frac{t + b^2}{\varepsilon} - b^2\beta^2 \frac{\varepsilon}{(t + b^2)^2} \right).$$

On the other hand, the point $Q_\varepsilon(\alpha, \beta)$ on the ellipse closest to $P_\varepsilon(\alpha, \beta)$ is given by

$$x_\varepsilon = \frac{a^2((a^2 - b^2)/a + \varepsilon\alpha)}{t + a^2}, \quad y_\varepsilon = \frac{b^2\varepsilon\beta}{t + b^2}.$$

It follows that

$$(3-10) \quad d_\varepsilon^2(\alpha, \beta) := d^2(P_\varepsilon, Q_\varepsilon) = \frac{t^2}{a^2} \left(\frac{a^2 - b^2 + a\varepsilon\alpha}{t + a^2} \right)^2 + \frac{t^2}{b^2} \left(\frac{b\varepsilon\beta}{t + b^2} \right)^2.$$

Using $t(0) = -b^2$, simplifications lead us to

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \frac{d_\varepsilon^2 - d_0^2}{\varepsilon} &= \frac{2ab^4\alpha}{a^2(a^2 - b^2)} - \lim_{\varepsilon \rightarrow 0^+} \left(\frac{(t + b^2)(a^2b^2 - a^2t + 2b^2t)}{\varepsilon(t + a^2)^2} - \beta^2 \frac{t^2\varepsilon}{(t + b^2)^2} \right) \\ &= \frac{2ab^4\alpha}{a^2(a^2 - b^2)} - \frac{2b^2}{a^2 - b^2} \lim_{\varepsilon \rightarrow 0} \frac{t + b^2}{\varepsilon} + \beta^2 b^4 \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon}{(t + b^2)^2} \\ &= \frac{2ab^4\alpha}{a^2(a^2 - b^2)} - \frac{2ab^2\alpha}{a^2 - b^2} = -\frac{2b^2\alpha}{a}. \end{aligned}$$

On the other hand, for $\varepsilon < 0$, the point $P_\varepsilon(\alpha, \beta)$ lies in the fourth quadrant. By symmetry, the distance between $P_\varepsilon(\alpha, \beta)$ and $Q_\varepsilon(\alpha, \beta)$ is the same as that between $P_{-\varepsilon}(-\alpha, \beta)$ and $Q_{-\varepsilon}(-\alpha, \beta)$. However, it is seen that

$$d^2(P_{-\varepsilon}(-\alpha, \beta), Q_{-\varepsilon}(-\alpha, \beta)) = d_{-\varepsilon}^2(-\alpha, \beta),$$

as defined in (3-10). Therefore,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^-} \frac{d^2(P_\varepsilon(\alpha, \beta), Q_\varepsilon(\alpha, \beta)) - d^2(P_0(\alpha, \beta), Q_0(\alpha, \beta))}{\varepsilon} &= \lim_{\varepsilon \rightarrow 0^-} \frac{d_{-\varepsilon}^2(-\alpha, \beta) - d_0^2(-\alpha, \beta)}{\varepsilon} \\ &= - \lim_{-\varepsilon \rightarrow 0^+} \frac{d_{-\varepsilon}^2(-\alpha, \beta) - d_0^2(-\alpha, \beta)}{-\varepsilon} \\ &= -\frac{2b^2\alpha}{a}, \end{aligned}$$

where the last equality follows from the right-hand derivative of d^2 , as computed previously.

When $\beta = 0$, we would like to compute $d_\varepsilon^2(\alpha, 0)$. If $\varepsilon > 0$, then

$$(3-11) \quad d_\varepsilon^2(\alpha, 0) = (b^2/a - \varepsilon\alpha)^2 = \frac{b^4}{a^2} - \frac{2b^2\alpha\varepsilon}{a} + \alpha^2\varepsilon^2.$$

On the other hand, if $\varepsilon < 0$ is sufficiently small, then there are two points on the ellipse closest to $P_\varepsilon(\alpha, 0) = ((a^2 - b^2)/a + \varepsilon\alpha, 0)$, with exactly one on the first quadrant, say Q_ε . Since the segment $P_\varepsilon Q_\varepsilon$ must be orthogonal to the tangent to the ellipse at Q_ε , we obtain the coordinates for Q_ε :

$$x_\varepsilon = \frac{a^2((a^2 - b^2)/a + \varepsilon\alpha)}{a^2 - b^2}, \quad y_\varepsilon^2 = b^2 \left(1 - \frac{x_\varepsilon^2}{a^2} \right), \quad y_\varepsilon > 0.$$

We may compute the distance

$$(3-12) \quad d_\varepsilon^2(\alpha, 0) := d^2(P_\varepsilon, Q_\varepsilon) = \frac{b^4}{a^2} - \frac{2b^2\alpha\varepsilon}{a} - \frac{b^2\alpha^2\varepsilon^2}{a^2 - b^2},$$

where $\varepsilon < 0$. Combining (3-11) and (3-12), we conclude that d^2 is differentiable at $P_0 = ((a^2 - b^2)/a, 0)$, a point in $\text{Cu}(N)$ but not in $\text{Se}(N)$. However, comparing the quadratic part of d^2 in (3-11)–(3-12), we conclude that d^2 is not twice differentiable at P_0 .

Theorem 3.32 *Let N be a closed embedded submanifold of a complete Riemannian manifold M . Let $d_N : M \rightarrow \mathbb{R}$ be the distance function with respect to N . If $f = d_N^2$, then its restriction to $M - \text{Cu}(N)$ is a Morse–Bott function, with N as the critical submanifold. Moreover, the gradient flow of f deforms $M - \text{Cu}(N)$ to N .*

Proof It follows from [Lemma A.2](#) that the map $\exp_v^{-1} : M - (\text{Cu}(N) \cup N) \rightarrow v - \{0\}$ is an (into) diffeomorphism and $d_N(q) = \|\exp_v^{-1}(q)\|$ and hence the distance function is of class C^∞ at $q \in M - (\text{Cu}(N) \cup N)$. Using Fermi coordinates (see [Proposition 3.5](#)), we have seen that the distance-squared function is smooth around N and therefore it is smooth on $M - \text{Cu}(N)$. By [Corollary 3.6](#), the Hessian of this function at N is nondegenerate in the normal direction. It is well known [[26](#), [Proposition 4.8](#)] that $\|\nabla d(q)\| = 1$ if d_N is differentiable at $q \in M$. Thus, for $q \in M - (\text{Cu}(N) \cup N)$, we have

$$(3-13) \quad \|\nabla f(q)\| = 2d_N(q)\|\nabla d_N(q)\| = 2d_N(q).$$

Let γ be the unique unit-speed N -geodesic that joins N to q , ie

$$\gamma : [0, d_N(q)] \rightarrow M, \quad \gamma(0) = p, \quad \gamma(d_N(q)) = q, \quad \|\gamma'\| = 1.$$

We may write $\nabla f(q) = \lambda \gamma'(d_N(q)) + w$, where w is orthogonal to $\gamma'(d_N(q))$. But

$$\begin{aligned} \langle \nabla f|_q, \gamma'(d_N(q)) \rangle &= \left. \frac{d}{dt} f(\gamma(d_N(q) + t)) \right|_{t=0} = \left. \frac{d}{dt} (d_N(q)^2 + 2d_N(q)t + t^2) \right|_{t=0} \\ &= 2d_N(q). \end{aligned}$$

Thus, $\lambda = 2d(q)$ and, combined with [\(3-13\)](#), we conclude that

$$\nabla f(q) = 2d_N(q)\gamma'(d_N(q)).$$

Therefore, the negative gradient flow line initialized at $q \in M - \text{Cu}(N)$ is given by

$$\eta(t) = \gamma(d_N(q)e^{-2t}).$$

These flow lines define a flow which deforms $M - \text{Cu}(N)$ to N in infinite time. □

The reader may choose to revisit the example of $\text{GL}(n, \mathbb{R})$ discussed in [Section 2.2](#) and treat it as a concrete illustration of the theorem above.

4 Applications to Lie groups

Due to classical results of Cartan, Iwasawa and others, we know that any connected Lie group G is diffeomorphic to the product of a maximally compact subgroup K and an Euclidean space. In particular, G deforms to K . For semisimple groups, this decomposition is stronger and is attributed to Iwasawa. The Killing form on the Lie algebra \mathfrak{g} is nondegenerate and negative-definite for compact semisimple Lie algebras.

For such a Lie group G , consider the Levi-Civita connection associated to the bi-invariant metric obtained from the negative of the Killing form. This connection coincides with the Cartan connection.

We consider two examples, both of which are noncompact and nonsemisimple. We prove that these Lie groups G deformation retract to maximally compact subgroups K via gradient flows of appropriate Morse–Bott functions. This requires a choice of a left-invariant metric which is right- K -invariant and a careful analysis of the geodesics associated with the metric. In particular, we provide a possibly new proof of the surjectivity of the exponential map for $U(p, q)$.

4.1 Invertible matrices with positive determinant

Let g be a left-invariant metric on $GL(n, \mathbb{R})$, the set of invertible matrices. Recall that a left-invariant metric g on a Lie group is determined by its restriction at the identity. For $A \in GL(n, \mathbb{R})$, consider the left multiplication map $l_A: GL(n, \mathbb{R}) \rightarrow GL(n, \mathbb{R})$, $B \mapsto AB$. This extends to a linear isomorphism from $M(n, \mathbb{R})$ to itself. Thus, the differential $(Dl_A)_I: T_I GL(n, \mathbb{R}) \rightarrow T_A GL(n, \mathbb{R})$ is an isomorphism and given by l_A itself. For $X, Y \in T_I GL(n, \mathbb{R})$,

$$g_I(X, Y) = g_A((Dl_A)_I X, (Dl_A)_I Y) = g_A(AX, AY).$$

We choose the left-invariant metric on $GL(n, \mathbb{R})$ generated by the Euclidean metric at I . Therefore,

$$g_{A^{-1}}(X, Y) = \langle AX, AY \rangle_I := \text{tr}((AX)^T AY) = \text{tr}(X^T A^T AY).$$

Note that this metric is right- $O(n, \mathbb{R})$ -invariant. We are interested in the distance between an invertible matrix A (with $\det(A) > 0$) and $SO(n, \mathbb{R})$. Since $SO(n, \mathbb{R})$ is compact, there exists $B \in SO(n, \mathbb{R})$ such that $d(A, B) = d_{SO(n, \mathbb{R})}(A)$.

Lemma 4.1 *If D is a diagonal matrix with positive diagonal entries $\lambda_1, \dots, \lambda_n$, then*

$$d_{SO(n, \mathbb{R})}(D) = d(D, I).$$

Moreover, I is the unique minimizer and the associated minimal geodesic is given by $\gamma(t) = e^{t \log D}$.

Proof Choose $B \in SO(n, \mathbb{R})$ satisfying $d(A, B) = d_{SO(n, \mathbb{R})}(A)$. Since, with respect to the left-invariant metric, $GL^+(n, \mathbb{R})$ is complete, there exists a minimal geodesic $\gamma: [0, 1] \rightarrow GL^+(n, \mathbb{R})$ joining B to D , ie

$$\gamma(0) = B, \quad \gamma(1) = D \quad \text{and} \quad \ell(\gamma) = d(D, B).$$

The first variational principle implies that $\gamma'(0)$ is orthogonal to $T_B\text{SO}(n, \mathbb{R})$. It follows from Martin and Neff [19, Section 2.1] that $\eta(t) = e^{tW}$ is a geodesic if W is a symmetric matrix. Moreover, $\eta'(0) = W$ is orthogonal to $T_I\text{SO}(n, \mathbb{R})$. As left translation is an isometry and isometry preserves geodesic, it follows that $\gamma(t) = Be^{tW}$ is a geodesic with $\gamma'(0)$ orthogonal to $T_B\text{SO}(n, \mathbb{R})$. By the defining properties of γ , $D = \gamma(1) = Be^W$. Since e^W is symmetric positive-definite, we obtain two polar decompositions of D , ie $D = ID$ and $D = Be^W$. By the uniqueness of the polar decomposition for invertible matrices, $B = I$ and $D = e^W$.

In order to compute $d(I, D)$, note that

$$e^W = D = e^{\log D},$$

where $\log D$ denotes the diagonal matrix with entries $\log \lambda_1, \dots, \log \lambda_n$. As W and $\log D$ are symmetric, and matrix exponential is injective on the space of symmetric matrices, we conclude that $W = \log D$. The geodesic is given by $\gamma(t) = e^{t \log D}$ and

$$(4-1) \quad d_{\text{SO}(n, \mathbb{R})}(D) = \|\gamma'(0)\|_I = \|\log D\|_I = \left(\sum_{i=1}^n (\log \lambda_i)^2 \right)^{1/2}.$$

Thus, the distance-squared function will be given by $\sum_{i=1}^n (\log \lambda_i)^2$. □

Now, for any $A \in \text{GL}^+(n, \mathbb{R})$, we can apply the SVD decomposition, ie $A = UDV^T$ with $\sqrt{A^T A} = VDV^T$ and $\log \sqrt{A^T A} = V(\log D)V^T$. Note that $U, V \in \text{SO}(n, \mathbb{R})$ and D is a diagonal matrix with positive entries. The left-invariant metric is right-invariant with respect to orthogonal matrices. Thus,

$$d_{\text{SO}(n, \mathbb{R})}(A) = d_{\text{SO}(n, \mathbb{R})}(D) = \|\log D\|_I,$$

where the last equality follows from the lemma (see (4-1)). As

$$\|\log D\|_I = \|V(\log D)V^T\|_I = \|\log \sqrt{A^T A}\|_I,$$

It follows from the arguments of the lemma and the metric being bi- $O(n, \mathbb{R})$ -invariant that

$$\gamma(t) = Ue^{t \log D}V^T$$

is a minimal geodesic joining UV^T to A , realizing $d_{\text{SO}(n, \mathbb{R})}(A)$. As the minimizer UV^T is unique, $\text{Se}(\text{SO}(n, \mathbb{R}))$ is empty, implying that $\text{Cu}(\text{SO}(n, \mathbb{R}))$ is empty as well. In fact, $UV^T = A\sqrt{A^T A}^{-1}$ and

$$(4-2) \quad \gamma(t) = Ue^{t \log D}V^T = UV^T V e^{t \log D} V^T = A\sqrt{A^T A}^{-1} e^{t \log \sqrt{A^T A}}.$$

If we compare (2-6) — the deformation of $GL(n, \mathbb{R})$ to $O(n, \mathbb{R})$ inside $M(n, \mathbb{R})$ — with (4-2), then, in both cases, an invertible matrix A deforms to $A\sqrt{A^T A}^{-1}$. Finally, observe that the normal bundle of $SO(n, \mathbb{R})$ is diffeomorphic to $GL^+(n, \mathbb{R})$.

4.2 Indefinite unitary groups

Let n be a positive integer with $n = p + q$. Consider the inner product on \mathbb{C}^n given by

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n) \rangle = z_1 \bar{w}_1 + \dots + z_p \bar{w}_p - z_{p+1} \bar{w}_{p+1} - \dots - z_n \bar{w}_n.$$

This is given by the matrix $I_{p,q}$ as

$$\langle w, z \rangle = \bar{w}^t I_{p,q} z = (\bar{w}_1 \ \dots \ \bar{w}_n) \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}.$$

Let $U(p, q)$ denote the subgroup of $GL(n, \mathbb{C})$ preserving this indefinite form, ie $\mathcal{A} \in U(p, q)$ if and only if $\mathcal{A}^* I_{p,q} \mathcal{A} = I_{p,q}$. In particular, $\det \mathcal{A}$ is a complex number of unit length. By convention, $I_{n,0} = I_n$ and $I_{0,n} = -I_n$, both of which correspond to $U(n, 0) = U(n) = U(0, n)$, the unitary group. In all other cases, the inner product is indefinite.

The group $U(1, 1)$ is given by matrices of the form

$$\mathcal{A} = \begin{pmatrix} \alpha & \beta \\ \lambda \bar{\beta} & \lambda \bar{\alpha} \end{pmatrix} \quad \text{with } \lambda \in S^1 \text{ and } |\alpha|^2 - |\beta|^2 = 1.$$

More generally, we shall use

$$\mathcal{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

to denote an element of $U(p, q)$. It follows from the definition that $\mathcal{A} \in U(p, q)$ if and only if

$$A^* A - C^* C = I_p, \quad A^* B - C^* D = 0_{p \times q}, \quad B^* B - D^* D = -I_q.$$

Observe that, if $Av = 0$, then

$$0 = A^* Av = C^* C v + v,$$

which implies that $C^* C$, a positive semidefnite matrix, has -1 as an eigenvalue unless $v = 0$. Therefore, A is invertible and the same argument works for D .

Lemma 4.2 *The intersection of $U(p + q)$ with $U(p, q)$ is $U(p) \times U(q)$. Moreover, if $\mathcal{A} \in U(p, q)$, then $\mathcal{A}^*, \sqrt{\mathcal{A}^* \mathcal{A}} \in U(p, q)$.*

Proof If $\mathcal{A} \in U(p) \times U(q)$, then

$$A^*A + C^*C = I_p, \quad B^*B + D^*D = I_q.$$

This implies that both B and C are zero matrices. If $\mathcal{A} \in U(p, q)$, then $\mathcal{A}^* = I_{p,q}\mathcal{A}^{-1}I_{p,q}$ and

$$\begin{aligned} (\mathcal{A}^*\mathcal{A})^* I_{p,q}(\mathcal{A}^*\mathcal{A}) &= (\mathcal{A}^*\mathcal{A})I_{p,q}(\mathcal{A}^*\mathcal{A}) = I_{p,q}\mathcal{A}^{-1}I_{p,q}\mathcal{A}I_{p,q}I_{p,q}\mathcal{A}^{-1}I_{p,q}\mathcal{A} \\ &= I_{p,q} = \mathcal{A}^* I_{p,q}\mathcal{A}. \end{aligned}$$

This also implies that $\mathcal{A}I_{p,q}\mathcal{A}^* = I_{p,q}$.

All the eigenvalues of $\mathcal{A}^*\mathcal{A}$ are positive. Moreover, if λ is an eigenvalue of $\mathcal{A}^*\mathcal{A}$ with eigenvector $\mathbf{v} = (v_1, \dots, v_p, v_{p+1}, \dots, v_n)$, then

$$I_{p,q}\mathbf{v} = \mathcal{A}^*\mathcal{A}I_{p,q}\mathcal{A}^*\mathbf{v} = \lambda(\mathcal{A}^*\mathcal{A}I_{p,q}\mathbf{v}),$$

which implies that λ^{-1} is also an eigenvalue with eigenvector

$$\mathbf{v}' = (v_1, \dots, v_p, -v_{p+1}, \dots, -v_n).$$

If $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an eigenbasis of $\mathcal{A}^*\mathcal{A}$ with (possibly repeated) eigenvalues $\lambda_1, \dots, \lambda_n$, then

$$\sqrt{\mathcal{A}^*\mathcal{A}}I_{p,q}\sqrt{\mathcal{A}^*\mathcal{A}}\mathbf{v}_j = \sqrt{\mathcal{A}^*\mathcal{A}}I_{p,q}\sqrt{\lambda_j}\mathbf{v}_j = \sqrt{\lambda_j}\sqrt{\mathcal{A}^*\mathcal{A}}\mathbf{v}'_j = \mathbf{v}'_j = I_{p,q}\mathbf{v}_j.$$

Thus, $\sqrt{\mathcal{A}^*\mathcal{A}}$ satisfies the defining relation for a matrix to be in $U(p, q)$. □

We may use the polar decomposition (for matrices in $GL(n, \mathbb{C})$) to write

$$\mathcal{A} = U|\mathcal{A}|, \quad \text{where } U = \mathcal{A}\sqrt{\mathcal{A}^*\mathcal{A}}^{-1} \text{ and } |\mathcal{A}| = \sqrt{\mathcal{A}^*\mathcal{A}},$$

where $U, |\mathcal{A}| \in U(p, q)$. For $U(1, 1)$, this decomposition takes the form

$$\begin{pmatrix} \alpha & \beta \\ \lambda\bar{\beta} & \lambda\bar{\alpha} \end{pmatrix} = \begin{pmatrix} \alpha/|\alpha| & 0 \\ 0 & \lambda\bar{\alpha}/|\alpha| \end{pmatrix} \begin{pmatrix} |\alpha| & |\alpha|\beta/\alpha \\ |\alpha|\beta/\bar{\alpha} & |\alpha| \end{pmatrix}.$$

The Lie algebra $\mathfrak{u}_{p,q}$ is given by matrices $X \in M_n(\mathbb{C})$ such that

$$X^*I_{p,q} + I_{p,q}X = 0.$$

This is a real Lie subalgebra of $M_{p+q}(\mathbb{C})$. It contains the subalgebras \mathfrak{u}_p and \mathfrak{u}_q as Lie algebras of the subgroups $U(p) \times I_q$ and $I_p \times U(q)$. Consider the inner product

$$\langle \cdot, \cdot \rangle : \mathfrak{u}_{p,q} \times \mathfrak{u}_{p,q} \rightarrow \mathbb{R}, \quad \langle X, Y \rangle := \text{tr}(X^*Y).$$

Lemma 4.3 *The inner product is symmetric and positive-definite.*

Proof Note that

$$\langle X, Y \rangle = \text{tr}(-I_{p,q} X I_{p,q} Y) = \text{tr}(-I_{p,q} Y I_{p,q} X) = \langle Y, X \rangle.$$

Since $\overline{\langle X, Y \rangle} = \langle Y, X \rangle$ due to the invariance of trace under transpose, we conclude that the inner product is real and symmetric. It is positive-definite as $\langle X, X \rangle = \text{tr}(X^* X) \geq 0$ and equality holds if and only if X is the zero matrix. \square

The Riemannian metric obtained by left translations of $\langle \cdot, \cdot \rangle$ will also be denoted by $\langle \cdot, \cdot \rangle$. We shall analyze the geodesics for this metric. The Lie algebra $\mathfrak{u}_p \oplus \mathfrak{u}_q$ of $U(p) \times U(q)$ consists of matrices

$$\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \quad \text{with } A + A^* = 0 \text{ and } D + D^* = 0.$$

Let \mathfrak{n} denote the orthogonal complement of $\mathfrak{u}_p \oplus \mathfrak{u}_q$ inside $\mathfrak{u}_{p,q}$. As \mathfrak{n} is of (complex) dimension pq and

$$\left\{ \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \mid B \in M_{p,q}(\mathbb{C}) \right\}$$

is contained in \mathfrak{n} , this is all of it. We may verify that

$$\begin{aligned} \left[\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \right] &= \begin{pmatrix} 0 & AB - BD \\ DB^* - B^*A & 0 \end{pmatrix} \in \mathfrak{n}, \\ \left[\begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}, \begin{pmatrix} 0 & C \\ C^* & 0 \end{pmatrix} \right] &= \begin{pmatrix} BC^* - CB^* & 0 \\ 0 & B^*C - C^*B \end{pmatrix} \in \mathfrak{u}_p \oplus \mathfrak{u}_q. \end{aligned}$$

Lemma 4.4 *Let γ be the integral curve, initialized at e , for a left-invariant vector field Y . This curve is a geodesic if $Y(e)$ belongs either to \mathfrak{n} or to $\mathfrak{u}_p \oplus \mathfrak{u}_q$.*

Proof The Levi-Civita connection ∇ is given by the Koszul formula

$$2\langle X, \nabla_Z Y \rangle = Z\langle X, Y \rangle + Y\langle X, Z \rangle - X\langle Y, Z \rangle + \langle Z, [X, Y] \rangle + \langle Y, [X, Z] \rangle - \langle X, [Y, Z] \rangle.$$

Putting $Z = Y$ and $Z = X$, two left-invariant vector fields, in the above, we obtain

$$\langle X, \nabla_Y Y \rangle = \langle Y, [X, Y] \rangle.$$

To prove our claim, it suffices to show that $\nabla_Y Y = 0$, ie $\langle Y, [X, Y] \rangle = 0$ for any X . Let us assume that $Y(e) \in \mathfrak{n}$. If $X(e) \in \mathfrak{n}$, then $[X(e), Y(e)] \in \mathfrak{u}_p \oplus \mathfrak{u}_q$, which implies that

$\langle Y(e), [X(e), Y(e)] \rangle = 0$. If $X(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$, then

$$\begin{aligned} \langle Y, [X, Y] \rangle &= \left\langle \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}, \begin{pmatrix} 0 & AB - BD \\ DB^* - B^*A & 0 \end{pmatrix} \right\rangle \\ &= \text{tr} \begin{pmatrix} B(DB^* - B^*A) & 0 \\ 0 & B^*(AB - BD) \end{pmatrix} \\ &= \text{tr}(BDB^* - BB^*A) + \text{tr}(B^*AB - B^*BD) \\ &= 0 \end{aligned}$$

by the cyclic property of trace. Thus, $\nabla_Y Y = 0$ if $Y(e) \in \mathfrak{n}$; a similar proof works if $Y(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$. □

Remark 4.5 An integral curve of a left-invariant vector field (also called one-parameter subgroups) need not be a geodesic in $U(p, q)$. For instance, if $X + Y$ is a left-invariant vector field given by $X(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$ and $Y(e) \in \mathfrak{n}$, then $\nabla_{X+Y}(X + Y) = 0$ if and only if $\nabla_X Y = \frac{1}{2}[X, Y]$ and $\nabla_Y X = \frac{1}{2}[Y, X]$. This happens if and only if the metric is bi-invariant, ie

$$\langle [X, Z], Y \rangle = \langle X, [Z, Y] \rangle.$$

This is not true; for instance, for $X(e) \in \mathfrak{u}_p \oplus \mathfrak{u}_q$ and linearly independent $Y(e), Z(e) \in \mathfrak{n}$, we get $\langle [X, Z], Y \rangle - \langle X, [Z, Y] \rangle \neq 0$.

Consider the matrix

$$Y = \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \in \mathfrak{n}.$$

Let $B = U\sqrt{B^*B}$ and $B^* = \sqrt{B^*B}U^*$ be polar decompositions, where U and U^* are partial isometries. It follows from direct computation that

$$\begin{aligned} e^Y &= \begin{pmatrix} I_p + \frac{1}{2!}BB^* + \frac{1}{4!}(BB^*)^2 + \dots & B + \frac{1}{3!}B(B^*B) + \frac{1}{5!}B(B^*B)^2 + \dots \\ B^* + \frac{1}{3!}(B^*B)B^* + \frac{1}{5!}(B^*B)^2B^* + \dots & I_q + \frac{1}{2!}B^*B + \frac{1}{4!}(B^*B)^2 + \dots \end{pmatrix} \\ &= \begin{pmatrix} \cosh(\sqrt{BB^*}) & U \sinh(\sqrt{B^*B}) \\ \sinh(\sqrt{B^*B})U^* & \cosh(\sqrt{B^*B}) \end{pmatrix}. \end{aligned}$$

It can be checked that

$$e^{\mathfrak{n}} \cap (U(p) \times U(q)) = \{I_n\}.$$

It is known that the nonzero eigenvalues of Y are the nonzero eigenvalues of $\sqrt{BB^*}$ and their negatives.

Theorem 4.6 For any element $A \in U(p, q)$, the associated matrix $\sqrt{A^*A}$ can be expressed uniquely as e^Y for $Y \in \mathfrak{n}$. Moreover, there is a unique way to express A as a product of a unitary matrix and an element of $e^{\mathfrak{n}}$, and it is given by the polar decomposition.

In order to prove the result, we discuss some preliminaries on logarithms of complex matrices. In general, there is no unique logarithm. However, the Gregory series

$$\log A = - \sum_{m=0}^{\infty} \frac{2}{2m+1} [(I - A)(I + A)^{-1}]^{2m+1}$$

converges if all the eigenvalues of $A \in M_n(\mathbb{C})$ have positive real part; see Higham [11, Section 11.3, page 273]. In particular, $\log A$ is well defined for Hermitian positive-definite matrices. This is often called the *principal logarithm* of A . This logarithm satisfies $e^{\log A} = A$. There is an integral form of the logarithm that applies to matrices without real or zero eigenvalues; it is given by

$$\log A = (A - I) \int_0^1 [s(A - I) + I]^{-1} ds.$$

Lemma 4.7 The inverse of $A^*A + I_n$ for $A \in U(p, q)$ is given by

$$[A^*A + I_n]^{-1} = \frac{1}{2} \begin{pmatrix} I_p & -A^{-1}B \\ -B^*(A^*)^{-1} & I_q \end{pmatrix}.$$

Proof Since A^*A has only positive eigenvalues, $A^*A + I_n$ has no kernel. We note that

$$A^*A + I_n = \begin{pmatrix} 2C^*C + 2I_p & 2A^*B \\ 2B^*A & 2B^*B + 2I_q \end{pmatrix} = \begin{pmatrix} 2A^*A & 2A^*B \\ 2B^*A & 2D^*D \end{pmatrix}.$$

The inverse matrix satisfies

$$\begin{pmatrix} 2A^*A & 2A^*B \\ 2B^*A & 2D^*D \end{pmatrix} \begin{pmatrix} E & F \\ F^* & G \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ 0 & I_q \end{pmatrix}.$$

As the matrices are Hermitian, the three constraints that E , F and G must satisfy (and are uniquely determined by) are

$$E = \frac{1}{2}(A^*A)^{-1} - A^{-1}BF^*, \quad G = \frac{1}{2}(D^*D)^{-1} - D^{-1}CF, \quad F = -A^{-1}BG.$$

We note that $E = \frac{1}{2}I_p$, $G = \frac{1}{2}I_q$ and $F = -\frac{1}{2}A^{-1}B$ satisfy the above equations. For instance,

$$\begin{aligned} \frac{1}{2}(A^*A)^{-1} - A^{-1}BF^* &= \frac{1}{2}(A^*A)^{-1} + \frac{1}{2}A^{-1}BB^*(A^*)^{-1} \\ &= \frac{1}{2}(A^*A)^{-1} + \frac{1}{2}A^{-1}(AA^* - I_p)(A^*)^{-1} = \frac{1}{2}I_p, \end{aligned}$$

where $BB^* = AA^* - I_p$ is a consequence of $\mathcal{A}^* \in U(p, q)$. Yet another consequence is $AC^* = BD^*$, which is equivalent to

$$A^{-1}B = (D^{-1}C)^*.$$

In a similar vein,

$$\begin{aligned} \frac{1}{2}(D^*D)^{-1} - D^{-1}CF &= \frac{1}{2}(D^*D)^{-1} + \frac{1}{2}D^{-1}CC^*(D^*)^{-1} \\ &= \frac{1}{2}(D^*D)^{-1} + \frac{1}{2}D^{-1}(DD^* - I_q)(D^*)^{-1} = \frac{1}{2}I_q, \end{aligned}$$

where $CC^* = DD^* - I_q$ is due to $\mathcal{A}^* \in U(p, q)$. □

Proof of Theorem 4.6 We use Gregory series expansion for computing the principal logarithm of $\mathcal{A}^*\mathcal{A}$ along with Lemma 4.7:

$$\begin{aligned} \log(\mathcal{A}^*\mathcal{A}) &= \sum_{m=0}^{\infty} \frac{2}{2m+1} \left[2 \begin{pmatrix} A^*A - I_p & A^*B \\ B^*A & D^*D - I_q \end{pmatrix} \frac{1}{2} \begin{pmatrix} I_p & -A^{-1}B \\ -B^*(A^*)^{-1} & I_q \end{pmatrix} \right]^{2m+1} \\ &= \sum_{m=0}^{\infty} \frac{2}{2m+1} \begin{pmatrix} 0 & A^{-1}B \\ B^*(A^*)^{-1} & 0 \end{pmatrix}^{2m+1}. \end{aligned}$$

We set $Y = \frac{1}{2} \log(\mathcal{A}^*\mathcal{A})$. It is clear that $Y \in \mathfrak{n}$ and $e^Y = \sqrt{\mathcal{A}^*\mathcal{A}}$. It is known that the exponential map is injective on Hermitian matrices. This implies the uniqueness of Y .

If $U_1e^{Y_1} = U_2e^{Y_2}$ are two decompositions of $\mathcal{A} \in U(p, q)$ with $U_i \in U(p) \times U(q)$ and $Y_i \in \mathfrak{n}$, then

$$e^{2Y_1} = e^{Y_1}U_1^*U_1e^{Y_1} = e^{Y_2}U_2^*U_2e^{Y_2} = e^{2Y_2}.$$

By the injectivity of the exponential map (on Hermitian matrices), we obtain $Y_1 = Y_2$, which implies that $U_1 = U_2$. □

We infer the following result (see Yakubovich and Starzhinskii [30, Lemma 1, page 211] for a different proof):

Corollary 4.8 *The exponential map $\exp: \mathfrak{u}_{p,q} \rightarrow U(p, q)$ is surjective.*

Proof Using the polar decomposition and Theorem 4.6,

$$A = A\sqrt{A^*A}^{-1}\sqrt{A^*A} = A\sqrt{A^*A}^{-1}e^Y.$$

Since the matrix exponential is surjective for $U(p) \times U(q)$, choose $Z \in \mathfrak{u}_p \oplus \mathfrak{u}_q$ such that $e^Z = A\sqrt{A^*A}^{-1}$. By the Baker–Campbell–Hausdorff formula, we may express e^Ze^Y as the exponential of an element in $\mathfrak{u}_{p,q}$. □

The distance from any matrix $\mathcal{A} \in U(p, q)$ to $U(p) \times U(q)$ is given by the length of the curve

$$\gamma(t) = \mathcal{A} \sqrt{\mathcal{A}^* \mathcal{A}}^{-1} e^{tY},$$

which can be computed (and simplified via left-invariance) as

$$\ell(\gamma) = \int_0^1 \|\gamma'(t)\|_{\gamma(t)} dt = \int_0^1 \|Y\| dt = \|Y\|.$$

Note that

$$\|Y\|^2 = \text{tr}(Y^* Y) = \text{tr}\left[\frac{1}{4}(\log(\mathcal{A}^* \mathcal{A}))^2\right].$$

Thus, the distance-squared function is given by

$$d_{U(p) \times U(q)}^2: U(p, q) \rightarrow \mathbb{R}, \quad \mathcal{A} \mapsto \frac{1}{4} \text{tr}[(\log(\mathcal{A}^* \mathcal{A}))^2].$$

Appendix A The continuity of the map (3-2)

Recall the statement of [Proposition 3.14](#):

Proposition A.1 *The map $s: S(v) \rightarrow [0, \infty)$, as defined in (3-2), is continuous.*

The proof relies on a characterization of $s(v)$.

Lemma A.2 *Let $u \in S_p(v)$. A positive real number T is $s(u)$ if and only if $\gamma_u: [0, T] \rightarrow M$ is an N -geodesic and at least one of the following holds:*

- (i) $\gamma_u(T)$ is the first focal point of N along γ_u .
- (ii) There exists $v \in S(v)$ with $v \neq u$ such that $\gamma_v(T) = \gamma_u(T)$.

Note that $\gamma_u(T)$ being a focal point of N along γ_u means that $(D \exp_v)(uT)$ is not of full rank, where \exp_v is the normal exponential, as defined in (3-1). When N is a point, this notion of focal points reduces to that of conjugate points.

In order to prove the lemma, we need the following observations:

Observation A [[26](#), Lemma 2.11, page 96] *Let N be a submanifold of M and $\gamma: [a, \infty) \rightarrow M$ a geodesic emanating perpendicularly from N . If $\gamma(b)$ is the first focal point of N along γ , then, for $t > b$, $\gamma|_{[a,t]}$ cannot be an N -geodesic, ie $\ell(\gamma|_{[a,t]}) > d_N(\gamma(t))$.*

Recall that a sequence $\{\gamma_n\}$ of geodesics, defined on closed intervals, is said to converge to a geodesic γ if $\gamma_n(0) \rightarrow \gamma(0)$ and $\gamma'_n(0) \rightarrow \gamma'(0)$. It follows from the continuity of the exponential map that, if $t_n \rightarrow t$, then $\gamma_n(t_n) \rightarrow \gamma(t)$.

Observation B Let γ_n be unit-speed N -geodesics joining $p_n = \gamma_n(0)$ to $q_n = \gamma_n(t_n)$. If γ_n converges to a geodesic γ and $t_n \rightarrow l$, then γ is a unit-speed N -geodesic joining $p = \lim_n p_n$ to $q := \gamma(l) = \lim_n \gamma_n(t_n)$.

Proof The unit normal bundle $S(v)$ is closed. Since $\gamma'_n(0) \rightarrow \gamma'(0)$, it follows that $\gamma'(0) \in S(v)$. Note that

$$d_N(q) = \lim_{n \rightarrow \infty} d_N(q_n) = \lim_{n \rightarrow \infty} d(p_n, q_n) = \lim_{n \rightarrow \infty} t_n = l = \ell(\gamma|_{[0,l]})$$

implies that γ is an N -geodesic. □

Proof of Lemma A.2 If $\gamma_u(t)$ is the first focal point of N along γ_u , then **Observation A** implies that γ_u cannot be minimal beyond this value. If (ii) holds, then we need to show that, for sufficiently small $\varepsilon > 0$, $\gamma_u|_{[0, T+\varepsilon]}$ is not minimal. Suppose, on the contrary, that γ_u is minimal beyond T . Take a minimal geodesic β joining $\gamma_v(T - \varepsilon)$ to $\gamma_u(T + \varepsilon)$. Observe that

$$2\varepsilon = d(\gamma_u(T + \varepsilon), \gamma_u(T)) + d(\gamma_v(T), \gamma_v(T - \varepsilon)) > d(\gamma_u(T + \varepsilon), \gamma_v(T - \varepsilon)).$$

If $p, q, r \in M$ are such that $d(p, q) + d(q, r) = d(p, r)$ and there exist shortest normal geodesics γ_1 and γ_2 joining p to q and q to r , respectively, then $\gamma_1 \cup \gamma_2$ is smooth at q and defines a shortest normal geodesic joining p to r . Therefore, we have

$$\ell(\gamma_v|_{[0, T-\varepsilon]} \cup \beta) = T - \varepsilon + d(\gamma_v(T - \varepsilon), \gamma_u(T + \varepsilon)) < T + \varepsilon = \ell(\gamma_u|_{[0, T+\varepsilon]}).$$

This contradiction establishes that $\gamma_u|_{[0, T+\varepsilon]}$ is not minimal.

For the converse, set $T = s(u)$ and observe that $\gamma_u|_{[0, T]}$ is an N -geodesic. Assuming that $q := \gamma_u(T)$ is not the first focal point of N along γ_u , we will prove that (ii) holds. Let $p = \gamma_u(0)$ and choose a neighborhood \tilde{U} of Tu in v such that $\exp_v|_{\tilde{U}}$ is a diffeomorphism. For sufficiently large n , $q_n := \gamma_u(T + 1/n) \in \exp_v(\tilde{U})$. Take N -geodesics γ_n parametrized by arc length joining p_n to q_n and set $u_n := \dot{\gamma}_n(0) \in S((T_{p_n}N)^\perp)$. Since $S((T_{p_n}N)^\perp)$ is compact, by passing to a subsequence, we may assume that u_n converges to $v \in S(N_p)$. By **Observation B**,

$$\gamma_v(T) = \lim_{n \rightarrow \infty} \gamma_{u_n}\left(T + \frac{1}{n}\right) = \gamma_u(T).$$

If $v = u$, then, for sufficiently large n , $d(p, q_n)u_n \in \tilde{U}$, whence

$$\left(T + \frac{1}{n}\right)u = d(p, q_n)u_n.$$

Taking absolute values on both sides implies $T + 1/n > d(p, q_n)$. This contradiction implies $v \neq u$. □

Proof of Proposition A.1 We will prove that $s(u_n) \rightarrow s(u)$ whenever $(p_n, u_n) \rightarrow (p, u)$ in the unit normal bundle $S(v)$. Let T be any accumulation point of the sequence $\{s(u_n)\}$ including ∞ . By **Observation B**, $\gamma_u|_{[0, T]}$ is an N -geodesic and hence $T \leq s(u)$. If $T = +\infty$, we are done. So let us assume that $T < +\infty$. From **Lemma A.2**, at least one of the following holds for infinitely many n :

- (i) $s(u_n)$ is the first focal point of N along γ_{u_n} .
- (ii) There exist $v_n \in S(N_{p_n})$ with $v_n \neq u_n$ and $\gamma_{u_n}(s(u_n)) = \gamma_{v_n}(s(u_n))$.

If (i) is true for infinitely many n , then choose infinitely many unit vectors $\{w_n\}$ which belong to the kernel $\ker(D \exp_v(s(u_n)u_n))$ and are contained in a compact subset of $S(v)$. Choose a convergent subsequence whose limit w is contained in $\ker(D \exp_v(Tu))$. Since $w \neq 0$, the rank of $D \exp_v(Tu)$ is less than $\dim M$. Thus, $\gamma_u(T)$ is the first focal point of N along γ_u and $T = s(u)$.

If (ii) is true for infinitely many n , then we may assume that $v_n \rightarrow v \in S(v)$. If $v \neq u$, then **Lemma A.2(ii)** holds for T , whence $T = s(u)$. If $v = u$, we claim that $\gamma_u(T)$ is the first focal point of N along γ_u . If not, then the map \exp_v is regular at $Tu \in v$ and hence the map

$$\Phi: v \rightarrow M \times M, \quad (p, u) \mapsto (p, \exp_v(p, u)),$$

is regular at Tu . Therefore, Φ is a diffeomorphism if restricted to an open neighborhood \tilde{U} of Tu in v . Since $v = u$, which implies, for sufficiently large n , $(p_n, s(u_n)u_n)$ and $(p_n, s(u_n)v_n)$ belong to \tilde{U} and are different. On the other hand, by assumption, $\Phi(s(u_n)u_n) = \Phi(s(u_n)v_n)$, which is a contradiction. Therefore, $\gamma_u(T)$ is the first focal point and $T = s(u)$. □

Appendix B Derivative of the square root map

Lemma B.1 Let A be a positive-definite matrix and $\psi: A \mapsto \sqrt{A}$. Then

$$D\psi_A(H) = \int_0^\infty e^{-t\sqrt{A}} H e^{-t\sqrt{A}} dt$$

for any symmetric matrix H .

Proof As $\psi(A) \cdot \psi(A) = A$, differentiating at A , we obtain

$$(B-1) \quad D\psi_A(H)\psi(A) + \psi(A)D\psi_A(H) = H.$$

(i) Given a positive-definite matrix A and a symmetric matrix H , we need to show that the equation

$$B\sqrt{A} + \sqrt{A}B = H$$

has a unique solution. For that, we will prove that the map

$$f : \text{Symm}_n \rightarrow \text{Symm}_n, \quad B \mapsto B\sqrt{A} + \sqrt{A}B,$$

is bijective, where Symm_n denotes the set of all $n \times n$ symmetric matrices. Equivalently, we will show that f is injective. Without loss of generality, we assume that \sqrt{A} is a diagonal matrix with positive entries t_1, t_2, \dots, t_n . Note that

$$\ker(f) = \{B \in \text{Symm}_n : B\sqrt{A} + \sqrt{A}B = 0\}.$$

Consider

$$\begin{aligned} B\sqrt{A} + \sqrt{A}B = 0 &\implies B \text{diag}(t_1, \dots, t_n) + \text{diag}(t_1, \dots, t_n)B = 0 \\ &\implies t_j b_{ij} + t_i b_{ij} = 0 \quad \text{for } 1 \leq i, j \leq n. \end{aligned}$$

Since $t_i > 0$ for $1 \leq i \leq n$, we see $b_{ij} = 0$. Therefore, f is injective.

(ii) For any positive-definite matrix X and for any symmetric matrix Y , the integral

$$(B-2) \quad \int_0^\infty e^{-tX} Y e^{-tX} dt$$

converges. We note that the eigenvalues of e^{-tX} are $e^{-t\lambda_j}$, where λ_j are the eigenvalues of X . Since X is a positive-definite matrix, each of the λ_j is positive. Without loss of generality, we assume that $\lambda = \lambda_1$ is the smallest eigenvalue of X . Then we have

$$e^{-t\lambda_j} \leq e^{-t\lambda} \implies \|e^{-tX}\| = e^{-t\lambda},$$

where $\|\cdot\|$ is the operator norm. Therefore, the operator norm of the integrand in (B-2) is bounded by $2e^{-t\lambda}\|Y\|$, which is an integrable function. Hence, the integral given by (B-2) converges.

(iii) $D\psi_A(H)$ satisfies (B-1). Observe that

$$\begin{aligned} &\left(\int_0^\infty e^{-t\sqrt{A}} \cdot H \cdot e^{-t\sqrt{A}} dt\right)\sqrt{A} + \sqrt{A}\left(\int_0^\infty e^{-t\sqrt{A}} \cdot H \cdot e^{-t\sqrt{A}} dt\right) \\ &= \int_0^\infty (e^{-t\sqrt{A}} \cdot H \cdot e^{-t\sqrt{A}}\sqrt{A} + \sqrt{A}e^{-t\sqrt{A}} \cdot H \cdot e^{-t\sqrt{A}}) dt \\ &= \int_0^\infty (e^{-t\sqrt{A}} H e^{-t\sqrt{A}})' dt = H. \end{aligned}$$

From (i), (ii) and the uniqueness of the derivative, the lemma is proved. □

Lemma B.2 *The map $g : M(n, \mathbb{R}) \rightarrow \mathbb{R}, A \mapsto \text{tr}(\sqrt{A^T A})$, is differentiable if and only if A is invertible.*

Proof Let A be an invertible matrix. We will prove that the function g is differentiable at A . Let \mathcal{P} be the set of all positive-definite matrices, which is an open subset of the set of all symmetric matrices \mathcal{S} . We will prove that the map

$$r: \mathcal{P} \rightarrow \mathcal{P}, \quad A \mapsto \sqrt{A},$$

is differentiable. Define a function

$$s: \mathcal{P} \rightarrow \mathcal{P}, \quad A \mapsto A^2.$$

We will show that s is a diffeomorphism and, from the inverse function theorem, r will be differentiable. In order to show that s is a diffeomorphism, we claim that, for $A \in \mathcal{P}$, $Ds_A: T_A\mathcal{P} \rightarrow T_{A^2}\mathcal{P}$ is injective. Note that \mathcal{P} is an open subset of a vector space \mathcal{S} and, therefore, $T_A\mathcal{P} \cong \mathcal{S} \cong T_{A^2}\mathcal{P}$. So take $B \in \mathcal{S}$ such that $Ds_A(B) = 0$. We will show that $B = 0$. Recall that $Ds_A(B) = AB + BA$. Now choose an orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of the eigenspace of A and let $Av_i = \lambda_i v_i$ ($\lambda_i > 0$). Then,

$$A(Bv_i) = -BAv_i = -B\lambda_i v_i = -\lambda_i(Bv_i),$$

which implies Bv_i is also an eigenvector of A with eigenvalue $-\lambda_i < 0$. Hence, $Bv_i = 0$, which implies $B = 0$.

For the converse, we will show that, if A is a singular matrix, then the map g is not directional differentiable. Let A be a singular matrix. Using the singular value decomposition, we write

$$A = U \begin{pmatrix} D & 0 \\ 0 & 0_k \end{pmatrix} V^T,$$

where D is an $(n - k) \times (n - k)$ diagonal matrix with positive entries. If

$$B = U \begin{pmatrix} 0_{n-k} & 0 \\ 0 & I_k \end{pmatrix},$$

then we claim that g is not differentiable in the direction of B . Since

$$\sqrt{(A + tB)^T(A + tB)} = V \begin{pmatrix} D & 0 \\ 0 & I_k|t| \end{pmatrix} V^T,$$

the limit

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{g(A + tB) - g(A)}{t} &= \lim_{t \rightarrow 0} \frac{1}{t} \left(\text{tr} \left(V \begin{pmatrix} D & 0 \\ 0 & I_k|t| \end{pmatrix} V^T \right) - \text{tr} \left(V \begin{pmatrix} D & 0 \\ 0 & 0_k \end{pmatrix} V^T \right) \right) \\ &= k \lim_{t \rightarrow 0} \frac{|t|}{t} \end{aligned}$$

does not exist and hence the function g is not differentiable. □

References

- [1] **A Banyaga, D Hurtubise**, *Lectures on Morse homology*, Kluwer Texts in the Mathematical Sciences 29, Kluwer Academic, Dordrecht (2004) [MR](#) [Zbl](#)
- [2] **M G Barratt, J Milnor**, *An example of anomalous singular homology*, Proc. Amer. Math. Soc. 13 (1962) 293–297 [MR](#) [Zbl](#)
- [3] **R L Bishop, R J Crittenden**, *Geometry of manifolds*, Pure and Applied Mathematics XV, Academic, New York (1964) [MR](#) [Zbl](#)
- [4] **MA Buchner**, *Simplicial structure of the real analytic cut locus*, Proc. Amer. Math. Soc. 64 (1977) 118–121 [MR](#) [Zbl](#)
- [5] **H Busemann**, *The geometry of geodesics*, Academic, New York (1955) [MR](#) [Zbl](#)
- [6] **A Daniilidis, R Deville, E Durand-Cartagena, L Rifford**, *Self-contracted curves in Riemannian manifolds*, J. Math. Anal. Appl. 457 (2018) 1333–1352 [MR](#) [Zbl](#)
- [7] **H Gluck, D Singer**, *Scattering of geodesic fields, I*, Ann. of Math. 108 (1978) 347–372 [MR](#) [Zbl](#)
- [8] **A Gray**, *Tubes*, 2nd edition, Progr. Math. 221, Birkhäuser, Basel (2004) [MR](#) [Zbl](#)
- [9] **J J Hebda**, *The local homology of cut loci in Riemannian manifolds*, Tohoku Math. J. 35 (1983) 45–52 [MR](#) [Zbl](#)
- [10] **J J Hebda**, *Cut loci of submanifolds in space forms and in the geometries of Möbius and Lie*, Geom. Dedicata 55 (1995) 75–93 [MR](#) [Zbl](#)
- [11] **N J Higham**, *Functions of matrices: theory and computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA (2008) [MR](#) [Zbl](#)
- [12] **M W Hirsch**, *Differential topology*, Graduate Texts in Math. 33, Springer (1994) [MR](#) [Zbl](#)
- [13] **J-i Itoh, S V Sabau**, *Riemannian and Finslerian spheres with fractal cut loci*, Differential Geom. Appl. 49 (2016) 43–64 [MR](#) [Zbl](#)
- [14] **J-i Itoh, C Vîlcu**, *Orientable cut locus structures on graphs*, preprint (2011) [arXiv 1103.3136](#)
- [15] **J-i Itoh, C Vîlcu**, *Every graph is a cut locus*, J. Math. Soc. Japan 67 (2015) 1227–1238 [MR](#) [Zbl](#)
- [16] **S Kobayashi**, *On conjugate and cut loci*, from “Studies in global geometry and analysis”, Math. Assoc. America, Englewood Cliffs, NJ (1967) 96–122 [MR](#)
- [17] **Y Li, L Nirenberg**, *The distance function to the boundary, Finsler geometry, and the singular set of viscosity solutions of some Hamilton–Jacobi equations*, Comm. Pure Appl. Math. 58 (2005) 85–146 [MR](#) [Zbl](#)
- [18] **C Mantegazza, A C Mennucci**, *Hamilton–Jacobi equations and distance functions on Riemannian manifolds*, Appl. Math. Optim. 47 (2003) 1–25 [MR](#) [Zbl](#)

- [19] **R J Martin, P Neff**, *Minimal geodesics on $GL(n)$ for left-invariant, right- $O(n)$ -invariant Riemannian metrics*, J. Geom. Mech. 8 (2016) 323–357 [MR](#) [Zbl](#)
- [20] **S B Myers**, *Connections between differential geometry and topology, I: Simply connected surfaces*, Duke Math. J. 1 (1935) 376–391 [MR](#) [Zbl](#)
- [21] **H Omori**, *A class of Riemannian metrics on a manifold*, J. Differential Geometry 2 (1968) 233–252 [MR](#) [Zbl](#)
- [22] **S Plotnick**, *Embedding homology 3–spheres in S^5* , Pacific J. Math. 101 (1982) 147–151 [MR](#) [Zbl](#)
- [23] **H Poincaré**, *Sur les lignes géodésiques des surfaces convexes*, Trans. Amer. Math. Soc. 6 (1905) 237–274 [MR](#)
- [24] **M M Postnikov**, *Geometry, VI: Riemannian geometry*, Encyclopaedia of Mathematical Sciences 91, Springer (2001) [MR](#) [Zbl](#)
- [25] **S V Sabau, M Tanaka**, *The cut locus and distance function from a closed subset of a Finsler manifold*, Houston J. Math. 42 (2016) 1157–1197 [MR](#) [Zbl](#)
- [26] **T Sakai**, *Riemannian geometry*, Translations of Mathematical Monographs 149, Amer. Math. Soc., Providence, RI (1996) [MR](#) [Zbl](#)
- [27] **V A Sharafutdinov**, *Complete open manifolds of nonnegative curvature*, Sibirsk. Mat. Zh. 15 (1974) 126–136 [MR](#) [Zbl](#) In Russian; translated with modifications as “Proof of soul theorem”, preprint (2006)
- [28] **H Singh**, *On the cut locus and the focal locus of a submanifold in a Riemannian manifold, II*, Publ. Inst. Math. (Beograd) 41(55) (1987) 119–124 [MR](#) [Zbl](#)
- [29] **F-E Wolter**, *Distance function and cut loci on a complete Riemannian manifold*, Arch. Math. (Basel) 32 (1979) 92–96 [MR](#) [Zbl](#)
- [30] **V A Yakubovich, V M Starzhinskii**, *Linear differential equations with periodic coefficients*, volume 1, Halsted, New York (1975) [MR](#) [Zbl](#)

Department of Mathematics and Statistics, Indian Institute of Science Education and Research
Kolkata, India

Department of Mathematics and Statistics, Indian Institute of Science Education and Research
Kolkata, India

somnath.basu@iiserkol.ac.in, sp17rs038@iiserkol.ac.in

Received: 4 June 2021 Revised: 15 February 2023

Staircase symmetries in Hirzebruch surfaces

NICKI MAGILL

DUSA MCDUFF

This paper continues the investigation of staircases in the family of Hirzebruch surfaces formed by blowing up the projective plane with weight b , that was started by Bertozzi, Holm Maw, McDuff, Mwakayoma, Pires and Weiler (2021). We explain the symmetries underlying the structure of the set of b that admit staircases, and show how the properties of these symmetries arise from a governing Diophantine equation. We also greatly simplify the techniques needed to show that a family of steps does form a staircase by using arithmetic properties of the accumulation function. There should be analogous results about both staircases and mutations for the other rational toric domains considered, for example, by Cristofaro-Gardiner, Holm, Mandini and Pires (2020) and by Casals and Vianna (2022).

53D05; 11D99

1 Introduction

1.1 Overview

This paper continues the investigation of the ellipsoidal embedding capacity function for the family of Hirzebruch surfaces H_b that was begun by Bertozzi, Holm, Maw, McDuff, Mwakayoma, Pires and Weiler [1]. Here (H_b, ω) is the one-point blowup $\mathbb{C}P^2(1) \# \overline{\mathbb{C}P^2}(b)$ of the complex projective plane with line class L of size 1 and exceptional divisor E_0 of size b . The capacity function $c_X : [1, \infty) \rightarrow \mathbb{R}$ for a general four-dimensional target manifold (X, λ) is defined by

$$c_X(z) := \inf\{\lambda \mid E(1, z) \xrightarrow{s} \lambda X\},$$

where $z \geq 1$ is a real variable, $\lambda X := (X, \lambda\omega)$, an ellipsoid $E(c, d) \subset \mathbb{C}^2$ is the set

$$E(c, d) = \left\{ (\zeta_1, \zeta_2) \in \mathbb{C}^2 \mid \pi \left(\frac{|\zeta_1|^2}{c} + \frac{|\zeta_2|^2}{d} \right) < 1 \right\},$$

and we write $E \xrightarrow{s} \lambda X$ if there is a symplectic embedding of E into λX .

It is straightforward to see that $c_{H_b}(z)$ is bounded below by the volume constraint function $V_b(z) = \sqrt{z/(1-b^2)}$, where $1-b^2$ is the appropriately normalized volume of H_b . Further, the function $z \mapsto c_{H_b}(z)$ is piecewise linear when not identically equal to the volume constraint curve. When its graph has infinitely many nonsmooth points (or steps) lying above the volume curve, we say that c_{H_b} has an *infinite staircase*.

It was proven by Cristofaro-Gardiner, Holm, Mandini and Pires [3] that when $b = 1/3$, ie in the case when H_b is monotone, the function c_{H_b} admits a staircase with outer steps at points $z = x_k/x_{k-1}$ that satisfy the recursion $x_{k+1} = 6x_k - x_{k-1}$ and accumulate at the fixed point $3 + 2\sqrt{2}$ of this recursion.¹ Another key result from this paper is [3, Theorem 1.8], stating that if c_{H_b} has an infinite staircase, then its accumulation point is at the point $z = \text{acc}(b)$, the unique solution > 1 of the following quadratic equation involving b :

$$(1.1.1) \quad z^2 - \left(\frac{(3-b)^2}{1-b^2} - 2 \right) z + 1 = 0.$$

Further, if H_b has an infinite staircase, then at $z = \text{acc}(b)$ the ellipsoid embedding function must equal the volume:

$$(1.1.2) \quad c_{H_b}(\text{acc}(b)) = \sqrt{\frac{\text{acc}(b)}{1-b^2}} =: V_b(\text{acc}(b)).$$

We say that H_b is *unobstructed* if $c_{H_b}(\text{acc}(b)) = V_b(\text{acc}(b))$. Thus if H_b has a staircase, it is unobstructed. However the converse does not hold: [1, Theorem 6] shows that, although $H_{1/5}$ is unobstructed, it has no staircase. As shown in Figure 1, the function $b \mapsto \text{acc}(b)$ decreases for $b \in [0, 1/3]$, with minimum value $a_{\min} := \text{acc}(1/3) = 3 + 2\sqrt{2}$, and then increases.

It turns out that the nature of the accumulation function plays a crucial role in our discussion. For example, as shown in Lemma 2.1.1, the properties of the pairs of rational numbers (b, z) with $z = \text{acc}(b)$ are a key to the symmetries of the problem. Other important consequences are collected in Section 2.2. Note also that the case $b = 0$ (that is, the case of $\mathbb{C}P^2$) was fully analyzed by McDuff and Schlenk [6]. Here there is a staircase, which is called the Fibonacci stairs because its numerics are governed by the Fibonacci numbers; see Figure 1. The new staircases that we have found for general H_b are all analogs of that one. However, as we explain in Example 2.3.7, it is perhaps better to consider our staircases to be offshoots of the $1/3$ -staircase.

¹This staircase actually consists of three intertwining strands that each satisfy this recursion; for details see Example 2.3.7. Further, conjecturally $b = 1/3$ is the only rational value of b at which H_b admits a staircase.

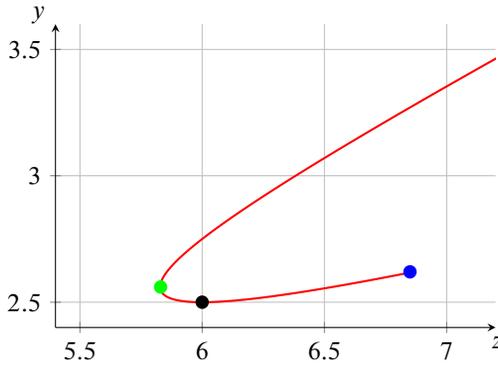


Figure 1: This shows the location of the accumulation point $(z, y) = (\text{acc}(b), V_b(\text{acc}(b)))$ for $0 \leq b < 1$. The blue point with $b = 0$ is at (τ^4, τ^2) and is the accumulation point for the Fibonacci stairs. The green point with $z = 3 + 2\sqrt{2}$ and $b = 1/3$ is the accumulation point for the stairs in $H_{1/3}$, and is the minimum of the function $b \mapsto \text{acc}(b)$. The black point with $z = 6$ and $b = 1/5$ is the place where $V_b(\text{acc}(b))$ takes its minimum.

Obstructions to embedding ellipsoids come from certain exceptional divisors in blowups of the target manifold. When $X = H_b$, these divisors live in $\mathbb{C}P^2 \# (N + 1)\overline{\mathbb{C}P}^2$ and their homology classes \mathbf{E} have the form

$$dL - mE_0 - \sum m_i E_i =: (d, m, \mathbf{m}),$$

where $\mathbf{m} := (m_1, \dots, m_N)$ with $m_1 \geq m_2 \geq \dots \geq m_N$. In the most relevant such classes, the tuple of coefficients \mathbf{m} consists of the (integral) weight expansion of a center point p/q (see Definition 2.1.5); correspondingly we say that \mathbf{E} is *perfect* and write $\mathbf{E} := (d, m, p, q)$. For z near p/q , the embedding obstruction is given by

$$(1.1.3) \quad \mu_{\mathbf{E},b}(z) = \begin{cases} qz/(d - mb) & \text{if } z \leq p/q, \\ p/(d - mb) & \text{if } z \geq p/q; \end{cases}$$

in particular, it has an outer corner (or step) at $z = p/q$. Since, as explained in [1, Section 2.1], $c_{H_b}(z)$ is the maximum over all exceptional classes \mathbf{E} of the obstruction functions $\mu_{\mathbf{E},b}(z)$, given $\mathbf{E} = (d, m, p, q)$ as above, we must have $c_{H_b}(z) \geq \mu_{\mathbf{E},b}(z)$ for $z \approx p/q$. We say that the function $\mu_{\mathbf{E},b}$ is

- *obstructive* at z if $\mu_{\mathbf{E},b}(z) > V_b(z)$, and
- *live* at z if $c_{H_b}(z) = \mu_{\mathbf{E},b}(z) > V_b(z)$.

Further, we call \mathbf{E} a *center-blocking class* if, for one of the two elements b of $\text{acc}^{-1}(p/q)$, the function $z \mapsto \mu_{\mathbf{E},b}(z)$ is obstructive at the center $z = p/q$, since in

this case it follows from (1.1.2) that the corresponding surface H_b has no staircase. As explained in [1, Lemma 38], it follows by continuity that for every center-blocking class E there is an open interval $J_E \subset [0, 1)$ that contains the appropriate² point of $\text{acc}^{-1}(p/q)$ and is a component of the set of b -values that are blocked by E .

The paper [1] found three families of center-blocking classes, B_n^U , B_n^L , and B_n^E for $n \geq 0$, together with six associated sequences of staircases. Each of these blocking classes B has an associated maximal blocked interval $J_B = (\beta_{B,\ell}, \beta_{B,u}) \subset (0, 1)$ consisting of points b that cannot admit a staircase because $\mu_{B,b}(\text{acc}(b)) > V_b(\text{acc}(b))$. However, it turns out that there are staircases at both endpoints of these intervals, which gives three staircase families, S^U , S^L , and S^E , with staircases indexed by $n \geq 0$ and ℓ or u , where the steps of staircases labeled ℓ (for “lower”) ascend, while those labeled u (for “upper”) descend. The Fibonacci stairs appear as $S_{\ell,0}^L$.

It was noted in [1, Corollary 60] that the centers p/q of the blocking and step classes for the family S^U are related to those of S^L and S^E by a fractional linear transformation, that we denote by either $p/q \mapsto (ap + bq)/(cp + dq)$ or $(p, q) \mapsto (ap + bq, cp + dq)$. In particular, the two families S^U, S^E are related by the *shift*

$$(1.1.4) \quad S: (p, q) \mapsto (6p - q, p)$$

that implements the recursion underlying the staircase at $1/3$; while the two families S^U and S^L are related by the *reflection*

$$(1.1.5) \quad R: (p, q) \mapsto (6p - 35q, p - 6q),$$

which fixes the point 7 and takes ∞ to 6 .

Our main result verifies a conjecture in [1], and can be informally stated as follows. (For more detail, see Theorems 1.2.4 and 1.2.6.)

Theorem 1.1.1 *For each $i \geq 1$ there are staircase families $(S^i)^\#(S^U)$ and $(S^i)^\#(S^L) = (S^i R)^\#(S^U)$, where the i -fold shift S^i and reflection R act on the centers of the blocking classes and staircase steps as above.*

Moreover, we will see in Proposition 1.2.2 that each staircase family is generated by its blocking classes together with two “seed” classes, a fact that makes it much easier to establish the effect of the symmetries on the staircase families. Note also that although the action of the symmetries S^i and $S^i R$ on the centers p/q of the classes is clear, the

²If $m/d > 1/3$ then this will be the larger element in $\text{acc}^{-1}(p/q)$, while if $m/d < 1/3$ it will be the smaller one; see [1, Definition 37]. By Lemma 2.2.13, there is no quasiperfect class with $m/d = 1/3$. If there is no possibility of confusion, we often simply call these classes blocking classes.

action on the other two coordinates (d, m) (that we call the *degree* coordinates) is much less obvious. In particular, this action is not compatible with composition. However, we will see in Section 3.4 that the action on degree can be understood because, as we explain below, the coefficients d and m of all the relevant classes are given in terms of p and q by a general formula (1.2.4).

Remark 1.1.2 In the setting considered by Usher [7], the target manifold $P(1, b)$ is the polydisc $B^2(1) \times B^2(b)$, or equivalently the product of two spheres of areas 1 and b . He finds a doubly indexed family of staircases $\mathcal{S}_{n,k} = (E_{i,n,k})_{i \geq 0}$, where i indexes the staircase steps, $n \geq 0$ indexes the intrinsic recursion $x_{i+1,n,k} = v_n x_{i,n,k} - x_{i-1,n,k}$ satisfied by the parameters of the perfect classes $E_{i,n,k}$ for $i \geq 0$ in $\mathcal{S}_{n,k}$, and k indexes a symmetry generated by so-called ‘‘Brahmagupta moves’’ that generate infinitely many families of staircases from a basic family $(\mathcal{S}_{n,0})_{n \geq 0}$. More precisely, in [7, Section 2.2.1], Usher finds a way to encode the parameters of the relevant perfect classes E by means of a triple (x, δ, ε) of integers that satisfy the Diophantine equation $x^2 - 2\delta^2 = 2 - \varepsilon^2$. Here the value of ε is related to the recursion variable n , and, if this is fixed, he shows that a (very!) classically known maneuver that goes from one solution of $x^2 - 2\delta^2 = N$ to another can be implemented in such a way that it preserves the set of perfect classes. Usher expressed this maneuver in arithmetic terms (multiplication by a unit in a number field). However, as we explain in Remark 2.2.6, when expressed in terms of the coordinates (p, q) , Usher’s basic symmetry is the same as ours, namely the transformation $(p, q) \mapsto (6p - q, p)$.

Usher’s setting is simpler than ours in that the function $b \mapsto \text{acc}(b)$ that specifies the accumulation point of any staircase for $P(1, b)$ is injective rather than two-to-one.³ Also the symmetry between the two classes $[pt \times S^2]$ and $[S^2 \times pt]$ allows the arithmetic properties of a general quasiperfect class to be encoded by means of variables that satisfy the equation $x^2 - 2\delta^2 = N$, while the corresponding equation in our setting is $x^2 - 8y^2 = k^2$ (see Lemma 2.1.1). Nevertheless, the two situations are very similar.

The work presented here leads to many interesting questions. Here are some of them.

- The picture developed here seems to make up the first level of an iterative ‘‘fractal’’ kind of structure for the Hirzebruch surfaces H_b . One might consider the family

³This statement is oversimplified in that one could well argue that the analog of our family H_b for $b \in [0, 1)$ is the family $P(1, b)$ for $b > 0$ with involution $b \mapsto 1/b$. However, $P(1, b)$ is symplectomorphic to a rescaling of $P(1, 1/b)$, so $c_{P(1,b)}$ is a rescaling of $c_{P(1,1/b)}$ and $\text{acc}(b) = \text{acc}(1/b)$. In our case, if $\text{acc}^{-1}(z) = \{b^+, b^-\}$, the two functions $c_{H_{b^+}}$ and $c_{H_{b^-}}$ can be very different, one with a staircase, and one without; see Figure 2.

of blocking classes B_n^U , for $n \geq 0$, extended by two seeds as in [Proposition 1.2.2](#), to be the backbone of the first level of this structure. This level also includes the associated staircase classes. We prove in [Proposition 2.2.9](#) that all the staircase classes are also center-blocking classes. Further, numerical evidence suggests that there are staircases whose steps have centers with 4–periodic continued fractions, indeed it seems with any even period. What seems to be the case is that each pair of adjacent ascending/descending staircases at level one shares a first step, and that this first step is a blocking class with associated 4–periodic staircases. Thus the backbone of the second level should consist of these shared steps, with associated 4–periodic staircases generated by appropriate seeds at level one. For more details, see Magill, McDuff and Weiler [\[5\]](#).

- It also would be very interesting to analyze Usher’s results using the current framework, to see if there are analogs of blocking classes, seeds and staircase families. One might be able to build a bridge between the two cases by thinking of a polydisc as a degenerate two-point blowup of $\mathbb{C}P^2$, and then looking at the ellipsoidal capacity function for the family of two-fold blowups of $\mathbb{C}P^2$ that join the two cases. This will also be the subject of future work.
- The recursive patterns behind the staircases for rational target manifolds X are related to almost toric structures and the transformations called mutations that appear for example in [\[3\]](#) and Casals and Vianna [\[2\]](#). It would be very interesting to know how the symmetries discussed here appear in those contexts.

Acknowledgements We thank Chao Li for help with Diophantine equations, and Tara Holm, Peter Sarnak and Morgan Weiler for useful discussions and comments. We also thank the referee for a very thorough reading that has helped to improve the accuracy and clarity of several arguments. Magill also thanks Tara Holm in her capacity as research advisor for introducing her to the subject and for support and encouragement along the way. Magill was supported by the NSF Graduate Research Grant DGE-1650441.

1.2 Main results

We now describe our main results in more detail.

In what follows it is important to distinguish purely numerical properties — such as those in [\(1.2.1\)](#) — from geometric properties that are needed to guarantee that a class E gives a live obstruction $\mu_{E,b}(z)$ at relevant values of b and z . As above we represent a

class $E = dL - mE_0 - \sum_i m_i E_i$, where (m_1, m_2, \dots) is the weight expansion of p/q , by the tuple (d, m, p, q) , and say that E is *quasiperfect* if and only if the Diophantine conditions

$$(1.2.1) \quad 3d = p + q + m, \quad d^2 - m^2 = pq - 1$$

hold. (As explained in [1, Section 2.1], these are equivalent to the conditions $c_1(E) = 1$ and $E \cdot E = -1$, where c_1 is the first Chern class of H_b .) We say that a quasiperfect class E is *perfect* if it is represented by an exceptional curve, which holds if and only if it reduces correctly by Cremona moves (see Lemma 4.1.1). Although a quasiperfect class is obstructive at its center $z = p/q$ for $b = m/d$ by [1, Lemma 15], the fact that this obstruction is live in a neighborhood of this (b, z) (and hence coincides with the capacity function in some range) follows from “positivity of intersections”, namely the fact that the intersection number of two different exceptional divisors is always nonnegative; see [1, Proposition 21]. This implies that the obstruction function $\mu_{E,b}(z)$ given by an exceptional class E is strictly larger than all other obstructions for $z \approx p/q$ and $b \approx m/d$. Hence in order to show that a given surface H_b actually has a staircase, we need to prove that the relevant staircase classes are exceptional classes. However, a large part of the following discussion is purely numerical. Notice also that because quasiperfect classes can be obstructive, it makes sense to consider *quasiperfect blocking classes*, ie tuples (d, m, p, q) such that $\mu_{E,b}(p/q) > V_b(p/q)$ for the appropriate $b \in \text{acc}^{-1}(p/q)$.

The coefficients $(d_\kappa, m_\kappa, p_\kappa, q_\kappa)$ of the step classes of the staircases we consider always satisfy a recursion of the form

$$(1.2.2) \quad x_{\kappa+1} = \nu x_\kappa - x_{\kappa-1}, \quad \kappa \geq \kappa_0,$$

for suitable recursion parameter ν and initial value κ_0 .⁴ Hence, given ν , each sequence of parameters $(x_\kappa)_{\kappa \geq i}$ (for $x = p, q, d, m$) is determined by two initial values x_{κ_0} and x_{κ_0+1} that are called *seeds*. It turns out that the relevant classes (d, m, p, q) can be extended to tuples

$$(1.2.3) \quad (d, m, p, q, t, \varepsilon), \quad t > 0, \varepsilon \in \{\pm 1\},$$

where the integer t is a function of p and q . Further, the tuple (p, q, t, ε) determines the degree variables (d, m) by the formulas

$$(1.2.4) \quad d := \frac{1}{8}(3(p + q) + \varepsilon t), \quad m := \frac{1}{8}((p + q) + 3\varepsilon t).$$

⁴Not all infinite staircases satisfy this recursion. The $b = 1/3$ staircase is given by a nonhomogenous recursion. Further, it is not known that all staircases must be given by some recursion.

Hence $|d - 3m| = t$, and $\varepsilon = 1$ if and only if $m/d > 1/3$. This point of view is explained in Section 2.2.

It is straightforward to check that both S and R_κ preserve t and hence $\varepsilon(d - 3m)$, while they both change the sign of ε . Further, for classes that give obstructions when $b > 1/3$ we have $\varepsilon = 1$, while $\varepsilon = -1$ if the classes are relevant for $b < 1/3$; see Example 3.4.7.

Our first main result is that all the numerical data of a staircase family such as S^U is determined by a family of quasiperfect classes $(\mathbf{B}^n)_{n \geq 0}$ together with two seeds $\mathbf{E}_{\ell, \text{seed}}$ and $\mathbf{E}_{u, \text{seed}}$. To explain this, we introduce the following language.

- A *prestaircase* \mathcal{S} is a sequence of tuples $\mathbf{E}_\kappa := ((d_\kappa, m_\kappa, p_\kappa, q_\kappa, t_\kappa, \varepsilon))_{\kappa \geq 0}$ that is defined recursively with recursion parameter ν , and that satisfy (1.2.4). Given such a sequence, the limits $a_\infty := \lim p_\kappa/q_\kappa$ and $b_\infty := \lim m_\kappa/d_\kappa$ always exist by Corollary 3.1.5. We say that \mathcal{S} is *perfect* if all the classes \mathbf{E}_κ are perfect, and that \mathcal{S} is *live* if the obstructions $\mu_{\mathbf{E}_\kappa, b_\infty}$ are live near $z = p_\kappa/q_\kappa$ for all sufficiently large κ and with b equal to the limiting value b_∞ . We will refer to a *staircase* as a live prestaircase.⁵ Thus if \mathcal{S} is live, H_{b_∞} has a staircase. This is a slight abuse of notation as not all staircases follow the recursive structure of a prestaircase, but all staircases considered in this paper are indeed prestaircases.

- A prestaircase \mathcal{S} is said to be *associated to a quasiperfect class*

$$\mathbf{B} = (d_{\mathbf{B}}, m_{\mathbf{B}}, p_{\mathbf{B}}, q_{\mathbf{B}})$$

if the following linear relation is satisfied by its step coefficients:

$$(1.2.5) \quad (3m_{\mathbf{B}} - d_{\mathbf{B}})d_\kappa = \begin{cases} (m_{\mathbf{B}} - q_{\mathbf{B}})p_\kappa + m_{\mathbf{B}}q_\kappa & \text{if } \mathcal{S} \text{ ascends,} \\ m_{\mathbf{B}}p_\kappa - (p_{\mathbf{B}} - m_{\mathbf{B}})q_\kappa & \text{if } \mathcal{S} \text{ descends.} \end{cases}$$

If in addition the prestaircase is perfect, then H_{b_∞} is unobstructed, ie $c_{H_b}(\text{acc}(b)) = V_b(\text{acc}(b))$, and it is shown in [1, Theorem 52] that the limits (b_∞, a_∞) are the parameters (b, z) of the appropriate endpoint of the blocked b -interval $J_{\mathbf{B}}$.⁶ Moreover, if there is both an ascending and a descending perfect prestaircase associated to \mathbf{B} then \mathbf{B} is a perfect blocking class. This means in particular that \mathbf{B} is obstructive for the b -value corresponding to its center.

⁵We do not insist that the classes in a staircase are perfect; however in all known cases they are perfect. Indeed the only way that we know of to prove that a staircase is live is first to show that it is perfect and then to show there are no “overshadowing classes”. See the beginning of Section 4 for more details.

⁶This follows very easily from the calculation in (2.2.8).

- A prestaircase family \mathcal{F} consists of a family of quasiperfect classes $(\mathbf{B}_n^{\mathcal{F}})_{n \geq 0}$ (called *preblocking classes*) together with ascending prestaircases $\mathcal{S}_{\ell,n}^{\mathcal{F}}$, for $n \neq 1$, and descending prestaircases $\mathcal{S}_{u,n}^{\mathcal{F}}$, for $n \neq 0$, where $\mathcal{S}_{\bullet,n}^{\mathcal{F}}$ is associated with $\mathbf{B}_n^{\mathcal{F}}$ for $\bullet = \ell, u$. The family \mathcal{F} is said to be *perfect* if all the classes in \mathcal{F} are perfect, and *live* if all the prestaircases $\mathcal{S}_{\bullet,n}^{\mathcal{F}}$, $\bullet = \ell, u$, are live.
- A prestaircase family is called a *staircase family* if it is live. Then the preblocking classes are perfect by [1, Theorem 52], and in all cases encountered here the step classes are also perfect.

Finally, we make the following definition. It follows from the formulas in Theorems 2.3.1 and 2.3.3 that, because S^i preserves order, prestaircase families that are obtained from \mathcal{S}^U by applying the shift S^i have preblocking classes whose centers ascend, while those that are obtained from \mathcal{S}^L by applying S^i have preblocking classes whose centers descend. It turns out that in all cases the adjacent preblocking class can be considered as part of the appropriate staircase; for example in \mathcal{S}^U the blocking class \mathbf{B}_{n-1}^U can be considered as a step in the ascending staircase $\mathcal{S}_{\ell,n}^U$, while \mathbf{B}_{n+1}^U can be considered as a step in the descending staircase $\mathcal{S}_{\ell,n}^U$. Clearly the numbering of this adjacent blocking class depends on whether the centers of these classes ascend or descend as n increases.

Definition 1.2.1 A prestaircase family \mathcal{F} is said to be *generated by the quasiperfect classes \mathbf{B}_n , for $n \geq 0$, and seeds $\mathbf{E}_{\ell,\text{seed}}$ and $\mathbf{E}_{u,\text{seed}}$* if the $\mathbf{B}_n = (d_n, m_n, p_n, q_n, t_n, \varepsilon)$ are its preblocking classes, and, for all n and $\bullet \in \{\ell, u\}$, the steps in the prestaircase $\mathcal{S}_{\bullet,n}^{\mathcal{F}}$ have recursion parameter $\nu = t_n$ and seeds

- $\mathbf{E}_{\ell,\text{seed}}, \mathbf{B}_{n-1}$ for $\bullet = \ell$ and $\mathbf{E}_{u,\text{seed}}, \mathbf{B}_{n+1}$ for $\bullet = u$, if the \mathbf{B}_n ascend;
- $\mathbf{E}_{\ell,\text{seed}}, \mathbf{B}_{n+1}$ for $\bullet = \ell$ and $\mathbf{E}_{u,\text{seed}}, \mathbf{B}_{n-1}$ for $\bullet = u$, if the \mathbf{B}_n descend.

In Propositions 3.2.2 and 3.2.6, we establish the following compact description of the numerical information that determines each of our staircase families. This information will make it much easier to understand the effect of the symmetries.

Proposition 1.2.2 *The staircase families \mathcal{S}^U and \mathcal{S}^L are generated by their blocking classes together with two seeds.*

Remark 1.2.3 As we explain in Example 2.3.7, all the classes (both blocking classes and seeds) that generate \mathcal{S}^U are directly related to the classes that generate the staircase at $b = 1/3$. Since the staircase classes in $H_{1/3}$ satisfy the recursion (1.2.2) with $\nu = 6$

given by S , one can see that the whole structure of the staircases so far discovered in the family of manifolds H_b stems from that of the staircase at $b = 1/3$.

We now explain the action of the family⁷ of transformations

$$(1.2.6) \quad \mathcal{G} := \{S^i R^\delta \mid \delta \in \{0, 1\}, i \geq 0\},$$

where S is the shift $p/q \mapsto (6p - q)/p$ and R is the reflection $z \mapsto (6z - 35)/(z - 6)$. The sequence of numbers

$$v_1 = \infty = \frac{1}{0}, \quad v_2 = S(v_1) = \frac{6}{1}, \quad \dots, \quad v_i = S^{i-1} v_1, \quad \dots$$

has limit $a_{\min} := 3 + 2\sqrt{2}$, which is the accumulation point of the $b = 1/3$ staircase. With $w_1 := 7$ and $w_i = S^{i-1}(w_1)$, we have the following interleaving family of numbers:

$$(1.2.7) \quad v_1 = \infty > w_1 = 7 > v_2 = 6 > w_2 = \frac{41}{7} > \dots > v_i > w_i > v_{i+1} > \dots \rightarrow a_{\min}.$$

The symmetry S acts as a shift, $S(v_i) = v_{i+1}$, $S(w_i) = w_{i+1}$, while R and its composition with powers of S are reflections. In particular $R_{v_i} := S^{i-2}RS^{-i+1}$ is a reflection that fixes v_i and interchanges w_{i-1} with w_i . As we show in (2.1.3), for each i the two b -values that correspond to the point $z = v_i$ are rational and have simple linear expressions in terms of the numerator and denominator p_i and q_i of $v_i = p_i/q_i$. Hence one might expect them to be relevant to the problem. Note that (except for one or two initial terms) the staircase steps in \mathcal{S}^U all lie in the interval $(7, \infty) = (w_1, v_1)$, where v_i and w_i are as in (1.2.7). Further, $S((w_1, v_1)) = (w_2, v_2)$ while $R((w_1, v_1)) = (v_2, w_1)$. We showed in [1] that R takes the blocking classes and staircase steps of \mathcal{S}^U to \mathcal{S}^L and that S takes \mathcal{S}^U to \mathcal{S}^E .

The next theorem gives a numerical description of the image of the staircase families \mathcal{S}^U by $T \in \mathcal{G}$. This extends [1, Corollary 60] where it is shown that S and R take the centers of the classes in \mathcal{S}^U to those of \mathcal{S}^E and \mathcal{S}^L respectively. This extended action is illustrated in Figure 2. Note that S preserves the direction (ascending or descending) of the centers of a family of classes, while R reverses it.

Theorem 1.2.4 *For each $T \in \mathcal{G}$ there is a corresponding prestaircase family $T^\#(\mathcal{S}^U)$, consisting of tuples $(d, m, p, q, t, \varepsilon)$ where*

- (p, q) are the image by T of the corresponding tuple in \mathcal{S}^U ;
- t is unchanged and ε transforms by the factor $(-1)^{i+\delta}$, where $T = S^i R^\delta$;

⁷ \mathcal{G} is not a semigroup because $S^i R S^j = S^{i-j} R \in \mathcal{G}$ only if $i \geq j$; see Lemma 2.1.3.

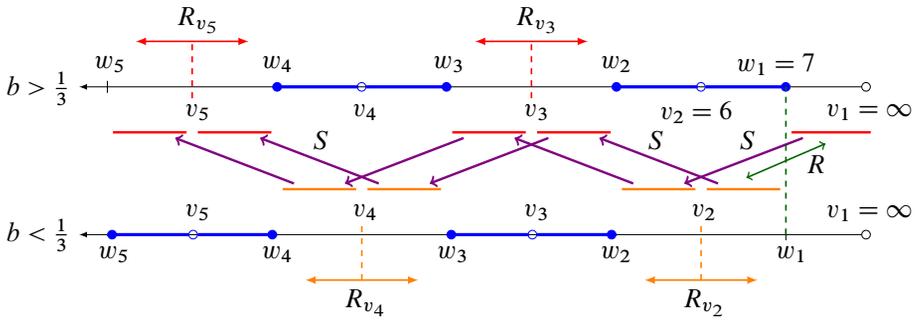


Figure 2: The centers of the staircase family S^U (resp. S^L) live in the right-most red (resp. orange) interval. Connecting these, in green, we have the reflection R about $w_1 = 7$ which sends the S^U to the S^L family. The shift S , purple, preserves orientation and takes staircase families with b value below (resp. above) $1/3$ to staircases with b value above (resp. below) $1/3$. By applying S iteratively to S^U and S^L , one obtains staircase families in the other red or orange half-intervals. In each of these half-intervals, there is a family of ascending and descending staircases whose accumulation points converge to some v_i . For each $i \geq 1$, each interval (w_i, w_{i-1}) admits a reflection R_{v_i} that fixes its center point v_i and interchanges the staircases in that interval. Finally, the blue intervals are blocked by a family of principal blocking classes with centers at the v_i .

- (d, m) are determined by (p, q, t, ε) according to (1.2.4). In particular, $\varepsilon(d - 3m)$ remains unchanged.

This prestaircase family is generated in the sense of Definition 1.2.1 by the images under T of the seeds and blocking classes of S^U . Further,

- If $T = S^i$ for some $i > 0$, then the centers of the preblocking classes of $T^\#(S^U)$ increase with n , and the prestaircase steps lie in the interval (w_{i+1}, v_{i+1}) and have $\varepsilon = (-1)^i$.
- If $T = S^i R$ for some $i \geq 0$, then the centers of the preblocking classes decrease with n , and the prestaircase steps lie in the interval (v_{i+1}, w_i) and have $\varepsilon = (-1)^{i+1}$.

In particular, $R^\#(S^U) = S^L$ while $S^\#(S^U) = S^E$.

For the proof see Section 3.3.

Remark 1.2.5 (i) When $b > 1/3$, the accumulation points of the staircases lie in the intervals (w_{2i+1}, w_{2i}) , while, for each i , the interval $[w_{2i}, w_{2i-1}]$ is blocked by

the principal blocking class $B_{v_{2i}}^P := (S^{2i-2})^\#(B_0^U)$ with center v_{2i} . Similarly, when $b < 1/3$, the accumulation points of the staircases lie in the intervals (w_{2i}, w_{2i-1}) , while, for each i , the interval $[w_{2i+1}, w_{2i}]$ is blocked by the principal blocking class $B_{v_{2i+1}}^P := (S^{2i-1})^\#(B_0^U)$ with center v_{2i+1} . See Figures 2 and 3 and Corollary 4.1.5.

(ii) As already noted, the action of a general symmetry $T = S^i R^\delta$ on the degree variable (d, m) of the blocking classes is determined by its action on (p, q) together with (1.2.4). We will see in Lemma 3.4.1 that the action of T on the (d, m) coordinates of the family of blocking classes B_n^U is given by a 2×2 integral matrix T_B^* . Because, as noted after (1.2.4), the linear function $d - 3m$ is invariant by T_B^* modulo sign, this matrix has eigenvector $(3, 1)^\vee$, with eigenvalue $(-1)^{i+\delta} \det(T_B^*)$. In general, the other eigenvector (with eigenvalue $(-1)^{i+\delta}$) has no obvious interpretation.

However, there are two cases in which it does. Indeed if T is the reflection R_{v_i} that fixes v_i , interchanging w_i with w_{i-1} , then it has two lifts to an action on degree depending on whether we take $b > 1/3$ or $b < 1/3$. As we show in Section 3.4, it turns out that only one of these actions on degree has order two, though they both have clear geometric interpretations. For example, if $i = 2$, so $v_2 = 6$, then $R_{v_2} = SR$ interchanges the centers of the blocking classes and seeds of the staircase families S^L and S^E . The lift $(R_{v_2})_B^*$ that acts on the degree components of the blocking classes has order two and eigenvector $(5, 1)^\vee$, since $1/5 = \text{acc}_L^{-1}(6)$ is the limit of the ratio of the degree components. On the other hand, the matrix $(R_{v_2})_P^* = (SR)_B^*$ takes the blocking classes of the staircase family S^U to those of $SR(S^U)$, fixing the shared principal blocking class B_0^U . Thus it has eigenvector $(3, 2)^\vee$, but, as we calculate in Example 3.4.7, does not have order two. For more details about the action of a general element $T \in \mathcal{G}$ on the degree components of the blocking classes, see Proposition 3.4.3.

Our second main result is that all the new staircases are live.

Theorem 1.2.6 *For each $T \in \mathcal{G}$ the prestaircase family $T^\#(S^U)$ is live, and hence is a staircase family.*

The proof that all the classes involved are perfect is given in Proposition 4.1.4. It is greatly eased by the discovery that their (d, m) components satisfy (1.2.4). The proof that the classes are live is given in Section 4. Although it is based on the methods developed in [1] that we explain at the beginning of Section 4, we have significantly simplified the proof by using new arithmetic arguments. Indeed, in Proposition 4.3.7 we establish a simple, widely applicable criterion for an arbitrary perfect prestaircase family to be live.

The paper [1] considered the following subsets of the b -parameter space $[0, 1)$:

$$\begin{aligned} \text{Stair} &:= \{b \mid H_b \text{ has a staircase}\} \subset [0, 1), \\ \text{Block} &:= \bigcup \{J_{\mathbf{B}} \mid \mathbf{B} \text{ is a blocking class}\} \subset [0, 1). \end{aligned}$$

Each blocking class \mathbf{B}_n^U defines an interval $J_{U,n} \subset (1/3, 1)$ of b -values that have no staircase because $\mu_{\mathbf{B}_n^U, b}(\text{acc}(b)) > V_b(\text{acc}(b))$. Further, the end points of these intervals lie in *Stair*. Similarly, the blocking classes in the staircase family $\mathcal{S}^L = R^\#(\mathcal{S}^U)$ define blocked intervals $J_{L,u} \subset (0, 1/3)$, whose endpoints lie in *Stair*.

There is an induced action of the symmetries in \mathcal{G} on the b -variable $[0, 1)$. This is easiest to describe for the shift S since this gives an injection $[a_{\min}, \infty) \rightarrow [a_{\min}, \infty)$ that fixes $a_{\min} := \text{acc}(1/3) = 3 + 2\sqrt{2}$. Hence, if we write

$$\text{acc}_L : [1/3, 1) \rightarrow [a_{\min}, \infty), \quad \text{acc}_U : [0, 1/3] \rightarrow [a_{\min}, \infty)$$

for the appropriate restriction of the function $b \mapsto \text{acc}(b)$ in (1.1.1), for each $k \geq 1$ we can define $(S^k)^*$ by

$$(1.2.8) \quad (S^k)^*(b) = \begin{cases} \text{acc}_L^{-1} \circ S^k \circ \text{acc}_L & \text{if } b \leq 1/3, \\ \text{acc}_U^{-1} \circ S^k \circ \text{acc}_U & \text{if } b \geq 1/3. \end{cases}$$

Since $(S^k)^*$ is conjugate to S^k , the assignment $k \mapsto (S^k)^*$ is a homomorphism; ie $(S^k)^* \circ (S^m)^* = (S^{k+m})^*$ for $k, m \geq 0$.

The reflection symmetries $S^k R$ are not defined on the whole z -interval $[a_{\min}, \infty)$ and so do not extend to a global action on b -variable. Instead, in view of Remark 1.2.5(ii), it is most natural to restrict the reflection $R_{v_i} := S^{2i-1} R$ that fixes v_i to the appropriate b -interval corresponding to $(v_{i+1}, v_{i-1}) \supset (w_i, w_{i-1})$. Thus we define

$$(1.2.9) \quad (R_{v_i})^*(b) = \begin{cases} \text{acc}_L^{-1} \circ R_{v_i}|_{(v_{i+1}, v_{i-1})} \circ \text{acc}_L & \text{if } i \text{ is even,} \\ \text{acc}_U^{-1} \circ R_{v_i}|_{(v_{i+1}, v_{i-1})} \circ \text{acc}_U & \text{if } i \text{ is odd.} \end{cases}$$

The following is an immediate consequence of Theorem 1.2.6 because the endpoints of the blocked b -intervals, $J_{U,n}$ and $J_{L,n}$, are taken by the function $b \mapsto \text{acc}(b)$ to the accumulation points of the corresponding staircases, upon which \mathcal{G} acts geometrically.

Corollary 1.2.7 *Let $J_{U,n}$ (resp. $J_{L,n}$) be the b -interval blocked by \mathbf{B}_n^U (resp. \mathbf{B}_n^L). For $J = J_{U,n}$ with $n \geq 0$, or $J_{L,n}$ with $n \geq 1$ and each $k \geq 0$, the interval $(S^k)^*(J)$ is a component of *Block*, and its endpoints are in *Stair*. Moreover, for each $i \geq 2$ the reflection $(R_{v_i})^*$ permutes those intervals $(S^k)^*(J)$ that lie entirely in (v_{i+1}, v_{i-1}) .*

Remark 1.2.8 (i) We conjecture that the action of the elements $T \in \mathcal{G}$ described above preserves the sets *Stair* and *Block*. This would hold if, for example, every interval

in *Block* was defined by a single blocking class with two associated staircases, and if these, plus the staircases at 0 and 1/3, were the only staircases. However, the proof of such a result seems out of reach at present.

(ii) Note that the map $(R_{v_i})^*$ has order two since it is defined to be the conjugate of a reflection. The statements in Remark 1.2.5(ii) about the transformations $(R_{v_i})^\sharp$ are rather different, since here we are concerned with the action of R_{v_i} on the degree variables (d, m) of the relevant blocking classes, and not on the conjugate to its action on the z -variable.

2 The accumulation function and its symmetries

In this section, we first discuss the arithmetic properties of the symmetries and related topics. In Section 2.2 we first give an alternative way to understand the coordinates (d, m, p, q) of a quasiperfect class, and then show that all such classes are center-blocking. This second result relies on the particular form of the accumulation function $b \mapsto \text{acc}(b)$. Finally, we describe the staircase families S^U and S^L and the staircase at $b = 1/3$ in the language used in [1].

2.1 The fundamental recursion

We have two sets of variables: the z variable on the domain $E(1, z)$ and the b variable on the target. They are related by the equation

$$(2.1.1) \quad z^2 - \left(\frac{(3-b)^2}{1-b^2} - 2 \right) z + 1 = 0.$$

Since $b \in [0, 1)$, one can check that this equation has two positive solutions that we denote by a and $1/a$, where $a := \text{acc}(b) > 1$. As illustrated in Figure 1, this function is in general two-to-one with a unique minimum $\text{acc}^{-1}(3 + 2\sqrt{2}) = 1/3$. We denote by⁸

$$\text{acc}_L^{-1}: (3 + 2\sqrt{2}, \frac{1}{2}(7 + 3\sqrt{5})] \rightarrow [0, \frac{1}{3}), \quad \text{acc}_U^{-1}: (3 + 2\sqrt{2}, \infty) \rightarrow (\frac{1}{3}, 1)$$

the corresponding inverses to the function $b \mapsto \text{acc}(b)$. Thus, with

$$\tau := a + \frac{1}{a} + 2 = \frac{(3-b)^2}{1-b^2} \geq 8,$$

we have

$$\text{acc}_L^{-1}(a) = \frac{3 - \sqrt{\tau^2 - 8\tau}}{\tau + 1}, \quad \text{acc}_U^{-1}(a) = \frac{3 + \sqrt{\tau^2 - 8\tau}}{\tau + 1}.$$

⁸Here and elsewhere, L (or ℓ) denotes “lower” while U (or u) denotes “upper”.

The minimum $\tau = 8$ is attained when $b = 1/3$. The corresponding equation

$$z^2 - 6z + 1 = 0$$

is therefore very special. It arises by taking the limit $\lim x_{\kappa+1}/x_{\kappa}$ of the recursion $x_{\kappa+1} = 6x_{\kappa} - x_{\kappa-1}$ that seems to play a central role in this staircase problem. For example, the steps of the staircase at $b = 1/3$ satisfy this recursion. Moreover, as the following lemma shows, certain properties of the function $b \mapsto \text{acc}(b)$ are invariant with respect to this recursion.

Lemma 2.1.1 (i) *If $\text{acc}(b) = p/q$ then*

$$(2.1.2) \quad b = \frac{3pq \pm (p+q)\sqrt{\sigma}}{p^2 + q^2 + 3pq}, \quad \text{where } \sigma := (p-3q)^2 - 8q^2.$$

In particular, b and $\text{acc}(b)$ are both rational if and only if $(p-3q)^2 - 8q^2 = k^2$ for some integer $k \geq 1$.

(ii) *The quantity $\sigma(p, q) := (p-3q)^2 - 8q^2 = p^2 + q^2 - 6pq$ is invariant under the transformation $(p, q) \mapsto (6p - q, p)$.*

(iii) *In particular, because $(p, q) = (6, 1)$ is a solution of $\sigma(p, q) = 1$, any successive pair in the sequence*

$$(y_1, y_2, y_3, y_4, \dots) = (1, 6, 35, 204, \dots)$$

gives another solution. Further, these are the only solutions for $\sigma = 1$.

(iv) *If $p = y_i$ and $q = y_{i-1}$ for some $i > 1$, we have*

$$(2.1.3) \quad \text{acc}_U^{-1}\left(\frac{p}{q}\right) = \frac{p+q+3}{3p+3q+1}, \quad \text{acc}_L^{-1}\left(\frac{p}{q}\right) = \frac{p+q-3}{3p+3q-1},$$

and so

$$\begin{aligned} \text{acc}_U^{-1}(6) &= \frac{5}{11}, & \text{acc}_U^{-1}\left(\frac{35}{6}\right) &= \frac{11}{31}, & \text{acc}_U^{-1}\left(\frac{204}{35}\right) &= \frac{121}{359}, & \dots &\searrow \frac{1}{3} \\ \text{acc}_L^{-1}(6) &= \frac{1}{5}, & \text{acc}_L^{-1}\left(\frac{35}{6}\right) &= \frac{19}{61}, & \text{acc}_L^{-1}\left(\frac{204}{35}\right) &= \frac{59}{179}, & \dots &\nearrow \frac{1}{3}. \end{aligned}$$

Proof Let $a = p/q > 1$ so that $a = \text{acc}(b)$ where $b = (3 \pm \sqrt{\tau^2 - 8\tau})/(\tau + 1)$, and $\tau = p/q + q/p + 2$. Since

$$\tau(\tau - 8) = \frac{(p^2 + q^2 + 2pq)(p^2 + q^2 - 6pq)}{p^2q^2},$$

and $p^2 + q^2 - 6pq = (p - 3q)^2 - 8q^2$, this formula for b simplifies to that in (2.1.2). The rest of (i) is clear. Next, note that (ii) holds because

$$(6p - q)^2 + p^2 - 6(6p - q)p = p^2 + q^2 - 6pq.$$

To prove (iii), note that given any solution (p, q) with $p > q$, one can use the reverse iteration $(p, q) \mapsto (q, 6q - p)$ to reduce to a solution with $p > q > 0$ and $6q - p \leq 0$. But the only such solution is $(6, 1)$. Finally, to see that the formula in (iv) for $\text{acc}_U^{-1}(p/q)$ is the same as that in (2.1.2) we must check that

$$(3pq + p + q)(3p + 3q + 1) = (3pq + p^2 + q^2)(p + q + 3) = (9pq + 1)(p + q + 3).$$

One can check that the third order terms on both sides are the same, and that the rest of the identity holds because $p^2 + q^2 = 6pq + 1$. The proof for $\text{acc}_L^{-1}(p/q)$ is similar. Thus $\text{acc}_U^{-1}(y_i/y_{i-1})$ decreases with limit $1/3$, while $\text{acc}_L^{-1}(y_i/y_{i-1})$ increases with limit $1/3$. □

Remark 2.1.2 Since $\sigma(p, q) = \sigma(q, p)$, (p, q) is a solution of the equation $\sigma(p, q) = 1$ if and only if (q, p) is. We always assume that $p > q$ so that the entries in the pairs $(p, q), S(p, q), S(S(p, q)), \dots$ increase; with the other convention they would decrease. Notice also that the Pell numbers $0, 1, 2, 5, 12, 29, 70, \dots$ that form such a basic element in the polydisc case considered in [4; 7] are closely related to the sequence $0, 1, 6, 35, \dots$ that is fundamental here; indeed the numbers $2y_i$ for $i \geq 0$ are precisely the even-placed Pell numbers.

Let $S := \begin{pmatrix} 6 & -1 \\ 1 & 0 \end{pmatrix}$ be the “shift” matrix that implements the recursion

$$(2.1.4) \quad x_{\kappa+1} = 6x_{\kappa} - x_{\kappa-1}, \quad S \begin{pmatrix} x_{\kappa} \\ x_{\kappa-1} \end{pmatrix} = \begin{pmatrix} x_{\kappa+1} \\ x_{\kappa} \end{pmatrix},$$

where the matrix $A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ acts on the z variables by the fractional linear transformation

$$(2.1.5) \quad z \mapsto \frac{az + b}{cz + d} =: Az$$

taking (by extension) ∞ to a/c . Starting with $y_0 = 0$ and $y_1 = 1$, we get the sequence $y_0 = 0, y_1 = 1, y_2 = 6, y_3 = 35, y_4 = 204, y_5 = 1189, y_6 = 6930, \dots,$

$$\frac{35}{6} = [5; 1, 5], \quad \frac{204}{35} = [5; 1, 4, 1, 5], \quad \frac{1189}{204} = [5; 1, 4, 1, 4, 1, 5], \quad \dots,$$

with general term $[5; \{1, 4\}^k, 1, 5]$; see Lemma 2.1.6. If we define

$$(2.1.6) \quad R := \begin{pmatrix} 6 & -35 \\ 1 & -6 \end{pmatrix} = \begin{pmatrix} y_2 & -y_3 \\ y_1 & -y_2 \end{pmatrix},$$

then $R^2 = \text{id}$ and $\det R = -1$, and we will see that the reflection R and shift S generate symmetries of our problem.

It is convenient to consider the following decreasing sequences of points in the interval $(3 + 2\sqrt{2}, \infty)$:

$$(2.1.7) \quad \begin{aligned} v_1 &:= \infty, & v_2 &:= 6, & v_3 &:= \frac{35}{6}, & v_j &:= \frac{y_j}{y_{j-1}}, \\ w_1 &:= 7, & w_2 &:= \frac{41}{7}, & w_k &:= \frac{y_{k+1} + y_k}{y_k + y_{k-1}}. \end{aligned}$$

These sequences interweave:

$$3 + 2\sqrt{2} < \dots < w_k < v_k < w_{k-1} < \dots < w_2 < v_2 < w_1 < v_1 = \infty.$$

Lemma 2.1.3 *Let v_j, w_j, S , and R be as above. Then:*

(i) *The following matrix identities hold:*

$$SR = RS^{-1}, \quad S^{-1}R = RS, \quad R \circ R = \text{Id}.$$

(ii) *The matrix*

$$S^k = \begin{pmatrix} y_{k+1} & -y_k \\ y_k & -y_{k-1} \end{pmatrix}$$

has determinant 1; ie

$$(2.1.8) \quad y_k^2 = y_{k+1}y_{k-1} + 1 = 6y_{k+1}y_k - y_{k+1}^2 + 1 \quad \text{for all } k \geq 1.$$

(iii) *With action as in (2.1.5), $S(v_j) = v_{j+1}$ and $S(w_j) = w_{j+1}$, for $j \geq 1$.*

(iv) *The matrices S and R generate the subgroup of $\text{PGL}(2, \mathbb{Z})$ that fixes the quadratic form $p^2 - 6pq + q^2$.*

Proof The proof of (i)–(iii) is straightforward. In particular the formula for S^k holds because S implements the recursion, and we also have $\det(S^k) = (\det(S))^k = 1$. Further, one can check that $A \in \text{GL}(2, \mathbb{Z})$ preserves the form $p^2 - 6pq + q^2$ if and only if

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{where } c = -b, \quad d = 6b + a.$$

Hence, if $\det(A) = 1$ then $a^2 + 6ab + b^2 = 1$, which implies by Lemma 2.1.1(iii) that when $a > -b > 0$ we must have $(a, b) = (y_{k+1}, -y_k)$ for some k , so $A = S^k$ for some $k \geq 1$. It follows similarly that the only other matrices A that preserve the form and have $\det(A) = 1$ have the form $\pm S^k$ for some $k \leq 0$. Further, if $\det(A) = -1$ then

because the matrix RA has determinant 1 and preserves the quadratic form, we must have $A = \pm S^k R$ for some $k \in \mathbb{Z}$. Thus (iv) holds. \square

Corollary 2.1.4 (i) For each $i \geq 1$, the restriction of

$$(2.1.9) \quad R_{v_i} := S^{i-2}RS^{-(i-1)} = S^{2i-3}R = \begin{pmatrix} y_{2i-1} & -y_{2i} \\ y_{2i-2} & -y_{2i-1} \end{pmatrix}$$

to the interval (v_{2i-1}, v_1) is a reflection that fixes v_i and interchanges the points w_{i+k}, w_{i-k-1} , and v_{i+k}, v_{i-k} , for $0 \leq k \leq i - 1$.

(ii) The restriction of R to the interval $(v_2, v_1) = (6, \infty)$ is a reflection that fixes $w_1 = 7$.

We end this subsection with a brief discussion of the weight expansion and continued fractions.

Definition 2.1.5 The (integral) *weight expansion* of a rational number $p/q \geq 1$ is a recursively defined, nonincreasing sequence of integers $W(p/q) = (W_1, W_2, \dots)$ defined as follows: $W_1 = q$ and $W_n \geq W_{n+1}$ for all n , and if $W_i > W_{i+1} = \dots = W_n$ (where we set $W_0 := p$), then

$$W_{n+1} = \begin{cases} W_{i+1} & \text{if } W_{i+1} + \dots + W_{n+1} = (n - i + 1)W_{i+1} < W_i, \\ W_i - (n - i)W_{i+1} & \text{otherwise.} \end{cases}$$

Thus $W(p/q)$ starts with $\lfloor p/q \rfloor$ copies of q (where $\lfloor p/q \rfloor$ is the largest integer $\leq p/q$), and ends with some number ≥ 2 of copies of 1. One can check that

$$\sum_{i \geq 1} W_i = p + q - 1, \quad \sum_{i \geq 1} W_i^2 = pq.$$

Using this, it is straightforward to check that equations (1.2.1), for a quasiperfect tuple (d, m, p, q) , imply that the corresponding class $\mathbf{E} = dL - mE_0 - \sum_i W_i E_i$ satisfies the conditions $c_1(\mathbf{E}) = 1$ and $\mathbf{E} \cdot \mathbf{E} = -1$, as claimed earlier. Moreover, the multiplicities ℓ_0, \dots, ℓ_k of the entries in $W(p/q)$ are the coefficients of the continued fraction expansion of p/q . Thus, if the distinct weights are $X_0 := q > X_1 > \dots > X_k = 1$ and we write

$$W(p/q) = (X_0^{\times \ell_0}, X_1^{\times \ell_1}, \dots, 1^{\times \ell_k}),$$

then

$$\frac{p}{q} = [\ell_0; \ell_1, \dots, \ell_k] := \ell_0 + \frac{1}{\ell_1 + \frac{1}{\ell_2 + \dots + \frac{1}{\ell_k}}}.$$

As we see from Theorems 2.3.1 and 2.3.3 below, the centers of the staircase steps have very regular continued fraction expansions that, as we now show, behave well under the symmetries.

Lemma 2.1.6 (i) *The shift $S = \begin{pmatrix} 6 & -1 \\ 1 & 0 \end{pmatrix}$ has the following effect on continued fraction expansions, where, for $x \geq 1$, $CF(x)$ denotes the continued fraction of x :*

- *If $z = p/q = [5 + k; CF(x)]$ for some $k \geq 0, x \geq 1$, then*

$$Sz = \frac{6p - q}{p} = [5; 1, 4 + k, CF(x)].$$

In particular, if $z > 6$ then $k \geq 1$ and the third entry in $CF(Sz)$ is at least 5, while if $z < 6$ then $k = 0$ and $x > 1$ and we have

$$S([5; CF(x)]) = [5; 1, 4, CF(x)].$$

(ii) *The reflection $R = \begin{pmatrix} 6 & -35 \\ 1 & -6 \end{pmatrix}$ has the following effect on continued fraction expansions:*

- *If $z = p/q = [6 + k; CF(x)]$ for some $k \geq 1$ and $x \geq 1$, then*

$$Rz = \frac{6p - 35q}{p - 6q} = [6; k, CF(x)].$$

Further, $R \circ R = \text{id}$.

(iii) *The quantity $p^2 - 6pq + q^2$ is invariant by both S and R .*

Proof If $z = [5 + k; CF(x)] = 5 + k + 1/x = ((5 + k)x + 1)/x$ then

$$Sz = \frac{(29 + 6k)x + 6}{(5 + k)x + 1} = 5 + \frac{(4 + k)x + 1}{(5 + k)x + 1},$$

while

$$[5; 1, 4 + k, CF(x)] = 5 + \frac{1}{1 + 1/(4 + k + 1/x)} = 5 + \frac{(4 + k)x + 1}{(5 + k)x + 1}.$$

This proves (i). The proof of (ii) is similar, and (iii) follows by an easy calculation. \square

2.2 Quasiperfect classes

We first explain the action of the symmetries on the quasiperfect classes, and then show in Proposition 2.2.9 that every quasiperfect class with center $> a_{\min} = 3 + 2\sqrt{2}$ is a center-blocking class.

As explained in (1.2.1), a quasiperfect class $E = dL - mE_0 - \sum_i m_i E_i$ is determined by a tuple (d, m, p, q) of positive integers (where (m_1, m_2, \dots) is the weight expansion of p/q as in Definition 2.1.5) that satisfies the conditions

$$(2.2.1) \quad 3d = p + q + m, \quad d^2 - m^2 = pq - 1.$$

We will continue to call these the *Diophantine conditions* on E as in [1]. If we use the first equation above to express m as a function of $d, p,$ and $q,$ the second equation is a quadratic in $d,$

$$(2.2.2) \quad 8d^2 - 6d(p + q) + p^2 + 3pq + q^2 - 1 = 0,$$

with solution

$$d = \frac{1}{8}(3(p + q) \pm \sqrt{p^2 - 6pq + q^2 + 8}).$$

Thus, if we define

$$(2.2.3) \quad t := \sqrt{p^2 - 6pq + q^2 + 8}, \quad \varepsilon := \pm 1,$$

the coefficients d and m in E are given by the formulas

$$(2.2.4) \quad d := \frac{1}{8}(3(p + q) + \varepsilon t), \quad m := \frac{1}{8}((p + q) + 3\varepsilon t)$$

in (1.2.4). In other words, modulo an appropriate choice of $\varepsilon,$ we can think of a quasiperfect class as an integer point on the quadratic surface X defined by (2.2.2), where we can use either the coordinates (d, p, q) or $(p, q, t).$ ⁹ Note, however, that the fact that $p, q,$ and t are integers does not imply that d and m are also.

We now show that a quasiperfect class $(d, m, p, q, t, \varepsilon)$ is uniquely determined by its center $p/q.$

Lemma 2.2.1 *For each integral solution (p, q, t) of the equation $t^2 = p^2 - 6pq + q^2 + 8,$ there are integers (d, m) satisfying*

$$(2.2.5) \quad d = \frac{1}{8}(3(p + q) + \varepsilon t), \quad m = \frac{1}{8}((p + q) + \varepsilon 3t)$$

for at most one value of $\varepsilon \in \{\pm 1\}.$

Proof If this is false there are positive integers $p, q, t, d_{\pm},$ and m_{\pm} such that (d_+, m_+) solve (2.2.5) for $\varepsilon = +1,$ while (d_-, m_-) solve it for $\varepsilon = -1.$ Then

$$d_+ + d_- = \frac{3}{4}(p + q), \quad d_+ - d_- = \frac{1}{4}t$$

⁹We are indebted to Peter Sarnak for explaining this point of view to us.

are integers. Since p and q cannot both be even they may be written as $4a + 1$ and $4b - 1$ (in some order), and, with $t := 4s$,

$$s^2 = a^2 - 6ab + b^2 + 2a - 2b + 1.$$

But also we need 8 to divide $3(p + q) + \varepsilon t$, which implies that $a + b + s$ is even. It is now easy to see that there are no integer solutions. \square

Corollary 2.2.2 *There is at most one quasiperfect class with center p/q . Conversely, for given (d, m) there is at most one quasiperfect class with these degree variables and $p/q > 1$.*

Proof The first claim follows immediately from Lemma 2.2.1. To prove the second, notice that d and m determine $p + q = 3d - m$ and $pq = d^2 - m^2 + 1$, which uniquely determines p and q modulo order. \square

We now discuss the effect of the symmetries on these classes. As always, we write

$$S = \begin{pmatrix} 6 & -1 \\ 1 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 6 & -35 \\ 1 & -6 \end{pmatrix},$$

where S is the shift and R is the reflection that fixes 7. Note that the action of S on p and q fixes t by Lemma 2.1.1(ii). It is also easy to check that R also fixes t . We now show that the action of these transformations act on the integer points of X extends to an action on the tuples $(d, m, p, q, t, \varepsilon)$. It follows from Lemma 2.1.3 that any element of the group generated by S and R can be written $S^i R^\delta$ for $i \in \mathbb{Z}$ and $\delta \in \{0, 1\}$.

Definition 2.2.3 Let $T = S^i R^\delta$ for $i \in \mathbb{Z}$ and $\delta \in \{0, 1\}$, and suppose that $(d, m, p, q, t, \varepsilon)$ is a tuple of integers that satisfy the identities in (2.2.3) and (2.2.4). Then we define

$$(2.2.6) \quad T^\sharp(d, m, p, q, t, \varepsilon) := (d', m', p', q', t, \varepsilon') = (d', m', T(p, q), t, (-1)^{i+\delta} \varepsilon)$$

where d' and m' are given by the formulas

$$d' := \frac{1}{8}(3(p' + q') + \varepsilon't), \quad m' := \frac{1}{8}((p' + q') + 3\varepsilon't), \quad \varepsilon' := (-1)^{i+\delta} \varepsilon.$$

The next lemma shows that this action of T preserves integrality.

Lemma 2.2.4 *For each $(d, m, p, q, t, \varepsilon)$ and $T = S^i R^\delta$ as above, $T^\sharp(d, m, p, q, t, \varepsilon)$ is also integral and satisfies the Diophantine conditions (2.2.1). Moreover, for all such T_1 and T_2 , $(T_1 T_2)^\sharp(d, m, p, q, t, \varepsilon) = (T_1)^\sharp(T_2)^\sharp(d, m, p, q, t, \varepsilon)$.*

Proof The above construction shows that every integral point (p, q, t) of X can be extended to a tuple $(d, m, p, q, t, \varepsilon)$ that satisfies the Diophantine conditions. In particular, since t is invariant under the action of S and T , the tuple $(d', m', p', q', t, \varepsilon')$ satisfies these conditions; however we do need to check that it is integral. Because $3(3(p' + q') + \varepsilon't) - ((p' + q') + 3\varepsilon't)$ is divisible by 8, it suffices to check that $(p' + q') + 3\varepsilon't$ is divisible by 8. Thus it suffices to check that if 8 divides $p + q + 3\varepsilon t$ for some $\varepsilon \in \{\pm 1\}$, and we set $(p', q') = S(p, q) = (6p - q, p)$, then 8 divides $p' + q' - 3\varepsilon t = 7p - q - 3\varepsilon t$. But this is immediate, since $7p - q - 3\varepsilon t = 8p - (p + q + 3\varepsilon t)$. A similar calculation proves that the action of R preserves integrality.

This proves the first claim. The second follows immediately from the fact that the action of S and R on the coordinates (p, q) is compatible with composition. \square

Remark 2.2.5 The tuples $(d, m, p, q, t, \varepsilon)$ that correspond to quasiperfect classes have positive entries with $p > q > 0$. As we shall see in [Example 2.3.7](#), the classes with $t = 1$ belong to the staircase at $b = 1/3$, while all other classes of interest have $t \geq 3$ and hence $p/q > a_{\min} := 3 + 2\sqrt{2}$ (so that $p^2 - 6pq + q^2 > 0$). Therefore the full subgroup of $\text{PGL}(2, \mathbb{Z})$ generated by S and R does not act on the staircases. This is why in [Theorem 1.2.4](#) we only consider the restriction of the action of the elements $S^i R^\delta$ for $i \geq 0$ to the classes with centers $p/q > 7$.

Remark 2.2.6 We now relate our description of the symmetries to that given by Usher in [\[7, Section 2.2.1\]](#). He denotes a quasiperfect class E in a blowup of $S^2 \times S^2$ by the tuple (a, b, c, d) , where a and b are the coefficients of the two lines (each with Chern class 2) and $(c, d) := (p, q)$ are the coordinates of its center. Thus the equations $c_1(E) = 1$ and $E \cdot E = -1$ become

$$2(a + b) = p + q, \quad 2ab = pq - 1.$$

The first equation implies that there are integers x, δ , and ε such that

$$(a, b, p, q) = \left(\frac{1}{2}(x + \varepsilon), \frac{1}{2}(x - \varepsilon), x + \delta, x - \delta\right).$$

With these variables, the second equation is then

$$(2.2.7) \quad x^2 - 2\delta^2 = 2 - \varepsilon^2.$$

In terms of the element $x + \delta\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$, this equation simply says that $x + \delta\sqrt{2}$ has norm $2 - \varepsilon^2$. He now considers symmetries of the form

$$x + \delta\sqrt{2} \mapsto x' + \delta'\sqrt{2} := (u + v\sqrt{2})(x + \delta\sqrt{2}) = ux + 2v\delta + (vx + u\delta)\sqrt{2},$$

where $u + v\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$ is an element of norm 1, ie $u^2 - 2v^2 = 1$. Then $(x', \delta', \varepsilon)$ is another solution of (2.2.7). Usher considers the symmetries given by

$$u + v\sqrt{2} = H_{2k} + P_{2k}\sqrt{2}$$

where H_{2k} and P_{2k} are respectively half-Pell and Pell numbers. When $k = 1$, $H_2 = 3$ and $P_2 = 2$, and we get the unit $u + v\sqrt{2} = 3 + 2\sqrt{2}$. Therefore $x' = 3x + 4\delta$ and $\delta' = 2x + 3\delta$, so $(p', q') := (x' + \delta', x' - \delta') = (5x + 7\delta, x + \delta)$. Substituting $x = \frac{1}{2}(p + q)$ and $\delta = \frac{1}{2}(p - q)$, we obtain the transformation

$$(p, q) \mapsto (p', q') = (6p - q, p).$$

More generally, Usher states that the transformation has formula

$$(x, \delta) \mapsto (H_{2k}x + 2P_{2k}\delta, P_{2k}x + H_{2k}\delta),$$

which, in terms of the (p, q) coordinates, translates to

$$(p, q) \mapsto ((H_{2k} + \frac{3}{2}P_{2k})p - \frac{1}{2}P_{2k}q, \frac{1}{2}P_{2k}p + (H_{2k} - \frac{3}{2}P_{2k})q).$$

To see that this map is the same as $(p, q) \mapsto S^k(p, q)$, notice that by Remark 2.1.2, the entries y_i of S^k have the form $\frac{1}{2}P_{2i}$. Therefore, we have to check a linear identity between the Pell numbers P_n and their half-companions H_n . But because of the recursion, such an identity holds if and only if it holds for two distinct values of n , and when $k = 2$ we have $(H_4, P_4) = (17, 12)$, which gives $(p, q) \mapsto (35p - 6q, 6p - q)$, as required.

Finally, we observe that Usher’s symmetries preserve ε which is related to the recursion variable for his staircases, and hence plays much the same role as our variable t .

The next lemma, taken from [1, Example 32] is the key to the proof that every quasiperfect class is center-blocking.

Lemma 2.2.7 *For all $b \in [0, 1)$ the graph of the function $z \mapsto (1 + z)/(3 - b)$ passes through the accumulation point $(\text{acc}(b), V_b(\text{acc}(b)))$. Moreover, for each b , the line lies above the volume curve when $z > \text{acc}(b)$.*

Proof Let $c(b) = (3 - b)^2/(1 - b^2) - 2$. We have

$$\begin{aligned} (2.2.8) \quad \frac{1 + \text{acc}(b)}{3 - b} &= \sqrt{\frac{\text{acc}(b)}{1 - b^2}} \\ &\iff (1 + \text{acc}(b))^2 = \text{acc}(b) \frac{(3 - b)^2}{1 - b^2} = \text{acc}(b)(c(b) + 2) \\ &\iff \text{acc}(b)^2 - c(b)\text{acc}(b) + 1 = 0, \end{aligned}$$

which holds by the definition of $\text{acc}(b)$ in (2.1.1). It follows that the two functions

$$(2.2.9) \quad b \mapsto V_b(\text{acc}(b)), \quad b \mapsto \frac{1 + \text{acc}(b)}{3 - b}$$

are the same.

It remains to note that for all $z \geq a_{\min} > 5$, the slope

$$\frac{d}{dz} \left(\sqrt{\frac{z}{1 - b^2}} \right) = \frac{1}{2} \frac{V_b(z)}{z} = \frac{1 + z}{2z(3 - b)}, \quad z := \text{acc}(b)$$

of the volume curve $V_b(z)$ at $z = \text{acc}(b)$ is smaller than $1/(3 - b)$ which is the slope of the line. □

Remark 2.2.8 (i) This special property of the function $z \mapsto (1 + z)/(3 - b)$ is the key reason why the linear relations in (1.2.5) imply that the staircase converges to an endpoint of the interval blocked by the associated blocking class; see the proof of [1, Theorem 52].

(ii) By [1, Example 32], when $5 < z < 6$ the function $z \mapsto (1 + z)/(3 - b)$ that occurs in Lemma 2.2.7 is the obstruction given by the class $E = 3L - E_0 - 2E_1 - E_{2\dots 6}$. As noticed by Tara Holm, all the other convex toric domains discussed in [3] seem to have classes that play a similar role.

Proposition 2.2.9 Every quasiperfect class $E := (d, m, p, q, t, \varepsilon)$ with $p/q > a_{\min} = 3 + 2\sqrt{2}$ is center-blocking.

Proof Define $\text{acc}_\varepsilon^{-1}$ to be acc_U^{-1} if $\varepsilon = 1$ and acc_L^{-1} if $\varepsilon = -1$.¹⁰ Then we must check that

$$\mu_{E,b} \left(\frac{p}{q} \right) = \frac{p}{d - mb} > V_b \left(\frac{p}{q} \right) = \frac{p + q}{q(3 - b)}, \quad b := \text{acc}_\varepsilon^{-1} \left(\frac{p}{q} \right),$$

where we have used (1.1.3) and (2.2.9). Thus we need

$$pq(3 - b) > (p + q)(d - mb),$$

or equivalently

$$3pq - d(p + q) > b(pq - m(p + q)).$$

By (2.1.2) we have $b = (3pq + \varepsilon(p + q)\sqrt{\sigma})/((p + q)^2 + pq)$, where $\sigma + 8 = t^2$. Thus we must check that

$$((p + q)^2 + pq)(3pq - d(p + q)) > (3pq + \varepsilon(p + q)\sqrt{\sigma})(pq - m(p + q)).$$

¹⁰By (2.2.4) we have $m/d > 1/3$ exactly when $\varepsilon = 1$. Moreover the condition $p/q > a_{\min}$ implies that $p/q = \text{acc}(b)$ for at least one b so it is in the domain of $\text{acc}_\varepsilon^{-1}$ for at least one value of ε .

By deleting the term $3p^2q^2$ from both sides, multiplying by 8 and substituting for d and m , we obtain the equivalent inequality

$$\begin{aligned} 24pq(p+q)^2 - (p+q)((p+q)^2 + pq)(3(p+q) + \varepsilon t) \\ > -3pq(p+q)(p+q + 3\varepsilon t) + \varepsilon(p+q)\sqrt{\sigma}(8pq - (p+q + 3\varepsilon t)(p+q)) \\ = -3pq(p+q)(p+q + 3\varepsilon t) + \varepsilon(p+q)\sqrt{\sigma}(-\sigma - 3\varepsilon t(p+q)), \end{aligned}$$

where the last equality uses the identity $8pq - (p+q)^2 = -\sigma$. If we take all the terms in this inequality that involve an even power of ε , and put them on the left-hand side, we obtain

$$\begin{aligned} 24pq(p+q)^2 - 3(p+q)^2((p+q)^2 + pq) + 3pq(p+q)^2 + 3(p+q)^2t\sqrt{\sigma} \\ = 3(p+q)^2(8pq - (p+q)^2 - pq + pq + \sigma + (t\sqrt{\sigma} - \sigma)) \\ = 3(t\sqrt{\sigma} - \sigma)(p+q)^2 =: A > 0, \end{aligned}$$

since $\sigma = p^2 - 6pq + q^2$ and $t^2 = \sigma + 8$. If we do the same with the coefficient of ε , we obtain

$$\begin{aligned} -t(p+q)((p+q)^2 + pq) + 9t(p+q)pq + (p+q)\sigma\sqrt{\sigma} \\ = t(p+q)(-p^2 - 3pq - q^2 + 9pq + \sigma) - (t - \sqrt{\sigma})(p+q)\sigma \\ = (\sqrt{\sigma} - t)(p+q)\sigma =: B. \end{aligned}$$

We need to check that $A > |B|$, which is equivalent to $3(p+q) > \sqrt{\sigma}$. Since this holds, the required inequality is established. □

Remark 2.2.10 Proposition 2.2.9 shows that every class that is defined by a tuple $(d, m, p, q, t, \varepsilon)$ with $p/q > a_{\min}$ as in (1.2.4) is in fact a center-blocking class. Thus all our stair steps are center-blocking. However, we have not been able to resolve the question of whether there is a blocking class \mathbf{B} of more general type that is not center-blocking. In this case, there would be a point $b_0 \in [0, 1)$ such that $\mu_{\mathbf{B}, b_0}(\text{acc}(b_0)) > V_{b_0}(\text{acc}(b_0))$. However, if I is the largest interval containing $\text{acc}(b_0)$ on which $\mu_{\mathbf{B}, b_0}$ is obstructive, and if $a \in I$ is the corresponding break point — see [1, Lemma 14] — then $\mu_{\mathbf{B}, b}$ would not block a , ie for both elements $b \in \text{acc}^{-1}(a)$ we would have $\mu_{\mathbf{B}, b}(a) \leq V_b(a)$. (For further discussion of blocking classes, see [1, Section 2.3].) We bypass this question here by restricting attention to (quasi)perfect center-blocking classes.

The following fact about blocking classes was pointed out to us by Morgan Weiler. It is somewhat surprising since we know from [1, Lemmma 15(iii)] that every quasi-perfect class $\mathbf{B} = (d, m, p, q)$ is obstructive when $b = m/d$ and $z = p/q$, ie we have

$\mu_{\mathbf{B},m/d}(p/q) > V_{m/d}(p/q)$. Thus it is natural to think that m/d would lie in the blocked interval $J_{\mathbf{B}}$, which is defined to be the maximal interval containing $\text{acc}_{\varepsilon}^{-1}(p/q)$ consisting of parameters b such that $\mu_{\mathbf{B},b}(\text{acc}(b)) > V_b(\text{acc}(b))$. However, we now show that this never happens.

Lemma 2.2.11 *Every quasiperfect class $\mathbf{B} = (d, m, p, q, t, \varepsilon)$ with $\varepsilon = 1$ (resp. $\varepsilon = -1$) has the property that $m/d > b$ (resp. $m/d < b$), for all b in the closure of the blocked b -interval $J_{\mathbf{B}}$.*

Proof We first argue by contradiction to show that $m/d \notin \text{cl}(J_{\mathbf{B}})$, and then finish the argument by considering the cases $b < 1/3$ and $b > 1/3$ separately.

Thus suppose that $m/d \in \text{cl}(J_{\mathbf{B}})$, and let $z_0 := \text{acc}(m/d)$. We first claim that $z_0 > p/q$. To see this note that by (2.1.1) z_0 is the unique solution > 1 of the equation

$$z_0 + \frac{1}{z_0} = \frac{(3 - m/d)^2}{1 - (m/d)^2} - 2,$$

and the function $z \mapsto z + 1/z$ increases when $z > 1$. Therefore, $z_0 > p/q$ exactly when $z_0 + 1/z_0 > p/q + q/p$. But

$$z_0 + \frac{1}{z_0} = \left(\frac{(3 - m/d)^2}{1 - (m/d)^2} - 2 \right) = \frac{p^2 + q^2 + 2}{pq - 1} > \frac{p}{q} + \frac{q}{p},$$

where the second equality uses the identities $d^2 - m^2 = pq - 1$ and $3d - m = p + q$. Thus $z_0 > p/q$.

If $m/d \in \text{cl}(J_{\mathbf{B}})$ then $z_0 = \text{acc}(m/d) \in \text{cl}(I_{\mathbf{B}}) = \text{acc}(\text{cl}(J_{\mathbf{B}}))$, so, by [1, Lemma 38(ii)] and [1, Lemma 16], $\mu_{\mathbf{B},m/d}(z_0)$ is given by the formula in (1.1.3). Thus we must have

$$\mu_{\mathbf{B},m/d}(z_0) = \frac{p}{d - m \cdot m/d} \geq V_{m/d}(z_0) = \frac{1 + z_0}{3 - m/d},$$

where the first equality holds by (1.1.3) and the fact that $z_0 > p/q$, the inequality holds because we assume $\mu_{\mathbf{B},m/d}$ is at least as large as the volume at z_0 , and the last equality holds by Lemma 2.2.7 and the fact that $z_0 = \text{acc}(m/d)$. Therefore $z_0 \leq z_1$, where z_1 satisfies the equation $p/(d - m \cdot m/d) = (1 + z_1)/(3 - m/d)$.

Next note that, by solving for z_1 and using the identities $d^2 - m^2 = pq - 1$ and $3d - m = p + q$, we obtain $z_1 = (p^2 + 1)/(pq - 1)$. It is now straightforward to verify that

$$z_0 + \frac{1}{z_0} = \frac{p^2 + q^2 + 2}{pq - 1} = z_1 + \frac{q^2 + 1}{pq - 1} > z_1 + \frac{pq - 1}{p^2 + 1} = z_1 + \frac{1}{z_1}.$$

But this is impossible since we saw above that $z_0 \leq z_1$. This completes the first step.

To finish the argument, notice that we proved above that $z_0 = \text{acc}(m/d) > p/q$, while Proposition 2.2.9 shows that p/q is always blocked by \mathbf{B} . Thus if $I_{\mathbf{B}} := \text{acc}(J_{\mathbf{B}})$ denotes the set of blocked z -values, we have $\text{acc}(m/d) > p/q$ where $p/q \in I_{\mathbf{B}}$. Therefore $\text{acc}(m/d)$ must be greater than all points in $\text{cl}(I_{\mathbf{B}})$. Since $b \mapsto \text{acc}(b)$ preserves orientation exactly if $b > 1/3$, this implies that $m/d > b$ for all $b \in \text{cl}(J_{\mathbf{B}})$ when $m/d > 1/3$ and $m/d < b$ for all $b \in \text{cl}(J_{\mathbf{B}})$ when $m/d < 1/3$. \square

We end this subsection with a few remarks about the case $b = 1/3$, which is the focal point of the shift S and separates the two regimes, $b > 1/3$ and $b < 1/3$. As far as we know, this is the unique rational value of b with a staircase.¹¹ Example 2.3.7 below describes the ascending staircase at $b = 1/3$. Although we have not managed to resolve the question of whether there is also a descending staircase at $b = 1/3$, we can make the following observation. Note that the proof uses the same idea as in Lemma 4.3.3.

Lemma 2.2.12 *If there is no descending staircase when $b = 1/3$, then there is an $\varepsilon > 0$ such that $c_{1/3}(z) = 3(1+z)/8$ for $3 + 2\sqrt{2} = a_{\min} < z < a_{\min} + \varepsilon$.*

Proof It follows from [1, Example 32] (also see Remark 2.3.8(ii)) that the obstruction $\mu_{E_1, b}$ given by the class $E_1 := 3L - E_0 - 2E_1 - E_{2\dots 6}$ is precisely $z \mapsto (1+z)/(3-b)$ when $5 < z < 6$. Thus for $z \in [a_{\min}, 6]$ we know that $c_{H_{1/3}}(z) \geq 3(1+z)/8$. If we do not have equality for $z \in (a_{\min}, a_{\min} + \varepsilon)$ and there is no staircase, then there must be a different obstruction curve $z \mapsto (A + Cz)/(d - m/3)$ that goes through the accumulation point $(a_{\min}, 3\sqrt{a_{\min}/8})$. Because a_{\min} is irrational, the equation $(A + Ca_{\min})/(d - m/3) = 3(1 + a_{\min})/8$ can hold only if $A = C$. But then the graphs of the two obstructions are lines of the form $z \mapsto \lambda(1+z)$, and hence coincide. \square

This following observation is also relevant because, for example, $d_{\mathbf{B}} - 3m_{\mathbf{B}}$ appears as the coefficient of d in the linear staircase relation (1.2.5), so that it would be awkward if it were ever zero.

Lemma 2.2.13 *There is no quasiperfect class $E = (d, m, p, q)$ with $m/d = 1/3$.*

Proof By (1.2.1), given any such E the positive integers $m, p,$ and q would have to satisfy $8m = p + q$ and $8m^2 = pq - 1$. Since $m \in \mathbb{Z}$ we must have $p + q \equiv_8 0$, so the second equation gives $p^2 \equiv_8 -1$. This is impossible because $(4k + 1)^2$ and $(4k + 3)^2$ are both congruent to 1 mod 8. \square

¹¹In particular, there should be no staircases at the points $b \in \text{acc}^{-1}(v_i)$. This was proved in [1, Theorem 6] for the case $b = 1/5 \in \text{acc}^{-1}(6)$, but the proof seems too elaborate to be easily generalized.

2.3 The known staircases

The paper [1] found three different families of blocking classes \mathbf{B}^U , \mathbf{B}^E , and \mathbf{B}^L and their associated staircases. For our purposes, \mathcal{S}^U and \mathcal{S}^L are the essential families, since applying powers of S to these staircases generate all the staircase families discussed here.

We now review the theorems in [1, Theorems 56 and 58] that define these two staircase families. In all cases the staircase steps $\mathbf{E}_{n,k}$ are quasiperfect classes given by tuples $(d_{n,k}, m_{n,k}, p_{n,k}, q_{n,k})$ that for $x = d, m, p, q$ satisfy the recursion described below. We call $p_{n,k}/q_{n,k}$ the center of the step, and use the staircase relation¹² (together with the linear Diophantine condition $3d = m + p + q$) to determine the entries d and m from knowledge of p and q .

Theorem 2.3.1 *The classes $\mathbf{B}_n^U = (n + 3, n + 2, 2n + 6, 1)$, for $n \geq 0$, with increasing centers, are perfect blocking classes, with the following associated staircases $\mathcal{S}_{\ell,n}^U$ and $\mathcal{S}_{u,n}^U$, where $\sigma_n = (2n + 5)(2n + 1)$ and $\text{end}_n = (2n + 4)$ or $(2n + 5, 2n + 2)$:*

- For each $n \geq 1$, $\mathcal{S}_{\ell,n}^U$ has limit point $a_{\ell,n,\infty}^U = [2n + 5; 2n + 1, \{2n + 5, 2n + 1\}^\infty]$, and its
 - **centers** $\{[2n + 5, 2n + 1]^k, \text{end}_n\}$, for $k \geq 0$, where $\text{end}_n = 2n + 4$ or $(2n + 5, 2n + 2)$,
 - **recursion** $x_{n,k+1} = (\sigma_n + 2)x_{n,k} - x_{n,k-1}$,
 - **relation** $(2n + 3)d_{n,k} = (n + 1)p_{n,k} + (n + 2)q_{n,k}$.
- For each $n \geq 0$, $\mathcal{S}_{u,n}^U$ has limit point $a_{u,n,\infty}^U = [2n + 7; \{2n + 5, 2n + 1\}^\infty]$, and has
 - **centers** $[2n + 7; \{2n + 5, 2n + 1\}^k, \text{end}_n]$,
 - **recursion** $x_{n,k+1} = (\sigma_n + 2)x_{n,k} - x_{n,k-1}$,
 - **relation** $(2n + 3)d_{n,k} = (n + 2)p_{n,k} - (n + 4)q_{n,k}$.

The limit points $a_{\bullet,n,\infty}^U$ form increasing unbounded sequences in $(6, \infty) = (v_2, v_1)$, while the corresponding b -values lie in $(5/11, 1)$, where $5/11 = \text{acc}_U^{-1}(6)$, and increase with limit 1.

Remark 2.3.2 (i) If ν is an integer ≥ 3 , the recursion $x_{k+1} = \nu x_k - x_{k-1}$ has a unique solution > 1 of the form $x_k = \alpha^k$ where $\alpha^2 - \nu\alpha + 1 = 0$. If the recursion has seeds x_0 and x_1 , the general rational solution can be written $X\alpha^k + \bar{X}\bar{\alpha}^k$ where $\bar{\alpha} := 1/\alpha$ is the other solution, $X \in \mathbb{Q}[\sqrt{\nu^2 - 4}]$, and for $X = a + b\sqrt{\nu^2 - 4}$ we define $\bar{X} := a - b\sqrt{\nu^2 - 4}$. Hence, if x_\bullet and y_\bullet both satisfy this recursion, the ratio x_\bullet/y_\bullet .

¹²In [1] we did not yet realize the role of the variable t .

converges to the quantity X/Y . It is thus straightforward to calculate quantities such as $a_{u,n,\infty}^U$ from knowledge of the recursion plus its seeds; see Lemma 3.1.4.

(ii) The staircases $S_{\bullet,n}^U$ are described above as having two intertwined strands, one for each end_n . We show in Lemma 3.2.1 that these classes may be combined into a single family with recursion variable $2n + 3$. This simpler description of the staircases clarifies their essential structure.

(iii) Notice that there is no ascending staircase in this family with $n = 0$ since the centers of its steps would be < 6 , the center of B_0^U . Of course, there is a staircase of this kind, but we view it here as the image of $S_{u,0}^U$ by the reflection $R_{v_2} = SR$, and so consider it a member of the staircase family $(SR)^\#(S^U)$.

(iv) Finally, note that staircases that are associated to a blocking class B and labeled with the subscript ℓ are always ascending, and converge to the lower end of the z -interval blocked by B , while those labeled u descend and converge to the upper endpoint of this z -interval.

There is a corresponding definition of the staircase family S^L .

Theorem 2.3.3 *The classes $B_n^L = (5n, n - 1, 12n + 1, 2n)$, for $n \geq 1$, with decreasing centers, are perfect and center-blocking, and have the following associated staircases $S_{\ell,n}^L$ and $S_{u,n}^L$ for $n \geq 1$, with σ_n and end_n as in Theorem 2.3.1:*

- $S_{\ell,n}^L$ is ascending, with limit point $a_{\ell,n,\infty}^L = [6; 2n + 1, \{2n + 5, 2n + 1\}^\infty]$, and has
 - **centers** $[6; 2n + 1, \{2n + 5, 2n + 1\}^k, \text{end}_n]$,
 - **recursion** $x_{n,k+1} = (\sigma_n + 2)x_{n,k} - x_{n,k-1}$, where $\sigma_n := (2n + 1)(2n + 5)$,
 - **relation** $(2n + 3)d_{n,k} = (n + 1)p_{n,k} - (n - 1)q_{n,k}$.
- $S_{u,n}^L$ is descending, with limit point $a_{u,n,\infty}^L = [6; 2n - 1, 2n + 1, \{2n + 5, 2n + 1\}^\infty]$ and has
 - **centers** $[6; 2n - 1, 2n + 1, \{2n + 5, 2n + 1\}^k, \text{end}_n]$,
 - **recursion** $x_{n,k+1} = (\sigma_n + 2)x_{n,k} - x_{n,k-1}$,
 - **relation** $(2n + 3)d_{n,k} = -(n - 1)p_{n,k} + (11n + 2)q_{n,k}$.

The limit points $a_{\bullet,n,\infty}^L$ (with $\bullet = \ell$ or u) form a decreasing sequence in $(6, 7) = (v_2, w_1)$ with limit 6, while the corresponding b -values lie in $(0, \frac{1}{5})$ and increase with limit $\frac{1}{5}$.

Remark 2.3.4 (i) It follows from Lemma 2.1.6(i)–(ii) that the symmetry R takes the centers both of the blocking classes and of the staircase steps for the family S^U into

those for \mathcal{S}^L . Notice that in the case of $\mathcal{S}_{\ell,n}^U$ it takes the step with label k to the step in $\mathcal{S}_{u,n}^U$ with label $k - 1$. Note that, with this choice of labeling, the image by R of the step with center $[2n + 7; 2n + 4]$ (which appears both as $\mathcal{S}_{\ell,n+1,0}^U$ and as $\mathcal{S}_{u,n,1}^U$) has no counterpart in the staircase $\mathcal{S}_{u,n+1}^L$, though it could be added to it.

(ii) **The Fibonacci stairs** The Fibonacci stairs are the ascending stairs that should be associated to $\mathbf{B}_0^L = R^\#(\mathbf{B}_0^U)$. However, such a class would have center at $R(6) = \infty$, and so it does not exist as a geometric obstruction. Nevertheless, if we ignore the first few steps, the steps of the Fibonacci stairs have precisely the form predicted by putting $n = 0$ in the formulas for $\mathcal{S}_{\ell,n}^L$; namely, they have

- **centers** $[6; 1, \{5, 1\}^k, \text{end}_0]$,
- **recursion** $x_{k+1} = 5x_k - x_{k-1}$,
- **relation** $3d_k = p_k + q_k$.

Moreover, although the class $E' := 3L(-0E_0) - 2E_1 - E_2 - \dots - E_7$ is not perfect, its obstruction $\mu_{E',0}(z)$ for $z \in (6, 7]$ is the function $z \mapsto (1+z)/3$, which goes through the point $(a_{0,\infty}, V_0(a_{0,\infty}))$ (where $a_{0,\infty} = \tau^4$) and equals $c_{H_0}(z)$ for $z \in [\tau^4, 7]$. Therefore this class E' plays the geometric role of the blocking class, and we consider these stairs as part of the family $\mathcal{S}^L = R^\#(\mathcal{S}^U)$. As we explain before Lemma 3.2.5, there is a different tuple with negative entries that plays the numeric role of the missing blocking class; below we denote this by \mathbf{B}_0^L .

Notice that all the other families $(S^i)^\#(\mathcal{S}^L)$, for $i \geq 1$, have an ascending staircase for $n = 0$ that is associated with the blocking class $(S^i)^\#(\mathbf{B}_0^L)$, which now has positive entries and so is geometric.

Lemma 2.2.11 shows that for every $n \geq 0$ the ratio $(n + 2)/(n + 3)$ does not lie in the b -interval $J_{\mathbf{B}_n^U}$ blocked by \mathbf{B}_n^U . The next lemma locates this point more precisely.

Lemma 2.3.5 *Let $(d_n, m_n) = (n + 3, n + 2)$ and $a_n = 2n + 6$ be the degree variables and center of the blocking class $\mathbf{B}_n^U = (n + 3, n + 2, 2n + 6, 1)$. Then for all $n \geq 0$, we have $\text{acc}(m_n/d_n) < a_{n+1}$ and $m_n/d_n \in J_{\mathbf{B}_{n+1}^U}$.*

Proof To see that $z_n := \text{acc}(m_n/d_n) < a_{n+1} = 2n + 8$, it suffices to check that $z_n + 1/z_n < 2n + 8 + 1/(2n + 8)$. But (2.1.1) implies that

$$\begin{aligned} z_n + \frac{1}{z_n} &= \frac{(3 - (n + 2)/(n + 3))^2}{1 - ((n + 2)/(n + 3))^2} - 2 = \frac{4n^2 + 24n + 39}{2n + 5} \\ &= 2n + 8 - \frac{2n + 1}{2n + 5} < 2n + 8 + \frac{1}{2n + 8}. \end{aligned}$$

A similar argument shows that $z_n > 2n + 7$. Therefore, [1, Lemma 16] implies that $\mu_{\mathbf{B}_{n+1}^U, m_n/d_n}(z_n)$ is given by the formula (1.1.3), so

$$\mu_{\mathbf{B}_{n+1}^U, m_n/d_n}(z_n) = \frac{z_n}{d_{n+1} - (m_n/d_n)m_{n+1}} = \frac{z_n}{2}.$$

On the other hand, it follows from [1, Lemma 38] that $m_n/d_n \in J_{\mathbf{B}_{n+1}^U}$ exactly when $\mu_{\mathbf{B}_{n+1}^U, m_n/d_n}(z_n) > V_{m_n/d_n}(z_n)$. Since $z_n = \text{acc}(m_n/d_n)$, we have

$$V_{m_n/d_n}(z_n) = \frac{1 + z_n}{3 - m_n/d_n}.$$

Thus we need $z_n/2 > (1 + z_n)(n + 3)/(2n + 7)$, which holds because $z_n > 2n + 6$. \square

Remark 2.3.6 We show in Corollary 4.2.6 that for all our staircases, whether ascending or descending, the ratios $(m_k/d_k)_{k \geq 1}$ decrease when $m_k/d_k > 1/3$ and increase when $m_k/d_k < 1/3$. (Since the staircase steps are defined recursively, Corollary 3.1.5 shows that this follows from the structure of the first two steps.) Further, Proposition 3.2.2 shows that the blocking class \mathbf{B}_n^U can be considered as a step in the ascending stair $\mathcal{S}_{\ell, n+1}^U$ associated to \mathbf{B}_{n+1}^U . Hence m_n/d_n is part of a decreasing sequence that limits on the lower endpoint of $J_{\mathbf{B}_{n+1}^U}$. Thus, once we have shown that $m_n/d_n < a_{n+1}$, this reasoning implies that $m_n/d_n \in J_{\mathbf{B}_{n+1}^U}$. More generally, an analog of Lemma 2.3.5 holds for any prestaircase family.

Example 2.3.7 (the staircase at $b = 1/3$) This staircase, discovered in [3], behaves in a different way from all other known staircases in H_b . It has three interwoven sequences,

$$E_{k,i} = (d_{k,i}, m_{k,i}, p_{i,k} = g_{k,i}, q_{i,k} = g_{k-1,i}), \quad i = 0, 1, 2,$$

where for each i the numbers $g_{k,i}$ satisfy the recursion

$$g_{k+1,i} = 6g_{k,i} - g_{k-1,i}$$

with seeds (ie initial values)

$$g_{0,0} = 1, \quad g_{1,0} = 2, \quad g_{0,1} = 1, \quad g_{1,1} = 4, \quad g_{0,2} = 1, \quad g_{1,2} = 5.$$

The centers $p_{i,k}/q_{i,k} := g_{k,i}/g_{k-1,i}$ have continued fractions $[5; \{1, 4\}^k, \text{end}_i]$, where the possible ends are $\text{end}_0 = 2$, $\text{end}_1 = (1, 3)$ and $\text{end}_2 = \emptyset$.

Normally one needs two seeds, say x_0 and x_1 , to fix an iteration of the form $x_{k+1} = \nu x_k - x_{k-1}$. However the variables p and q for the $1/3$ -staircase have the form x_k

and x_{k-1} , and so are part of a single recursive sequence. Thus we can consider that each of the three strands of the $1/3$ -staircase has a single seed $E_{\text{seed}} = (d, m, p, q)$, namely

$$E_{\text{seed},0} = (1, 0, 2, 1), \quad E_{\text{seed},1} = (2, 1, 4, 1), \quad E_{\text{seed},2} = (2, 0, 5, 1),$$

and that the rest of the staircase comes from the images of the (p, q) components of these three classes under the action of $S : (p, q) \mapsto (6p - q, p)$, with (d, m) determined by a modification of (2.2.4) as follows. We have

$$(2.3.1) \quad d_{k,i} = \frac{1}{8}(3(g_{k,i} + g_{k-1,i}) + \varepsilon_{k,i}t_i), \quad m_{k,i} = \frac{1}{8}(g_{k,i} + g_{k-1,i} + 3\varepsilon_{k,i}t_i),$$

where $\varepsilon_{k,i} = (-1)^{k+i}$, while t_i is constant with respect to k and equal to

$$t_i = \sqrt{g_{1,i}^2 + g_{0,i}^2 - 6g_{1,i}g_{0,i} + 8} = \begin{cases} 1 & \text{if } i = 0, 1, \\ 2 & \text{if } i = 2. \end{cases}$$

Note that $\varepsilon_{k,i} = 1$ if $m_{k,i}/d_{k,i} > 1/3$ and $= -1$ otherwise.

Note that, in contrast with the case of S^U and S^L discussed in Remark 2.3.2(ii) above, it is *not* possible to combine these three interwoven sequences into a single recursive sequence; for example, there is no constant c such that for all k

$$g_{k,2} = cg_{k,1} - g_{k,0}.$$

Further, the relation between the $d, p,$ and q variables in this staircase is significantly different from the homogenous linear relation satisfied by the other staircases. By [1, Proposition 49] the latter relation implies that the ratios $m_{k,i}/d_{k,i}$ are monotonic and converge as $k \rightarrow \infty$ so quickly that the step classes are themselves center-blocking classes. On the other hand, classes with centers $< a_{\min}$ cannot be center-blocking, and the ratios $m_{k,i}/d_{k,i}$ lie on both sides of $1/3$. Another important distinction between this staircase and the others is the fact that the t -variable is fixed for each strand. The point here is that in all cases the variable called t is fixed by the shift S . This shift implements the recursion of the $1/3$ -staircase, while it takes every other staircase to a different one.

Finally, we remark that the three seed classes $E_{\text{seed},i}$, for $i = 0, 1, 2$, have rather different natures. The first two have the form of the blocking classes

$$B_n^U := (n + 3, n + 2, 2n + 6, 1), \quad n \geq 0,$$

and indeed are given by taking $n = -2, -1$ in this formula. One might consider that the B_n^U , for $n \geq -2$, form a single family of classes, that behave differently according

as to where the center $2n + 6$ lies in relation to a_{\min} . If the center is $< a_{\min}$ then iteration by S gives a staircase for $b = 1/3 = \text{acc}^{-1}(a_{\min})$, while if the center is $> a_{\min}$ then we get a family of blocking classes that block b -intervals that converge to $1/3$, but do not form a staircase since they are not live when $b = 1/3$. (To see this, note that $t = |d - 3m|$ is fixed by S , while by [1, Lemma 15] a perfect class E with center p/q is obstructive at its center for a given value of b if and only if $|bd - m| < \sqrt{1 - b^2}$. Hence a class that is live at $b = 1/3$ must have $|t| < 3$, while B_n^U has $t = 2n + 3$.)

The third seed $E_{\text{seed},2} = (2, 0, 5, 1)$ with $(t, \varepsilon) = (2, -1)$ is significantly different. Both it and $E_{\text{seed},0} = (1, 0, 2, 1)$ are steps in the Fibonacci staircase at $b = 0$, but the iteration that gives this sequence is not S but rather $x_{\kappa+1} = 3x_{\kappa} - x_{\kappa-1}$. All the other steps in the Fibonacci staircase have centers $> a_{\min}$, and are blocking classes by [1, Lemma 41]. On the other hand, because $S(1, 1) = (5, 1)$, the strand formed by $E_{\text{seed},2}$ and its iterates under S can be extended one place backwards by the tuple

$$(2.3.2) \quad E_{-1,2} := (1, 1, 1, 1) \quad \text{with } (t, \varepsilon) = (2, 1).$$

It turns out that the sequence of classes formed by $E_{-1,2}$, $E_{\text{seed},2}$, and the images of $E_{\text{seed},2}$ by S^i have an important role to play in generating the staircase families; see for example Corollaries 3.2.3 and 3.2.7 and Lemma 3.3.3.

Remark 2.3.8 (the role of low degree classes) (i) The three seeds of the $1/3$ -staircase are given by the only exceptional curves in H_b with degree at most two. These classes have centers $< a_{\min} = 3 + 2\sqrt{2}$, while all the other classes of interest here have centers $> a_{\min}$.

(ii) There is one exceptional divisor of degree three, that gives rise to three potentially interesting exceptional divisors in H_b that we will label by the coefficient of E_0 , namely¹³

$$E_0 := 3L - 2E_1 - E_{2\dots 7}, \quad E_1 := 3L - E_0 - 2E_1 - E_{2\dots 6}, \quad E_2 := 3L - 2E_0 - E_{1\dots 6}.$$

Each of these classes has a special role to play:

- E_0 substitutes geometrically (but not numerically) for the blocking class for the Fibonacci stairs (see Remark 2.3.4(ii)).
- E_1 is a (nonperfect) class with breakpoint 6. This class is discussed extensively in [1, Example 32]. It is obstructive at $z = 6$ for

$$\text{acc}_L^{-1}(6) = \frac{1}{5} < b < \frac{5}{11} = \text{acc}_U^{-1}(6),$$

¹³Here we use the shorthand of [6], where $E_{i\dots j} := \sum_{k=i}^j E_k$.

and for this range of b we have

$$\mu_{E,b}(z) = \frac{1+z}{3-b} \quad \text{for } z \in (5, 6), \quad \mu_{E,b}(z) = \frac{7}{3-b} \quad \text{for } z \geq 6.$$

The obstruction given by this class is discussed further in [Lemma 2.2.7](#). Its properties turn out to be a crucial element in our proofs.

- E_2 is the perfect blocking class B_0^U with center 6.

3 Structure of the prestaircase families

We begin by explaining the recursion that underlies our staircases. In [Section 3.2](#), we then describe the staircase families S^U and S^L in the new language, and show that they are indeed generated by their blocking classes together with two seeds, as claimed in [Proposition 1.2.2](#).

In [Section 3.3](#), we prove [Theorem 1.2.4](#). This establishes that for each $T = S^i R^\delta$, $T^\#(S^U)$ is a prestaircase family. This entails examining how the seeds and blocking classes are transformed under S . Furthermore, this implies that each prestaircase in $T^\#(S^U)$ is associated to a blocking class, an essential feature of a prestaircase family. In [Section 3.4](#), we compute how $S^\#$ and $R_{v_i}^\#$ act on the degree coordinates (d, m) of the blocking classes.

3.1 The single recursion

The following lemma shows that we can consider each staircase in S^U and S^L to be a single family of classes satisfying the recursion $x_{\kappa+1} = (2n + 3)x_\kappa - x_{\kappa-1}$ with parameter $2n + 3$ instead of a double family of classes as in [Theorems 2.3.1](#) and [2.3.3](#) that each satisfy the recursion $x_{k+1} = ((2n + 1)(2n + 5) + 2)x_k - x_{k-1}$ with parameter $(2n + 1)(2n + 5)$. Note that an increase of k by 1 corresponds to an increase of κ by 2.

Lemma 3.1.1 (i) *Let $m \geq 1$ be any integer, and consider*

$$\begin{aligned} \frac{p_1}{q_1} &= [m; 2n + 4], & \frac{p_2}{q_2} &= [m; 2n + 5, 2n + 2], \\ \frac{p_3}{q_3} &= [m; 2n + 5, 2n + 1, 2n + 4], & \frac{p_4}{q_4} &= [m; 2n + 5, 2n + 1, 2n + 5, 2n + 2]. \end{aligned}$$

Then for $x_\bullet = p_\bullet, q_\bullet$ we have $x_{\kappa+1} = (2n + 3)x_\kappa - x_{\kappa-1}$ for $\kappa = 1, 2$.

(ii) If x_\bullet is a sequence such that $x_{\kappa+1} = \nu x_\kappa - x_{\kappa-1}$ for all κ , then

$$x_{\kappa+2} = (\nu^2 - 2)x_\kappa - x_{\kappa-2} \quad \text{for all } \kappa.$$

In particular, if $\nu = 2n + 3$ then $\nu^2 - 4 = (2n + 1)(2n + 5)$.

(iii) If y_κ and z_κ satisfy the recursion $x_{\kappa+1} = \nu x_\kappa - x_{\kappa-1}$, then

$$y_{\kappa+2}z_\kappa - z_{\kappa+2}y_\kappa = \nu(y_{\kappa+1}z_\kappa - z_{\kappa+1}y_\kappa).$$

Proof Item (i) by direct calculation. Item (ii) holds because $\nu x_{\kappa-1} = x_\kappa + x_{\kappa-2}$, so

$$x_{\kappa+2} = \nu x_{\kappa+1} - x_\kappa = \nu(\nu x_\kappa - x_{\kappa-1}) - x_\kappa = (\nu^2 - 1)x_\kappa - \nu x_{\kappa-1} = (\nu^2 - 2)x_\kappa - x_{\kappa-2}.$$

Finally, (iii) holds because

$$y_{\kappa+2}z_\kappa - z_{\kappa+2}y_\kappa = \det \begin{pmatrix} z_\kappa & y_\kappa \\ z_{\kappa+2} & y_{\kappa+2} \end{pmatrix} = \det \begin{pmatrix} z_\kappa & y_\kappa \\ \nu z_{\kappa+1} & \nu y_{\kappa+1} \end{pmatrix}. \quad \square$$

We next show that this recursion extends naturally to the triples $(p_\kappa, q_\kappa, t_\kappa)$ that parametrize quasiperfect classes as in Definition 2.2.3, provided that the initial seeds satisfy the compatibility condition (3.1.3) that is given below. Further, (3.1.2) below states that the tuple (p, q, t) is the coordinate of a point on the surface X defined by (2.2.2). Thus the next lemma gives compatibility conditions on the seeds (p_0, q_0, t_0) and (p_1, q_1, t_1) under which the recursion with parameter ν extends to the integral points of X .

It is convenient to use matrix notation with

$$(3.1.1) \quad A := \begin{pmatrix} -1 & 3 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{x} := \begin{pmatrix} p \\ q \\ t \end{pmatrix}.$$

Thus the matrix A is symmetric, and $X = \{\mathbf{x} \mid \mathbf{x}^T A \mathbf{x} = 8\}$.

Lemma 3.1.2 Suppose that \mathbf{x}_0 and \mathbf{x}_1 are integral vectors that satisfy the following conditions for some integer $\nu > 0$:

$$(3.1.2) \quad \mathbf{x}_i^T A \mathbf{x}_i = 8, \quad i = 0, 1,$$

$$(3.1.3) \quad \mathbf{x}_1^T A \mathbf{x}_0 = 4\nu.$$

Then the vectors $\mathbf{x}_2 := \nu \mathbf{x}_1 - \mathbf{x}_0$, \mathbf{x}_1 also satisfy these conditions for the given ν .

Proof Since A is symmetric,

$$(\nu \mathbf{x}_1 - \mathbf{x}_0)^T A (\nu \mathbf{x}_1 - \mathbf{x}_0) = \nu^2 \mathbf{x}_1^T A \mathbf{x}_1 - 2\nu \mathbf{x}_1^T A \mathbf{x}_0 + \mathbf{x}_0^T A \mathbf{x}_0 = 8$$

exactly when $8v^2 = 2v\mathbf{x}_1^T A \mathbf{x}_0$, which holds by (3.1.3). Further, (3.1.3) holds for $v\mathbf{x}_1 - \mathbf{x}_0, \mathbf{x}_1$ because, by (3.1.3),

$$(v\mathbf{x}_1 - \mathbf{x}_0)^T A \mathbf{x}_1 = 8v - (\mathbf{x}_0)^T A \mathbf{x}_1 = 4v. \quad \square$$

Corollary 3.1.3 Any two integral triples $\mathbf{x}_i = (p_i, q_i, t_i)$ for $i = 0, 1$ that satisfy (3.1.2) and (3.1.3) for a given v can be extended to a sequence \mathbf{x}_i for $i \geq 0$ using recursion parameter v , where each successive pair satisfies these conditions. Further, for fixed $\varepsilon \in \{\pm 1\}$, the corresponding quantities

$$d_i = \frac{1}{8}(3(p_i + q_i) + \varepsilon t_i), \quad m_i = \frac{1}{8}(p_i + q_i + 3\varepsilon t_i)$$

of (2.2.4) also satisfy this recursion and hence are integers, provided that they are integers for $i = 0, 1$.

We end this section with some useful formulas about recursively defined sequences. The following is an adaptation of [1, Lemma 47].

Lemma 3.1.4 Let x_κ for $\kappa \geq 0$ be a sequence of integers that satisfy the recursion

$$(3.1.4) \quad x_{\kappa+1} = vx_\kappa - x_{\kappa-1}, \quad v \geq 3,$$

and let

$$\lambda = \frac{1}{2}(v + \sqrt{\sigma}) \in \mathbb{Q}[\sqrt{\sigma}]$$

be the larger root of the equation $x^2 - vx + 1 = 0$, where $\sigma = v^2 - 4 > 1$. Then there is a number $X \in \mathbb{Q}[\sqrt{\sigma}]$ such that

$$(3.1.5) \quad x_\kappa = X\lambda^\kappa + \bar{X}\bar{\lambda}^\kappa,$$

where $\overline{a + b\sqrt{\sigma}} := (a - b\sqrt{\sigma})$, so that $\lambda\bar{\lambda} = 1$. Further, if we write $X = X' + X''\sqrt{\sigma}$, then

$$(3.1.6) \quad X' = \frac{x_0}{2}, \quad X'' = \frac{2x_1 - vx_0}{2\sigma}.$$

Proof If the monomials $x_\kappa = c^\kappa$ satisfy the recursion then we must have $c^2 - vc + 1 = 0$, so that $c = \frac{1}{2}(v \pm \sqrt{v^2 - 4}) = \frac{1}{2}(v \pm \sqrt{\sigma})$. Let λ be the larger solution so that $\bar{\lambda}$ is the smaller one, and we have $\lambda\bar{\lambda} = 1$. Since (3.1.4) has a unique solution once given the seeds x_0 and x_1 , it follows that for each choice of constants A and B , the numbers

$$(3.1.7) \quad x_\kappa := A\lambda^\kappa + B\bar{\lambda}^\kappa$$

form the unique solution with

$$x_0 = A + B, \quad x_1 = A\lambda + B\bar{\lambda}.$$

Then $A, B \in \mathbb{Q}[\sqrt{\sigma}]$. Notice also that $\sqrt{\sigma}$ is irrational since $\sigma = \nu^2 - 4$ is never a perfect square when $\nu \geq 3$. It follows easily that $x_0, x_1 \in \mathbb{Q}$ only if we also have $B := \bar{A}$.

Thus we have

$$x_0 = X + \bar{X} = X' + X''\sqrt{\sigma} + X' - X''\sqrt{\sigma} = 2X',$$

which gives $X' = x_0/2$, and

$$x_1 = (X' + X''\sqrt{\sigma}) \cdot \frac{1}{2}(\nu + \sqrt{\sigma}) + (X' - X''\sqrt{\sigma}) \cdot \frac{1}{2}(\nu - \sqrt{\sigma}) = X'\nu + X''\sigma,$$

which gives $X'' = (2x_1 - \nu x_0)/(2\sigma)$, as claimed □

Corollary 3.1.5 *Let x_κ and y_κ , for $\kappa \geq 0$, be increasing sequences that both satisfy (3.1.4) for some $\nu \geq 3$ and are such that x_κ and y_κ are positive for $\kappa \geq 1$. Suppose further that at most one of x_0 and y_0 are zero. Then:*

- (i) *The ratios $(x_\kappa/y_\kappa)_{\kappa \geq 1}$ form a monotonic sequence that is strictly increasing if $x_1y_0 - x_0y_1 > 0$ and is strictly decreasing if $x_1y_0 - x_0y_1 < 0$.*
- (ii) *In all cases, $\lim_{\kappa \rightarrow \infty} x_\kappa/y_\kappa$ exists and equals X/Y , where X and Y are the constants defined in (3.1.6).*
- (iii) *Provided that $x_1y_0 - x_0y_1 \neq 0$, the limit X/Y is irrational.*

Proof Note first that

$$x_{\kappa+1}y_\kappa - x_\kappa y_{\kappa+1} = (6x_\kappa - x_{\kappa-1})y_\kappa - x_\kappa(6y_\kappa - y_{\kappa-1}) = x_\kappa y_{\kappa-1} - x_{\kappa-1} y_\kappa$$

is independent of κ . Hence the sequence x_κ/y_κ for $\kappa \geq 1$ is monotonic,¹⁴ and the quantity $x_1y_0 - x_0y_1$ determines whether it is increasing or decreasing. Part (ii) is an immediate consequence of the formula in (3.1.7), and the fact that $\lambda > 1$.

To prove the third claim, suppose first that $2x_1 > \nu x_0$. Then $X = X' + X''\sqrt{\sigma}$ is irrational since the coefficient X'' of $\sqrt{\sigma}$ does not vanish, and $\sigma = \nu^2 - 4$ is not a perfect square. Similarly, if $2y_1 > \nu y_0$ then Y is irrational. Further,

$$\frac{X}{Y} = \frac{(X' + X''\sqrt{\sigma})(Y' - Y''\sqrt{\sigma})}{(Y' + Y''\sqrt{\sigma})(Y' - Y''\sqrt{\sigma})} = \frac{(X'Y' - X''Y''\sigma) + (X''Y' - X'Y'')\sqrt{\sigma}}{(Y')^2 - (\sigma Y'')^2}$$

is irrational unless $X''Y' - X'Y'' = 0$. But formula (3.1.6) implies that $X''Y' - X'Y''$ is a multiple of $x_1y_0 - x_0y_1$ and so is nonzero by hypothesis.

¹⁴If $x_0/y_0 \leq 0$, then the sequence is only monotonic for $\kappa \geq 1$. This will be the case for the seeds of S_u^U computed in Lemma 3.2.1(ii)

To deal with the possibility that X'' or Y'' equals 0, notice that the limit ratio X/Y does not depend on which pair of terms $k_0, k_0 + 1$ we take as the initial terms (though the values of X and Y do depend on this choice). Further, because by hypothesis $\nu \geq 3$ and $x_0 < x_1$, we have $x_2 = \nu x_1 - x_0 > \nu x_1/2$, so $2x_2 > \nu x_1$. Similarly, $2y_2 > \nu y_1$. Therefore if we define the quantities X and Y as in (3.1.6) but starting with $\kappa = 1$ then the above argument shows that the ratio X/Y is irrational. \square

3.2 Generating the known staircase families

We now prove Proposition 1.2.2; its proof is contained in Corollaries 3.2.3 and 3.2.7. In all cases considered in this paper, the recursion variable $\nu = 2n + 3$ for the prestaircase $S_{\bullet, n}^{\mathcal{F}}$ in a staircase family \mathcal{F} is a linear function of n , and we will enumerate the terms x_κ in our recursively defined sequences such that the value of x_κ is a polynomial of degree κ in n . In particular, x_0 is a constant. In contrast, the staircase sequences were enumerated in Theorems 2.3.1 and 2.3.3 via a number k that counted the iterations of the repeated pair $\{2n + 5, 2n + 1\}$ in the continued fraction expansion of p_k/q_k .

- Lemma 3.2.1** (i) *If $(p_0, q_0, t_0) = (1, 1, 2)$ and $(p_1, q_1, t_1) = (\nu + 1, 1, \nu - 2)$ for any $\nu \geq 3$, then the identities in Lemma 3.1.2 hold. Further, with $(d_0, m_0) = (1, 1)$ and $(d_1, m_1) = (\frac{1}{2}(\nu + 1), \frac{1}{2}(\nu - 1))$, the identities in (2.2.4) hold with $\varepsilon = 1$.*
- (ii) *If $(p_0, q_0, t_0) = (-5, -1, 2)$ and $(p_1, q_1, t_1) = (\nu + 5, 1, \nu + 2)$ for any $\nu \geq 1$, then the identities in Lemma 3.1.2 hold. Further, with $(d_0, m_0) = (-2, 0)$ and $(d_1, m_1) = (\frac{1}{2}(\nu + 5), \frac{1}{2}(\nu + 3))$, the identities in (2.2.4) hold with $\varepsilon = 1$.*
- (iii) *If we define the triple $(p_\kappa, q_\kappa, t_\kappa)$ by the recursion $x_{\kappa+1} = \nu x_\kappa - x_{\kappa-1}$ for $\kappa \geq 1$, then the ratios p_κ/q_κ form an increasing sequence in case (i) and a decreasing sequence in case (ii).*

Proof The claims in (i) and (ii) hold by an easy computation. To prove (iii), it suffices by Corollaries 3.1.3 and 3.1.5 to check the first two terms. But in case (i) we have $p_1/q_1 > p_0/q_0$, while in case (ii), $p_1/q_1 = \nu + 5 > p_2/q_2 = (\nu^2 + 5\nu + 5)/(\nu + 1)$. \square

Proposition 3.2.2 *All the classes involved in the staircase family S^U can be extended to tuples $(d, m, p, q, t, \varepsilon)$ with $\varepsilon = 1$. Moreover, the staircase $S_{\bullet, n}^U$ has recursion parameter $2n + 3$, which is the t -coefficient in B_n^U , and it can be extended to have the seeds described in Lemma 3.2.1. More precisely:*

- The ascending staircase $\mathcal{S}_{\ell,n}^U$ for $n \geq 1$ has initial step given by the tuple $(d, m, p, q, t) = (1, 1, 1, 1, 2)$ and next step (at $\kappa = 1$) given by the $(t$ -extended) blocking class

$$\mathbf{B}_{n-1}^U = (n + 2, n + 1, 2n + 4, 1, 2n + 1).$$

- The descending staircase $\mathcal{S}_{u,n}^U$ for $n \geq 1$ has initial step given by the tuple $(d, m, p, q, t) = (-2, 0, -5, -1, 2)$ and next step (at $\kappa = 1$) given by the $(t$ -extended) blocking class

$$\mathbf{B}_{n+1}^U = (n + 4, n + 3, 2n + 8, 1, 2n + 5).$$

Corollary 3.2.3 The family \mathcal{S}^U is generated in the sense of [Definition 1.2.1](#) by its blocking classes $\mathbf{B}_n^U = (n + 3, n + 2, 2n + 6, 1, 2n + 3)$ together with the seeds

$$\mathbf{E}_{\ell,\text{seed}}^U = (1, 1, 1, 1, 2), \quad \mathbf{E}_{u,\text{seed}}^U = (-2, 0, -5, -1, 2), \quad \varepsilon = 1.$$

Moreover, for all staircases in this family, whether ascending or descending, the ratios m_κ/d_κ decrease.

Proof of Proposition 3.2.2 It is straightforward to check that the d and m coordinates in $\mathbf{B}_n^U := (n + 3, n + 2, 2n + 6, 1)$ for $n \geq 0$ are given by the formulas in [\(2.2.4\)](#) with

$$(3.2.1) \quad (p, q, t) = (2n + 6, 1, 2n + 3), \quad \varepsilon = 1.$$

By [Lemma 3.1.1](#), if for each $n \geq 1$ we enumerate the ascending staircase $\mathcal{S}_{\ell,n}^U$ as a single staircase that is indexed by the degree κ of p_\bullet as a function of n , then this staircase has

- recursion parameter $\nu = 2n + 3$,
- initial steps with centers $p_1/q_1 = (2n + 4)/1$ and
$$\frac{p_2}{q_2} = [2n + 5; 2n + 2] = \frac{(2n + 5)(2n + 2) + 1}{2n + 2},$$
- linear relation $(2n + 3)d_\kappa = (n + 1)p_\kappa + (n + 2)q_\kappa$.

The linear relation implies that the d values of the first two steps are $d_1 = n + 2$ and $d_2 = (n + 1)(2n + 5)$. Note also that the class with center at p_1/q_1 is precisely \mathbf{B}_{n-1}^U .

By the definition of t in [\(2.2.3\)](#), we have that $t_1 = 2n + 1 = \nu - 2$ and that d_1 is as predicted by [\(2.2.4\)](#) with $\varepsilon = 1$. Further if we take $(p_0, q_0, t_0) = (1, 1, 2)$ as in [Lemma 3.2.1](#)(i), then

$$(p_2, q_2) = (\nu p_1 - p_0, \nu q_1 - q_0).$$

Therefore we may think of the tuple $(d_2, m_2, p_2, q_2, t_2)$ as given by a recursion with $\nu = 2n + 3$ and $\varepsilon = 1$, with initial terms

$$(3.2.2) \quad (p_0, q_0, t_0) = (1, 1, 2), \quad (d_0, m_0) = (1, 1).$$

Note that the entries d_0, p_0 , and q_0 also satisfy the linear relation

$$(2n + 3)d_\kappa = (n + 1)p_\kappa + (n + 2)q_\kappa.$$

Therefore, as in [Corollary 3.1.3](#), all subsequent terms in this staircase must have degree coefficients (d, m) given by [\(2.2.4\)](#). Thus, as claimed, for each n this sequence is generated by the tuple $(d, m, p, q, t) = (1, 1, 1, 1, 2)$ together with the appropriate blocking class.

Similarly, by [Lemma 3.1.1](#), the descending staircase classes for $\mathcal{S}_{u,n}^U$ with $n \geq 0$, when indexed by κ have

- recursion parameter $\nu = 2n + 3$,
- initial steps with centers

$$\frac{p_2}{q_2} = [2n + 7; 2n + 4] = \frac{(2n + 7)(2n + 4) + 1}{2n + 4},$$

$$\frac{p_3}{q_3} = [2n + 7; 2n + 5, 2n + 2] = \frac{(2n + 7)((2n + 5)(2n + 2) + 1) + 2n + 2}{(2n + 5)(2n + 2) + 1},$$

- linear relation $(2n + 3)d_\kappa = (n + 2)p_\kappa - (n + 4)q_\kappa$.

The linear relation implies that the corresponding d values are $d_2 = 2n^2 + 11n + 14$ and $d_3 = 4n^3 + 28n^2 + 60n + 38$. Note that we can add the blocking class $\mathbf{B}_{n+1}^U = (n + 4, n + 3, 2n + 8, 1)$ to the staircase as the step for $\kappa = 1$ because

$$p_1 := (2n + 3)p_2 - p_3 = 2n + 8, \quad q_1 := (2n + 3)q_2 - q_3 = 1,$$

and the appropriate linear relation $(2n + 3)(n + 4) = (n + 2)(2n + 8) - (n + 4)$ holds.

In fact, this staircase has the form of [Lemma 3.2.1\(ii\)](#), with initial tuples

$$(3.2.3) \quad (p_0, q_0, t_0) = (-5, -1, 2), \quad (p_1, q_1, t_1) = (2n + 8, 1, 2n + 5).$$

The next entry in the recursive sequence is then

$$(p_2, q_2, t_2) = ((2n + 8)(2n + 3) + 5, 2n + 4, (2n + 5)(2n + 3) - 2),$$

which has the same values for p_2 and q_2 , as does the first staircase step given above. The formulas in [\(2.2.4\)](#) give

$$(3.2.4) \quad (d_0, m_0) = (-2, 0), \quad (d_1, m_1) = (n + 4, n + 3)$$

with $\varepsilon = 1$. Further, the linear relation $(2n + 3)d_\kappa = (n + 2)p_\kappa - (n + 4)q_\kappa$ holds in both cases. Thus, again, this staircase is generated by the initial tuple $(d, m, p, q, t) = (-2, 0, -5, -1, 2)$ together with the appropriate blocking class. \square

Proof of Corollary 3.2.3 The first claim is an immediate consequence of Proposition 3.2.2. By Corollary 3.1.5 to check the second claim we must check that $m_1d_0 - m_0d_1 < 0$ for all staircases. For the descending staircases, this is immediate because $m_0 = 0$ and $d_0 < 0$ while $m_1 > 0$. For the ascending staircases, this holds because $m_0 = d_0 = 1$ while $m_1 < d_1$. \square

Remark 3.2.4 (i) The parameters $(d, m, p, q) = (1, 1, 1, 1)$ of the initial step in the ascending staircases do correspond to those of an exceptional class in H_b , albeit one that is not live for the relevant z values (which are all $> 3 + 2\sqrt{2} = a_{\min}$). Thus its parameters are both geometrically and numerically meaningful. (See also the discussion concerning (2.3.2).) On the other hand, the parameters $(-2, 0, -5, -1)$ of the initial step in the descending staircases are negative and so, though numerically meaningful, have no immediate interpretation in terms of a point $p/q \in (1, \infty)$. Instead, they parametrize the ray $\{(-5, -1)\lambda \mid \lambda > 0\}$ in \mathbb{R}^2 . Though we do not do this here, one could think of the symmetries in terms of their action on these rays; see Remark 3.3.6.

(ii) For each $n \geq 1$, the ascending staircase $S_{\ell, n+1}^U$ with recursion parameter $t_B = 2n + 5$ has the same second step as the descending staircase $S_{u, n}^U$ with $t_B = 2n + 3$. Indeed, the formulas given above show that this step has center $(d, m, p, q, t, 1)$ with p and q given by

$$\frac{p}{q} = [2n + 7; 2n + 4] = [2(n + 1) + 5; 2(n + 1) + 2]$$

and with $t = (2n + 3)(2n + 5) - 2$.

There is a similar story for the staircase family S^L , except that now there is no geometric blocking class for $n = 0$, and the tuple $(d, m, p, q, t) = (0, -1, 1, 0, 3)$ that replaces it has no obvious geometric meaning. Nevertheless, we define $B_0^L := (0, -1, 1, 0, 3)$ for the current purposes. (See Remark 2.3.4(ii).)

Here is the appropriate numerical lemma.

Lemma 3.2.5 (i) If $(p_0, q_0, t_0) = (5, 1, 2)$ and $(p_1, q_1, t_1) = (6v - 5, v - 1, v + 2)$ for any $v > 1$, then the identities in Lemma 3.1.2 hold. Further, with $(d_0, m_0) = (2, 0)$ and $(d_1, m_1) = (\frac{5}{2}(v - 1), \frac{1}{2}(v - 3))$, the identities in (2.2.4) hold with $\varepsilon = -1$.

- (ii) If $(p_0, q_0, t_0) = (-29, -5, 2)$ and $(p_1, q_1, t_1) = (6v - 29, v - 5, v - 2)$ for any $v \geq 3$, then the identities in Lemma 3.1.2 hold. Further, with $(d_0, m_0) = (-13, -5)$ and $(d_1, m_1) = (\frac{5}{2}(v - 5), \frac{1}{2}(v - 7))$, the identities in (2.2.4) hold with $\varepsilon = -1$.
- (iii) If we define the triple $(p_\kappa, q_\kappa, t_\kappa)$ by the recursion $x_{\kappa+1} = vx_\kappa - x_{\kappa-1}$ for $\kappa \geq 1$, then the ratios p_κ/q_κ form an increasing sequence in case (i) and a decreasing sequence (for $\kappa \geq 2$) in case (ii).

Proof This is left to the reader. □

Proposition 3.2.6 All the classes involved in the staircase family S^L can be extended to tuples $(d, m, p, q, t, \varepsilon)$ with $\varepsilon = -1$. Moreover, the staircase $S_{\bullet,n}^L$ has recursion parameter $2n + 3$, which is the t -coefficient in B_n^L , and it can be extended to have the seeds described in Lemma 3.2.5. More precisely:

- The ascending staircase $S_{\ell,n}^L$ for $n \geq 0$ has initial step given by $(d, m, p, q, t) = (2, 0, 5, 1, 2)$ and with the next step (at $\kappa = 1$) given by the blocking class B_{n+1}^L with $t = 2n + 5$.
- The descending staircase $S_{u,n}^L$ for $n \geq 2$ has initial step given by $(d, m, p, q, t) = (-13, -5, -29, -5, 2)$ and with the next step (at $\kappa = 1$) given by the blocking class B_{n-1}^L with $t = 2n + 1$. When $n = 1$ there is the same initial step, and the step at $\kappa = 1$ is given by the tuple $B_0^L := (0, -1, 1, 0, 3)$.

Corollary 3.2.7 The family S^L is generated in the sense of Definition 1.2.1 by its blocking classes $B_n^L := (5n, n - 1, 12n + 1, 2n, 2n + 3)$ for $n \geq 1$, together with the seeds

$$E_{\ell,seed}^L = (2, 0, 5, 1, 2), \quad E_{u,seed}^L = (-13, -5, -29, -5, 2), \quad \varepsilon = -1,$$

where in the definition of $S_{u,1}^L$ we take the tuple B_0^L to be $(0, -1, 1, 0, 3)$. Further, for all staircases in the family with $n > 0$, both ascending and descending, the ratios m_κ/d_κ increase.

Proof of Proposition 3.2.6 The blocking classes $B_n^L := (5n, n - 1, 12n + 1, 2n)$ for $n \geq 1$ are given by the formulas in (2.2.4), with

$$(3.2.5) \quad (p, q, t) = (12n + 1, 2n, 2n + 3), \quad \varepsilon = -1.$$

By Lemma 3.1.1, if for each $n \geq 0$ we enumerate the ascending staircase $S_{\ell,n}^L$ as a single staircase that is indexed by the degree κ of p_\bullet as a function of n , then this

staircase has

- recursion parameter $\nu = 2n + 3$,
- initial steps p_κ/q_κ with centers

$$\frac{p_2}{q_2} = [6; 2n + 1, 2n + 4] = \frac{24n^2 + 62n + 34}{4n^2 + 10n + 5},$$

$$\frac{p_3}{q_3} = [6; 2n + 1, 2n + 5, 2n + 2] = \frac{(24n^2 + 98n + 121)2n + 89}{(4n^2 + 16n + 19)2n + 13},$$

- linear relation $(2n + 3)d_\kappa = (n + 1)p_\kappa - (n - 1)q_\kappa$.

The linear relation implies that $(d_2, m_2) = (10n^2 + 25n + 13, 2n^2 + 3n)$.

One can easily check that the values (p_2, q_2) and (p_3, q_3) given above agree with those in the recursive sequence with $\nu = 2n + 3$ and initial terms

$$(d_0, m_0, p_0, q_0, t_0) = (2, 0, 5, 1, 2),$$

$$(d_1, m_1, p_1, q_1, t_1) = (5(n + 1), n, 12n + 13, 2n + 2, 2n + 5).$$

Hence because the tuple for $\kappa = 0, 1$ satisfies (2.2.4) with $\varepsilon = -1$, all classes in the staircase have this form. This establishes the claims concerning $\mathcal{S}_{\ell,n}^L$.

By Lemma 3.1.1, if for each $n \geq 1$ we enumerate the descending staircase $\mathcal{S}_{u,n}^L$ as a single staircase that is indexed by the degree κ of p_\bullet as a function of n , then this staircase has

- recursion parameter $\nu = 2n + 3$,
- initial steps with centers

$$\frac{p_3}{q_3} = [6; 2n - 1, 2n + 1, 2n + 4] = \frac{48n^3 + 100n^2 + 22n - 1}{8n^3 + 16n^2 + 2n - 1},$$

$$\frac{p_4}{q_4} = [6; 2n - 1, 2n + 1, 2n + 5, 2n + 2] = \frac{96n^4 + 344n^3 + 320n^2 + 50n + 1}{16n^4 + 56n^3 + 48n^2 + 2n - 2},$$

- linear relation $(2n + 3)d_\kappa = -(n - 1)p_\kappa + (11n + 2)q_\kappa$.

Again one can check by direct computation that this sequence is generated by the tuples

$$(p_0, q_0, t_0) = (-29, -5, 2),$$

$$(p_1, q_1, t_1) = (12n - 11, 2n - 2, 2n + 1),$$

$$(p_2, q_2, t_2) = (24n^2 + 14n - 4, 4n^2 + 2n - 1, 4n^2 + 8n + 1),$$

which have the form described in Lemma 3.2.5(ii) with $\nu = 2n + 3$, $\varepsilon = -1$. Moreover, formula (2.2.4) gives the following values for d and m with $\varepsilon = -1$:

$$(d_0, m_0) = (-13, -5), \quad (d_1, m_1) = (5n - 5, n - 2). \quad \square$$

Proof of Corollary 3.2.7 The first claim is a consequence of Proposition 3.2.6. By Corollary 3.1.5, to check the second claim we must check that $m_1d_0 - m_0d_1 > 0$. For the descending staircases, we have $m_0 = -5, d_0 = -13, m_1 = n - 2, d_1 = 5n - 5$ and $n \geq 1$, so $m_1d_0 = -13(n - 2) > -5(5n - 5) = -25n + 5 = m_0d_1$. For the ascending staircases with $n \geq 1$, this holds because $m_0 = 0$ while $m_1, d_0 > 0$. \square

3.3 Proof of Theorem 1.2.4

We now prove Theorem 1.2.4 stating that each symmetry $T \in \mathcal{G}$ transforms the staircase family \mathcal{S}^U into another pre staircase family $T^\#(\mathcal{S}^U)$; in particular $T = R$ interchanges the families \mathcal{S}^U and \mathcal{S}^L .

We know from Definition 2.2.3 and Lemma 2.2.4 how $T = S^i R^\delta$ acts on quasiperfect classes. The corresponding definition for staircase families is as follows.

Definition 3.3.1 Given $T = S^i R^\delta \in \mathcal{G}$, we define $T^\#(\mathcal{S}^U)$ to be the collection of preblocking classes $(T^\#(\mathbf{B}_n^U))_{n \geq 0}$ together with the seeds $T^\#(\mathbf{E}_{\text{seed}, \ell}^U)$ and $T^\#(\mathbf{E}_{\text{seed}, u}^U)$.

We first prove our earlier claim that the reflection R takes the family \mathcal{S}^U to the family \mathcal{S}^L . Recall from Propositions 3.2.2 and 3.2.6 that all the classes in \mathcal{S}^U have $\varepsilon = 1$ while all those in \mathcal{S}^L have $\varepsilon = -1$.

Lemma 3.3.2 (i) *The map $R^\#$ takes the blocking classes and seeds of the staircase family \mathcal{S}^U together with all the associated staircase steps to the corresponding elements in the family \mathcal{S}^L . Further, $\mathbf{E}_{u, \text{seed}}^U = -\mathbf{E}_{\ell, \text{seed}}^L$, where for a class $\mathbf{E} = (d, m, p, q, t, \varepsilon)$ we define $-\mathbf{E} := (-d, -m, -p, -q, t, -\varepsilon)$.*

(ii) *Moreover, $S^\#(\mathbf{E}_{\ell, \text{seed}}^U) = \mathbf{E}_{\ell, \text{seed}}^L$ and $S^\#(\mathbf{E}_{u, \text{seed}}^U) = \mathbf{E}_{u, \text{seed}}^L$.*

Proof We already noted in [1, Corollary 60] that the reflection R takes the step classes in \mathcal{S}^U to those of \mathcal{S}^L . By Corollaries 3.2.3 and 3.2.7, the seeds of these families are

$$\mathbf{E}_{\ell, \text{seed}}^U, \mathbf{E}_{u, \text{seed}}^U : (d, m, p, q, t, \varepsilon) = (1, 1, 1, 1, 2, 1), (-2, 0, -5, -1, 2, 1)$$

$$\mathbf{E}_{\ell, \text{seed}}^L, \mathbf{E}_{u, \text{seed}}^L : (d, m, p, q, t, \varepsilon) = (2, 0, 5, 1, 2, -1), (-13, -5, -29, -5, 2, -1)$$

Therefore $\mathbf{E}_{u, \text{seed}}^U = -\mathbf{E}_{\ell, \text{seed}}^L$ as claimed. Moreover, because $R(1, 1) = (-29, -5)$ and $R(5, 1) = (-5, -1)$, we find that

$$R^\#(\mathbf{E}_{\ell, \text{seed}}^U) = \mathbf{E}_{u, \text{seed}}^L, \quad R^\#(\mathbf{E}_{u, \text{seed}}^U) = \mathbf{E}_{\ell, \text{seed}}^L.$$

i	-1	0	1	2	3	4	5	6
g_i	1	1	5	29	169	985	5741	33 461
m_i		1	0	5	24	145	840	4901
d_i		1	2	13	74	433	2522	14 701

Table 1

Since R is a reflection that interchanges the ascending and descending staircases, this reversal of seeds is to be expected. Note also that the centers of the blocking classes B_n^L descend rather than ascend, so this reversal is also consistent with Definition 1.2.1 that explains how the blocking classes and seeds generate a staircase. Thus R takes the full structure of the family S^U to that of S^L .

The proof of (ii) is straightforward, and is left to the reader. □

In order to show that for arbitrary $T \in \mathcal{G}$ the seeds and preblocking classes $T^\#(S^U)$ define a staircase family, we must show that the appropriate linear relations (1.2.5) hold. This proof is largely based on analyzing the seed classes.

As already noted, modulo sign, the seeds $E_{\bullet, \text{seed}}^U$ and $E_{\bullet, \text{seed}}^L$ for the staircase families S^U and S^L are classes that appear in the third strand (the one with $i = 2$) of the staircase at $b = 1/3$; see Example 2.3.7. This strand is generated by the recursion S with action on (d, m) given by (2.3.1), and hence consists of the classes

$$(3.3.1) \quad \begin{aligned} E_i &:= (d_i, m_i, g_i, g_{i-1}, 2, (-1)^i), \quad i \geq 0, \\ d_i &= \frac{1}{8}(3(g_i + g_{i-1}) + 2(-1)^i), \quad m_i = \frac{1}{24}(3(g_i + g_{i-1}) + 18(-1)^i), \end{aligned}$$

with values given in Table 1. Note that $g_i = y_{i+1} - y_i$ for all i . Further, the ε coordinate alternates between the value $+1$ and -1 , while $t = 2$ for all i .

Lemma 3.3.3 *For all $i \geq 0$, the families $(S^{i+1})^\#(S^U)$ and $(S^i R)^\#(S^U)$ have the same lower seeds and the same upper seeds; namely*

$$E_{\ell, \text{seed}}^{(S^{i+1})^\#(S_\ell^U)} = E_{\ell, \text{seed}}^{(S^i R)^\#(S_u^U)} = E_{i+1}, \quad E_{u, \text{seed}}^{(S^{i+1})^\#(S_u^U)} = E_{u, \text{seed}}^{(S^i R)^\#(S_\ell^U)} = -E_{i+2},$$

where E_i is as in (3.3.1). Furthermore, the extended action $R_{v_{i+2}}^\#$ of the reflection $R_{v_{i+2}}$ that is defined in (2.2.6) interchanges the lower and upper seeds of these families.

Proof The first statement is an easy consequence of Lemma 3.3.2(ii) and the fact that $S^\#(E_i) = E_{i+1}$ for all i .

For the second statement, note that by Corollary 2.1.4(i),

$$(R_{v_{i+2}})^\# = (S^{2i+1}R)^\# = (S^iRS^{-(i+1)})^\#.$$

The upper seed of \mathcal{S}_u^U is $-E_1$; thus to compute the lower seed of $(R_{v_{i+2}})^\#((S^{i+1})^\#(\mathcal{S}_u^U))$ we are considering

$$(S^iRS^{-(i+1)})^\#((S^{i+1})^\#(-E_1)) = (S^i)^\#R^\#(-E_1) = (S^i)^\#(E_1) = E_{i+1},$$

where the first equality follows from Lemma 2.2.4. The upper seed can be computed similarly. □

Recall from the discussion concerning (1.2.5) that, in order for a staircase S with steps $E := (d, m, p, q, t, \varepsilon)$ to be associated to the preblocking class

$$B = (d_B, m_B, p_B, q_B, t_B, \varepsilon),$$

we require that for each step the following linear relation holds:

$$(3.3.2) \quad (3m_B - d_B)d = \begin{cases} (m_B - q_B)p + m_Bq & \text{if } S \text{ ascends,} \\ m_Bp - (p_B - m_B)q & \text{if } S \text{ descends.} \end{cases}$$

Lemma 3.3.4 (i) *The identities in (3.3.2) may be rewritten as*

$$(3.3.3) \quad m_Bm = d_Bd - q_Bp \quad \text{if } S \text{ ascends,}$$

$$(3.3.4) \quad m_Bm = d_Bd - p_Bq \quad \text{if } S \text{ descends.}$$

(ii) *Given any two classes E' and E'' , (3.3.3) holds for the pair $(E, B) = (E', E'')$ if and only if (3.3.4) holds for the pair $(E, B) = (E'', E')$.*

(iii) *Equation (3.3.3) holds for the pair (E, B) if and only if (3.3.4) holds for the pair $(S^\#(E), B)$. Further, (3.3.4) holds for the pair (E, B) if and only if (3.3.3) holds for the pair $(E, S^\#(B))$.*

(iv) *If both (3.3.3) and (3.3.4) hold for (E, B) then they both hold for $(S^\#(E), S^\#(B))$.*

Proof To prove (i), rewrite the term $3m_Bd$ on the left-hand side as $m_B(p + q + m)$ and simplify.

Proving (ii) is also straightforward: if $E' := (d', m', p', q')$ and $E'' := (d'', m'', p'', q'')$, then both equations reduce to the claim that $m'm'' = d'd'' - p'q''$.

Now consider (iii). Let $E = (d, m, p, q, t, \varepsilon)$, so that

$$S^\#(E) = (d', m', (6p - q), p, t, -\varepsilon),$$

$$d' = \frac{1}{8}(3(7p - q) - \varepsilon t), \quad m' = \frac{1}{8}((7p - q) - 3\varepsilon t).$$

To prove the claim about the pair $(S^\sharp(E), \mathbf{B})$, we want to show that the equation $m_{\mathbf{B}}(p + q + 3t\varepsilon) = d_{\mathbf{B}}(3(p + q) + t\varepsilon) - 8q_{\mathbf{B}}p$ holds if and only if

$$m_{\mathbf{B}}((7p - q) - 3\varepsilon t) = d_{\mathbf{B}}(3(7p - q) - \varepsilon t) - 8p_{\mathbf{B}}p$$

also holds. But by adding the two equations, we obtain the linear Diophantine identity

$$m_{\mathbf{B}}8p = d_{\mathbf{B}}24p - (q_{\mathbf{B}} + p_{\mathbf{B}})8p,$$

which always holds. This proves the first claim in (iii).

The second claim in (iii) now follows from the symmetry relation in (ii). We have

$$\begin{aligned} (3.3.4) \text{ holds for } (E, \mathbf{B}) &\iff (3.3.3) \text{ holds for } (\mathbf{B}, E) && \text{(by (ii))} \\ &\iff (3.3.4) \text{ holds for } (S^\sharp(\mathbf{B}), E) \\ &\iff (3.3.3) \text{ holds for } (E, S^\sharp(\mathbf{B})) && \text{(by (ii)).} \end{aligned}$$

Thus (iii) holds, and a similar argument proves (iv). □

Proof of Theorem 1.2.4 We must show that for each $T \in \mathcal{G}$ the seeds and preblocking classes in $T^\sharp(S^U)$ form a prestaircase family in the sense of Definition 1.2.1. Here, we define the preblocking classes and seeds of $T^\sharp(S^U)$ to be the images of the preblocking classes and seeds in S^U by the formula in (2.2.6). Lemma 2.2.4 shows that these are quasiperfect classes. Further, Lemma 3.3.3 shows that the lower and upper seeds of both families $(S^{i+1})^\sharp(S^U)$ and $(S^i R)^\sharp(S^U)$ are E_{i+1} and $-E_{i+2}$.

Therefore, it remains to check that each preblocking class \mathbf{B}_n is associated both to the ascending prestaircase with seeds $E_{\ell, \text{seed}}, \mathbf{B}_{n-1}$ and recursion parameter $t_{\mathbf{B}_n}$, and to the descending prestaircase with seeds $E_{u, \text{seed}}, \mathbf{B}_{n+1}$ and recursion parameter $t_{\mathbf{B}_n}$. Thus each prestaircase must satisfy the appropriate linear relation (1.2.5). Because the steps in the staircases of a staircase family \mathcal{F} with seeds $E_{\ell, \text{seed}}^{\mathcal{F}}, E_{u, \text{seed}}^{\mathcal{F}}$ and preblocking classes $\mathbf{B}_n^{\mathcal{F}}$ are defined recursively by Corollary 3.1.3, it suffices to check this in the cases

$$\begin{aligned} (E, \mathbf{B}) &= (E_{\ell, \text{seed}}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), (\mathbf{B}_{n-1}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), & n \geq 1, & \text{ if steps } \uparrow, \mathbf{B}_n^{\mathcal{F}} \uparrow \\ &= (E_{u, \text{seed}}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), (\mathbf{B}_{n+1}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), & n \geq 0, & \text{ if steps } \downarrow, \mathbf{B}_n^{\mathcal{F}} \uparrow, \\ (E, \mathbf{B}) &= (E_{\ell, \text{seed}}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), (\mathbf{B}_{n+1}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), & n \geq 1, & \text{ if steps } \uparrow, \mathbf{B}_n^{\mathcal{F}} \downarrow \\ &= (E_{u, \text{seed}}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), (\mathbf{B}_{n-1}^{\mathcal{F}}, \mathbf{B}_n^{\mathcal{F}}), & n \geq 0, & \text{ if steps } \downarrow, \mathbf{B}_n^{\mathcal{F}} \downarrow. \end{aligned}$$

It is shown in [1] that the prestaircases in the families S^U and S^L are staircases, and their seeds, blocking classes and steps are described in Corollaries 3.2.3 and 3.2.7.

Therefore the needed identities hold when $\mathcal{F} = S^U$ (with ascending blocking classes) and S^L (with descending blocking classes). Thus it suffices to show that if \mathcal{F} satisfies the appropriate set of identities, then its image $S^\sharp(\mathcal{F})$ does as well.

For clarity, let us first suppose that \mathcal{F} has ascending blocking classes, and denote its seeds by $E_{\ell, \text{seed}}^{\mathcal{F}}, E_{u, \text{seed}}^{\mathcal{F}} = -S^\sharp(E_{\ell, \text{seed}}^{\mathcal{F}})$ and its preblocking classes by $B_n^{\mathcal{F}}$ for $n \geq 0$. Then we know that for all n ,

$$(E_{\ell, \text{seed}}^{\mathcal{F}}, B_n^{\mathcal{F}}) \text{ and } (B_{n-1}^{\mathcal{F}}, B_n^{\mathcal{F}}) \text{ satisfy (3.3.3),}$$

$$(E_{u, \text{seed}}^{\mathcal{F}}, B_n^{\mathcal{F}}) \text{ and } (B_{n+1}^{\mathcal{F}}, B_n^{\mathcal{F}}) \text{ satisfy (3.3.4).}$$

In particular, adjacent blocking classes $B_n^{\mathcal{F}}$ and $B_{n+1}^{\mathcal{F}}$ satisfy both relations (3.3.3) and (3.3.4). Hence Lemma 3.3.4(iv) shows that their images by S^\sharp also satisfy both relations. Further, since

$$E_\ell := S^\sharp(E_{\ell, \text{seed}}^{\mathcal{F}}) = E_{\ell, \text{seed}}^{S^\sharp(\mathcal{F})} = -E_{u, \text{seed}}^{\mathcal{F}},$$

Lemma 3.3.4(iii) shows that the pair

$$(S^\sharp(E_{\ell, \text{seed}}^{\mathcal{F}}), S^\sharp(B_n^{\mathcal{F}})) = (-E_{u, \text{seed}}^{\mathcal{F}}, S^\sharp(B_n^{\mathcal{F}}))$$

satisfies (3.3.3) because (3.3.4) holds for $(E_{u, \text{seed}}^{\mathcal{F}}, B_n^{\mathcal{F}})$. (Note that the $-$ sign in front of $E_{u, \text{seed}}^{\mathcal{F}}$ is immaterial.) In turn, applying Lemma 3.3.4(ii), we find that (3.3.3) for $(E_{u, \text{seed}}^{\mathcal{F}}, S^\sharp(B_n^{\mathcal{F}}))$ implies (3.3.4) for $(S^\sharp(B_n^{\mathcal{F}}), E_{u, \text{seed}}^{\mathcal{F}})$, and hence, by Lemma 3.3.4(iii), (3.3.3) for $(S^\sharp(B_n^{\mathcal{F}}), S^\sharp(E_{u, \text{seed}}^{\mathcal{F}}))$ and therefore (3.3.4) for $(S^\sharp(E_{u, \text{seed}}^{\mathcal{F}}), S^\sharp(B_n^{\mathcal{F}}))$. This completes the proof in this case.

The argument when \mathcal{F} has descending blocking classes is essentially the same, and is left to the reader. Finally note that claims (i) and (ii) in the theorem follow from the known behavior of S^U and S^L , and the fact that $S(v_i) = v_{i+1}$ and $S(w_i) = w_{i+1}$. \square

Remark 3.3.5 (i) As the elements $T \in \mathcal{G}$ bring the seeds to seeds and the blocking classes to blocking classes by formula (2.2.6), Lemma 2.2.4(ii) implies composition behaves nicely on the staircase families, namely $T_1^\sharp T_2^\sharp(S^U) = (T_1 T_2)^\sharp(S^U)$. This is simply because the formulas for $T^\sharp(p, q, t)$ are compatible with composition, and the degree coordinates (d, m) are determined by the values of p and q . We will see in Section 3.4 that while T^\sharp acts linearly on (p, q) (namely, via products of S and R), there is no general linear map on the degree coordinates (d, m) that respects composition.

(ii) The paper [1] established the existence of a third set of staircases, there called S^E , and showed that the centers of its blocking classes and staircase steps are the

images via S of the centers of the corresponding classes in \mathcal{S}^U . The interested reader can check that, just as in the proofs of Propositions 3.2.2 and 3.2.6, \mathcal{S}^E forms a staircase family with seeds $S^\#(E_{\ell, \text{seed}}^U)$, $S^\#(E_{u, \text{seed}}^U)$ and blocking classes $S^\#(E_n^U)$. Thus $\mathcal{S}^E = S^\#(\mathcal{S}^U)$, and so has the same seeds as \mathcal{S}^L by Lemma 3.3.2.

Now the E -family has an ascending family of blocking classes in the interval $(w_2, v_2) = (41/7, 6)$, while the L -family has a descending family of blocking classes in the interval (v_2, w_1) . We showed in [1, Corollary 60] that the centers of the blocking classes and steps in \mathcal{S}^E are mapped to those in \mathcal{S}^L by the reflection $R_{v_2} = RS^{-1} = SR$ that fixes v_2 and interchanges $w_2 = 41/7$ and $w_1 = 7$. Moreover, by Lemma 3.3.3, $R_{v_2}^\#$ interchanges the seeds of these two staircase families \mathcal{S}^E and \mathcal{S}^L . We show in Lemma 3.4.5 below that the action of $R_{v_2}^\#$ on the (d, m, p, q) components of those blocking classes of \mathcal{S}^E and \mathcal{S}^L with centers in (w_2, w_1) decomposes as the product $(R_{v_2})_{\mathbf{B}}^* \times R_{v_2} : \mathbb{Z}^2 \times \mathbb{Z}^2 \rightarrow \mathbb{Z}^2 \times \mathbb{Z}^2$ of two matrices of order 2. This is a rather special situation that, we explain in the discussion before Proposition 3.4.3, does have a nice geometric interpretation.

Remark 3.3.6 The sequence g_i is recursively generated by $g_{i+1} = 6g_i - g_{i-1}$ giving us terms in the sequence as i increases. Rewriting this as $g_{i-1} = 6g_i - g_{i+1}$, we can solve for terms in the sequence as i decreases. In particular, from the data in Table 1, we can solve for terms in this sequence with negative indices. Geometrically, we want to consider the sequence $\{g_i/g_{i-1}\}$, the center of these classes. Computing a few terms in these sequences, we get

i	-4	-3	-2	-1	0	1	2	3
g_i	169	29	5	1	1	5	29	169
g_i/g_{i-1}	$\frac{169}{985}$	$\frac{29}{169}$	$\frac{5}{29}$	$\frac{1}{5}$	1	5	$\frac{29}{5}$	$\frac{169}{29}$

Thus the sequence $\{g_i\}$ reflects on itself, namely $g_i = g_{-(i+1)}$. Note that both the centers of E_i and $-E_i$ are elements of this sequence $\{g_i/g_{i-1}\}$ when $i \geq 0$. In particular, only terms with $i \geq 0$ appear as centers of the seed classes.

As S implements the recursion by 6, $S(g_i/g_{i-1}) = g_{i+1}/g_i$, so S always shifts this sequence one step to the right. Now, R is the reflection of this sequence about $5 = g_1/g_0$; indeed we can compute that $R(g_i/g_{i-1}) = g_{2-i}/g_{1-i}$. (Note: R is not the reflection about 1 which would map each element to its reciprocal.) Hence, R takes centers of seeds to terms in the sequence, but not necessarily to valid centers

since $g_{2-i}/g_{1-i} < 1$ when $i > 2$. But applying S will shift g_{2-i}/g_{1-i} to the right. In particular, the reflection $R_{v_i} = S^{2i-3}R$ has just the right number of shifts to move the image under R of the seeds of $S^{i-1}(S^U)$ to the seeds of $S^{i-2}(S^L)$. In other words, this reflection interchanges the lower and upper seeds of these families. As we show in Lemma 3.4.5 it also interchanges their blocking classes.

3.4 Action of the symmetries on blocking class degree

Although the results in this section are not needed for the proof of our main results, they throw light on the geometric nature of the symmetries. To simplify the language needed, we will assume it is already known that all classes in $T^\sharp(S^U)$ are perfect, a result that is proved in Proposition 4.1.4. In particular, this means that all the preblocking classes in $T^\sharp(S^U)$ are in fact blocking classes, and so we will talk about blocking classes rather than preblocking classes.

The first step is to derive a formula for the action of T on the degree coordinates (d, m) of the blocking classes in S^U . Because the formula (2.2.4) for d and m in terms of p and q is affine, we might expect this action to be affine and to depend on the t -variable (or equivalently on n). However, it turns out to be linear and independent of t and n , and in fact is given by a 2×2 matrix (with integer entries) that we denote by $T_{\mathbf{B}}^*$. Note that $T_{\mathbf{B}}^*$ gives the action on the degree coordinates of blocking classes, and in particular does not describe the action on the degrees of the staircase steps. For convenience, we denote the vector with components d and m by $(d, m)^\vee$.

- Lemma 3.4.1** (i) For each $T = S^i R^\delta \in \mathcal{G}$, there is a 2×2 matrix $T_{\mathbf{B}}^*$ such that for each blocking class $\mathbf{B}_n^U = (d_n, m_n, p_n, q_n, t, 1)$ with $4n \geq 0$, in S^U , the corresponding blocking class in $T^\sharp(\mathbf{B}_n^U)$ has degree coefficients $T_{\mathbf{B}}^*((d_n, m_n)^\vee)$.
- (ii) In all cases, $T_{\mathbf{B}}^*$ has eigenvector $(3, 1)^\vee$ with eigenvalue $(-1)^{i+\delta} \det(T_{\mathbf{B}}^*)$.

Proof Let $T^\sharp(d, m, p, q, t, 1) = (d', m', p', q', t, \varepsilon')$, where $\mathbf{B}_n^U = (d, m, p, q, t, 1)$. By (2.2.4), $3m' - d' = t\varepsilon'$ where t is constant under T^\sharp and

$$(3.4.1) \quad \varepsilon' t = (-1)^{i+\delta} t = (-1)^{i+\delta} (3m - d) = (-1)^{i+\delta} (p - 3q),$$

where the last equality holds because $\mathbf{B}_n^U = (n + 3, n + 2, 2n + 6, 1, 2n + 3, 1)$. Therefore,

$$(3.4.2) \quad \begin{aligned} d' &= \frac{1}{8}(3(p' + q') + \varepsilon' t) = \frac{1}{8}(3(p' + q') + (-1)^{i+\delta}(p - 3q)), \\ m' &= \frac{1}{8}((p' + q') + 3\varepsilon' t) = \frac{1}{8}((p' + q') + 3(-1)^{i+\delta}(p - 3q)). \end{aligned}$$

Since p' and q' are linear functions of p and q , and the coefficients of the B_n^U satisfy $p = 2d$ and $q = d - m$, the degree coefficients d' and m' are linear functions of d and m . This proves (i).

To prove (ii), let $A = T_B^* = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Because the transformation A preserves the linear function $d - 3m$ modulo the sign factor $(-1)^{i+\delta}$, the transpose matrix A^T has eigenvector $(1, -3)^\vee$ with eigenvalue $(-1)^{i+\delta}$. This implies that the transformation A preserves the subspace orthogonal to $(1, -3)^\vee$, and hence has eigenvector $(3, 1)^\vee$ with some eigenvalue λ . Further, the matrices A and A^T have the same eigenvalues, λ_1 and λ_2 , where $\lambda_1 + \lambda_2 = \text{Tr}(A) = a + d$. We know that one eigenvalue, say λ_1 , is $(-1)^{i+\delta}$. If $\lambda_1 = \lambda = (-1)^{i+\delta}$ then the identities $A(3, 1)^\vee = \lambda(3, 1)^\vee$ and $A^T(1, -3)^\vee = \lambda(1, -3)^\vee$ imply that $a + d = 2\lambda$. The identity $\text{Tr}(A) = a + d = 2\lambda$ shows that the other eigenvalue is also $\lambda = (-1)^{i+\delta}$. Therefore $\det(A) = 1$ and $\lambda = (-1)^{i+\delta} \det(T_B^*)$ as claimed. On the other hand, if $\lambda \neq \lambda_1 = (-1)^{i+\delta}$, then $\det(T_B^*) = \lambda_1 \lambda_2$, so $\lambda = \lambda_2 = (-1)^{i+\delta} \det(T_B^*)$ in this case as well. \square

Remark 3.4.2 It would be more correct (though also more cumbersome) to denote the map T_B^* by $T_{U,B}^*$ since its formula depends on the domain staircase S^U via the identity $3m - d = p - 3q$ in (3.4.1) used to express $\varepsilon't$ as a function of p and q . Each staircase family has an analogous, but different, identity of this kind. For example, the blocking classes $(5n, n - 1, 12n + 1, 2n)$ satisfy $3m - d = -3p + 17q$. Correspondingly, the assignment $T \rightarrow T_B^*$ is not compatible with composition in general, though it is in a few special cases. For example, even though the reflection R in $w_1 = 7$ has order 2, R_B^* does not have determinant ± 1 and $(R_B^*)^2 \neq \text{id}$; see Proposition 3.4.3(ii). Further, we show in Example 3.4.4 that $(S^4)_B^* \neq ((S^2)_B^*)^2$.

We now show that the formula for T_B^* does have understandable features. Observe that the blocking classes B_n^U have centers at $(p_n, q_n) = (2n + 6, 1)$ and degree components (d, m) ,

$$(d_n, m_n) = (3 + n, 2 + n) = (3, 2) + (1, 1)n, \quad \lim_{n \rightarrow \infty} \frac{m_n}{d_n} = 1.$$

This is no accident: we should expect the limits of the sequences (m_n/d_n) and (p_n/q_n) to correspond via the function acc_U^{-1} . However, this is a degenerate case since both limits lie on the boundary of the domain of this function. Let us apply the same analysis to the staircase families $S^L = R^\#(S^U)$ and $S^E = S^\#(S^U)$. As illustrated in Figures 2 and 3, both these families have steps in the interval $(w_2, w_1) = (41/7, 7)$. Further, the

centers of the blocking classes in \mathcal{S}^E increase with limit $v_2 = 6$, while those of \mathcal{S}^L decrease, also with limit v_2 . Thus, in both cases, the ratios (m_n/d_n) of the degree coefficients of the blocking classes converge to $\text{acc}_L^{-1}(6) = 1/5$. Correspondingly, we show below that in both cases $T = R, S$ the matrix $T_{\mathbf{B}}^*$ takes the vector $(1, 1)^\vee$ that gives the coefficient of n in (d, m) to $(5, 1)^\vee$. This observation generalizes as follows.

Proposition 3.4.3 (i) *The matrix $(S^i)_{\mathbf{B}}^*$ has integer entries and is determined by the following properties:*

- $(S^i)_{\mathbf{B}}^*((3, 2)^\vee) = \frac{1}{8}(3(y_{i+2} + y_{i+1}) + 3(-1)^i, y_{i+2} + y_{i+1} + 9(-1)^i)^\vee$.
- $(S^i)_{\mathbf{B}}^*((1, 1)^\vee) = ((s_i, r_i)^\vee)$, where

$$(3.4.3) \quad s_i = \frac{1}{4}(3y_{i+1} + 3y_i + (-1)^i), \quad r_i = \frac{1}{4}(y_{i+1} + y_i + 3(-1)^i).$$

In particular $r_i/s_i = \text{acc}_\varepsilon^{-1}(v_{i+1})$, where $\varepsilon = (-1)^i$ and we define

$$\text{acc}_\varepsilon^{-1} := \begin{cases} \text{acc}_U^{-1} & \text{if } \varepsilon = +1, \\ \text{acc}_L^{-1} & \text{if } \varepsilon = -1. \end{cases}$$

Further, $\det((S^i)_{\mathbf{B}}^*) = (-1)^i(y_{i+1} - y_i)$.

(ii) *When $T = R$ we have $R_{\mathbf{B}}^* = \begin{pmatrix} -10 & 15 \\ -3 & 4 \end{pmatrix}$.*

(iii) *The matrix $(S^i R)_{\mathbf{B}}^*$ has integer entries and is determined by the following properties:*

- $(S^i R)_{\mathbf{B}}^*((3, 2)^\vee) = \frac{1}{8}(3(y_{i+1} + y_i) - 3(-1)^i, y_{i+1} + y_i - 9(-1)^i)^\vee$.
- $(S^i R)_{\mathbf{B}}^*((1, 1)^\vee) = ((s_{i+1}, r_{i+1})^\vee)$, where s_{i+1} and r_{i+1} are as in (3.4.3).

Further, $\det((S^i R)_{\mathbf{B}}^*) = (-1)^i(y_{i+2} - y_{i+1})$.

Proof The matrix $(S^i)_{\mathbf{B}}^*$ is obviously determined by the images of the vectors $(3, 2)^\vee$ and $(1, 1)^\vee$, and we first check that these images are as stated. Recall the sequence

$$y_0, y_1, y_2, y_3, \dots = 0, 1, 6, 35, \dots, \quad \text{where } S(y_i, y_{i-1}) = (y_{i+1}, y_i),$$

and write

$$(p_n, q_n)^\vee = (6, 1)^\vee + 2n(1, 0)^\vee = (y_2, y_1)^\vee + 2n(y_1, y_0)^\vee.$$

Then

$$S^i((p_n, q_n)^\vee) = (y_{i+2}, y_{i+1})^\vee + 2n(y_{i+1}, y_i)^\vee,$$

and because $t = 2n + 3$ is fixed by S , we can derive the formula for $(S^i)_B^* ((3, 2)^\vee)$ and $(S^i)_B^* ((1, 1)^\vee)$ by looking at the constant term and coefficient of n in (2.2.6). Thus, if X is the matrix with columns $(3, 2)^\vee$ and $(1, 1)^\vee$, we have $(S^i)_B^* X = \frac{1}{8}A$, where

$$A = \begin{pmatrix} 3w_{i+1} + 3\varepsilon & 6w_i + 2\varepsilon \\ w_{i+1} + 9\varepsilon & 2w_i + 6\varepsilon \end{pmatrix}, \quad w_i = y_{i+1} + y_i, \quad \varepsilon := (-1)^i.$$

It follows from Lemma 2.2.4 that the entries of $(S^i)_B^* X$ are integers. Hence, because the matrix X with columns $(3, 2)^\vee$ and $(1, 1)^\vee$ has determinant $+1$, the entries of $(S^i)_B^*$ are also integers. Further,

$$\begin{aligned} \det A &= \varepsilon(18w_{i+1} + 6w_i - 54w_i - 2w_{i+1}) \\ &= 16\varepsilon(y_{i+2} + y_{i+1} - 3(y_{i+1} + y_i)) \\ &= 64\varepsilon(y_{i+1} - y_i), \end{aligned}$$

where the last equality holds because $y_{i+2} = 6y_{i+1} - y_i$. Therefore $\det((S^i)_B^*)$ is as claimed. This proves (i).

Claim (ii) follows from the fact that R_B^* takes the degree components $(4, 3)$ and $(5, 4)$ of B_1^U and B_2^U to the corresponding components $(5, 0)$ and $(10, 1)$ of B_1^L and B_2^L ; see (3.2.5). Note that

$$R_B^* \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -10 & 15 \\ -3 & 4 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 5 \\ -1 & 1 \end{pmatrix}.$$

Finally, (iii) follows by arguing as in (i), noting that R interchanges the pairs $(6, 1) = (y_2, y_1)$ and $(1, 0) = (y_1, y_0)$. Therefore

$$S^i R((6, 1)^\vee) = ((y_{i+1}, y_i)^\vee), \quad S^i R((1, 0)^\vee) = ((y_{i+2}, y_{i+1})^\vee),$$

so, as before, (2.2.6) implies that the action on the corresponding degree coordinates $(3, 2)^\vee$ and $(1, 1)^\vee$ is

$$\begin{aligned} (S^i R)_B^* ((3, 2)^\vee) &= \frac{1}{8}((3(y_{i+1} + y_i) + 3(-1)^{i+1}, y_{i+1} + y_i + 9(-1)^{i+1}))^\vee, \\ (S^i R)_B^* ((1, 1)^\vee) &= \frac{1}{4}((3(y_{i+2} + y_{i+1}) + (-1)^{i+1}, y_{i+2} + y_{i+1} + 3(-1)^{i+1}))^\vee. \end{aligned}$$

These formulas are consistent with the fact that by Proposition 3.2.6 the blocking classes $B_0^L = R^\#(B_0^U)$ and $B_1^L = R^\#(B_1^U)$, are $(0, -1, 1, 0)$ and $(5, 0, 13, 2)$. The determinant calculation can be checked as before. □

Example 3.4.4 As examples of the above formulas, we have

$$S_B^* = \begin{pmatrix} 5 & 0 \\ 2 & -1 \end{pmatrix}, \quad (S^2)_B^* = \begin{pmatrix} 28 & 3 \\ 9 & 2 \end{pmatrix}, \quad (S^3)_B^* = \begin{pmatrix} 164 & 15 \\ 55 & 4 \end{pmatrix}, \quad (S^4)_B^* = \begin{pmatrix} 955 & 90 \\ 318 & 31 \end{pmatrix}.$$

Note that the second column of these matrices coincides with the degree components of a corresponding principal¹⁵ blocking class $(S^{i-2})^*(\mathbf{B}_0^U)$. However, these matrices do *not* give the action of S^i on the seed classes, even when i is even. For example, the lower seeds of S^U and $(S^2)^\sharp(S^U)$ are

$$\mathbf{E}_{\ell, \text{seed}}^U = (1, 1, 1, 1, 2, 1), \quad \mathbf{E}_{\ell, \text{seed}}^{(S^2)^\sharp(S^U)} = (13, 5, 29, 5, 2, 1).$$

Because these matrices $(S^i)_{\mathbf{B}}^*$ do not give the action on the seeds, they also do not act on the degrees of the staircase steps.

We could compute similar matrices $(T)_{\bullet, \text{seed}}^*$ that would take the degrees of $\mathbf{E}_{\bullet, \text{seed}}^U$ to the degrees of $T^\sharp(\mathbf{E}_{\bullet, \text{seed}}^U)$. However, as above, $(T)_{\bullet, \text{seed}}^*$ would also not respect composition because there is no analog of the identity $t = 3m - d = ap + bq$ in (3.4.1) for the seeds. (This is easy to check using the fact that $t = 2$ by (2.3.1).) Thus we only present the formulas for the degree coordinates generated by the recursion S by (3.3.1); also see Lemma 3.3.3 and Remark 3.3.6.

By Lemma 2.1.3 and Corollary 2.1.4 we can also write

$$S^{i-1}R = (S^{i-1}RS^{-i})S^i = R_{v_{i+1}}S^i,$$

where $R_{v_{i+1}}$ is a reflection that fixes the point v_{i+1} and interchanges the steps of the two staircase families $(S^i)^\sharp(S^U)$ and $(S^{i-1}R)^\sharp(S^U)$ with steps in the short interval (w_{i+1}, w_i) around v_{i+1} ; see Figure 2. Let us denote by $(R_{v_{i+1}})_{\mathbf{B}}^*$ the matrix that takes the blocking class degrees in $(S^{i-1}R)^\sharp(S^U)$ to those of $(S^i)^\sharp(S^U)$; see Figure 3. Note that to be consistent with our interpretation of R that takes S^U to S^L we take the p/q coordinates of the domain staircase family of this reflection to lie *above* the fixed point.¹⁶ Further, when i is even (resp. odd), $(R_{v_{i+1}})_{\mathbf{B}}^*$ acts on blocking classes of staircases with $b_\infty > 1/3$ (resp. $b_\infty < 1/3$). For example, $(R_{v_2})_{\mathbf{B}}^*$ takes the blocking classes of S^L to those of $S^E = S^\sharp(S^U)$.

Lemma 3.4.5 *The matrix*

$$(R_{v_{i+1}})_{\mathbf{B}}^* := (S^i)_{\mathbf{B}}^* \circ ((S^{i-1}R)_{\mathbf{B}}^*)^{-1}$$

has eigenvectors $(s_i, r_i)^\vee$ and $(3, 1)^\vee$ with corresponding eigenvalues 1 and -1 . Thus it has order two, and hence also takes the degree components of the blocking classes in $(S^i)^\sharp(S^U)$ to those of $(S^{i-1}R)^\sharp(S^U)$.

¹⁵See Lemma 3.4.6; these are well defined for $i - 2 \geq 0$ and need appropriate interpretation when $i = 1$.

¹⁶This choice is relevant because in general the degree component $T_{\mathbf{B}}^*$ of a reflection T^\sharp does not have order two; see Lemma 3.4.6.

Proof Proposition 3.4.3 shows that

$$(S^{i-1}R)_B^*((1, 1)^\vee) = (s_i, r_i) = (S^i)_B^*((1, 1)^\vee)$$

and $\det((S^{i-1}R)^*) = -\det((S^i)^*)$. The result now follows from Lemma 3.4.1(ii). \square

There turns out to be a second natural action of the reflections R_{v_j} for $j \geq 0$ on blocking class degree corresponding to its action on the other component of $[0, 1/3) \cup (1/3, \infty)$. Indeed, as illustrated in Figure 3 and Remark 1.2.5(i), this reflection acts in two ways on our staircases depending on whether one takes $b > 1/3$ or $b < 1/3$. The action discussed above (with $j = i + 2 \geq 2$) interchanges the blocking classes of the two staircase families $(S^{i+1})^\#(S^U)$ and $(S^i R)^\#(S^U) = (S^i)^\#(S^L)$, the centers of whose blocking classes converge to v_{i+2} while, as we will see, the second action of $R_{v_{i+2}}$ fixes the center v_{i+2} of the blocking class $(S^i)^\#(B_0^U)$ and takes the ascending blocking classes in $(S^i)^\#(S^U)$ with centers in $[w_{i+1}, v_{i+1}]$ to the descending blocking classes in $(S^{i+1}R)^\#(S^U) = (S^{i+1})^\#(S^L)$ with centers in $[v_{i+3}, w_{i+2}]$. Note, when i is odd (resp. even), the principal blocking class $(S^i)^\#(B_0^U)$ blocks a b -region with $b < 1/3$ (resp. $b > 1/3$).

We will denote the matrix that gives this second action on degree by $(R_{v_{i+1}})_P^*$. It is again defined by its action on relevant blocking classes, but the family used is different than before.

Notice that the matrix $(R_{v_{i+1}})_P^*$ obtained this way does not have order two, so that it matters that we define it via (2.2.6) applied to its action on the blocking classes in $(S^i)^\#(S^U)$. Besides the eigenvector $(3, 1)^\vee$, its second eigenvector (with eigenvalue 1) must be given by the degree components of the class $(S^i)^\#(B_0^U)$. Note also that the resulting matrix $(R_{v_{i+2}})_P^*$ is not in general integral.

We call the blocking classes $(S^i)^\#(B_0^U)$, $i \geq 0$, *principal blocking classes*. The next result spells out their main properties.

Lemma 3.4.6 (i) *The principal blocking class $(S^i)^\#(B_0^U)$ has components*

$$\left(\frac{3}{8}(y_{i+2} + y_{i+1} + (-1)^i), \frac{1}{8}(y_{i+2} + y_{i+1} + 9(-1)^i), y_{i+2}, y_{i+1}, 3, (-1)^i\right).$$

(ii) *The transformation $(R_{v_{i+2}})_P^\# : (S^i)^\#(S^U) \rightarrow (S^{i+1})^\#(S^L)$ acts on the degrees of the blocking classes by the matrix*

$$(R_{v_{i+2}})_P^* := (S^{i+1}R)_B^*((S^i)_B^*)^{-1}$$

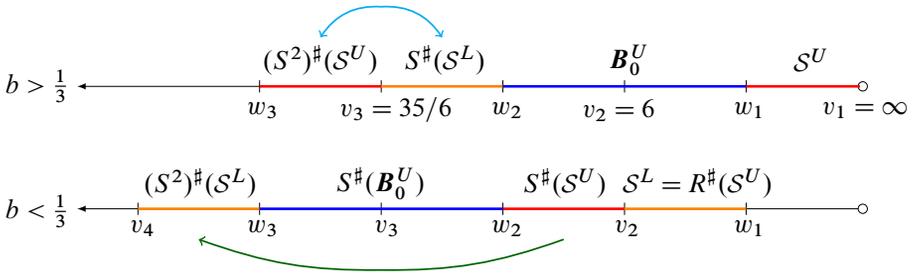


Figure 3: Here we illustrate the different actions of the reflection R_{v_3} that acts on the z axis by fixing v_3 . The reflection $(R_{v_3})^\sharp$, represented by the light blue arrow, interchanges the upper and lower staircase families of $(S^2)^\sharp(S^U)$ and $S^\sharp(S^L)$, and the corresponding matrix $(R_{v_3})^\sharp_{\mathbf{B}}$ acts as a reflection on the degrees of the blocking classes associated to these staircases. On the other hand, on the lower number line, we have the principal blocking class $S^\sharp(\mathbf{B}_0^U)$ with center v_3 blocking the corresponding blue interval with $b < 1/3$. On either side of this blocked region live the centers of the blocking classes of $(S^2)^\sharp(S^L)$ and $S^\sharp(S^U)$. The matrix $(R_{v_3})^*_{\mathbf{P}}$ maps the degree coordinates of $S^\sharp(S^U)$ to those of $(S^2)^\sharp(S^L)$. This is represented by the green arrow. Note that this is not a reflection, so the arrow only goes in one direction.

with determinant given by $-(y_{i+3} - y_{i+2}) / (y_{i+2} - y_{i+1})$. Its eigenvalues are $\det((R_{v_{i+2}})^*_{\mathbf{P}})$ and 1, with corresponding eigenvectors given by $(3, 1)^\vee$ and the degree components of $(S^i)^\sharp(\mathbf{B}_0^U)$.

Proof The formula for the degree components of the principal blocking class $\mathbf{B} := (S^i)^\sharp(\mathbf{B}_0^U)$ follows from (2.2.6). The claim in (ii) follows from Proposition 3.4.3. \square

Example 3.4.7 The reflection R_{v_2} that fixes 6 has two extensions to an action on blocking classes,

$$(3.4.4) \quad (R_{v_2})^*_{\mathbf{B}} = \begin{pmatrix} 4 & -15 \\ 1 & -4 \end{pmatrix}, \quad (R_{v_2})^*_{\mathbf{P}} = \begin{pmatrix} -59 & 90 \\ -20 & 31 \end{pmatrix}.$$

Here $(R_{v_2})^*_{\mathbf{B}}$, with determinant -1 , interchanges the blocking classes of $S^E = S(S^U)$ with those of S^L . Both of these families have $b_\infty < 1/3$ and their accumulation points limit to v_2 . On the other hand, $(R_{v_2})^*_{\mathbf{P}}$, with determinant -29 , fixes the degree components $(3, 2)$ of \mathbf{B}_0^U and takes the blocking classes of S^U to those of $(S^2R)^\sharp(S^U)$. Both of these staircases have $b_\infty > 1/3$ and lie on different sides of the interval blocked by the principal blocking class \mathbf{B}_0^U ; see Figure 3. Thus $(R_{v_2})^*_{\mathbf{B}} = (S)^*_{\mathbf{B}} \circ ((R)_{\mathbf{B}})^{-1}$, and one can check that $(R_{v_2})^*_{\mathbf{P}} = (SR)^*_{\mathbf{B}}$.

4 The prestaircases are live

This section completes the proof of [Theorem 1.2.6](#). We first show that the prestaircase families defined in [Section 3](#) are perfect prestaircase families. Establishing that all staircase classes are perfect implies that each such class $E_\kappa = (d_\kappa, m_\kappa, p_\kappa, q_\kappa, t_\kappa, \varepsilon)$ is live at its center p_κ/q_κ when $b = m_\kappa/d_\kappa$. Therefore, by [\(1.1.3\)](#), the capacity function $c_{H_{m_\kappa/d_\kappa}}$ takes the value $p_\kappa d_\kappa / (d_\kappa^2 - m_\kappa^2)$ at the point p_κ/q_κ , and this implies by continuity¹⁷ that the limiting value b_∞ is unobstructed, ie $c_{H_{b_\infty}}(a_\infty) = V_{b_\infty}(a_\infty)$, where $a_\infty = \lim(p_\kappa/q_\kappa)$.

What we have to show is that, at least for sufficiently large κ , the class E_κ remains live at the limiting b value $b_\infty := \lim m_\kappa/d_\kappa$. In [\[1\]](#) we established this in two steps, first showing that if the ratios m_κ/d_κ satisfy a bound such as that in [\(4.2.1\)](#) below then, by the positivity of the intersections of exceptional classes, the degree of any class E such that $\mu_{E, b_\infty}(p_\kappa/q_\kappa) \geq \mu_{E_\kappa, b_\infty}(p_\kappa/q_\kappa)$ is bounded above by a constant that is independent of κ . This means that there can be only finitely many such classes. In particular, there must be one class E_{ov} that dominates infinitely many of the steps. This is possible only if the obstruction $\mu_{E_{ov}, b_\infty}(z)$ given by this class goes through the accumulation point $(a_\infty, V_{b_\infty}(a_\infty))$ of the staircase.¹⁸ We call such a class an *overshadowing class* because its obstruction $\mu_{E_{ov}, b_\infty}(z)$ overshadows the staircase steps so that they cease to be visible at $b = b_\infty$.

In [\[1\]](#) we were able to find rather good bounds for the degree of a potentially overshadowing class, and hence could show that they do not exist by a case by case analysis. This method is not feasible here since the bounds on the degree of any overshadowing class of a staircase in the family $(S^i R^\delta)^\#(S^U)$ increase too rapidly with i . However it turns out that we can exploit the fact that the obstruction $\mu_{E_{ov}, b_\infty}(z)$ goes through the accumulation point to obtain powerful arithmetic information about the degree components d_{ov} and m_{ov} of E_{ov} , which is enough to rule out the existence of such a class. This argument hinges on the results of [Lemma 2.2.7](#), namely, that the following two functions are the same:

$$(4.0.1) \quad b \mapsto V_b(\text{acc}(b)), \quad b \mapsto \frac{1 + \text{acc}(b)}{3 - b}$$

¹⁷This holds because, for any quasiperfect class $E = (d, m, p, q)$, we have that $\mu_{E, b}(p/q) \leq V_b(p/q)(\sqrt{1 + 1/(d^2 - m^2)})$ and here $d_\kappa \rightarrow \infty$; see [\[1, Lemma 15\]](#).

¹⁸Indeed, we know that the accumulation point is unobstructed, so E_{ov} cannot be obstructive at this point, and if (for a descending staircase) the obstruction μ_{E_{ov}, b_∞} crossed the volume curve to the right of the accumulation point then it could at best overshadow only a finite number of steps.

4.1 The prestaircase classes are perfect

We first prove that all the classes in the families $T^\sharp(\mathcal{S}^U)$ are perfect, that is, they are exceptional classes, and then use this fact in Corollary 4.1.5 to gain information about the z -intervals that are blocked by the blocking classes.

We use the following recognition principle, which is explained for example in [6, Proposition 1.2.12].

Lemma 4.1.1 *An integral class*

$$E := dL - \sum_{i=1}^N n_i E_i$$

in the N -fold blowup $\mathbb{C}P^2 \# N\overline{\mathbb{C}P}^2$ represents an exceptional divisor if and only if it may be reduced to E_1 by repeated application of Cremona transformations.

Here, a Cremona transformation is a composition of the transformation

$$c_{xyz} \left(dL - \sum_{i=1} n_i E_i \right) = (d + \delta_{xyz})L - \sum_{i \in \{x,y,z\}} (n_i + \delta_{x,y,z}) E_i - \sum_{i \notin \{x,y,z\}} n_i E_i,$$

where $\delta_{xyz} = d - n_x - n_y - n_z$, and a reordering operation. Writing E in coordinates $(d; n_1, \dots, n_N)$, c_{xyz} adds δ_{xyz} to the coordinates d, n_x, n_y and n_z , and the reordering can reorder any of the n_i . Because Cremona moves are reversible, to verify E_k is exceptional, we just need to show it reduces to some other E_j that we know to be exceptional. We say E_k and E_j are Cremona equivalent if one can be reduced to the other.

Our staircase classes are quasiperfect and hence have the form $(d, m, W(p/q))$ where $W(p/q) = (W_1, \dots, W_N)$ is the integral weight expansion of p/q . We denote such a class by the tuple (d, m, p, q) , and consider the Cremona moves as acting on the corresponding sequence $(d, m, W_1, W_2, W_3, \dots)$. Thus, for example,

$$\begin{aligned} c_{012}((d, m, W_1, W_2, W_3, \dots)) \\ = (2d - m - W_1 - W_2, d - (W_1 + W_2), d - (m + W_2), d - (m + W_1), W_3, \dots). \end{aligned}$$

Here is the key lemma.

Lemma 4.1.2 *Suppose (d, m, p, q) with $p/q > 5$ satisfies $3d = m + p + q$, and that d and m are defined by (2.2.4). Then (d, m, p, q) is Cremona equivalent to $S^\sharp(d, m, p, q)$.*

Proof Let $S^\sharp(d, m, p, q, \varepsilon) = (D, M, P, Q, -\varepsilon)$. By Lemma 2.1.6, when $p/q = [5 + k, CF(x)]$ we have $S(p/q) = [5; 1, 4 + k, CF(x)]$. Thus, the integral weight expansion of $S(p/q)$ will always have five more integers than the integral weight expansion of p/q . As $S(p/q) = P/Q = (6p - q)/p$, we have that

$$\begin{aligned} W(P/Q) &= W((6p - q)/p) = (p^{\times 5}, p - q) \sqcup W((p - q)/q) \\ &= (Q^{\times 5}, P - 5Q) \sqcup W((p - q)/q), \end{aligned}$$

where $W((p - q)/q)$ by definition is $W(p/q)$ with the first entry q removed. Thus, to reduce the class (D, M, P, Q) to the class (d, m, p, q) , it suffices to show that we can reduce

$$(D, M, Q^{\times 5}, P - 5Q) \sqcup W((p - q)/q) \quad \text{to} \quad (d, m, q) \sqcup W((p - q)/q).$$

Note that in this reduction we must get rid of five terms because as mentioned $W(S(p/q))$ has five more terms than $W(p/q)$.

Next, observe that

$$\begin{aligned} c_{256}c_{234}c_{016}c_{345}c_{012}(D; Q^{\times 5}, M, P - 5Q) \\ = (8D - 3(M + P); 0^{\times 2}, 3D - M - 2P + 5Q, 0^{\times 3}, 3D - 2M - P). \end{aligned}$$

This can be seen by direct computation, where the zeros come from the linear Diophantine condition $3D - M - P - Q = 0$. The first three steps give the two zeros in positions 1 and 2, the fourth step results in the two zeros in positions 4 and 5, and the fifth step results in one zero in position 6. Furthermore, since $P = 6p - q$ and $Q = p$, it follows from (2.2.4) that

$$D = \frac{1}{8}(3(7p - q) - \varepsilon t), \quad M = \frac{1}{8}((7p - q) - 3\varepsilon t).$$

Performing these substitutions, we get

$$\begin{aligned} 8D - 3(M + P) &= \frac{1}{8}(3(p + q) + \varepsilon t) = d, \\ 3D - M - 2P + 5Q &= q, \\ 3D - 2M - P &= \frac{1}{8}(p + q + 3\varepsilon t) = m. \end{aligned}$$

We conclude that

$$c_{256}c_{234}c_{016}c_{345}c_{012}(D; Q^{\times 5}, M, P - 5Q) = (d; 0^{\times 2}, q, 0^{\times 3}, m).$$

Thus these five Cremona moves and an appropriate reordering reduces

$$(D, M, Q^{\times 5}, P - 5Q) \sqcup W((p - q)/q) \quad \text{to} \quad (d, m, q) \sqcup W((p - q)/q).$$

Hence, the class (D, M, P, Q) is Cremona equivalent to (d, m, p, q) , as claimed. \square

In [1], it was shown that both S^U and S^L are perfect, so it is enough to show that $S^\#$ preserves Cremona equivalence, but there is an equally nice argument that $R^\#$ preserves Cremona equivalence.

Lemma 4.1.3 *Suppose (d, m, p, q) are such that d and m are defined by (2.2.4), and $p/q > 7$. Then (d, m, p, q) is Cremona equivalent to $R^\#(d, m, p, q)$.*

Proof Let $R^\#(d, m, p, q) = (D, M, P, Q)$. Assume $p/q = [6 + k; CF(x)]$ for some $k \geq 1$ and $x \geq 1$. Then

$$W(p/q) = (q^{6+k}, p - (6 + k)q, \dots).$$

By Lemma 2.1.6(ii), $R(p/q) = (6p - 35q)/(p - 6q) = [6, k, CF(x)]$, and thus

$$W(P/Q) = (Q^{\times 6}, P - 6Q^{\times k}, Q - k(P - 6Q), \dots) = (p - 6q^{\times 6}, q^{\times k}, p - (6 + k)q, \dots).$$

Only the first 6 terms of the weight expansions $W(P/Q)$ and $W(p/q)$ differ. Thus, to show that (D, M, P, Q) is Cremona equivalent to (d, m, p, q) , we need to consider the degree coordinates and the first 6 terms of the weight sequence for each.

We use the notation \mapsto_{ijk} to represent applying c_{ijk} to the previous tuple. Applying two Cremona moves to (D, M, P, Q) gives

$$(4.1.1) \quad \begin{aligned} (D, M, Q^{\times 6}) &\mapsto_{123} (2D - 3Q, M, D - 2Q^{\times 3}, Q^{\times 3}) \\ &\mapsto_{456} (4D - 9Q, M, D - 2Q^{\times 3}, 2D - 5Q^{\times 3}). \end{aligned}$$

Applying three Cremona moves to (d, m, p, q) gives:

$$\begin{aligned} (d, m, q^{\times 6}) &\mapsto_{012} (2d - m - 2q, d - 2q, d - m - q^{\times 2}, q^{\times 4}) \\ &\mapsto_{034} (3d - 2(m + 2q), 2d - m - 4q, d - m - q^{\times 4}, q^{\times 2}) \\ &\mapsto_{156} (5d - 3(m + 3q), 2d - m - 4q, 3d - 2(m + 3q), d - m - q^{\times 3}, 2d - m - 4q^{\times 2}), \end{aligned}$$

which we can reorder to get

$$(4.1.2) \quad (5d - 3(m + 3q), 3d - 2(m + 3q), 2d - m - 4q^{\times 3}, d - m - q^{\times 3}).$$

We claim this reordered tuple is precisely (4.1.1). To see this, we write each term in each of the tuples in terms of p and q . We are going to assume that for (d, m, p, q) , $\varepsilon = 1$. By (2.2.4), Lemma 2.1.6(iii), and the definition of R , we can write (d, m, p, q) and (D, M, P, Q) completely in terms of p and q ,

$$\begin{aligned} (d, m, p, q) &= \left(\frac{1}{8}(3(p + q) + t), \frac{1}{8}(p + q + 3t), p, q\right), \\ (D, M, P, Q) &= \left(\frac{1}{8}(21p - 123q - t), \frac{1}{8}(7p - 41q - 3t), 6p - 35q, p - 6q\right). \end{aligned}$$

Now, we expand all of the entries in (4.1.2) and (4.1.1) in terms of p and q to get the equalities

$$\begin{aligned} 5d - 3(m + 3q) &= \frac{1}{2}(3p - 15q - t) = 4D - 9Q, \\ 3d - 2(m + 3q) &= \frac{1}{8}(7p - 41q - 3t) = M, \\ 2d - m - 4q &= \frac{1}{8}(5p - 27q - t) = D - 2Q, \\ d - m - q &= \frac{1}{4}(p - 3q - t) = 2D - 5Q. \end{aligned} \quad \square$$

Proposition 4.1.4 For each $T \in \mathcal{G}$, the classes in the prestaircase family $T^\#(S^U)$ are perfect, that is, they are exceptional classes.

Proof By [1, Section 3.4], this holds for S^U and S^L . Hence this is an immediate consequence of Lemma 4.1.2. □

Here is a typical corollary. For simplicity we only consider the principal blocking classes mentioned in Lemma 3.4.6, but a similar argument applies to all blocking classes that have associated perfect staircases.

Corollary 4.1.5 For each $i \geq 0$, the z -interval blocked by the principal blocking class $(S^i)^\#(B_0^U)$ contains $[w_{i+2}, w_{i+1}]$.

Proof It follows from Remark 2.3.4(i) and Theorem 1.2.4 that the prestaircases $(S^{i+1}R)^\#(S_0^U)$, $(S^i)^\#(S_0^U)$ are associated to $B = (S^i)^\#(B_0^U)$. Because they consist of perfect classes, we know from [1, Lemma 27 and Theorem 52] that their z -limit points $\alpha_{B,\ell}$ and $\alpha_{B,u}$ are unobstructed and that the interval $(\alpha_{B,\ell}, \alpha_{B,u})$ lying between them is precisely the z -interval blocked by B . Thus it suffices to show that

$$\alpha_{B,\ell} < w_{i+2} < w_{i+1} < \alpha_{B,u}.$$

Since S preserves order and takes w_i to w_{i+1} for all i , we only need check this for $i = 0$, and this can be done either by direct evaluation or by comparing the continued fraction expansions of these quantities. □

4.2 Recognizing staircases

We proved the following staircase recognition theorem in [1, Theorem 51].

Theorem 4.2.1 Let $S = (E_\kappa)$ be a perfect prestaircase, let λ be as in Lemma 3.1.4, and denote by D, M, P, Q the constants X defined by (3.1.6), where $x_\kappa = d_\kappa, m_\kappa, p_\kappa, q_\kappa$ respectively. Suppose in addition that at least one of the following conditions holds:

(i) There is an $r/s > 0$ such that $M/D < r/s$,

$$\frac{m_\kappa^2 - 1}{d_\kappa m_\kappa} < \frac{M}{D} < \frac{s + m_\kappa(rd_\kappa - sm_\kappa)}{r + d_\kappa(rd_\kappa - sm_\kappa)} \quad \text{for all } \kappa \geq \kappa_0,$$

and there is no overshadowing class at $(z, b) = (P/Q, M/D)$ of degree $d' < s/(r - sb_\infty)$ and with $m'/d' > r/s$.

(ii) There is an $r/s > 0$ such that $M/D > r/s$,

$$\frac{m_\kappa(sm_\kappa - rd_\kappa) - s}{d_\kappa(sm_\kappa - rd_\kappa) - r} < \frac{M}{D} < \frac{m_\kappa}{d_\kappa} \quad \text{for all } \kappa \geq \kappa_0,$$

and there is no overshadowing class at $(z, b) = (P/Q, M/D)$ of degree $d' < s/(sb_\infty - r)$ and with $m'/d' < r/s$.

Then \mathcal{S} is live, and it is a staircase for $H_{M/D}$ that accumulates at P/Q .

The next result states the estimates that we must establish in order to apply the above theorem.

Lemma 4.2.2 Consider a prestaircase with classes $(d_\kappa, m_\kappa; q_\kappa \mathbf{w}(p_\kappa/q_\kappa))$, where the ratios $b_\kappa := m_\kappa/d_\kappa$ have limit b_∞ , and let the constants D, D', D'', M, M', M'' and σ be as in (3.1.6), with $x_\kappa = d_\kappa, m_\kappa$ respectively.

(i) Suppose that $M\bar{D} - \bar{M}D \neq 0$. Then the b_κ are strictly increasing if and only if

$$M\bar{D} - \bar{M}D = 2\sqrt{\sigma}(M''D' - M'D'') = \frac{m_1d_0 - m_0d_1}{\sqrt{\sigma}} > 0,$$

and otherwise they are strictly decreasing.

(ii) If $m_1d_0 - d_1m_0 > 0, b_\infty < r/s \leq 1$, and

$$(4.2.1) \quad |m_1d_0 - d_1m_0| \leq \sqrt{\sigma} \frac{sD - rM}{|rD - sM|},$$

then there is a κ_0 such that

$$\frac{m_\kappa}{d_\kappa} \leq b_\infty = \frac{M}{D} \leq \frac{s + m_\kappa(rd_\kappa - sm_\kappa)}{r + d_\kappa(rd_\kappa - sm_\kappa)} \quad \text{for } \kappa \geq \kappa_0.$$

(iii) If $m_1d_0 - d_1m_0 < 0, b_\infty > r/s > 0$, and (4.2.1) holds, then there is a κ_0 such that

$$\frac{m_\kappa(sm_\kappa - rd_\kappa) - s}{d_\kappa(sm_\kappa - rd_\kappa) - r} \leq b_\infty = \frac{M}{D} \leq \frac{m_\kappa}{d_\kappa} \quad \text{for } \kappa \geq \kappa_0.$$

Proof This is a reformulation of [1, Lemma 67] in which (i) incorporates the calculation

$$M''D' - M'D'' = \frac{m_1d_0 - d_1m_0}{2\sigma}$$

that follows from (3.1.6). □

In [1, Section 3], an r/s was carefully chosen in order to reduce the number of potential overshadowing classes that needed to be ruled out. For our purposes, we simply need to know an r/s exists since we use an arithmetic argument to rule out overshadowing classes. The following corollary shows that some r/s does indeed exist.

Corollary 4.2.3 *There exists r/s such that either condition (ii) or (iii) in Lemma 4.2.2 is satisfied.*

Proof For condition (ii), assume $m_1d_0 - d_1m_0$ is positive. As r/s approaches b_∞ from the right,

$$\frac{sD - rM}{|rD - sM|} = \frac{1 - (r/s)b_\infty}{r/s - b_\infty}$$

approaches infinity. Hence, there always exists some $b_\infty < r/s$ such that condition (ii) is satisfied. A similar argument applies for condition (iii). □

Remark 4.2.4 The statement in Lemma 4.2.2 is not identical to that in [1, Lemma 67] because we have changed notation, now indexing by κ rather than by k . We now explain the relation between the two statements. In [1, Remark 68(i)], we calculated that

$$M''D' - M'D'' = \frac{1}{2(2n + 3)\sqrt{\sigma}}(m_1d_0 - m_0d_1).$$

However, there the staircases were divided into two parts according to the different endings, and (as in Lemma 3.1.1) the recursion parameter was $\sigma + 2$, where

$$\sigma = (2n + 1)(2n + 5) = v^2 - 4.$$

This reformulation should not change the limiting values M and D , and hence M' , M'' , D' , and D'' . However in our current notation $m_{k=1}$ and $d_{k=1}$ are denoted by $m_{\kappa=2}$ and $d_{\kappa=2}$, since the staircases in [1] are indexed by k while in the current paper we combine the two strands and index via κ , which is (approximately) $2k$. Thus if we index by κ and take $2n + 3 = v$ we have

$$m_2d_0 - m_0d_2 = (vm_1 - m_0)d_0 - m_0(vd_1 - d_0) = v(m_1d_0 - m_0d_1),$$

$$\frac{1}{2(2n + 3)\sqrt{\sigma}}(m_2d_0 - m_0d_2) = \frac{1}{2\sqrt{\sigma}}(m_1d_0 - m_0d_1),$$

which is consistent with the identities in Lemma 4.2.2.

By Corollary 4.2.3, we do not need to estimate r/s by computing $m_1d_0 - d_1m_0$. However, Lemma 4.2.2(i) shows that the sign of $m_1d_0 - d_1m_0$ determines whether the terms $b_\kappa = m_\kappa/d_\kappa$ strictly increase or decrease for a prestaircase, which is important for the overshadowing argument.

While the quantities m_κ and d_κ do depend on both i and n , we suppress those indices for simplicity.

- Lemma 4.2.5** (i) For $S^i(S_{\ell,n}^U)$, $m_1d_0 - m_0d_1 = -\varepsilon(2ny_i + y_{i+1})$, where $\varepsilon = (-1)^i$.
 (ii) For $S^i(S_{u,n}^U)$, $m_1d_0 - m_0d_1 = -\varepsilon(2ny_{i+1} + y_{i+2})$, where $\varepsilon = (-1)^i$.
 (iii) For $S^i(S_{\ell,n}^L)$, $m_1d_0 - m_0d_1 = -\varepsilon(2ny_{i+1} + y_i)$, where $\varepsilon = (-1)^{i+1}$.
 (iv) For $S^i(S_{u,n}^L)$, $m_1d_0 - m_0d_1 = -\varepsilon(2ny_{i+2} + y_{i+1})$, where $\varepsilon = (-1)^{i+1}$.

Proof Let \bullet be one of (U, ℓ) , (U, u) , (L, ℓ) , or (L, u) . For two seeds $(p_0, q_0, t_0, \varepsilon)$ and $(p_1, q_1, t_1, \varepsilon)$ of $S^i(S_{\bullet,n})$, the degree formulas from (2.2.4) imply that

$$(4.2.2) \quad m_1d_0 - m_0d_1 = \frac{1}{8}\varepsilon((p_0 + q_0)t_1 - (p_1 + q_1)t_0).$$

Denote the seeds of the initial staircase in S^U or S^L by $(p_0^\bullet, q_0^\bullet, t_0^\bullet)$ and $(p_1^\bullet, q_1^\bullet, t_1^\bullet)$, and note that

$$S^i = \begin{pmatrix} y_{i+1} & -y_i \\ y_i & -y_{i-1} \end{pmatrix},$$

where $y_{i-1} = 6y_i - y_{i+1}$. Using the invariance of t_0^\bullet and t_1^\bullet under S , we obtain that

$$m_1d_0 - m_0d_1 = \frac{1}{8}\varepsilon((t_1^\bullet(p_0^\bullet + q_0^\bullet) - t_0^\bullet(p_1^\bullet + q_1^\bullet))y_{i+1} + (t_1^\bullet(p_0^\bullet - q_0^\bullet) - t_0^\bullet(p_1^\bullet - q_1^\bullet)) + 6(t_0^\bullet q_1^\bullet - t_1^\bullet q_0^\bullet)y_i).$$

Then, by substituting the relevant formulas for the seeds from Lemma 3.2.1 and Lemma 3.2.5, we get the desired results.

For example, for (ii), we have

$$(4.2.3) \quad (p_{0,u}^U, q_{0,u}^U, t_{0,u}^U) = (-5, -1, 2), \quad (p_{1,u}^U, q_{1,u}^U, t_{1,u}^U) = (2n + 8, 1, 2n + 5)$$

giving

$$m_1d_0 - m_0d_1 = -\varepsilon((2n + 6)y_{i+1} - y_i) = 2ny_{i+1} + y_{i+2},$$

as claimed. □

Corollary 4.2.6 For a prestaircase $T^\sharp(S^U)$, if $b_\infty > 1/3$, then m_κ/d_κ strictly decreases. If $b_\infty < 1/3$, then m_κ/d_κ strictly increases.

Proof Lemma 4.2.5 implies that $m_1d_0 - m_0d_1$ always has the opposite sign to ε . As ε is positive if and only if $b_\infty > 1/3$, it follows that $m_1d_0 - m_0d_1 < 0$ if and only if $b_\infty > 1/3$. Now use Lemma 4.2.2(i). \square

Finally, we prove an estimate that will be useful below.

Lemma 4.2.7 *Let \mathcal{S} be one of the descending stairs $S^i(S_{u,n}^U)$ for all $i, n \geq 0$ except $(i, n) = (0, 0)$ or one of the stairs $S^i(S_{u,n}^L)$ for all $i \geq 0$ and $n \geq 1$. Denote the steps of \mathcal{S} by $(d_\kappa, m_\kappa, p_\kappa, q_\kappa, t_\kappa, \varepsilon)$ for $\kappa \geq 0$, and let $b_\infty = \lim m_\kappa/d_\kappa$. Then there is a κ_0 such that*

$$(4.2.4) \quad \frac{p_\kappa q_\kappa}{p_\kappa + q_\kappa} > \frac{d_\kappa - m_\kappa b_\infty}{3 - b_\infty}, \quad \kappa \geq \kappa_0.$$

Proof Since $(3d_\kappa - m_\kappa)p_\kappa q_\kappa = (p_\kappa + q_\kappa)(d_\kappa^2 - m_\kappa^2 + 1)$, we have

$$\begin{aligned} \frac{p_\kappa + q_\kappa}{d_\kappa} &= \left(3 - \frac{m_\kappa}{d_\kappa}\right) p_\kappa q_\kappa - (p_\kappa + q_\kappa) \left(d_\kappa - \frac{m_\kappa^2}{d_\kappa}\right) \\ &= (3 - b_\infty) p_\kappa q_\kappa - (p_\kappa + q_\kappa) (d_\kappa - m_\kappa b_\infty) - \left(\frac{m_\kappa}{d_\kappa} - b_\infty\right) (p_\kappa q_\kappa - m_\kappa (p_\kappa + q_\kappa)). \end{aligned}$$

The inequality (4.2.4) is equivalent to the claim that the sum of the first two terms on the right-hand side is positive. Therefore (4.2.4) will hold provided that

$$\frac{p_\kappa + q_\kappa}{d_\kappa} > \left(\frac{m_\kappa}{d_\kappa} - b_\infty\right) (m_\kappa (p_\kappa + q_\kappa) - p_\kappa q_\kappa), \quad \kappa \geq \kappa_0.$$

Since

$$m_\kappa (p_\kappa + q_\kappa) - p_\kappa q_\kappa = m_\kappa (3d_\kappa - m_\kappa) - (d_\kappa^2 - m_\kappa^2 + 1),$$

it suffices to show that

$$\frac{3d_\kappa - m_\kappa}{d_\kappa} > \left(\frac{m_\kappa}{d_\kappa} - b_\infty\right) (m_\kappa (3d_\kappa - m_\kappa) - (d_\kappa^2 - m_\kappa^2 + 1)), \quad \kappa \geq \kappa_0.$$

Since by Lemma 3.1.4 we have $m_\kappa = M\lambda^\kappa + \bar{M}\lambda^{-\kappa}$ and $d_\kappa = D\lambda^\kappa + \bar{D}\lambda^{-\kappa}$ for some $\lambda > 1$, one easily checks that this will hold if

$$(4.2.5) \quad \frac{3D - M}{|3M - D|} > |\bar{M}D - \bar{D}M| = \frac{|m_1d_0 - d_1m_0|}{\sqrt{\sigma}}, \quad \sigma = \nu^2 - 4,$$

where ν is the recursion parameter of \mathcal{S} .

Claim 1 Equation (4.2.5) holds when $\mathcal{S} = (S^i)^\#(S_{u,n}^U)$ for all $i, n \geq 0$ except $i = n = 0$.

Proof First consider the case $\mathcal{S} = S_{u,n}^U$. When $X = P, Q, T$, Lemma 3.1.4 implies that $X = (1/2\sqrt{\sigma})(2x_1 - x_0(\nu - \sqrt{\sigma}))$, where $\nu = 2n + 3$ is the recursion variable,

and $\sigma = v^2 - 4 = (2n + 1)(2n + 5)$. The seeds for this staircase are given in (4.2.3). Since $v - \sqrt{\sigma} > 0$ and the seed class has $p_0 + q_0 < 0$ while $t_0 > 0$, we have

$$\begin{aligned} \frac{3D - M}{|3M - D|} &= \frac{P + Q}{T} = \frac{2(p_1 + q_1) - (p_0 + q_0)(v - \sqrt{\sigma})}{2t_1 - t_0(v - \sqrt{\sigma})} \\ &> \frac{p_1 + q_1}{t_1} = \frac{2n + 9}{2n + 5} \\ &> \frac{m_0d_1 - m_1d_0}{\sqrt{\sigma}} = \frac{2n + 6}{\sqrt{(2n + 1)(2n + 5)}} \end{aligned}$$

provided that $(2n + 9)^2(2n + 1) > (2n + 6)^2(2n + 5)$. But this holds unless $n = 0, 1, 2, 3$. If $n = 1, 2, 3$ then one can check by direct calculation that (4.2.5) holds. Moreover, although it does not hold when $n = 0$ one can check that in this case we have

$$(4.2.6) \quad \frac{3D - M}{|3M - D|} > \frac{34}{35} \frac{m_0d_1 - m_1d_0}{\sqrt{\sigma}} = \frac{y_4}{y_2y_3} \frac{6}{\sqrt{\sigma}}.$$

Now suppose that $\mathcal{S} = (S^i)^\#(S^U_{u,n})$. Notice that σ and v are invariant under the shift. The quantity $|3M - D|$ is also invariant under the shift since $3M - D$ is the limit of $(3m_k - d_k)\lambda^{-k} = \varepsilon t_k \lambda^{-k}$. To consider how much the right-hand side of (4.2.5) increases under iterations of the shift, we again use Lemma 4.2.5. Because $y_{i+2} < 6y_{i+1}$ for all $i > 0$, we can estimate

$$\frac{2ny_{i+1} + y_{i+2}}{2ny_1 + y_2} = \frac{2ny_{i+1} + y_{i+2}}{2n + 6} < y_{i+1} \quad \text{for all } i > 0, n \geq 0.$$

Therefore the result will hold in the case $n > 0$ if we show that when we apply S^i the quantity $3D - M$ increases by a factor of at least y_{i+1} . But by Lemma 2.1.3, $3D - M = P + Q$ is taken by S^i to $(y_{i+1} + y_i)P - (y_i + y_{i-1})Q$. Therefore we need $(y_{i+1} + y_i)P - (y_i + y_{i-1})Q > y_{i+1}(P + Q)$, or equivalently

$$y_i P > (y_{i+1} + y_i + y_{i-1})Q = 7y_i Q.$$

But this holds because P/Q is the accumulation point of a staircase in S^U and so satisfies $P/Q > 7$. Indeed $w_1 = 7$ is blocked by B_0^U by Corollary 4.1.5.

In the case $n = 0$, the right-hand side of (4.2.5) increases by the factor $y_{i+2}/6$. Therefore it suffices to show that when we apply S^i the quantity $3D - M$ increases by a factor of at least $(35/34)(y_{i+2}/6)$. Since $(35/34)(y_{i+2}/6) \leq y_{i+1}$ when $i \geq 2$, the previous argument applies to show that this holds. In the case $i = 1$, we must check that

$$7P - Q > \frac{35}{34} \frac{35}{6}(P + Q),$$

which holds because $P/Q = [7; \{5, 1\}^\infty] > [7; 6] = 43/6$. □

Claim 2 Equation (4.2.5) holds when $\mathcal{S} = (S^i)^\#(S_{u,n}^L)$, $n \geq 1$ and $i \geq 0$.

Proof Let $\mathcal{S} = S_{u,n}^L$. The initial seeds are now $(p_{0,u}^L, q_{0,u}^L, t_{0,u}^L) = (-29, -5, 2)$ and $(p_{1,u}^L, q_{1,u}^L, t_{1,u}^L) = (12n - 11, 2n - 2, 2n + 1)$. Further $v = 2n + 3$ so that $\sigma = \sqrt{(2n + 1)(2n + 5)}$ as before. If we simplify the inequality as before and use the result in Lemma 4.2.5(iv), it follows that it suffices to check that

$$\frac{p_1 + q_1}{t_1} > \frac{m_1 d_0 - m_0 d_1}{\sqrt{\sigma}} = \frac{2ny_2 + y_1}{\sqrt{\sigma}} = \frac{12n + 1}{\sqrt{\sigma}}.$$

But this holds for all $n \geq 4$. For the cases $n = 1, 2, 3$, we directly compute both sides of (4.2.5) to verify the inequality holds.

Next consider the case $i > 0$. We argue as before, noting that for each n the accumulation point P/Q is larger than the center of the blocking class B_n^L , which is $(12n + 1)/2n$ by Theorem 2.3.3. By Lemma 4.2.5(iv), when we apply S^i the right-hand side of (4.2.5) grows by the factor

$$\frac{2ny_{i+2} + y_{i+1}}{2ny_2 + y_1} = \frac{2ny_{i+2} + y_{i+1}}{12n + 1}, \quad n \geq 1.$$

Therefore it suffices to check that

$$\begin{aligned} (y_{i+1} + y_i)P - (y_i + y_{i-1})Q &> \frac{2ny_{i+2} + y_{i+1}}{12n + 1}(P + Q) \\ &= \left(y_{i+1} - \frac{2n}{12n + 1}y_i\right)(P + Q). \end{aligned}$$

Because $y_{i+1} + y_i + y_{i-1} = 7y_i$, this simplifies to

$$\frac{14n + 1}{12n + 1}y_i P > y_i \left(7 - \frac{2n}{12n + 1}\right)Q.$$

Thus we need $P/Q > (82n + 7)/(14n + 1)$. But this holds for all $n \geq 1$ since $P/Q > (12n + 1)/2n$. □

The proof of the lemma is now complete. □

4.3 Overshadowing classes

To utilize the staircase recognition Theorem 4.2.1, it remains to show there are no overshadowing classes. Here, we not only show that there are no overshadowing classes for $S^i R^\delta(S^U)$, but prove a more general result about overshadowing classes. Namely, given a general perfect prestaircase family with recursion parameter ≥ 3 , if $b_\infty < 1/3$

and m_κ/d_κ strictly increases (resp. $b_\infty > 1/3$ and m_κ/d_κ strictly decreases), then the staircase family is live provided only that the staircase is not overshadowed by the obstruction from the exceptional class $E_1 := 3L - E_0 - 2E_1 - E_{2\dots 6}$. Recall from Remark 2.3.8(ii) and Lemma 2.2.7 that when $b \in (1/5, 5/11)$ this obstruction is given by the formula $z \mapsto (1+z)/(3-b)$, and hence always passes through the accumulation point $(\text{acc}(b), V_b(\text{acc}(b)))$. Hence, it could overshadow the descending staircases for these b . Since the S^U staircases have $b > 5/11$ while the S^L staircases have $b < 1/5$, we only need be concerned about their images under a shift S^i . The next lemma shows that these staircases are not overshadowed in this way, because the slope of any overshadowing class must be greater than the slope of the obstruction $\mu_{E_1, b}$.

Lemma 4.3.1 *Let S be any descending prestaircase in one of the families $(S^i R^\delta)^\#(S^U)$ where $\delta \in \{0, 1\}$ and $i > 0$ or $i = 0$ and $n \geq 1$. Then the slope of an overshadowing class must be larger than $1/(3 - b_\infty)$.*

Proof The slope of an overshadowing class must be larger than the slope of the line segments from the accumulation point (z_∞, b_∞) to the outer corners

$$\left(\frac{p_\kappa}{q_\kappa}, \frac{p_\kappa}{d_\kappa - m_\kappa b_\infty} \right)$$

of the staircase. Therefore we must check that

$$\frac{p_\kappa/(d_\kappa - m_\kappa b_\infty) - (1 + z_\infty)/(3 - b_\infty)}{p_\kappa/q_\kappa - z_\infty} > \frac{1}{3 - b_\infty}.$$

When we simplify z_∞ cancels from the inequality, and we get

$$(3 - b_\infty)p_\kappa q_\kappa > (p_\kappa + q_\kappa)(d_\kappa - m_\kappa b_\infty),$$

which was proved in Lemma 4.2.7. □

We now use an arithmetic argument to rule out the existence of any other overshadowing classes. The following lemma establishes properties about the common divisors of (d, m, p, q) needed for the argument.

Lemma 4.3.2 *Given a quasiperfect class (d, m, p, q) , assume there is an integer k such that km/p and kd/p (resp. km/q and kd/q) are both integers. Then $p|k$ (resp. $q|k$).*

Proof Assume km/p and kd/p are both integers. Since $g := \text{gcd}(m, d)$ is an integral combination of m and d , this implies $kg/p \in \mathbb{Z}$. But because $q = 3d - p - m$ and $\text{gcd}(p, q) = 1$, we must have $\text{gcd}(p, d, m) = \text{gcd}(p, g) = 1$. Therefore $p|k$ and, similarly, $q|k$. □

Suppose that at (z_∞, b_∞) there is an overshadowing class $E_{\text{ov}} = (d_{\text{ov}}, m_{\text{ov}}, \mathbf{m})$ for some prestaircase. Recall from Corollary 3.1.5 that the limit points z_∞ and b_∞ are irrational. By [6, Proposition 2.3.2], there are integers $A \geq 0$ and $C \geq 0$ such that

$$\mu_{E_{\text{ov}},b}(z) = \frac{A + Cz}{d_{\text{ov}} - m_{\text{ov}}b}, \quad z \approx z_\infty.$$

Lemma 4.3.3 *Let \mathcal{S} be a descending prestaircase with irrational accumulation point z_∞ that is associated to a perfect blocking class $\mathbf{B} = (d_{\mathbf{B}}, m_{\mathbf{B}}, p_{\mathbf{B}}, q_{\mathbf{B}}, t_{\mathbf{B}})$. Suppose that $E_{\text{ov}} = (d_{\text{ov}}, m_{\text{ov}}, \mathbf{m})$ is an overshadowing class, and denote by*

$$\mu_{E_{\text{ov}},b}(z) = \frac{A + Cz}{d_{\text{ov}} - m_{\text{ov}}b}$$

the corresponding obstruction. Assume the slope $C/(d_{\text{ov}} - m_{\text{ov}}b_\infty)$ of $\mu_{E_{\text{ov}},b_\infty}(z)$ is $> 1/(3 - b_\infty)$. Then there is a positive integer k such that

- (i) *If $b_\infty > 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} > 1/3$, then $m_{\text{ov}}/d_{\text{ov}} < 1/3$ and $d_{\text{ov}} - 3m_{\text{ov}} = kt_{\mathbf{B}}$.*
- (ii) *If $b_\infty < 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} < 1/3$, then $m_{\text{ov}}/d_{\text{ov}} > 1/3$ and $3m_{\text{ov}} - d_{\text{ov}} = kt_{\mathbf{B}}$.*

Proof Note first that the function $\mu_{E_{\text{ov}},b_\infty}$ must obstruct an interval $(z_\infty, z_\infty + \varepsilon)$ to the right of the limit point and hence have break point $a_{\text{ov}} > z_\infty$. Further, the condition on the slope implies that $C > A$.

Next note that $V_{b_\infty}(z_\infty)$ is given by the expressions

$$\frac{1 + z_\infty}{3 - b_\infty} = \frac{p_{\mathbf{B}}}{d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty} = \frac{A + Cz_\infty}{d_{\text{ov}} - m_{\text{ov}}b_\infty},$$

where the first equality holds by (4.0.1) and the second holds by [1, Lemma 16]. The first equality implies

$$\begin{aligned} (1 + z_\infty)(d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty) &= p_{\mathbf{B}}(3 - b_\infty) \\ \Rightarrow z_\infty &= \frac{p_{\mathbf{B}}(3 - b_\infty) - (d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty)}{d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty} = \frac{b_\infty(m_{\mathbf{B}} - p_{\mathbf{B}}) + 3p_{\mathbf{B}} - d_{\mathbf{B}}}{d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty}, \end{aligned}$$

while the second gives

$$\begin{aligned} (A + Cz_\infty)(d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty) &= p_{\mathbf{B}}(d_{\text{ov}} - m_{\text{ov}}b_\infty) \\ \Rightarrow z_\infty &= \frac{p_{\mathbf{B}}(d_{\text{ov}} - b_\infty m_{\text{ov}}) - A(d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty)}{C(d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty)} = \frac{b_\infty(Am_{\mathbf{B}} - pm_{\text{ov}}) + p_{\mathbf{B}}d_{\text{ov}} - Ad_{\mathbf{B}}}{C(d_{\mathbf{B}} - m_{\mathbf{B}}b_\infty)}. \end{aligned}$$

Therefore,

$$b_\infty(Am_{\mathbf{B}} - p_{\mathbf{B}}m_{\text{ov}}) + p_{\mathbf{B}}d_{\text{ov}} - Ad_{\mathbf{B}} = C(b_\infty(m_{\mathbf{B}} - p_{\mathbf{B}}) + 3p_{\mathbf{B}} - d_{\mathbf{B}}).$$

All quantities here are integers except for b_∞ which is irrational because z_∞ is. Therefore the coefficients of b_∞ must be equal. Thus we have

$$Am_{\mathbf{B}} - p_{\mathbf{B}}m_{\text{ov}} = C(m_{\mathbf{B}} - p_{\mathbf{B}}), \quad p_{\mathbf{B}}d_{\text{ov}} - Ad_{\mathbf{B}} = C(3p_{\mathbf{B}} - d_{\mathbf{B}}).$$

We can solve these equations for m_{ov} and d_{ov} to get

$$(4.3.1) \quad \begin{aligned} m_{\text{ov}} &= \frac{(A - C)m_{\mathbf{B}} + Cp_{\mathbf{B}}}{p_{\mathbf{B}}} = C + (A - C)m_{\mathbf{B}}/p_{\mathbf{B}}, \\ d_{\text{ov}} &= \frac{3p_{\mathbf{B}}C + (A - C)d_{\mathbf{B}}}{p_{\mathbf{B}}} = 3C + (A - C)d_{\mathbf{B}}/p_{\mathbf{B}}. \end{aligned}$$

As m_{ov} and d_{ov} are integers, $(A - C)m_{\mathbf{B}}/p_{\mathbf{B}}$ and $(A - C)d_{\mathbf{B}}/p_{\mathbf{B}}$ are both integers. Then, by Lemma 4.3.2, $p_{\mathbf{B}} \mid (A - C)$. This proves $k = (C - A)/p_{\mathbf{B}}$ is an integer.

Since $C - A > 0$, we also have $k > 0$. Since $m_{\mathbf{B}}/d_{\mathbf{B}} > 1/3$ by assumption, the formulas in (4.3.1) then imply that $3m_{\text{ov}} < d_{\text{ov}}$.

We can compute

$$d_{\text{ov}} - 3m_{\text{ov}} = 3C - kd_{\mathbf{B}} - 3C + 3km_{\mathbf{B}} = k(3m_{\mathbf{B}} - d_{\mathbf{B}}) = kt_{\mathbf{B}}.$$

This proves (i). The proof for (ii) follows similarly. □

We prove a similar result for ascending staircases.

Lemma 4.3.4 *Let \mathcal{S} be an ascending prestaircase with irrational accumulation point z_∞ that is associated to a perfect blocking class $\mathbf{B} = (d_{\mathbf{B}}, m_{\mathbf{B}}, p_{\mathbf{B}}, q_{\mathbf{B}}, t_{\mathbf{B}})$. Suppose that $\mathbf{E}_{\text{ov}} = (d_{\text{ov}}, m_{\text{ov}}, \mathbf{m})$ is an overshadowing class, and denote by*

$$\mu_{\mathbf{E}_{\text{ov}}, b}(z) = \frac{A + Cz}{d_{\text{ov}} - m_{\text{ov}}b}$$

the corresponding obstruction. Then there is a positive integer k such that

- (i) *If $b_\infty > 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} > 1/3$, then $m_{\text{ov}}/d_{\text{ov}} < 1/3$ and $d_{\text{ov}} - 3m_{\text{ov}} = kt_{\mathbf{B}}$.*
- (ii) *If $b_\infty < 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} < 1/3$, then $m_{\text{ov}}/d_{\text{ov}} > 1/3$ and $3m_{\text{ov}} - d_{\text{ov}} = kt_{\mathbf{B}}$.*

Proof Notice first that we must have $A - C > 0$ since if $A \leq C$ the slope of the obstruction $\mu_{\mathbf{E}_{\text{ov}}, b_\infty}$ is at least that of the line $z \mapsto (1 + z)/(3 - b_\infty)$ which is $< V_{b_\infty}(z)$ for $z < z_\infty$, and so is not obstructive for $z < z_\infty$, while the overshadowing class is obstructive.

Next, as in Lemma 4.3.3 we consider the equalities

$$\frac{1 + z_\infty}{3 - b_\infty} = \frac{q_{\mathbf{B}} z_\infty}{d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty} = \frac{A + C z_\infty}{d_{\text{ov}} - m_{\text{ov}} b_\infty},$$

where the second equality follows from [1, Lemma 16]. This implies

$$\begin{aligned} (1 + z_\infty)(d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty) &= q_{\mathbf{B}} z_\infty (3 - b_\infty) \implies z_\infty = \frac{d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty}{q_{\mathbf{B}}(3 - b_\infty) - (d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty)}, \\ (A + C z_\infty)(d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty) &= q_{\mathbf{B}} z_\infty (d_{\text{ov}} - m_{\text{ov}} b_\infty) \\ \implies z_\infty &= \frac{A(d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty)}{q_{\mathbf{B}}(d_{\text{ov}} - m_{\text{ov}} b_\infty) - C(d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty)}. \end{aligned}$$

Hence we must have

$$A(q_{\mathbf{B}}(3 - b_\infty) - (d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty)) = q_{\mathbf{B}}(d_{\text{ov}} - m_{\text{ov}} b_\infty) - C(d_{\mathbf{B}} - m_{\mathbf{B}} b_\infty).$$

As before, the coefficients of b_∞ on both sides must agree, which gives

$$(4.3.2) \quad d_{\text{ov}} = 3A - (A - C)d_{\mathbf{B}}/q_{\mathbf{B}}, \quad m_{\text{ov}} = A - (A - C)m_{\mathbf{B}}/q_{\mathbf{B}}.$$

Since $A - C > 0$ and d_{ov} and m_{ov} are integers, Lemma 4.3.2 implies $k = (A - C)/q_{\mathbf{B}}$ is a positive integer.

Furthermore, these formulas prove that if $b_\infty > 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} > 1/3$ then $d_{\text{ov}} > 3m_{\text{ov}}$. We can again compute

$$d_{\text{ov}} - 3m_{\text{ov}} = k t_{\mathbf{B}}.$$

This proves (i). The proof for (ii) follows similarly. □

We use the results of Lemmas 4.3.3 and 4.3.4 to rule out overshadowing classes for a large set of prestaircases, which implies that none of the prestaircases considered in this paper have overshadowing classes.

Lemma 4.3.5 *Let S be a prestaircase with irrational accumulation point z_∞ that is associated to a perfect blocking class $\mathbf{B} = (d_{\mathbf{B}}, m_{\mathbf{B}}, p_{\mathbf{B}}, q_{\mathbf{B}}, t_{\mathbf{B}})$ such that $t_{\mathbf{B}} \geq 3$. Suppose that either $b_\infty > 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} > 1/3$ and the m_κ/d_κ strictly decrease or $b_\infty < 1/3$ and $m_{\mathbf{B}}/d_{\mathbf{B}} < 1/3$ and the m_κ/d_κ strictly increase. Assume further that if S is descending any overshadowing class must have slope $> 1/(3 - b_\infty)$. Then the prestaircase S has no overshadowing classes at all.*

Proof These conditions, with Lemma 4.3.1, ensure that the conditions in either Lemma 4.3.3 or Lemma 4.3.4 hold for S . For a class $\mathbf{E}_{\text{ov}} = (d_{\text{ov}}, m_{\text{ov}}, m_{\text{ov}})$ to be

overshadowing, it must be obstructive for $b = b_\infty$ at its break point a_{ov} . By [1, Lemma 15] this is possible only if $|d_{\text{ov}}b_\infty - m_{\text{ov}}| < 1$. Further, for each staircase there is a positive integer k such that $|d_{\text{ov}} - 3m_{\text{ov}}| = t_{\mathbf{B}}k$. Thus we have

$$3 > 3|b_\infty d_{\text{ov}} - m_{\text{ov}}| > |d_{\text{ov}} - 3m_{\text{ov}}| = t_{\mathbf{B}}k \geq 3k,$$

where the second inequality holds because Lemmas 4.3.3 and 4.3.4 imply that either

$$\frac{m_{\text{ov}}}{d_{\text{ov}}} < \frac{1}{3} < b_\infty \quad \text{or} \quad b_\infty < \frac{1}{3} < \frac{m_{\text{ov}}}{d_{\text{ov}}}.$$

As k must be positive, no such d_{ov} and m_{ov} can exist. □

Corollary 4.3.6 *For each $T \in \mathcal{G}$, the prestaircase family $T^\#(S^U)$ has no overshadowing classes.*

Proof Because $t_{\mathbf{B}} = 2n + 3$ in all cases, the results of Corollary 4.2.6 and Lemma 4.3.1 show that all the staircases in $T^\#(S^U)$ except for $S_{u,0}^U$ satisfy the conditions in Lemma 4.3.5 and hence have no overshadowing classes. The proof for $S_{u,0}^U$ is given in [1, Example 70]. □

Now, we establish a straightforward way to check if a perfect prestaircase is live based on the above overshadowing arguments.

Proposition 4.3.7 *Let \mathcal{S} be a perfect prestaircase with irrational accumulation point z_∞ associated to a blocking class \mathbf{B} with recursion parameter $t_{\mathbf{B}} \geq 3$. Suppose that either $b_\infty > 1/3$ and the m_κ/d_κ strictly decrease, or $b_\infty < 1/3$ and the m_κ/d_κ strictly increase. Assume further that if \mathcal{S} is descending the slope of any overshadowing class must be $> 1/(3 - b_\infty)$. Then \mathcal{S} is a staircase, namely it is perfect and live.*

Proof By Corollary 4.2.3 and Lemma 4.3.5, the perfect prestaircase satisfies the conditions of Theorem 4.2.1 implying that \mathcal{S} is live. □

A consequence of this proposition is Theorem 1.2.6.

Corollary 4.3.8 *For each $T \in \mathcal{G}$, $T^\#(S^U)$ is live.*

Proof For all staircases except $S_{u,0}^U$ (which was shown to be live in [1]), the conditions of Proposition 4.3.7 are satisfied since $t_{\mathbf{B}} = 2n + 3$, and Corollary 4.2.6 and Lemma 4.3.1 hold. □

References

- [1] **M Bertozzi, T S Holm, E Maw, D McDuff, G T Mwakyoma, A R Pires, M Weiler**, *Infinite staircases for Hirzebruch surfaces*, from “Research directions in symplectic and contact geometry and topology” (B Acu, C Cannizzo, D McDuff, Z Myer, Y Pan, L Traynor, editors), Assoc. Women Math. Ser. 27, Springer (2021) 47–157 [MR](#) [Zbl](#)
- [2] **R Casals, R Vianna**, *Full ellipsoid embeddings and toric mutations*, *Selecta Math.* 28 (2022) art. id. 61 [MR](#) [Zbl](#)
- [3] **D Cristofaro-Gardiner, T S Holm, A Mandini, A R Pires**, *On infinite staircases in toric symplectic four-manifolds*, preprint (2020) [arXiv 2004.13062](#)
- [4] **D Frenkel, D Müller**, *Symplectic embeddings of 4–dim ellipsoids into cubes*, *J. Symplectic Geom.* 13 (2015) 765–847 [MR](#) [Zbl](#)
- [5] **N Magill, D McDuff, M Weiler**, *Staircase patterns in Hirzebruch surfaces*, preprint (2022) [arXiv 2203.06453](#)
- [6] **D McDuff, F Schlenk**, *The embedding capacity of 4–dimensional symplectic ellipsoids*, *Ann. of Math.* 175 (2012) 1191–1282 [MR](#) [Zbl](#)
- [7] **M Usher**, *Infinite staircases in the symplectic embedding problem for four-dimensional ellipsoids into polydisks*, *Algebr. Geom. Topol.* 19 (2019) 1935–2022 [MR](#) [Zbl](#)

Mathematics Department, Cornell University
Ithaca, NY, United States

Mathematics Department, Barnard College
New York, NY, United States

nm627@cornell.edu, dusa@math.columbia.edu

Received: 1 August 2021 Revised: 14 June 2022

Geometric triangulations of a family of hyperbolic 3–braids

BARBARA NIMERSHIEM

We construct topological triangulations for complements of $(-2, 3, n)$ –pretzel knots and links with $n \geq 7$. Following a procedure outlined by Futer and Guéritaud, we use a theorem of Casson and Rivin to prove the constructed triangulations are geometric. Futer, Kalfagianni and Purcell have shown (indirectly) that such braids are hyperbolic. The new result here is a direct proof.

57K32

1 Introduction

A knot or link in S^3 is *hyperbolic* if its complement admits a complete hyperbolic structure. A braid on n strands, which is represented by a word in the braid group, $w \in B_n$, is hyperbolic if its braid closure, L_w , is. Hyperbolic braids formed from three strands have been characterized by Futer, Kalfagianni and Purcell [7], a characterization that depends on their representation in the braid group. B_3 has two generators, σ_1 and σ_2 shown in Figure 1, and one relation, $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$. The square of this element is commonly denoted by C , which is central. In [7, Theorem 5.5], Futer, Kalfagianni and Purcell have shown that $S^3 - L_w$ is hyperbolic if and only if w is conjugate to $C^k\sigma_1^{p_1}\sigma_2^{-q_1} \dots \sigma_1^{p_s}\sigma_2^{-q_s}$ with $k \in \mathbb{Z}$ and p_i, q_i , and s positive integers, and w is not conjugate to $\sigma_1^{p_0}\sigma_2^{q_0}$ for some integers p_0 and q_0 (closures of such braids are either unknots, torus knots, or connected sums of torus knots). The proof is by contradiction; they show the required hyperbolic structures exist without constructing associated triangulations. In this paper, we construct geometric triangulations for braids with $k = 2$, $s = 1$, $p_1 \geq 1$, and $q_1 = 1$, thus offering a direct proof of the “if” direction of their theorem in these cases.

Let L_p be the braid closure of $C^2\sigma_1^p\sigma_2^{-1}$ with $p \geq 1$, which has one component if p is odd and two if p is even. The complement of L_p is homeomorphic to the complement

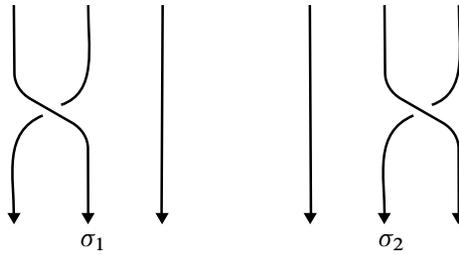


Figure 1: The generators of B_3 , the braid group on three strands.

of a $(-2, 3, p+6)$ -pretzel (Corollary 5.6), so proving $C^2\sigma_1^p\sigma_2^{-1}$ is hyperbolic will cover those pretzel knots and links as well. To show $C^2\sigma_1^p\sigma_2^{-1}$ is hyperbolic, we will first decompose its complement into ideal tetrahedra whose faces are identified in pairs (Section 2). These topological triangulations (one for each p) are very much in the spirit of the canonical decompositions first described for 2-bridge knot and link complements by Sakuma and Weeks in [19]. It turns out that, as in the 2-bridge case detailed by Futer in [9, Appendix A], our triangulations are not just topological; they are also *geometric* (though not canonical). In other words, each ideal tetrahedron can be given a hyperbolic shape of positive volume and, when the faces are paired by hyperbolic isometries, the resulting hyperbolic structure on $S^3 - L_p$ is metrically complete.

To prove that our topological triangulations are geometric, we will employ the “Casson–Rivin program” laid out by Futer and Guéritaud in [5]—their survey of Casson and Rivin’s technique for finding the hyperbolic structure on a 3-manifold whose boundary consists of tori. Guéritaud has implemented this program for punctured torus bundles and 4-punctured sphere bundles [9] and Futer for 2-bridge knots [9, Appendix A], whereas Guéritaud and Schleimer have applied it to layered solid tori and Dehn fillings [10] and Ham and Purcell used the same procedure to determine geometric triangulations on highly twisted links [11].

Geometric triangulations can lead to better understandings of the geometry of finite-volume hyperbolic 3-manifolds, including simpler proofs. However, it is not known whether every such 3-manifold admits a geometric triangulation. Currently the list of infinite families of finite-volume hyperbolic 3-manifolds that have been shown to admit geometric triangulations is short: punctured torus bundles and 4-punctured sphere bundles [9], 2-bridge knots and links [9, Appendix A], and certain Dehn fillings of fully augmented 2-bridge knots and links [11]. The main result of this paper is that the $(-2, 3, n)$ -pretzel knots and links with $n \geq 7$ can be added to the list.

Organization

After constructing the topological triangulations in [Section 2](#), we offer definitions and a basic outline of the Casson–Rivin program in [Section 3](#). The first step of the program is carried out in [Section 4](#), where we show that the space of positive angle structures for our triangulations ([Definition 3.1](#)) is nonempty. The final step—arguing that the critical point of the volume functional ([Definition 3.3](#)) is a positive angle structure—is presented in [Section 5](#) with some comments about extending the construction in [Section 6](#).

Acknowledgements

The explicit triangulations appearing here, as well as several more in the infinite family, were confirmed using the software packages Regina [\[3\]](#) and SnapPy [\[4\]](#) and checked against the census of veering triangulations created by Giannopolous, Schleimer and Segerman [\[8\]](#). See the Regina data file [\[15\]](#) for the triangulations constructed in [Section 2](#).

To make the graphics more accessible to those with color vision deficiencies, I have used two of Paul Tol’s qualitative color schemes—*bright* (Figures [3](#) and [14](#)) and *light* (Figures [4](#) and [13](#)). His palettes are mathematically designed to appear distinct to all [\[21\]](#).

David Futer kindly gave feedback on an early draft of the [appendix](#) for which I am grateful. I especially want to thank Bill Dunbar for many helpful conversations. There is no aspect of this paper—the structure, the notation, the figures, the tables, the exposition, the results—that has not been improved by his insights, careful reading, and self-described pickiness. I am also grateful for several useful suggestions offered by the referee. That said, any remaining errors are mine.

2 Ideal triangulations

Let $X_p = S^3 - L_p$, the complement of the closure of the braid $C^2\sigma_1^p\sigma_2^{-1}$. In this section, we construct a triangulation of X_p . In other words, we decompose the space into tetrahedra with face pairings. The decomposition will be developed using ideas from the [appendix](#), which contains a slight revisualization of the geometric triangulations of 2–bridge link complements presented by Futer in [\[9, Appendix A\]](#). There he uses the fact that a 2–bridge link can be described as a 4–braid “closed up” with a clasp at each

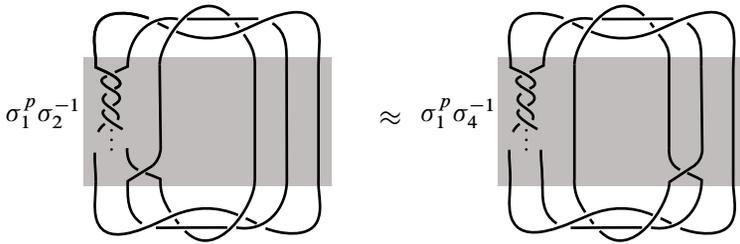


Figure 2: The closure of $C^2\sigma_1^p\sigma_2^{-1}$ is the 6–braid $\sigma_1^p\sigma_2^{-1}$ — or equivalently $\sigma_1^p\sigma_4^{-1}$ — “closed up” with a three–stranded full twist on both the top and the bottom.

end. The 4–braid lives in a product region, $S^2 \times I$, and its complement in this region is also a product, $S \times I$, where S is a 4–punctured sphere. In [9, Appendix A], the product region, $S \times I$, appears between two nested pillowcases and Futer showed that the region can be triangulated using a sequence of layers of ideal tetrahedra. In the appendix, we position the product region vertically. As a result, the faces of Futer’s layered tetrahedra can be easily seen in the braid complement, and the bottom of one layer is identified to the top of the next according to the half–twist between them. (See, for example, Figures 12 and 13.) The full link complement can be obtained by identifying the top of the product region to itself in a way that forms the clasp needed at the top and similarly for the bottom (Figure 14). These identifications amount to capping off each end of the $S \times I$ with a 3–ball from which the clasp has been removed.

Analogously, we observe that the closure of $C^2\sigma_1^p\sigma_2^{-1}$ can be formed from a 6–braid “closed up” with a three–stranded full twist on both the top and the bottom. In this context, the 6–braid $\sigma_1^p\sigma_2^{-1}$ can be replaced by $\sigma_1^p\sigma_4^{-1}$ because σ_2^{-1} can slide past the central full twist C on the bottom to become σ_4^{-1} as indicated in Figure 2. In this section, we will triangulate the complement, X_p , by placing ideal tetrahedra in the product region containing the 6–braid $\sigma_1^p\sigma_4^{-1}$ and identifying the topmost layer to itself in a way that forms a full twist on three strands and similarly at the bottom. We begin with the challenge of how to identify a 6–punctured sphere to itself to form a full twist.

2.1 Forming a three–stranded full twist

Let S_0 be the top 6–punctured sphere. Label the punctures 0 through 5 and divide S_0 into eight triangles as shown in Figure 3(a). Think of the punctures — and the strands descending from them — as living in the plane of the page, so $\Delta 025$ is in front of the

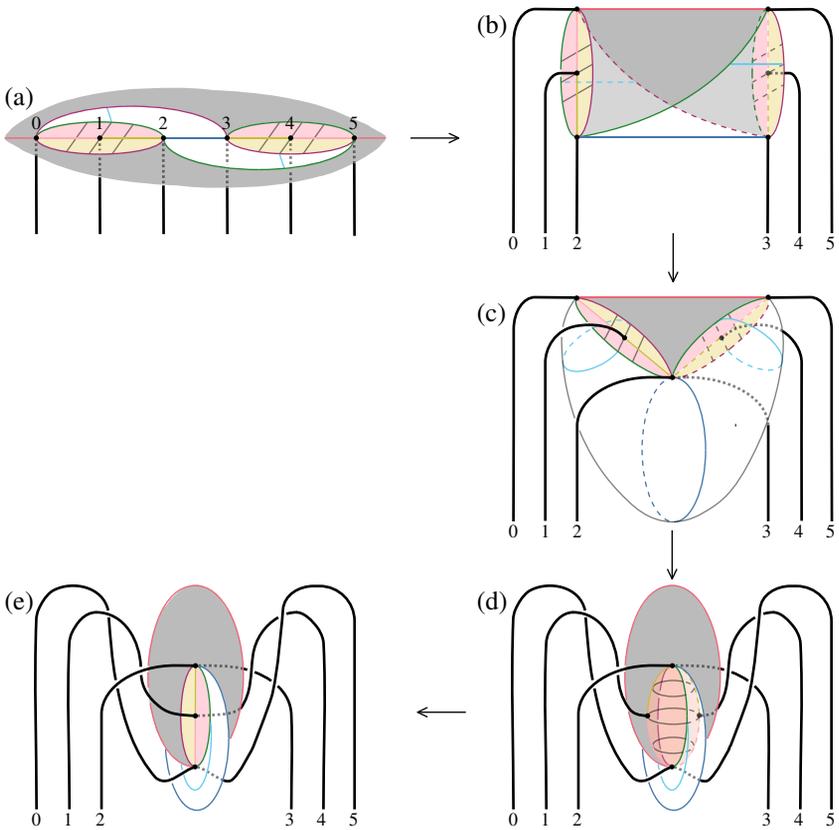


Figure 3: Forming a full twist C .

page and $\triangle 035$ is behind the page. Their shared edge, 05 , passes through the point at infinity. The colors indicate the identifications on S_0 that will yield a full twist:

gray $\triangle 025$ is identified to $\triangle 035$.

white $\triangle 023$ is identified to $\triangle 523$.

yellow $\triangle 012$ in the front is identified to $\triangle 543$ in the front.

pink $\triangle 012$ in the back is identified to $\triangle 543$ in the back.

To see that identifying S_0 to itself in this way is the same as attaching a 3–ball with a full twist removed, first fold S_0 along the 23 edge into a cylindrical pillow shown in Figure 3(b). As we go step-by-step through the identifications, we will follow the punctures (their paths are drawn as strands) and see they sweep out a full twist. The extra markings — cyan on the white triangles and gray on the yellow and pink ones — will help us follow the triangles through each step.

- $\Delta 025$ is identified to $\Delta 035$: [Figure 3\(c\)](#) shows the result of bringing the 2 and 3 strands together to identify the gray triangles. Placing this triangle in the plane of the page means the 23 (blue) edge forms a belt around a new pillow with half in front, as the 1 and 2 strands are, and half behind as the 3 and 4 strands are. The left side of the pillow consists of $\Delta 023$ and both $\Delta 012$'s (appearing in the front). The right side is formed by $\Delta 523$ and both $\Delta 543$'s (in the back). This diagram is analogous to the third step in [Figure 14](#) in the [appendix](#).
- $\Delta 023$ is identified to $\Delta 523$: By pushing the 0 and 5 strands into the pillow until they meet, $\Delta 023$ can be identified to $\Delta 523$ as shown in [Figure 3\(d\)](#). The 0 strand goes behind 1 and 2 and the 5 strand passes in front of 3 and 4. The result is not the same as the final step in [Figure 14](#), because of the presence of the two $\Delta 012$'s and the two $\Delta 543$'s. While these triangles themselves are not identified yet, their 02 edges and 53 edges (dark red and green) have come together, and they form a sphere with the two pink triangles in front and the two yellow in back. The interlocked gray and white faces meet along the dark red and green longitudes. These faces appear behind the sphere on the left and in front on the right.
- The $\Delta 012$'s are identified to the $\Delta 543$'s: Bringing strands 1 and 4 together and identifying each $\Delta 012$ to its corresponding $\Delta 543$ collapses the sphere. The result, shown in [Figure 3\(e\)](#), is the desired positive full twist on three strands (read right to left, the positive direction for the top of the 3–braid with a counterclockwise orientation).

To obtain the triangles for the 6–punctured sphere at the bottom of the product region, rotate S_0 about the horizontal line through the punctures, so now the strands of the braid go up, and use the same identifications. For example, at the bottom $\Delta 023$ is in the front and is identified to $\Delta 523$ in the back.

Having capped off the ends, we turn our attention to placing ideal tetrahedra in the product region, determining their face pairings, and thus defining triangulations τ_p on the manifolds X_p for all p . We start by constructing a triangulation of X_1 upon which the other triangulations will be built.

2.2 A foundational ideal triangulation

To specify the triangulation of X_1 , we will place tetrahedra in a product region from which the 6–braid, $\sigma_1\sigma_4^{-1}$, has been removed. The left diagram in [Figure 4](#) shows this region with the braid in the plane of the page, except near the crossings.

The method of the [appendix](#) suggests using three layers of tetrahedra (one on either side of the two crossings). Label the top, middle, and bottom layers Δ_t , Δ_m , and Δ_b (with Δ_t split into Δ_{tt} and Δ_{tb}). The six ideal tetrahedra that will triangulate the complement are shown in [Figure 4](#): t'_1 and t'_2 in Δ_t , m_1 and m_2 in Δ_m , and b_1 and b'_2 in Δ_b , where a prime indicates that the tetrahedron is behind the plane of the page; the labels on the ideal vertices come from the labels on their punctures (0 through 5 when read left to right); and the “top” faces of a tetrahedron are above the “bottom” faces relative to the product region.

The edges of the tetrahedra are drawn both in the braid complement on the left and in flattened versions in the center and on the right. On the left, layers are expanded by making two copies of the edges where the top of a tetrahedron meets the bottom, allowing the faces of the tetrahedra and their remaining edges to be seen more easily. The flattened versions show just one copy of each edge.

The tetrahedra t'_1 and t'_2 share the face 025, shaded orange on the left, and their flattened versions are drawn separately. The pair m_1 and m_2 share only an edge (thickened in the center diagram), so their interiors do not overlap when flattened, and they can both be seen when drawn side-by-side. The same is true for b_1 and b'_2 .

The left figure contains additional segments that divide the boundaries of each layer into triangles. These curves appear gray in the diagram and will be useful in determining the face pairings for this initial triangulation, which we shall refer to as $\hat{\tau}_1$.

When comparing to the [appendix](#), the reader should be concerned that the pairs of tetrahedra do not fill the entire layer. In Δ_m , for example, the tetrahedra are only in front of the plane of the page. Nonetheless, if we slide the faces of these six tetrahedra around the link — up and down the braid and through the full twists at the top and bottom — they *will* fill out the entire product region. For example, face 235 on the bottom of t'_1 can slide down the braid through Δ_{tb} and Δ_m and past the σ_4^{-1} half-twist to be identified with the 245 face of b'_2 . We denote such faces and face pairings by

$$t'_1(235) \sim b'_2(245),$$

taking care that the order of the vertices indicates the correct match. Images of the intersection of this triangle’s path with each layer’s boundary are shaded pink in the diagram. They make use of the aforementioned additional gray segments and show that the back 235 region of Δ_m is indeed covered.

As indicated in [Figure 4](#), t'_1 sits on top of t'_2 with their 025 faces identified. The remaining faces on the top of the tetrahedra in Δ_t and those on the bottom of the

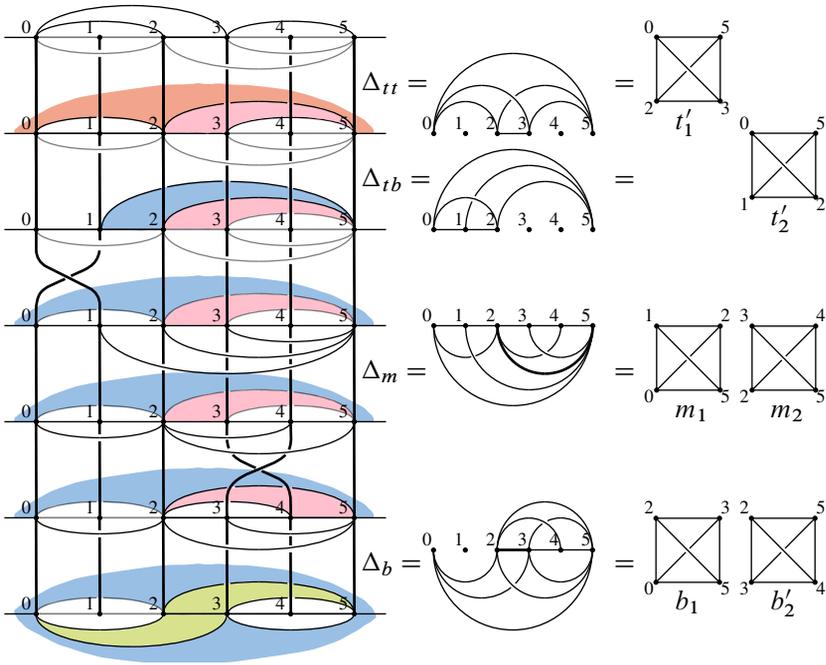


Figure 4: A triangulation of the complement of the closure of $C^2\sigma_1\sigma_2^{-1}$.

tetrahedra in Δ_b are triangles appearing in Section 2.1 and are identified by the full-twist identifications described there. For example, the face $b_1(023)$ on the bottom of Δ_b is the triangle $\Delta 023$, which is identified to $\Delta 523$, which is the 523 face of b'_2 , another face on the bottom of Δ_b , so $b_1(023) \sim b'_2(523)$. These faces are shaded green.

Sometimes the full-twist identifications will combine with sliding along the braid. To see such a combination, look at triangles shaded blue, starting with the 125 face of t'_2 . After traveling through the half-twist that interchanges 0 and 1 , $t'_2(125)$ becomes $\Delta 025$ on the top of Δ_m still in the back. Then it can travel through Δ_m and Δ_b all the way to $\Delta 025$ on the bottom of Δ_b in the back. This triangle is identified to $\Delta 035$ in the front, ie the 035 face of b_1 , so $t'_2(125) \sim b_1(035)$.

The remaining face pairings of $\hat{\tau}_1$ are obtained by sliding the remaining faces along the braid, twisting through half-twists and using full-twist identifications at the top and bottom as needed. The reader may enjoy tracing the long and twisted journey of $t'_2(012)$ to its mate, $m_1(021)$. All twelve face pairings of $\hat{\tau}_1$ are listed in Table 1.

For reasons explained in Section 4.1, the triangulation $\hat{\tau}_1$ is not quite what we want for our triangulation of X_1 . Instead, we will simplify $\hat{\tau}_1$ by applying Pachner moves — also

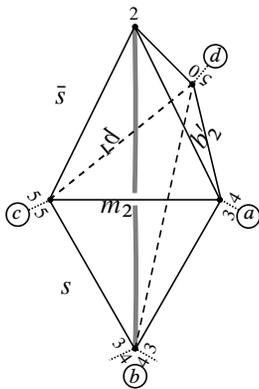
$t'_1(023) \sim m_2(523)$	$t'_2(125) \sim b_1(035)$
$t'_1(025) \sim t'_2(025)$	$m_1(025) \sim b_1(025)$
$t'_1(035) \sim m_1(125)$	$m_2(345) \sim b'_2(354)$
$t'_1(235) \sim b'_2(245)$	$m_2(234) \sim b'_2(243)$
$t'_2(015) \sim m_1(105)$	$m_2(245) \sim b_1(235)$
$t'_2(012) \sim m_1(021)$	$b_1(023) \sim b'_2(523)$

Table 1: Face pairings for $\hat{\tau}_1$, an initial triangulation of X_1 .

known as bistellar flips, introduced in [16]. Table 1 shows the following identifications of the 24 edge of m_2 :

$$24 \text{ in } m_2 \sim 23 \text{ in } b_1 \sim 23 \text{ in } b'_2 \sim_{\text{back to } 24} \text{ in } m_2.$$

Whenever three tetrahedra surround an edge, we can perform a 3-2 Pachner move that replaces them with two tetrahedra sharing a face. Figure 5 shows how to replace m_2 , b_1 , and b'_2 with \bar{s} on the top and s on the bottom. Label the vertices of s with $abcd$ (circled in the diagram), where a represents 3 in m_2 and 4 in b'_2 ; b represents 3 in b_1 and b'_2 and 4 in m_2 ; c represents 5 in m_2 and b_1 ; and d represents 0 in b_1 and 5 in b'_2 . The surviving faces of m_2 , b_1 , and b'_2 are renamed as follows: $m_2(235) \mapsto \bar{s}(2ac)$, $m_2(345) \mapsto s(abc)$, $b_1(025) \mapsto \bar{s}(d2c)$, $b_1(035) \mapsto s(dbc)$, $b'_2(245) \mapsto \bar{s}(2ad)$, and $b'_2(345) \mapsto s(bad)$. Making these substitutions and adding $\bar{s}(acd) \sim s(acd)$ yields the face pairings listed in Figure 5.



$t'_1(023) \sim \bar{s}(c2a)$
$t'_1(025) \sim t'_2(025)$
$t'_1(035) \sim m_1(125)$
$t'_1(235) \sim \bar{s}(2ad)$
$t'_2(015) \sim m_1(105)$
$t'_2(012) \sim m_1(021)$
$t'_2(125) \sim s(dbc)$
$m_1(025) \sim \bar{s}(d2c)$
$s(abc) \sim s(bda)$
$\bar{s}(acd) \sim s(acd)$

Figure 5: A 3-2 move eliminates the 24 edge of m_2 , which is also the 23 edge in b_1 and b'_2 , and yields the indicated face pairings.

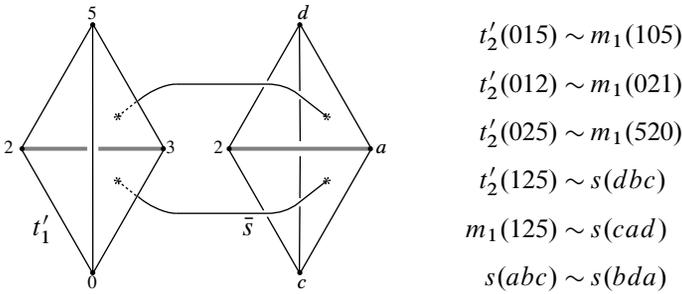


Figure 6: A 2-0 move eliminates t'_1 and \bar{s} , stacked as indicated, and yields face pairings for τ_1 , a triangulation of X_1 .

Next, observe that the 23 edge of t'_1 is identified to the $2a$ edge of \bar{s} , which is taken back to the 23 edge of t'_1 . This very short cycle of edges means t'_1 and \bar{s} are identified across an edge, so we can perform a 2-0 Pachner move to collapse t'_1 and \bar{s} . This move identifies $t'_1(025)$ with $\bar{s}(c2d)$ and $t'_1(035)$ with $\bar{s}(cad)$. See Figure 6, which shows how t'_1 and \bar{s} are stacked before they are collapsed and eliminated. The 2-0 move results in a triangulation of X_1 comprising three tetrahedra — t'_2 , m_1 , and s — whose face pairings are in Figure 6. Define τ_1 to be this triangulation.

2.3 Adding layers to extend to $p = 2$ and 3

It will be more instructive to first extend the construction of the previous section to $p = 3$. To triangulate the complement of the closure of $C^2\sigma_1^3\sigma_2^{-1}$, where there are four half-twists in the braid, begin with five layers: Δ_t , Δ_m , and Δ_b and two layers inserted between Δ_t and Δ_m , labeled Δ_{w_1} and Δ_{w_2} . Place two tetrahedra in each layer as indicated in Figure 7. Though Δ_t is depicted slightly differently (as one layer with the shared face of t'_1 and t'_2 , 025 , positioned in the middle), all tetrahedra in Δ_t , Δ_m , and Δ_b are named and situated exactly as in Figure 4. The new tetrahedra, those in the Δ_{w_i} layers, will be called w_i and w'_i , where a prime still indicates the tetrahedron is behind the plane of the page.

The face pairings are determined in the same manner as they were for $p = 1$, but now over half of the twenty face pairings result from moving through a half-twist from the bottom of one layer to the top of the next. These appear in the left column of Table 2. The middle column contains the face pairings that come directly from Table 1 and are unaffected by adding the w -layers. The final column contains more complicated face pairings that do not occur in Table 1 because these triangles hit tetrahedra in the w -layers before encountering their mates from Table 1. For example, $t'_1(035)$ is

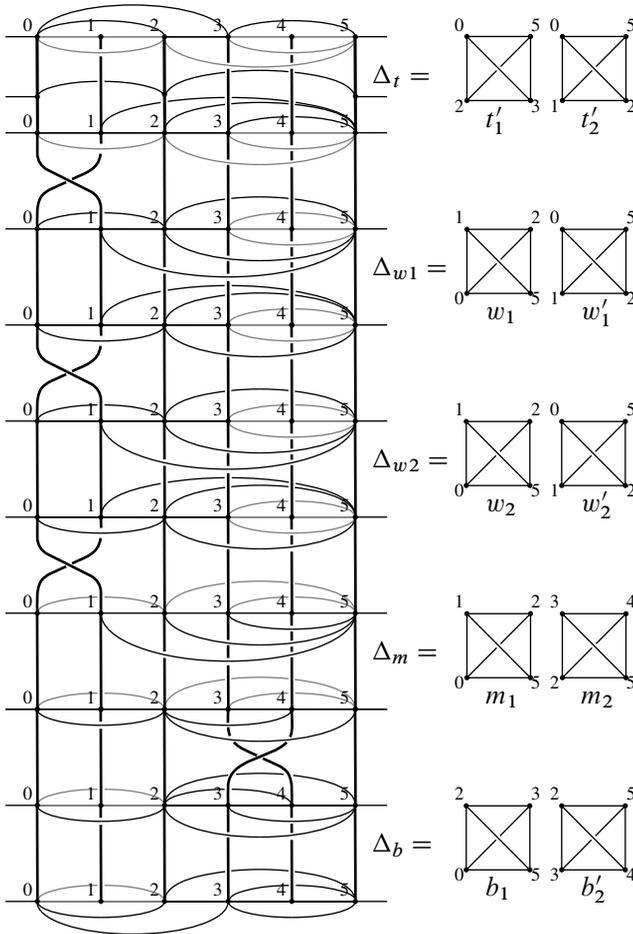


Figure 7: A triangulation of the complement of the closure of $C^2\sigma_1^3\sigma_2^{-1}$, with Δ_t depicted as one layer with the shared face of t'_1 and t'_2 , 025, positioned in the middle.

identified to Δ_{025} in the front, which, after passing through the half-twist, becomes Δ_{125} , a top face of w_1 , so $t'_1(035)$ hits w_1 before it gets to m_1 . We will call this initial triangulation $\hat{\tau}_3$, and, as in the $p = 1$ case, will perform simplifying Pachner moves on $\hat{\tau}_3$ to create τ_3 . The shadings in Table 2 will help us determine the resulting identifications.

Again, three tetrahedra in $\hat{\tau}_3$ surround the 24 edge of m_2 (which is identified to the 23 edges of b_1 and b'_2), allowing for the same 3-2 move shown in Figure 5. Thus m_2 , b_1 , and b'_2 are replaced by \bar{s} and s . As before, we can then perform a 2-0 move collapsing

$t'_2(015) \sim w_1(105)$		
$t'_2(125) \sim w'_1(025)$		
$w_1(025) \sim w_2(125)$		
$w_1(012) \sim w'_2(102)$	$t'_1(023) \sim m_2(523)$	$t'_1(035) \sim w_1(125)$
$w'_1(015) \sim w_2(105)$	$t'_1(025) \sim t'_2(025)$	$m_2(345) \sim w'_1(201)$
$w'_1(125) \sim w'_2(025)$	$t'_1(235) \sim b'_2(245)$	$b_1(035) \sim w'_2(125)$
$w_2(025) \sim m_1(125)$	$t'_2(012) \sim m_1(021)$	$b'_2(345) \sim w_2(201)$
$w'_2(015) \sim m_1(105)$	$b_1(023) \sim b'_2(523)$	
$m_1(025) \sim b_1(025)$		
$m_2(234) \sim b'_2(243)$		
$m_2(245) \sim b_1(235)$		

Table 2: Face pairings for $\hat{\tau}_3$, an initial triangulation of X_3 .

t'_1 and \bar{s} because they are identified across the edge $23 = 2a$. Again, $t'_1(025)$ and $t'_1(035)$ are identified with $\bar{s}(c2d)$ and $\bar{s}(cad)$ as in Figure 6. Let τ_3 be the resulting triangulation of X_3 , consisting of the seven remaining tetrahedra: t'_2 , w_1 , w'_1 , w_2 , w'_2 , m_1 , and s .

The face pairings in the darkest cells in Table 2 involve only the four replaced or collapsed tetrahedra (m_2 , b_1 , b'_2 , t'_1), so none of them will survive the Pachner moves defining τ_3 . The face pairings in the unshaded cells, on the other hand, do not involve any of these tetrahedra, and are, thus, unaffected by the two Pachner moves.

The remaining face pairings of τ_3 come from those in the light gray cells in Table 2. They involve the faces of m_2 , b_1 , b'_2 , and t'_1 that survive the Pachner moves. In other words, these faces do appear in τ_3 . They are just relabeled by the Pachner moves. Consider first $m_1(025) \sim b_1(025)$. Under the 3-2 move shown in Figure 5, $b_1(025)$ becomes $\bar{s}(d2c)$, which is identified by the 2-0 move in Figure 6 to $t'_1(520)$. This face is paired with $t'_2(520)$, so, after the two Pachner moves, $m_1(025)$ is identified with $t'_2(520)$, which we choose to write as $t'_2(025) \sim m_1(520)$.

The final four faces to be relabeled — $t'_1(035)$, $m_2(345)$, $b_1(035)$, and $b'_2(345)$ — are all faces that belong to the new tetrahedra s . More specifically, the 2-0 move identifies $t'_1(035)$ with $\bar{s}(cad)$, which is the new face introduced by the 3-2 move, the face shared by \bar{s} and s . Thus, after the two Pachner moves, $t'_1(035)$ becomes $s(cad)$. The other three faces form the rest of s as indicated in Figure 5 and are thus the faces $s(abc)$, $s(dbc)$,

$t'_2(015) \sim w_1(105)$		
$t'_2(125) \sim w'_1(025)$		
$w_1(025) \sim w_2(125)$		$s(cad) \sim w_1(125)$
$w_1(012) \sim w'_2(102)$	$t'_2(012) \sim m_1(021)$	$s(abc) \sim w'_1(201)$
$w'_1(015) \sim w_2(105)$	$t'_2(025) \sim m_1(520)$	$s dbc) \sim w'_2(125)$
$w'_1(125) \sim w'_2(025)$		$s(bad) \sim w_2(201)$
$w_2(025) \sim m_1(125)$		
$w'_2(015) \sim m_1(105)$		

Table 3: Face pairings for τ_3 , a triangulation of X_3 .

and $s(bad)$, respectively. Table 3 contains all identifications of τ_3 , where the first column still lists face pairings that result from moving from the bottom of one layer to the top of the next and the final column now shows how the new tetrahedron, s , glues in.

The reader likely anticipates that we will create the desired triangulation of X_p in two steps: first form an initial triangulation, $\hat{\tau}_p$, consisting of the tetrahedra in Figure 4 together with some w -layers between Δ_t and Δ_m , and then perform two Pachner moves — a correct guess that will be made explicit in the next section. Before moving on, we make some observations about the $p = 3$ construction and address the $p = 2$ case.

Figure 7 shows that t'_2 is situated and labeled in exactly the same way as w'_1 and w'_2 , so the labels on the bottom faces of t'_2 , 015 and 125, are the same as those on w'_1 and w'_2 . Furthermore, as the face $t'_2(015)$ slides through the half-twist, it rotates to the front, matching with $w_1(105)$, just as the 015 face of w'_1 matches with $w_2(105)$. Similarly, as the face $t'_2(125)$ slides through the half-twist, it stays in the back, matching with $w'_1(025)$, just as the 125 face of w'_1 matches with $w'_2(025)$. In these identifications, t'_2 is acting as a tetrahedron labeled w'_0 would. A similar analysis at the bottom of the w -layers shows that the top faces of m_1 glue to the bottom faces of w_2 in the same way those in a tetrahedron labeled w_3 would.

Using this relabeling, we make several observations about the face pairings in τ_3 .

- By relabeling t'_2 as w'_0 and m_1 as w_3 , the unshaded face pairings in the first column of Table 2 can be summarized as follows:

$$\begin{aligned}
 w'_i(015) \sim w_{i+1}(105) \quad \text{for } i = 0, 1, 2, & \quad w'_i(125) \sim w'_{i+1}(025) \quad \text{for } i = 0, 1, \\
 w_i(025) \sim w_{i+1}(125) \quad \text{for } i = 1, 2, & \quad w_i(012) \sim w'_{i+1}(102) \quad \text{for } i = 1.
 \end{aligned}$$

$w'_0(015) \sim w_1(105)$		$s(cad) \sim w_1(125)$
$w'_0(125) \sim w'_1(025)$	$w'_0(012) \sim w_2(021)$	$s(abc) \sim w'_1(201)$
$w_1(025) \sim w_2(125)$	$w'_0(025) \sim w_2(520)$	$s(dbc) \sim w'_1(125)$
$w'_1(015) \sim w_2(105)$		$s(bad) \sim w_1(201)$

Table 4: Face pairings for τ_2 , a triangulation of X_2 .

These inter- w identifications of $\hat{\tau}_3$ are not affected by Pachner moves involving m_2 , b_1 , b'_2 , and t'_1 , so they are also face pairings of τ_3 , as shown in the first column of Table 3.

- The top faces of w'_0 are not included in the identifications listed above, and neither are the bottom faces of w_3 . These faces are identified to each other. More specifically, after the Pachner moves, $w'_0(012) \sim w_3(021)$ and $w'_0(025) \sim w_3(520)$. Both of these pairings also occur in τ_1 (see the table in Figure 6) and were, thus, unaffected by the addition of the w -layers. These face pairings appear in the second column of Table 3.
- The face pairings in the third column of Table 3, those involving s , can be characterized as follows: two of the faces of s are identified to top faces in the first w -layer, $w_1(125)$ and $w'_1(201)$, and the other two are identified to bottom faces in the last w -layer, $w_2(201)$ and $w'_2(125)$.

Properly interpreted, these observations also apply when $p = 2$. Let the triangulation τ_2 of X_2 , the complement of the closure of $C^2\sigma_1^2\sigma_2^{-1}$, consist of w'_0 , w_1 , w'_1 , w_2 and s . Because there is only one w -layer, it is both the first w -layer and the last w -layer, and the final observation tells us that the faces of s are identified to four faces in Δ_{w_1} . Harvesting the remaining face pairings from the first two observations, we obtain the face pairings for τ_2 given in Table 4.

2.4 An ideal triangulation for any p

Guided by the previous examples along with Figure 7, form an initial triangulation, $\hat{\tau}_p$, of X_p , the complement of the closure of $C^2\sigma_1^p\sigma_2^{-1}$, that consists of $2p + 4$ tetrahedra: t'_1 and w'_0 (née t'_2) in Δ_t ; w_p (née m_1) and m_2 in Δ_m ; b_1 and b'_2 in Δ_b ; and w_i and w'_i in Δ_{wi} where $i = 1, \dots, p - 1$. The face pairings of $\hat{\tau}_p$ are determined by sliding faces along the braid and using the full-twist identifications at the top and bottom of the product region as described in Sections 2.2 and 2.3. As before, the addition of the w -layers does not affect identifications amongst m_2 , b_1 , b'_2 , and t'_1 , so the Pachner moves depicted in Figures 5 and 6 can be applied. Define τ_p to be the resulting

$$\begin{aligned}
 w'_i(015) &\sim w_{i+1}(105) && \text{for } i = 0, \dots, p-1 \\
 w'_i(125) &\sim w'_{i+1}(025) && \text{for } i = 0, \dots, p-2 \\
 w_i(025) &\sim w_{i+1}(125) && \text{for } i = 1, \dots, p-1 \\
 w_i(012) &\sim w'_{i+1}(102) && \text{for } i = 1, \dots, p-2 \\
 w'_0(012) &\sim w_p(021) \\
 w'_0(025) &\sim w_p(520) \\
 s(cad) &\sim w_1(125) \\
 s(abc) &\sim w'_1(201) \\
 s(dbc) &\sim w'_{p-1}(125) \\
 s(bad) &\sim w_{p-1}(201)
 \end{aligned}$$

Table 5: Face pairings for τ_p , a triangulation of X_p for $p > 1$.

triangulation. As with τ_2 , the observations about τ_3 from the previous section apply to τ_p , yielding the face pairings listed in Table 5 for $p > 1$. (For $p = 1$, see Figure 6.)

The triangulation $\hat{\tau}_p$ had $2p + 4$ tetrahedra. The 3-2 Pachner move reduced this number by one and the 2-0 Pachner move further reduced the number of tetrahedra by two, so τ_p has $2p + 1$ tetrahedra and thus $4p + 2$ face pairings. There are exactly $4p + 2$ identifications in Table 5 (and six in Figure 6). Thus, the definition of the triangulation τ_p is complete.

Note that most of the tetrahedra in τ_p (all but s) are visible in the complement of the 6–braid $\sigma_1^P \sigma_4^{-1}$. Using this visualization can help us quickly deduce relationships between the tetrahedra. For example, Figure 8 shows how $w'_0, w_1, w'_1, \dots, w_{p-1}, w'_{p-1}$ and w_p are situated in the braid complement, and all but six face pairings of τ_p can be derived directly from following their faces through the half-twists in the braid.

2.5 Equivalence classes of edges in the triangulation

Knowing the relationships between the edges of τ_p will prove helpful because their equivalence classes form the edges of X_p and the (internal) dihedral angles along these edges will dictate its geometry. In determining equivalence classes, we can use Figure 8 or Table 5, whichever is easier.

As an example, we will determine all edges identified to the 01 edge of w'_0 . In the braid picture, $w'_0(01)$ slides down through the first σ_1 half-twist to be identified with $w_1(10)$ and $w'_1(10)$, so the equivalence class of $w'_0(01)$, $[w'_0(01)]$, contains the 01 edge

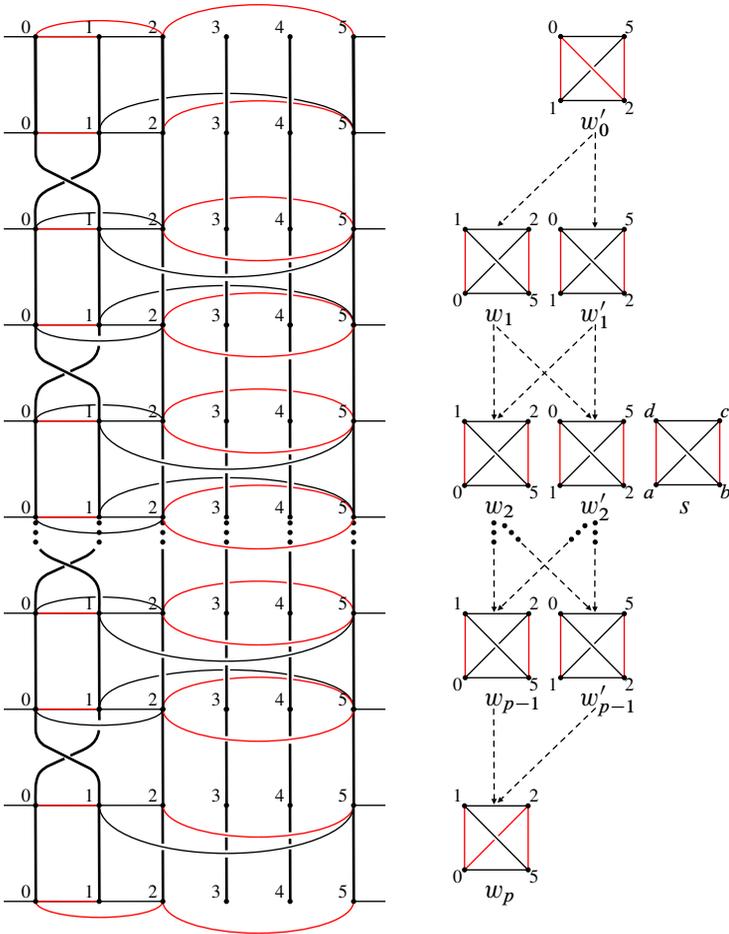


Figure 8: The tetrahedra of τ_p with $[w'_0(01)]$ colored red. All but s are shown in the braid complement on the left. The dashed arrows on the right indicate that a bottom face of the originating tetrahedron is identified through a half-twist to a top face of the target tetrahedron. Among faces unpaired by half-twists, the top faces of w'_0 are identified to the bottom faces of w_p and s is glued to the top faces of w_1 and w'_1 and the bottom faces of w_{p-1} and w'_{p-1} as in Table 5.

of w_1 and of w'_1 . Continuing to move the 01 edge through more half-twists shows the 01 edges of w_i and w'_i belong to $[w'_0(01)]$ for all i . Two edges of s also appear in this equivalence class, which can be seen most easily in Table 5: $w'_1(01) \sim s(bc)$ and $w_{p-1}(01) \sim s(ad)$. These edges of s , $s(bc)$ and $s(ad)$, are also identified to $w'_{p-1}(25)$ and $w_1(25)$, respectively, which can themselves be slid along the braid to show that

the 25 edges of each w_i and w'_i are in $[w'_0(01)]$. We have not yet considered face pairings between w'_0 and w_p . The first in Table 5, $w'_0(012) \sim w_p(021)$ shows $w_p(02)$ and $w'_0(02)$ are also in $[w'_0(01)]$. Based on how the tetrahedra are positioned in the braid complement, 02 is the top edge in w'_0 and the bottom edge in w_p , which allows us to describe $[w'_0(01)]$ as follows.

Observation 1 *The equivalence class $[w'_0(01)]$ contains a total of $4p + 4$ edges: the 01 edges and 25 edges of each w_i and w'_i ; $s(bc)$ and $s(ad)$; the top edge of w'_0 and the bottom edge of w_p . These edges are colored red in Figure 8.*

Observation 2 *A similar analysis shows that X_p has two edges with degree 5. The edge $[w'_0(05)]$ arises from*

$$w'_0(05) \sim w_1(15) \sim s(cd) \sim w'_{p-1}(51) \sim w_p(05) \sim_{\text{back to } w'_0} w'_0(05)$$

and, when $p > 1$, the edge $[w'_0(12)]$ from

$$w'_0(12) \sim w'_1(02) \sim s(ba) \sim w_{p-1}(20) \sim w_p(21) \sim_{\text{back to } w'_0} w'_0(12).$$

When $p = 1$, $w'_1(02)$ and $w_{p-1}(20)$ are undefined and are replaced by $s(db)$ and $s(ac)$.

It turns out that the remaining $2p - 2$ edges have degree 4.

3 Angle structures and the Casson–Rivin program

We will prove that the triangulation τ_p of X_p defined in Section 2.4 is geometric by applying the Casson–Rivin program. This section provides some definitions, useful results, and a basic outline of the program. For more details and a beautiful exposition, including an elementary proof of the Casson–Rivin theorem, the reader is encouraged to read Futer and Guéritaud’s paper [5].

3.1 Angle structures

To move from a topological triangulation to a geometric one, we need a way to impose a hyperbolic structure on each tetrahedron where the face pairings are hyperbolic isometries. A start is to assign angles to the edges of the triangulation.

Definition 3.1 *An angle structure on an ideal triangulation τ of a 3–manifold is an assignment of angles to the edges of the tetrahedra in τ satisfying the following conditions:*

- (1) Angles assigned to opposite edges of a tetrahedron are equal, meaning it is enough to specify three angles θ_{3i-2} , θ_{3i-1} , and θ_{3i} , for each tetrahedron $T_i \in \tau$.
- (2) $\theta_{3i-2} + \theta_{3i-1} + \theta_{3i} = \pi$ for all i .
- (3) The sum of the angles surrounding an edge of the 3-manifold equals 2π .

This paper utilizes both *taut angle structures*, where $\theta_k \in \{0, \pi\}$ for all k , and *positive angle structures*, where $\theta_k > 0$ for all k . Note that condition (2) of the definition implies that a positive angle structure must have all angles strictly between 0 and π . In this case, the three angles that sum to π specify a *positively oriented ideal hyperbolic tetrahedron*. The tetrahedron is isometric to one whose vertices are at 0, 1, ∞ , and a complex number z with $\text{Im}(z) > 0$, where 0, 1, and z form a Euclidean triangle having the assigned angles. If $\text{Im}(z) = 0$, the tetrahedron is said to be *degenerate*, and, if $\text{Im}(z) < 0$, *negatively oriented*. If an angle structure θ assigns the same angles to two tetrahedra, T_1 and T_2 , then we will say they are *isometric with respect to θ* and write $T_1 \cong_{\theta} T_2$.

We use $\mathcal{A}(\tau)$ to denote the space of all positive angle structures on a triangulation τ and note that, if τ has n tetrahedra, $\mathcal{A}(\tau) \subset (0, \pi)^{3n}$ is a convex polytope with compact closure. As such, the points obtained through coordinate-by-coordinate averaging of elements of $\mathcal{A}(\tau)$ are again in $\mathcal{A}(\tau)$. Also, real-valued functions defined on $\overline{\mathcal{A}(\tau)}$ attain their extreme values. See [5, Proposition 3.2] for more details about the space of angle structures.

As will be described in Section 3.3, the initial step in the Casson–Rivin program is to argue that $\mathcal{A}(\tau)$ is nonempty, ie that the triangulation admits a positive angle structure. We accomplish this in our case by first showing (in Section 4.1) that τ_p admits a special type of taut angle structure — a veering angle structure. The notion of “veering”, first introduced by Agol [1], has several equivalent formulations. The simplest to apply to our triangulations is from [12].

Definition 3.2 A *veering angle structure* on an ideal triangulation τ of an oriented 3-manifold M is a taut angle structure meeting an additional condition: The edges of M can be colored red and blue in such a way that, within each tetrahedron,

- two of the edges with angle 0 are red and two are blue (the color on an angle π edge can be either red or blue);
- when viewed from any of the four ideal vertices, the π -angled edge is followed in the counterclockwise direction by a blue edge and then a red edge.

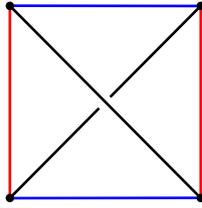


Figure 9: A tetrahedron in a veering triangulation.

Each tetrahedron in a veering triangulation appears as the one in Figure 9, where the π –angled edges on the diagonal can be red or blue. If there is also a consistent “upward” orientation on the four faces (for example out of the page in Figure 9), the angle structure is *transverse taut*. A triangulation that admits a veering angle structure is said to be *veering*.

Positive angle structures give geometric structures on the tetrahedra in a triangulation. In turn, these structures determine the face-pairing isometries. Condition (3) of Definition 3.1 guarantees that, under these isometries, the tetrahedra fit together to fill the space around the edges of the manifold, because the sum of dihedral angles around any edge is 2π . However, there is no guarantee that the result is metrically complete. In fact, if a 3–manifold has a complete hyperbolic structure of finite volume, Mostow–Prasad rigidity ensures that the structure is unique up to isometry [14; 17], so one should not expect completeness for an arbitrary point in $\mathcal{A}(\tau)$. It turns out that the independent work of Casson and Rivin [18], which forms the basis of the Casson–Rivin program in [5] and outlined in Section 3.3, connects the question of completeness to the volume associated to an angle structure in $\mathcal{A}(\tau)$, ie the sum of the volumes of the corresponding hyperbolic tetrahedra.

3.2 Volumes of angle structures

Recall — see [13], for example — that the volume of an ideal hyperbolic tetrahedron T , with dihedral angles θ_1 , θ_2 and θ_3 , is given by

$$V(T) = \mathbb{J}(\theta_1) + \mathbb{J}(\theta_2) + \mathbb{J}(\theta_3),$$

where the Lobachevsky function, \mathbb{J} , is defined as

$$\mathbb{J}(x) = - \int_0^x \log |2 \sin t| dt.$$

Definition 3.3 For a point $\theta \in \mathcal{A}(\tau)$, which specifies angles for the n tetrahedra $T_i \in \tau$ and has all $\theta_i \in (0, \pi)$, define the *volume associated to θ* , $\mathcal{V}(\theta)$, by

$$\mathcal{V}(\theta) = \sum_{i=1}^n V(T_i) = \sum_{i=1}^{3n} \mathbb{J}(\theta_i)$$

and note that $\mathcal{V}(\theta) > 0$ for $\theta \in \mathcal{A}(\tau)$.

As summarized in [5], \mathbb{J} is well-defined and continuous on \mathbb{R} . Thus, the definition of \mathcal{V} can be extended to the closure of $\mathcal{A}(\tau)$, which may include points with coordinates equal to 0 or π where the integral itself is improper. If $\theta \in \overline{\mathcal{A}(\tau)}$, then $\mathcal{V}(\theta) \geq 0$.

Straightforward tetrahedra-by-tetrahedra calculations show that for a positive angle structure θ and a tangent vector $\vec{v} \in T_\theta \mathcal{A}(\tau)$,

$$(1) \quad \frac{\partial \mathcal{V}}{\partial \vec{v}} = \sum_{i=1}^{3n} -v_i \log \sin \theta_i \quad \text{and} \quad \frac{\partial^2 \mathcal{V}}{\partial \vec{v}^2} < 0.$$

See for example [5, Lemma 5.3].

Because \mathcal{V} is strictly concave down on $\mathcal{A}(\tau)$, any critical point in $\mathcal{A}(\tau)$ is unique and an absolute maximum. In fact, the first and second partials (1) are enough to prove the same for a maximal point on the boundary.

Proposition 3.1 *Whenever $\mathcal{A}(\tau) \neq \emptyset$, the point at which the volume functional $\mathcal{V}: \overline{\mathcal{A}(\tau)} \rightarrow [0, \infty)$ attains its maximum is unique.*

Proof If there is a critical point α in $\mathcal{A}(\tau)$, the strict concavity of \mathcal{V} guarantees that α is unique and that $\mathcal{V}(\alpha)$ is the maximum volume. If, on the other hand, there is no critical point on the interior, the compactness of $\overline{\mathcal{A}(\tau)}$ ensures that there is a point $\alpha \in \overline{\mathcal{A}(\tau)} - \mathcal{A}(\tau)$ with $\mathcal{V}(\alpha)$ maximal. To show such an angle structure is unique, we first examine its coordinates. Because $\alpha \notin \mathcal{A}(\tau)$, there is at least one tetrahedron in τ with an assigned angle of 0 or π . The maximality of α further constrains the angles of such a tetrahedron as noted in the following lemma.

Lemma 3.2 *When $\mathcal{V}(\alpha)$ is a maximum, any tetrahedron assigned an angle of 0 or π must have angles of 0, 0, and π in some order.*

Proof By condition (2) of Definition 3.1, the angles in a tetrahedron must sum to π , so this statement is obviously true if one angle in a tetrahedron is assigned π or if two

angles are assigned 0. We now argue, as also noted in [9, Proposition 7.1] and [10, Proposition 7], that assigning exactly one 0 angle contradicts the maximality of α .

Consider, for example, a tetrahedron, $T_0 \in \tau$, with nonnegative angles 0, θ_0 , and $\pi - \theta_0$. For any $0 < \lambda < 1$, define a family of tetrahedra $T(t)|_{t \geq 0}$ with angles $\theta_1 = t$, $\theta_2 = \theta_0 - \lambda t$, and $\theta_3 = \pi - (t + \theta_0 - \lambda t)$, so $T(0) = T_0$ and, as t increases, the angles of the tetrahedron change — θ_1 is increased, while θ_2 and θ_3 are decreased. The first partial in (1) restricted to this family of tetrahedra yields

$$\left. \frac{dV(T(t))}{dt} \right|_{t=0^+} = \sum_{i=1}^3 \left. -\frac{d\theta_i}{dt} \log \sin \theta_i \right|_{t=0^+} = +\infty.$$

In other words, increasing the solitary 0 angle by a small amount would increase T_0 's volume (and thus the total volume \mathcal{V}) by a much larger amount, violating the maximality of α . Thus, if a maximal angle structure assigns any angles in $\{0, \pi\}$ to a tetrahedron, two of its angles are assigned 0 and one π , thus proving the lemma. We call a tetrahedron with angles in $\{0, \pi\}$ flat and observe that its volume is 0. \square

Returning to the proof of Proposition 3.1, we will use the constraints Lemma 3.2 imposes on the coordinates of a maximal point $\alpha \in \overline{\mathcal{A}(\tau)} - \mathcal{A}(\tau)$ to prove that α is unique. Let β be in $\overline{\mathcal{A}(\tau)} - \mathcal{A}(\tau)$ with $\mathcal{V}(\beta) = \mathcal{V}(\alpha)$ and let $\mu \in [0, \pi]^{3n}$ be the point whose coordinates are determined by averaging those of α and β . Then $\mu \in \overline{\mathcal{A}(\tau)}$, because the set $\overline{\mathcal{A}(\tau)}$ is convex. We will examine the volume of each tetrahedron T under the angle structure assigned by μ and will denote this volume by $\mathcal{V}(\mu|_T)$.

Case 1 If α and β assign the same angles to a tetrahedron T , then their average μ does as well, so $\mathcal{V}(\mu|_T)$ is equal to $\mathcal{V}(\alpha|_T)$ (and also $\mathcal{V}(\beta|_T)$).

Case 2 If α and β assign different angles to a tetrahedron T and both $\alpha|_T$ and $\beta|_T$ are flat, then, because α and β are maximum points, the angles must be 0, 0, and π (in different orders), and the coordinates of $\mu|_T$ are 0, $\frac{\pi}{2}$, and $\frac{\pi}{2}$ in some order. In this case, $\mathcal{V}(\mu|_T)$ equals 0 just as $\mathcal{V}(\alpha|_T)$ and $\mathcal{V}(\beta|_T)$ do.

Case 3 If α and β assign different angles to a tetrahedron T and at least one of $\alpha|_T$ or $\beta|_T$ is not flat, we will argue that $\mathcal{V}(\mu|_T)$ is strictly greater than the average of $\mathcal{V}(\alpha|_T)$ and $\mathcal{V}(\beta|_T)$. Without loss of generality suppose $\alpha|_T$ is not flat. Define $\gamma(t)$ to be the line segment joining $\alpha|_T$ and $\beta|_T$,

$$\gamma(t) = t\alpha|_T + (1 - t)\beta|_T \quad \text{for } t \in [0, 1].$$

Let $V(t)$ be the volume of the tetrahedron whose angles are given by $\gamma(t)$. As a restriction of \mathcal{V} , the function V is continuous on $[0, 1]$. Also, because $\alpha|_T$ is not flat, γ is in $(0, \pi)^3$ for $0 < t \leq 1$, so V is differentiable on $(0, 1]$ and by the second partial in (1), $V'' < 0$. Applying the mean value theorem to V yields points $c_1 \in (0, \frac{1}{2})$ and $c_2 \in (\frac{1}{2}, 1)$ with

$$V'(c_1) = 2(V(\frac{1}{2}) - V(0)), \quad V'(c_2) = 2(V(1) - V(\frac{1}{2}))$$

and negative second derivatives, so $V'(c_1) > V'(c_2)$, which means¹

$$2V(\frac{1}{2}) > V(1) + V(0).$$

Because $V(\frac{1}{2}) = \mathcal{V}(\mu|_T)$, $V(1) = \mathcal{V}(\alpha|_T)$, and $V(0) = \mathcal{V}(\beta|_T)$, this inequality implies $\mathcal{V}(\mu|_T) > \frac{1}{2}(\mathcal{V}(\alpha|_T) + \mathcal{V}(\beta|_T))$.

In all cases, the volume of a tetrahedron T computed using μ , the average of the angle structures, is greater than or equal to the average of the volumes of T computed using the angle structures α and β , so summing the individual volumes over all tetrahedra in τ yields

$$\mathcal{V}(\mu) \geq \frac{1}{2}(\mathcal{V}(\alpha) + \mathcal{V}(\beta)) = \mathcal{V}(\alpha)$$

with equality only when there are no tetrahedra in case 3. But $\mathcal{V}(\alpha)$ is the maximum volume, so $\mathcal{V}(\mu) \leq \mathcal{V}(\alpha)$, meaning $\mathcal{V}(\mu) = \mathcal{V}(\alpha)$ and there are no case 3 tetrahedra. The only way α and β can assign different angles to a tetrahedron T is if $\alpha|_T$ and $\beta|_T$ are both flat (case 2). In this case, $\mathcal{V}(\mu) = \mathcal{V}(\alpha)$, so μ is also maximal, but, as we have seen, the coordinates of $\mu|_T$ are $0, \frac{\pi}{2}$, and $\frac{\pi}{2}$, which is not allowed for maximal angle structures. Therefore, α and β cannot assign different angles to any of the tetrahedra and $\beta = \alpha$, which is the unique maximum point. □

Positive angle structures that maximize the total volume play an important role in the Casson–Rivin program.

3.3 The Casson–Rivin program

In their proof of Casson and Rivin’s theorem, Futer and Guéritaud relate the critical point of the volume (if it exists) and the complete hyperbolic structure (if it exists). More specifically, they show that the derivative of the volume vanishes in every direction

¹If neither $\alpha|_T$ nor $\beta|_T$ is flat, this inequality follows immediately from the strict concavity of V . We appeal to the mean value theorem, because, in general, strict concavity does not extend to the boundary of $\mathcal{A}(\tau)$.

exactly when Thurston’s system of gluing equations guaranteeing completeness (given in [20]) is satisfied. In doing so, they prove their main result, a theorem based on independent work of Rivin and Casson usually cited as [18].

Casson–Rivin theorem [5, Theorem 1.2] *Let M be an orientable 3–manifold with boundary consisting of tori, and let τ be an ideal triangulation of M . Then a point $\theta \in \mathcal{A}(\tau)$ corresponds to a complete hyperbolic metric on the interior of M if and only if θ is a critical point of the functional $\mathcal{V}: \mathcal{A}(\tau) \rightarrow [0, \infty)$.*

In our situation, the manifold M is the complement of an open neighborhood of the braid closure, L_p . We seek a complete hyperbolic metric on X_p , which is homeomorphic to the interior of M and is triangulated by τ_p .

The Casson–Rivin theorem will enable us to conclude τ_p is geometric as follows. In Section 4.1, we establish that the space of positive angle structures on our triangulations, $\mathcal{A}(\tau_p)$, is nonempty. Consequently, the volume functional $\mathcal{V}: \overline{\mathcal{A}(\tau_p)} \rightarrow [0, \infty)$ will attain its maximum. Then, in Section 5, we show that the maximum point of \mathcal{V} must belong to $\mathcal{A}(\tau_p)$. Then, by the Casson–Rivin theorem, there is a positive angle structure on our triangulation τ_p that yields the complete hyperbolic structure on the link complement, X_p . Thus, the constructed triangulations are geometric and $C^2\sigma_1^p\sigma_2^{-1}$ is hyperbolic for $p \geq 1$.

3.4 Symmetries of angle structures

When arguing that the maximum of the volume cannot appear on the boundary of $\mathcal{A}(\tau_p)$, a certain symmetry of the space of positive angle structures will be helpful.

Definition 3.4 Any symmetry, ρ , of a triangulation, τ , induces a map on the space of positive angle structures by assigning the angles of T to $\rho(T)$ for all $T \in \tau$. When such an assignment results in a positive angle structure for all points in $\mathcal{A}(\tau)$, we say ρ induces a symmetry of the space of positive angle structures, which we will also denote by ρ . More specifically, given a symmetry $\rho: \tau \rightarrow \tau$ with $\rho(T_i) = T_j$, let $\theta \in \mathcal{A}(\tau)$ and define the coordinates of $\rho\theta$ by $\rho\theta_{3j-2} = \theta_{3i-2}$, $\rho\theta_{3j-1} = \theta_{3i-1}$, and $\rho\theta_{3j} = \theta_{3i}$. The angles of $\rho\theta$ are positive and meet conditions (1) and (2) of Definition 3.1, so ρ induces a symmetry of $\mathcal{A}(\tau)$ whenever condition (3) is also met.

Observation 3 A symmetry $\rho: \mathcal{A}(\tau) \rightarrow \mathcal{A}(\tau)$ rearranges the angles of θ triple by triple, so $\mathcal{V}(\theta) = \mathcal{V}(\rho\theta)$.

4 The space of positive angle structures is nonempty

We now return to the triangulations, τ_p , defined in [Section 2.4](#). Our immediate goal is to complete the first step of the Casson–Rivin program by showing that $\mathcal{A}(\tau_p)$ is nonempty, ie that τ_p admits a positive angle structure. We will do so by showing that τ_p has a veering angle structure, which can be deformed to a positive angle structure [[6](#); [12](#)]. We then use the veering structure to describe an explicit coordinate system for $\mathcal{A}(\tau_p)$ and conclude the section by writing the edge equations in these coordinates and using them to describe a useful symmetry.

4.1 The triangulations are veering

While the initial triangulations $\hat{\tau}_p$ were not veering, the Pachner moves eliminated the obstructions and:

Proposition 4.1 *Each triangulation τ_p is veering.*

Proof With the exception of s , the tetrahedra in τ_p have an upward orientation induced by the product region containing the 6–braid, which is how the non- s tetrahedra of τ_p are flattened in [Figure 8](#). Flatten s as indicated in the same figure. In each tetrahedron, assign an angle of π to the diagonal edges and an angle of 0 to the others. Use red to color the 01 and 25 edges of each w_i and w'_i ; the edges $s(bc)$ and $s(ad)$; and the top diagonal of w'_0 along with the bottom diagonal of w_p . (These are exactly the edges colored red in [Figure 8](#).) Color the remaining edges blue. We claim this coloring forms a veering angle structure on τ_p , which is also transverse taut.

By [Observation 1](#), the edges in the equivalence class $[w'_0(01)]$ are exactly the edges colored red. Thus, the blue edges compose the remaining equivalence classes, and the colorings are consistent with the face pairings that form M .

Next we show that the assignment of angles forms a taut angle structure. The angle assignments themselves guarantee conditions (1) and (2) of [Definition 3.1](#) as well as the requirement that taut angle structures have angles in $\{0, \pi\}$. Only diagonals are assigned an angle of π , so, if there are exactly two diagonals in each equivalence class, condition (3) will also be met (the angle sum around each edge is 2π).

The class $[w'_0(01)]$, whose members are listed in [Observation 1](#), contains exactly two diagonals, $w'_0(02)$ and $w_p(02)$. The degree 5 edges of X_p are completely described in

Observation 2. The edges $[w'_0(05)]$ and $[w'_0(12)]$ each contain exactly two diagonals: the top diagonal $w_1(15)$ is matched with the bottom diagonal $w'_{p-1}(51)$ and similarly for $w'_1(02)$ and $w_{p-1}(20)$ (or $s(db)$ and $s(ac)$ when $p = 1$, which completes this case).

To check the remaining $2p - 2$ edges of X_p for $p > 1$, ie those of degree 4, we examine $w_{i-1}(02)$ and $w'_{i-1}(15)$, the bottom diagonals of w_{i-1} and w'_{i-1} in the braid complement in **Figure 8**. The edge $w_{i-1}(02)$ bounds both $w_{i-1}(012)$ and $w_{i-1}(025)$. After passing through the σ_1 half-twist, the first of these faces is rotated to the back and identified with $w'_i(102)$, whereas the second stays in front and is identified with $w_i(125)$. These faces share the edge 12, which, after another σ_1 is twisted to the back and identified to $w'_{i+1}(02)$. A similar argument applies to $w'_{i-1}(15)$. Symbolically,

$$(2) \quad \begin{aligned} w_{i-1}(02) \sim w_i(12) = w'_i(12) \sim w'_{i+1}(02) & \quad \text{for } i = 2, \dots, p - 2, \\ w'_{i-1}(15) \sim w_i(05) = w'_i(05) \sim w_{i+1}(15) & \quad \text{for } i = 1, \dots, p - 1. \end{aligned}$$

The only diagonals are the 02 and 15 edges, so the edge classes, $[w_i(12)]$ and $[w_i(05)]$, contain exactly two π angles as required. The remaining degree-4 edges contain the diagonals of s , $s(ac)$ and $s(bd)$. The identifications in **Table 5** show that their equivalence classes contain exactly one other diagonal:

$$(3) \quad \begin{aligned} s(ac) \sim w'_1(21) = w_1(21) \sim w'_2(20), \\ s(bd) \sim w'_{p-1}(21) = w_{p-1}(21) \sim w_{p-2}(20). \end{aligned}$$

With exactly two diagonals in each equivalence class, the angle structure is taut. Each tetrahedron appears as in **Figure 9**, and the gluings of **Table 5** identify bottom faces to top faces, so the “upward” orientations are consistent, forming a transverse taut angle structure, which is also layered. □

Corollary 4.2 *The triangulation τ_p admits a positive angle structure.*

Proof In their main result, [12, Theorem 1.5], Hodgson, Rubinstein, Segerman and Tillmann prove that veering triangulations admit positive angle structures (which they call strict angle structures). A constructive proof showing how to deform a veering angle structure to a positive angle structure has also been given by Futer and Guéritaud [6, Theorem 1.3]. □

4.2 Coordinates for the space of positive angle structures

The veering structure on τ_p allows us to introduce convenient coordinates for a point θ in $\mathcal{A}(\tau_p)$. For T in τ_p , let $\theta_D T$ denote the angle assigned to the diagonals of T , $\theta_R T$

the angle assigned to the red edges, and $\theta_B T$ to the blue edges. Thus, we can write $\theta \in \mathcal{A}(\tau_p) \subset (0, \pi)^{3(2p+1)}$ as

$$(\theta_R w'_0, \theta_B w'_0, \theta_D w'_0, \dots, \theta_R w'_{p-1}, \theta_B w'_{p-1}, \theta_D w'_{p-1}, \\ \theta_R w_1, \theta_B w_1, \theta_D w_1, \dots, \theta_R w_p, \theta_B w_p, \theta_D w_p, \theta_{Rs}, \theta_{Bs}, \theta_{Ds}).$$

Observation 4 Using this notation, the edge identifications in Observations 1 and 2 and equations (2) and (3) together with condition (3) of Definition 3.1 yield the following edge equations:

$$\left(\sum_{i=0}^{p-1} 2\theta_R w'_i\right) + \left(\sum_{i=1}^p 2\theta_R w_i\right) + 2\theta_{Rs} + \theta_D w'_0 + \theta_D w_p = 2\pi, \\ \theta_B w'_0 + \theta_D w_1 + \theta_{Bs} + \theta_D w'_{p-1} + \theta_B w_p = 2\pi, \\ \theta_B w'_0 + \theta_D w'_1 + \theta_{Bs} + \theta_D w_{p-1} + \theta_B w_p = 2\pi, \\ \theta_D w_{i-1} + \theta_B w_i + \theta_B w'_i + \theta_D w'_{i+1} = 2\pi \quad \text{for } i = 2, \dots, p-2, \\ \theta_D w'_{i-1} + \theta_B w_i + \theta_B w'_i + \theta_D w_{i+1} = 2\pi \quad \text{for } i = 1, \dots, p-1, \\ \theta_{Ds} + \theta_B w'_1 + \theta_B w_1 + \theta_D w'_2 = 2\pi, \\ \theta_{Ds} + \theta_B w'_{p-1} + \theta_B w_{p-1} + \theta_D w_{p-2} = 2\pi.$$

These equations hold for $p > 1$ as long as each term is defined. (Recall that there are no tetrahedra labeled w_0 or w'_p .)

The first two equations also hold when $p = 1$, and, after substituting $s(db)$ and $s(ac)$ for the undefined terms $w'_1(02)$ and $w_{p-1}(20)$ as in Observation 2, so does the third:

$$\theta_B w'_0 + \theta_{Ds} + \theta_{Bs} + \theta_{Ds} + \theta_B w_1 = 2\pi.$$

Observation 5 Several pairs of edge equations above share angles coming from the blue edges. These angles will cancel when one equation is subtracted from the other, yielding the following equalities for the diagonals:

$$\theta_D w_1 - \theta_D w'_1 = \theta_D w_3 - \theta_D w'_3 = \theta_D w_5 - \theta_D w'_5 = \dots, \\ \theta_D w_2 - \theta_D w'_2 = \theta_D w_4 - \theta_D w'_4 = \theta_D w_6 - \theta_D w'_6 = \dots.$$

In addition,

$$\theta_D w_{p-2} - \theta_D w'_{p-2} = \theta_D w_p - \theta_{Ds}, \\ \theta_{Ds} - \theta_D w'_0 = \theta_D w_2 - \theta_D w'_2, \\ \theta_D w_{p-1} - \theta_D w'_{p-1} = \theta_D w_1 - \theta_D w'_1.$$

These equations hold when $p > 1$ and each term is defined. The final equation, which is derived by taking the difference between the equations for the degree-5 edges implies that, if p is odd (so $p - 1$ is even), then all of the differences listed above are equal. If p is even, all differences, $\theta_D w_j - \theta_D w'_j$, with j even are also equal to $\theta_D s - \theta_D w'_0$ and $\theta_D w_p - \theta_D s$, but not necessarily to $\theta_D w_1 - \theta_D w'_1$, etc.

The corresponding equation for $p = 1$ is

$$\theta_D s - \theta_D w'_0 = \theta_D w_1 - \theta_D s.$$

4.3 A symmetry of the space of angle structures

Consider the closure of $C^2 \sigma_1^p \sigma_2^{-1}$ as shown on the right of Figure 2. Moving the lone half-twist up so that it occurs halfway along the σ_1 half-twists reveals an order 2 symmetry of X_p — rotate about a horizontal line through the lone half-twist. This involution induces a symmetry on the triangulation τ_p — rotate the tetrahedra in Figure 8 about a horizontal line and take s to itself. This symmetry will respect angle structures:

Proposition 4.3 *An involution $\iota: \tau_p \rightarrow \tau_p$ that fixes s and takes w'_i to w_{p-i} by matching up their R , B , and D edges induces a symmetry of $\mathcal{A}(\tau_p)$.*

Proof Let $\theta \in \mathcal{A}(\tau_p)$. According to Definition 3.4, to verify that $\iota\theta$ is also in $\mathcal{A}(\tau_p)$, we need to check that the sum of the dihedral angles at each edge of X_p is 2π , when the angle measure is given by $\iota\theta$. Thus, it is enough to verify that $\iota\theta$ satisfies the edge equations listed in Observation 4 where

$$\iota\theta_\star s = \theta_\star s, \quad \iota\theta_\star w'_i = \theta_\star w_{p-i}, \quad \iota\theta_\star w_i = \theta_\star w'_{p-i}$$

for $\star = R, B$, and D .

The first equation describes the angle sum around the red edge of X_p , which consists of all R edges and the red diagonals of w'_0 and w_p . The involution ι permutes these edges, thereby permuting the terms of the sum:

$$\begin{aligned} & \left(\sum_{i=0}^{p-1} 2\iota\theta_R w'_i \right) + \left(\sum_{i=1}^p 2\iota\theta_R w_i \right) + 2\iota\theta_R s + \iota\theta_D w'_0 + \iota\theta_D w_p \\ &= \left(\sum_{i=1}^p 2\theta_R w_i \right) + \left(\sum_{i=0}^{p-1} 2\theta_R w'_i \right) + 2\theta_R s + \theta_D w_p + \theta_D w'_0 \\ &= 2\pi, \end{aligned}$$

so the first edge equation is satisfied by $\iota\theta$. The terms in the next two sums, those for the degree-5 edges, are also permuted by ι , so those equations are satisfied by $\iota\theta$ as well.

Something a little different happens with the degree-4 edges (only present when $p > 1$). To confirm the final equation of [Observation 4](#), for example, we need to examine

$$\iota\theta_D s + \iota\theta_B w'_{p-1} + \iota\theta_B w_{p-1} + \iota\theta_D w_{p-2},$$

but this sum is just

$$\theta_D s + \theta_B w_1 + \theta_B w'_1 + \theta_D w'_2,$$

which equals 2π by the penultimate equation in [Observation 4](#). In this instance, the involution takes the terms in one sum to the terms in another, effectively permuting the edge equations themselves. The same happens for the remaining degree-4 equations. \square

Corollary 4.4 *Let $\iota: \tau_p \rightarrow \tau_p$ be as in [Proposition 4.3](#) and let k be such that $p = 2k$ if p is even and $p = 2k + 1$ if p is odd. Positive angle structures $\theta \in \mathcal{A}(\tau_p)$ with the property that $\iota\theta = \theta$ have $w'_i \cong_\theta w_{p-i}$ and thus satisfy a simpler list of edge equations:*

$$(4) \quad \left(\sum_{i=1}^p 4\theta_R w_i \right) + 2\theta_R s + 2\theta_D w_p = 2\pi,$$

$$(5) \quad 2\theta_B w_p + 2\theta_D w_1 + \theta_B s = 2\pi,$$

$$(6) \quad 2\theta_B w_p + 2\theta_D w_{p-1} + \theta_B s = 2\pi,$$

$$(7) \quad \theta_D w_{i-1} + \theta_B w_i + \theta_B w_{p-i} + \theta_D w_{p-(i+1)} = 2\pi \quad \text{for } i = 2, \dots, k,$$

$$(8) \quad \theta_D w_{p-(i-1)} + \theta_B w_i + \theta_B w_{p-i} + \theta_D w_{i+1} = 2\pi \quad \text{for } i = 1, \dots, k,$$

$$(9) \quad \theta_D s + \theta_B w_{p-1} + \theta_B w_1 + \theta_D w_{p-2} = 2\pi.$$

In addition,

$$(10) \quad \theta_D w_p = \theta_D s \quad \text{and} \quad \theta_D w_i = \theta_D w_{p-i} \quad \text{for } i = 1, \dots, k.$$

These equations hold for all $p \geq 1$, with one exception. When $p = 1$, (6) reads: $2\theta_B w_1 + 2\theta_D s + \theta_B s = 2\pi$.

Proof By assumption, $\iota\theta_\star w'_i = \theta_\star w'_i$ for $\star = R, B$, and D , and, by definition, $\iota\theta_\star w'_i$ also equals $\theta_\star w_{p-i}$, so $\theta_\star w'_i = \theta_\star w_{p-i}$, and the tetrahedra w'_i and w_{p-i} have the same angle assignments under θ , so $w'_i \cong_\theta w_{p-i}$. This isometry allows us to obtain (4)–(9), by replacing w'_i with w_{p-i} in the edge equations of [Observation 4](#) and eliminating redundancies.

To obtain (10) for $p = 1$, replace w'_0 with w_1 in the $p = 1$ equation of Observation 5, which shows

$$\theta_D s - \theta_D w_1 = \theta_D w_1 - \theta_D s,$$

so $\theta_D w_1 = \theta_D s$. If $p > 1$, (5) and (6) imply that $\theta_D w_1$ equals $\theta_D w_{p-1}$, which, because of the isometry, equals $\theta_D w'_1$, so the difference $\theta_D w_1 - \theta_D w'_1$ equals 0 as do all odd-index differences listed in Observation 5. Replacing w'_i with w_{p-i} yields (10) for odd i . The remaining equations are derived differently depending on the parity of p . As noted in Observation 5, when p is odd, all differences are equal (in this case, equal to 0), so, after replacing w'_i with w_{p-i} , (10) holds for all i when p is odd. To see that (10) also holds when p and i are even, note that

$$\theta_D w_p - \theta_D s = \theta_D w_2 - \theta_D w'_2 = \theta_D w_{p-2} - \theta_D w'_{p-2} = \theta_D s - \theta_D w'_0 = \theta_D s - \theta_D w_p$$

so $\theta_D s = \theta_D w_p$. Thus, all differences in Observation 5 with even index are also 0 and replacing w'_i with w_{p-i} results in (10). □

5 The maximal volume occurs at a positive angle structure

Corollary 4.2 shows that the space of positive angle structures on each triangulation τ_p is nonempty, so we have accomplished the first step of the Casson–Rivin program described by Futer and Guéritaud in [5] and summarized in Section 3.3. In this section, we complete the program, proving the main result.

Theorem 5.1 *Let L_p be the closure of the 3–braid $C^2\sigma_1^p\sigma_2^{-1}$ and X_p be its complement in the 3–sphere. Then there is an ideal triangulation of X_p that is geometric.*

Proof Let τ_p be the triangulation defined in Section 2.4. By Corollary 4.2, $\mathcal{A}(\tau_p) \neq \emptyset$, which means the volume functional \mathcal{V} attains its maximum on the compact set $\overline{\mathcal{A}(\tau_p)}$, say at the point α . A maximal point has two important properties. The first was proved in Lemma 3.2.

Property 1 *If α is maximal and assigns an angle in $\{0, \pi\}$ to a tetrahedron, then the tetrahedron must be flat, ie its angles are 0, 0, and π in some order.*

The second important property derives from a symmetry of the coordinates of a maximal point. Let the involution $\iota: \tau_p \rightarrow \tau_p$ be as in Proposition 4.3, so ι fixes s and takes w'_i to w_{p-i} . Then ι induces a symmetry on $\overline{\mathcal{A}(\tau_p)}$, which rearranges the coordinates of an angle structure tetrahedron by tetrahedron, so, just as in Observation 3, $\mathcal{V}(\iota\alpha) = \mathcal{V}(\alpha)$.

Thus, if α is maximal, so is $\iota\alpha$, but maximal points are unique by [Proposition 3.1](#), so $\iota\alpha = \alpha$, and [Corollary 4.4](#) together with [Definition 3.1](#) guarantee the following:

Property 2 *If α is maximal, then the angles assigned by α belong to $[0, \pi]$, sum to π within a tetrahedron, and are determined by the angles assigned to w_1, w_2, \dots, w_p and s , which must satisfy (4)–(10).*

Our goal is to use these properties to show that the maximal point, α , is in $\mathcal{A}(\tau_p)$, not its boundary. Then we can apply the Casson–Rivin theorem from [Section 3.3](#) [[5](#), [Theorem 1.2](#)] and conclude that α corresponds to the complete hyperbolic structure on X_p , thus proving [Theorem 5.1](#). We will show that α is in $\mathcal{A}(\tau_p)$ by showing that points in $\mathcal{B} = \overline{\mathcal{A}(\tau_p)} - \mathcal{A}(\tau_p)$ will never maximize the volume. We do so by proving the following:

Proposition 5.2 *Any $\beta \in \mathcal{B}$ satisfying Properties 1 and 2 must have $\mathcal{V}(\beta) = 0$.*

Proof Let $\beta \in \mathcal{B}$ satisfy both Properties 1 and 2. Because β is in the boundary of the space of positive angle structures, there is a tetrahedron in τ_p to which β assigns angles from the set $\{0, \pi\}$. By [Property 1](#), this tetrahedron is flat (has angles 0, 0, and π). Using [Property 2](#), we can conclude that β must assign π to at least one of the edges of w_1, w_2, \dots, w_p or s . We now explore which edges can have such assignments and determine the resulting volume.

Observe that, because all angle assignments are nonnegative, assigning an angle measure of π to any R edge of a w tetrahedron violates (4) of [Corollary 4.4](#) and thus [Property 2](#). Therefore, no $\beta_R w_i$ can equal π . We will use this result often, so we record it as a lemma and follow with another useful lemma.

Lemma 5.3 *If $\beta \in \mathcal{B}$ satisfies both Properties 1 and 2, $\beta_R w_i$ cannot equal π for $i = 1, \dots, p$. □*

Lemma 5.4 *If $\beta \in \mathcal{B}$ satisfies both Properties 1 and 2, and either w_p or s is flat, then all tetrahedra are flat, so $\mathcal{V}(\beta) = 0$.*

Proof If a tetrahedron is flat, one of its edges is assigned an angle of π . By [Lemma 5.3](#), $\beta_R w_p$ cannot equal π . However, $\beta_R s$ could equal π , and, if this is the case, (4) forces $\beta_R w_i$ to equal 0 for all $i = 1, \dots, p$, so, by [Property 1](#), the non- s tetrahedra are also flat. Because, $\beta_D w_p$ also appears in (4), a similar argument shows that if $\beta_D w_p = \pi$, all tetrahedra are flat. By (10), $\beta_D w_p = \beta_D s$, so the same holds if $\beta_D s = \pi$. Thus, it only remains to check what happens if the angle at a B edge of w_p or s equals π .

- If $\beta_B w_p = \pi$, (5) implies that $\beta_B s = 0$, so, by Property 1, there is a π angle at either the R or D edge of s , and, in either case, as observed above, all tetrahedra are flat.
- If $\beta_B s = \pi$, then $\beta_D s = 0$, so $\beta_D w_p$ also equals 0—see (10)—and, by Property 1, either $\beta_R w_p$ or $\beta_B w_p$ equals π . Because of Lemma 5.3, $\beta_B w_p$ must be π , so, by the previous case, all tetrahedra are flat.

Therefore, if either w_p or s has an assigned angle of π , $\mathcal{V}(\beta) = 0$. □

Returning to the proof of Proposition 5.2, recall that $\beta \in \mathcal{B}$ must assign π to at least one of the edges of w_1, w_2, \dots, w_p or s . Lemmas 5.3 and 5.4 cover the R edges and the tetrahedra w_p and s , so the only edges left to consider are the B and D edges of w_1, w_2, \dots, w_{p-1} . We start with the B edges and first examine the case when $\beta_B w_1 = \pi$.

If $\beta_B w_1 = \pi$, then $\beta_D w_1$ equals 0 and, by (10) with $i = 1$, so does $\beta_D w_{p-1}$. Property 1 implies that in w_{p-1} one of the other angles must be π , but, because of Lemma 5.3, it cannot be the angle at the R edge. Consequently, $\beta_B w_{p-1}$ equals π and (9) implies $\beta_D s$ must equal 0, so s is flat and, by Lemma 5.4, $\mathcal{V}(\beta) = 0$.

Having shown that if β assigns the angle π to a B edge of w_1 , then $\mathcal{V}(\beta) = 0$, we now consider the B edges of w_j with $1 < j \leq k$. By repeatedly applying (10) and (7) together with Property 1 and Lemma 5.3, we will push the flatness of w_j all the way down to w_1 , allowing us to once again conclude $\mathcal{V}(\beta) = 0$. In particular, if $\beta_B w_j = \pi$, then $\beta_D w_j$ equals 0 and, by (10) with $i = j$, so does $\beta_D w_{p-j}$. There is a π angle in w_{p-j} (Property 1) and it cannot occur at the R edge (Lemma 5.3), so $\beta_B w_{p-j}$, will equal π . With both $\beta_B w_j$ and $\beta_B w_{p-j}$ equal to π , (7) with $i = j$ implies that $\beta_D w_{j-1} = 0$, so, by (10) with $i = j - 1$, $\beta_D w_{p-(j-1)} = 0$ also. Another application of Property 1 and Lemma 5.3 shows both $\beta_B w_{j-1}$ and $\beta_B w_{p-(j-1)}$ are equal to π and we can apply (7) with $i = j - 1$ to conclude $\beta_D w_{j-2} = 0$. Continuing in this manner and eventually applying (7) with $i = 2$ when both $\beta_B w_2$ and $\beta_B w_{p-2}$ are equal to π , allows us to conclude that $\beta_D w_1 = 0$. Thus (by Property 1 and Lemma 5.3), $\beta_B w_1 = \pi$, so $\mathcal{V}(\beta) = 0$.

It remains to examine the B edges of w_j where $j > k$. If $\beta_B w_j$ equals π , then $\beta_D w_j = 0$, so applying (10) with $i = p - j$, Property 1, and Lemma 5.3 shows that $\beta_B w_{p-j} = \pi$. Because $p - j \leq k$, this case has already been covered. Therefore, if any of the B edges of w_1, w_2, \dots, w_{p-1} have angle equal to π , then $\mathcal{V}(\beta) = 0$.

The proof for the D edges is a little simpler, because there is no need to appeal to Lemma 5.3. If $\beta_D w_1 = \pi$, then, by (10) with $i = 1$, $\beta_D w_{p-1} = \pi$, so both $\beta_B w_1$ and $\beta_B w_{p-1}$ are equal to 0. In this situation, applying (9) yields $\beta_D s = \pi$, so s is flat and, by Lemma 5.4, $\mathcal{V}(\beta) = 0$. If $j \leq k$ and $\beta_D w_j$ equals π , then, by (10) with $i = j$, so does $\beta_D w_{p-j}$, and both $\beta_B(w_j)$ and $\beta_B(w_{p-j})$ equal 0, so (7) with $i = j$ implies $\beta_D w_{j-1} = \pi$. Continuing in this manner shows that if $\beta_D w_j$ equals π , so does $\beta_D w_1$, and thus $\mathcal{V}(\beta) = 0$. If $j > k$ and $\beta_D w_j = \pi$, then, by (10) with $i = p - j$, $\beta_D w_{p-j} = \pi$ with $p - j \leq k$, a previous case. \square

Any point maximizing the volume functional must satisfy Properties 1 and 2. By proving Proposition 5.2, we have shown that any point, β , on the boundary of $\mathcal{A}(\tau_p)$ satisfying these properties has $\mathcal{V}(\beta) = 0$. Therefore, the maximal point α must be on the interior of $\mathcal{A}(\tau_p)$, and, by the Casson–Rivin theorem (Section 3.3 and [5, Theorem 1.2]), α corresponds to the complete hyperbolic structure on X_p . Thus, the triangulation τ_p is geometric, which concludes the proof of Theorem 5.1. \square

Corollary 5.5 *In the braid group, $C^2\sigma_1^p\sigma_2^{-1}$ is not conjugate to $\sigma_1^{p_0}\sigma_2^{q_0}$ for integers p_0 and q_0 .*

Proof Closures of braids of the form $\sigma_1^{p_0}\sigma_2^{q_0}$ are not hyperbolic, but Theorem 5.1 shows that the closure of $C^2\sigma_1^p\sigma_2^{-1}$ is. \square

Corollary 5.6 *The complements of the $(-2, 3, n)$ -pretzel knots and links admit geometric triangulations for $n \geq 7$.*

Proof Recall that the braid group has the single relation $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$ and that C , which is this element’s square, is central. Conjugate words in the braid group yield equivalent braid closures, so, using the relation and the centrality of C , L_p , the closure of $C^2\sigma_1^p\sigma_2^{-1}$ with $p \geq 1$, is also the closure of

$$\begin{aligned} (\sigma_2\sigma_1\sigma_2)(\sigma_1\sigma_2\sigma_1)C\sigma_1^p\sigma_2^{-1} &\sim \sigma_2\sigma_1\sigma_2C\sigma_1^{p+2} \\ &= \sigma_1\sigma_2C\sigma_1^{p+3} \\ &\sim \sigma_2(\sigma_1\sigma_2\sigma_1)(\sigma_1\sigma_2\sigma_1)\sigma_1^{p+4} \\ &= \sigma_1\sigma_2\sigma_1\sigma_1\sigma_1\sigma_2\sigma_1^{p+5} \\ &\sim \sigma_1^3\sigma_2\sigma_1^{p+6}\sigma_2; \end{aligned}$$

here \sim denotes the equivalence relation of conjugacy.

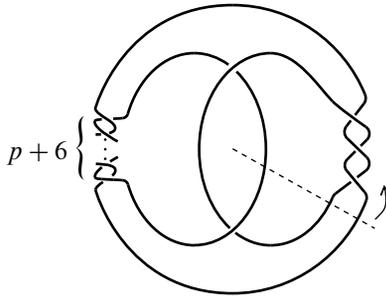


Figure 10: The $(-(p+6), 2, -3)$ –pretzel is also the closure of the braid $\sigma_1^3\sigma_2\sigma_1^{p+6}\sigma_2$. Start at the dashed line and read the braid counterclockwise (strands numbered from the outside in).

As indicated in Figure 10, the closure of $\sigma_1^3\sigma_2\sigma_1^{p+6}\sigma_2$ is a pretzel. Using our definitions, σ_1 and σ_2 generate left-handed (negative) twists, so L_p is a pretzel with $-(p + 6)$, 2, and -3 half-twists, or equivalently — after rolling the $p + 6$ half-twists to the right — L_p is the $(2, -3, -(p+6))$ –pretzel, and X_p is its complement. By Theorem 5.1, X_p admits a geometric triangulation. But the complement of the $(-2, 3, p+6)$ –pretzel is homeomorphic (via a reflection) to X_p , so it also has a geometric triangulation. \square

The braid relation $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$ and the centrality of $C = (\sigma_1\sigma_2)^3$ can also be used to show that L_p is equivalent to the braid closure of $\sigma_1^{p+1}(\sigma_1\sigma_2)^5$:

$$C^2\sigma_1^p\sigma_2^{-1} = \sigma_1^p(\sigma_1\sigma_2)^6\sigma_2^{-1} = \sigma_1^p(\sigma_1\sigma_2)^5\sigma_1 \sim \sigma_1^{p+1}(\sigma_1\sigma_2)^5.$$

In this form, L_p is a T-link as defined by Birman and Kofman in [2], where they show that T-links are in one-to-one correspondence with Lorenz links.

Corollary 5.7 *The complements of the T-links/Lorenz links formed as closures of the braids $\sigma_1^{p+1}(\sigma_1\sigma_2)^5$ admit geometric triangulations for $p > 1$.* \square

6 Extending the construction

Having constructed geometric triangulations for the braid closure of $C^2\sigma_1^p\sigma_2^{-1}$, it is natural to ask whether this construction can be extended to cover more cases. For example, hyperbolic T-links on three strands must be equivalent to closures of braids of the form $C^k\sigma_1^p\sigma_2^{-1}$ with $p > 0$ and $k \geq 2$, so adding more full twists to the construction would provide triangulations of all possible hyperbolic T-links on three strands. However, a useful combinatorial triangulation is not obvious.

Another possible extension would be to add more σ_2^{-1} half-twists. The construction of $\hat{\tau}_p$ in Section 2.4 can be extended, without difficulty, to $C^2\sigma_1^p\sigma_2^{-q}$ for all $q > 0$, and, in each case, there is also a sequence of Pachner moves that results in a veering triangulation. Thus, the first step of the Casson–Rivin program can be carried out. However, the final push for geometricity (the analog of Proposition 5.2) would require additional arguments. After completing $C^2\sigma_1^p\sigma_2^{-q}$, it may be possible to further extend the construction to $C^2\sigma_1^{p_1}\sigma_2^{-q_1}\cdots\sigma_1^{p_s}\sigma_2^{-q_s}$, moving closer to the general form in the Futer–Kalfagianni–Purcell characterization of hyperbolic 3–braids [7] stated in Section 1.

Appendix Visualizing triangulations of 2–bridge link complements

Sakuma and Weeks constructed topological triangulations of 2–bridge link complements in [19], which Futer showed were geometric [9, Appendix A]. Futer’s approach follows that of Guéritaud, whose development of geometric triangulations of once-punctured torus bundles and 4–punctured sphere bundles forms the bulk of [9]. This appendix presents another way to visualize the triangulations of 2–bridge link complements.

Futer’s description of these triangulations is based on the fact that, with the exception of the unknot and the trivial link with two components, 2–bridge links can be constructed from certain 4–braids whose ends are connected. More specifically, these 4–braids run between nested pillowcases; are formed by a sequence of R and L moves on the strands; and are closed off by adding crossing strands inside the inner pillowcase and outside the outer one. Please see [9, Appendix A] for the details and a very clear exposition.

As Futer mentions, his 2–bridge link diagrams can be isotoped so the pillowcases are horizontal; that is, they are perpendicular to the plane of the page and contain the point at infinity. The result is a diagram in which the braid strands run vertically between the bounding pillowcases, constrained to the plane of the page, except near the crossings. See Figure 11 for a comparison of the defining R and L moves that occur between two pillowcases.

Just as in Futer [9, Appendix A], the 4–braid lives in a product region, $S^2 \times I$, and its complement in the region is also a product, $S \times I$, where S is a 4–punctured sphere. As an example (left side of [9, Figure 18]), Futer uses the 2–bridge link $K(\Omega)$ where

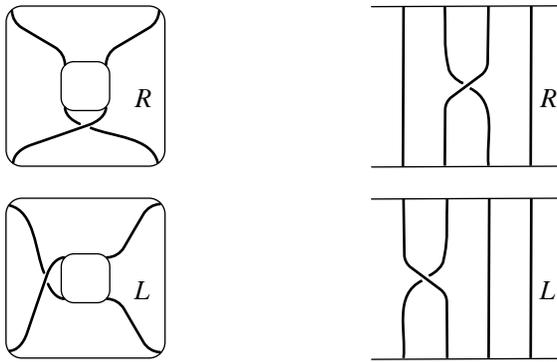


Figure 11: The actions of R and of L on strands between pillowcases. The left side is [9, Figure 14]. The right side is a version with vertical strands.

$\Omega = R^3L^2R$, so we will too. The left side of Figure 12 shows how to visualize $K(\Omega)$ — the thickened curves — in a vertical product region. (The thinner curves will be explained later as will the right side.)

Following Futer’s techniques, we triangulate the product region using a sequence of layers of ideal tetrahedra where the top of one layer is identified to the bottom of the next. Adopting his notation, the layers in Figure 12 are labeled Δ_i and they each contain ideal tetrahedra T_i and T'_i . The layer Δ_i is bounded by 4–punctured spheres, below by S_i and above by S_{i+1} , with S_{i+1} in Δ_i identified to S_{i+1} in Δ_{i+1} by passing through a half-twist. Numbering the punctures 0–3 will help in describing the resulting face pairings.

The right side of Figure 12 shows the tetrahedra forming the complement of $K(\Omega)$. These tetrahedra also appear in the left side of the figure. The thin curves are their edges. To see how this works, examine the bottom layer, redrawn in Figure 13. The first diagram of Figure 13 (in the top left) shows an expanded version of the tetrahedra that compose Δ_1 , one in which there are two copies of each edge in the plane of the page. This expansion makes it easier to see the triangles in the bounding punctured spheres, S_1 and S_2 , and how these triangles — faces of the tetrahedra T_1 and T'_1 — live in the link complement. The coloring indicates how the triangles in the lower S_2 pillowcase are identified to those in the upper one after passing through the R half-twist (our version of [9, Figure 15]):

- Δ_{013} in the front (blue) is half twisted to Δ_{023} in the front;
- Δ_{123} in the front (green) is half twisted to Δ_{213} in the back;

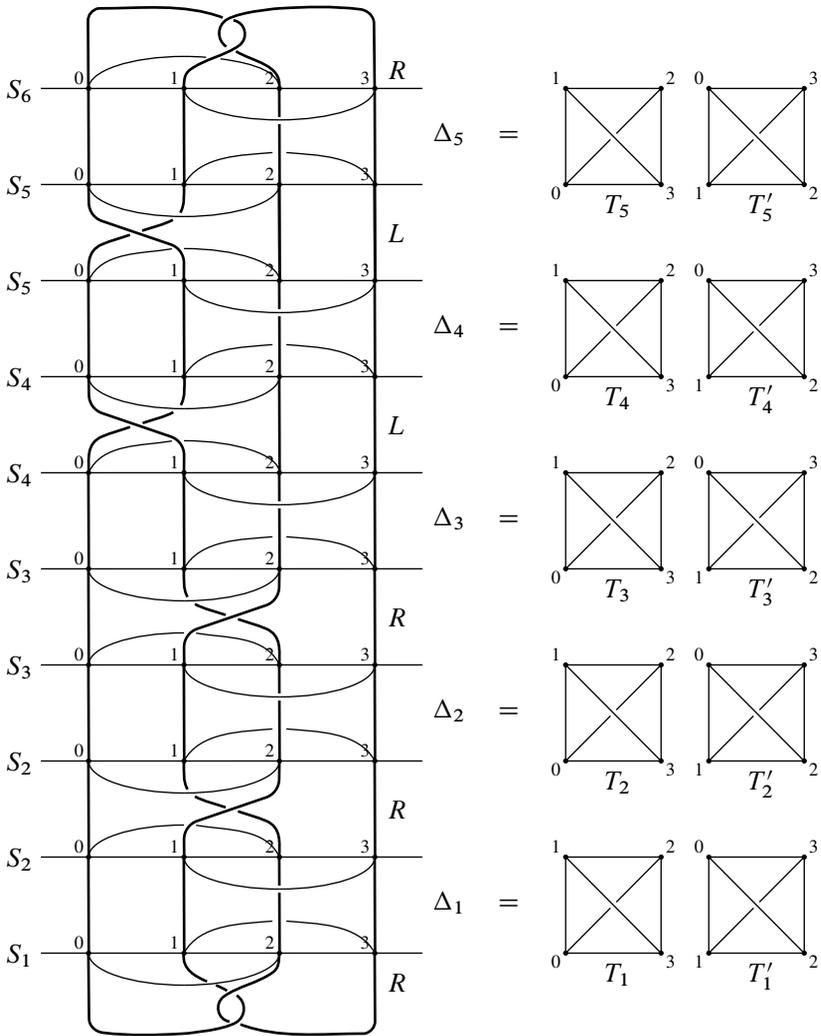


Figure 12: The link $K(\Omega)$ where $\Omega = R^3L^2R$. Compare the left side to the left side of [9, Figure 18]. This is also similar to the top of [19, Figure II.3.3].

- Δ_{023} in the back (orange) is half twisted to Δ_{013} in the back; and
- Δ_{012} in the back (pink) is half twisted to Δ_{021} in the front.

The face pairings defined by passing through the L half-twist can be determined in the same manner, using, for example, the pillowcase S_4 in Figure 12.

The second diagram in Figure 13 shows a collapsed version of Δ_1 , one consisting of two tetrahedra identified along the edges 01, 12, 23, and 30, which contains the point at

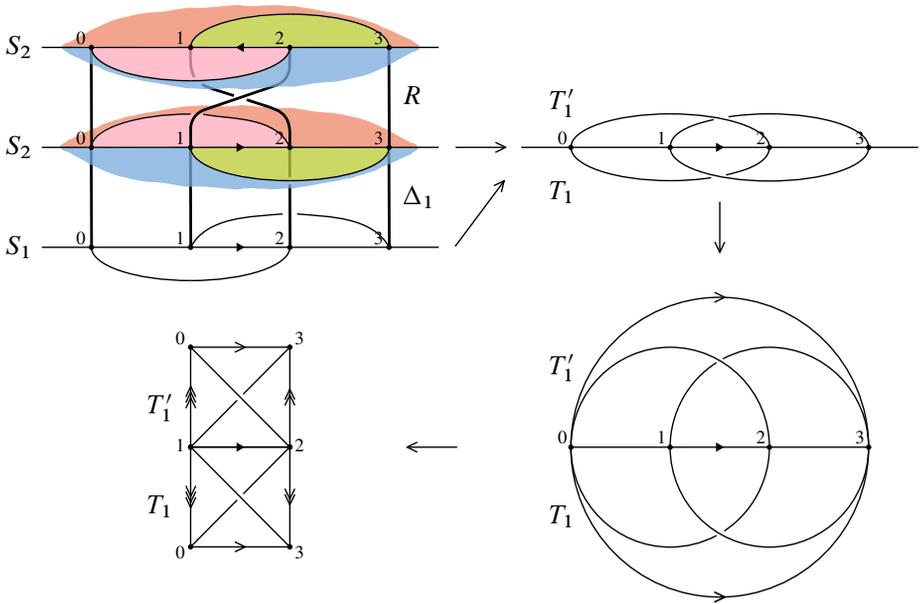


Figure 13: The layer Δ_1 . Compare the first figure to [9, Figure 15] and the last figure to [9, Figure 16].

infinity. T_1 is in front of the page with faces 013 and 123 on top. T'_1 is behind the page with faces 012 and 023 on top. The next two diagrams also show T_1 and T'_1 , using arrows to indicate edge identifications. The last one is our version of the two tetrahedra in [9, Figure 16], which can be separated to form the pair of tetrahedra on the bottom right in Figure 12.

In each layer, the expanded version of the pair of tetrahedra on the left of Figure 12 can be similarly associated with the flattened versions on the right. Thus, the tetrahedra that triangulate the product region appear in our visualization and the half-twists show how to obtain the face pairings between them.

All that remains to specify a triangulation of the link complement is to show how to close off the top and bottom. This amounts to saying how the triangles in the bounding pillowcases are identified to themselves to form clasps: If Ω starts with R , identify bottom layer triangles $\Delta_{012} \sim \Delta_{013}$ and $\Delta_{023} \sim \Delta_{123}$. If Ω starts with L , identify bottom layer triangles $\Delta_{012} \sim \Delta_{312}$ and $\Delta_{023} \sim \Delta_{013}$. The top layer triangles are identified in the same way. For example, if Ω ends with R , $\Delta_{012} \sim \Delta_{013}$ and $\Delta_{023} \sim \Delta_{123}$.

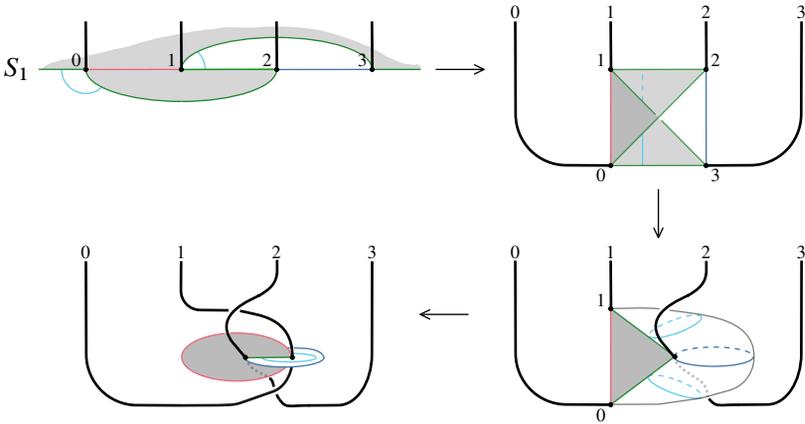


Figure 14: Forming a clasp. Compare to [9, Figure 17; 19, Figure II.2.7].

It should not be surprising that these identifications yield diagrams looking very much like [9, Figure 17; 19, Figure II.2.7]. For example, the steps forming the bottom clasp of $K(\Omega)$, where $\Omega = R^3 L^2 R$, are shown in Figure 14. While making the specified identifications, we can follow the punctures (their paths are drawn as strands) and see that they sweep out a clasp. Folding S_1 down along the 12 edge forms a pillow. Bringing the 2 and 3 strands together identifies the gray triangles, $\triangle 012$ and $\triangle 013$. Placing this triangle in the plane of the page means the blue 23 edge forms a belt around a new pillow with half in the front, like the 2 strand, and half behind as the 3 strand is. The top of the pillow is a cone formed by $\triangle 123$ and the bottom by $\triangle 023$, the white triangles with the cyan markings. Pushing the 1 and 0 strands into the pillow flattens the cones and identifies $\triangle 123$ to $\triangle 023$, thus completing the identifications.

Because the strands form a clasp, identifying S_1 to itself in such a way is equivalent to attaching a 3–ball with the needed clasp removed. Doing the same at the top results in a triangulation of the 2–bridge link complement. This triangulation is the same as Sakuma and Weeks’ as described by Futer and is one in which the faces of the tetrahedra are easy to see in the braid diagram, as are the face pairings induced by the half-twists.

References

- [1] **I Agol**, *Ideal triangulations of pseudo-Anosov mapping tori*, from “Topology and geometry in dimension three” (W Li, L Bartolini, J Johnson, F Luo, R Myers, JH Rubinstein, editors), Contemp. Math. 560, Amer. Math. Soc., Providence, RI (2011) 1–17 [MR](#) [Zbl](#)

- [2] **J Birman, I Kofman**, *A new twist on Lorenz links*, J. Topol. 2 (2009) 227–248 [MR](#) [Zbl](#)
- [3] **B A Burton, R Budney, W Pettersson**, et al., *Regina: software for low-dimensional topology* (1999) Available at <https://regina-normal.github.io>
- [4] **M Culler, N M Dunfield, M Goerner, J R Weeks**, *SnapPy: a computer program for studying the geometry and topology of 3–manifolds* (2016) Available at <http://snappy.computop.org>
- [5] **D Futer, F Guéritaud**, *From angled triangulations to hyperbolic structures*, from “Interactions between hyperbolic geometry, quantum topology and number theory” (A Champanerkar, O Dasbach, E Kalfagianni, I Kofman, W Neumann, N Stoltzfus, editors), Contemp. Math. 541, Amer. Math. Soc., Providence, RI (2011) 159–182 [MR](#) [Zbl](#)
- [6] **D Futer, F Guéritaud**, *Explicit angle structures for veering triangulations*, Algebr. Geom. Topol. 13 (2013) 205–235 [MR](#) [Zbl](#)
- [7] **D Futer, E Kalfagianni, J S Purcell**, *Cusp areas of Farey manifolds and applications to knot theory*, Int. Math. Res. Not. 2010 (2010) 4434–4497 [MR](#) [Zbl](#)
- [8] **A Giannopoulos, S Schleimer, H Segerman**, *A census of veering structures* (2019) Available at <https://math.okstate.edu/people/segerman/veering.html>
- [9] **F Guéritaud**, *On canonical triangulations of once-punctured torus bundles and two-bridge link complements*, Geom. Topol. 10 (2006) 1239–1284 [MR](#) [Zbl](#) With appendices by D Futer
- [10] **F Guéritaud, S Schleimer**, *Canonical triangulations of Dehn fillings*, Geom. Topol. 14 (2010) 193–242 [MR](#) [Zbl](#)
- [11] **S L Ham, J S Purcell**, *Geometric triangulations and highly twisted links*, Algebr. Geom. Topol. 23 (2023) 1399–1462 [MR](#) [Zbl](#)
- [12] **C D Hodgson, J H Rubinstein, H Segerman, S Tillmann**, *Veering triangulations admit strict angle structures*, Geom. Topol. 15 (2011) 2073–2089 [MR](#) [Zbl](#)
- [13] **J Milnor**, *Hyperbolic geometry: the first 150 years*, Bull. Amer. Math. Soc. 6 (1982) 9–24 [MR](#) [Zbl](#)
- [14] **G D Mostow**, *Strong rigidity of locally symmetric spaces*, Annals of Mathematics Studies 78, Princeton Univ. Press (1973) [MR](#) [Zbl](#)
- [15] **B Nimershiem**, Regina data file (2021) Available at <https://tinyurl.com/ReginaFileC2sigma1psigma2inv>
- [16] **U Pachner**, *P.L. homeomorphic manifolds are equivalent by elementary shellings*, European J. Combin. 12 (1991) 129–145 [MR](#) [Zbl](#)
- [17] **G Prasad**, *Strong rigidity of Q –rank 1 lattices*, Invent. Math. 21 (1973) 255–286 [MR](#) [Zbl](#)

- [18] **I Rivin**, *Euclidean structures on simplicial surfaces and hyperbolic volume*, Ann. of Math. 139 (1994) 553–580 [MR](#) [Zbl](#)
- [19] **M Sakuma, J Weeks**, *Examples of canonical decompositions of hyperbolic link complements*, Japan. J. Math. 21 (1995) 393–439 [MR](#) [Zbl](#)
- [20] **W P Thurston**, *Geometry and topology of three-manifolds*, lecture notes, Princeton University (1980) Available at <https://url.msp.org/gt3m>
- [21] **P Tol**, *Colour schemes and templates: qualitative colour schemes* (2021) Available at <https://personal.sron.nl/~pault/>

*Department of Mathematics, Franklin & Marshall College
Lancaster, PA, United States*

barbara.nimershiem@fandm.edu

<https://www.fandm.edu/directory/barbara-nimershiem.html>

Received: 1 September 2021 Revised: 28 May 2022

Beta families arising from a v_2^9 self-map on $S/(3, v_1^8)$

EVA BELMONT
KATSUMI SHIMOMURA

We show that v_2^9 is a permanent cycle in the 3–primary Adams–Novikov spectral sequence computing $\pi_*(S/(3, v_1^8))$, and use this to conclude that the families $\beta_{9t+3/i}$ for $i = 1, 2$, $\beta_{9t+6/i}$ for $i = 1, 2, 3$, $\beta_{9t+9/i}$ for $i = 1, \dots, 8$, $\alpha_1\beta_{9t+3/3}$, and $\alpha_1\beta_{9t+7}$ are permanent cycles in the 3–primary Adams–Novikov spectral sequence for the sphere for all $t \geq 0$. We use a computer program by Wang to determine the additive and partial multiplicative structure of the Adams–Novikov E_2 page for the sphere in relevant degrees. The $i = 1$ cases recover previously known results of Behrens and Pemmaraju and the second author. The results about $\beta_{9t+3/3}$, $\beta_{9t+6/3}$ and $\beta_{9t+9/8}$ were previously claimed by the second author; the computer calculations allow us to give a more direct proof. As an application, we determine the image of the Hurewicz map $\pi_*S \rightarrow \pi_*\text{tmf}$ at $p = 3$.

55Q45, 55Q51, 55T25

1 Introduction

Miller, Ravenel and Wilson [9, Theorem 2.6] showed that the 2–line of the Adams–Novikov E_2 page for the sphere is generated by classes $\beta_{i/j,k}$ for i, j, k satisfying certain conditions. At the prime 3, the β elements with $i \leq 9$ are

$$\begin{array}{ll} \beta_i & \text{for } i = 1, 2, 4, 5, 7, 8, \quad \beta_{3/j} \text{ and } \beta_{6/j} \quad \text{for } j = 1, 2, 3, \\ \beta_{9/j} & \text{for } j = 1, \dots, 9, \quad \beta_{9/3,2}, \end{array}$$

where we write $\beta_{i/j} := \beta_{i/j,1}$ and $\beta_i := \beta_{i/1}$. They have order 3 except for $\beta_{9/3,2}$, which satisfies $3\beta_{9/3,2} = \beta_{9/3}$. Of these, the permanent cycles are

$$(1) \quad \begin{array}{l} \beta_1, \beta_2, \beta_{3/2}, \beta_3, \beta_5, \beta_{6/3}, \beta_{6/2}, \beta_6, \\ \beta_7 + c\beta_{9/9} \quad \text{for some } c \in \{\pm 1\}, \quad \beta_{9/j} \quad \text{for } 1 \leq j \leq 8. \end{array}$$

For $i \leq 7$ and $\beta_7 + c\beta_{9/9}$, these assertions can be read off the exhaustive calculation of Ravenel [14, Table A3.4] of $\pi_n S_3^\wedge$ in stems $n \leq 108$; see also Oka [11] for many of the survival results and Shimomura [15] for the nonsurvival results. The element $\beta_7 + c\beta_{9/9}$ is an Arf invariant class (an odd-primary analogue of the $p = 2$ Kervaire invariant classes), discussed in Ravenel [13, page 439]; the survival of the Arf invariant classes is not known in general at $p = 3$. The survival of $\beta_{9/j}$ for $j \leq 8$ is a consequence of Theorem 5.1, which does not depend on prior knowledge about this element, but we do not claim originality for this result.

These results arise from exhaustive calculations in tractable stems, but it is possible to prove results about β elements outside the range of feasible computation. One strategy is as follows. Suppose $\beta_{i/j}$ is a permanent cycle. It is v_1 -power-torsion; that is, there exists a type 2 complex V such that $\beta_{i/j}$ factors as $S \xrightarrow{f_1} V \xrightarrow{f_2} S$ (we omit degree shifts for clarity of notation). If v_2^t is a v_2 self-map on V , then we may construct elements of $\pi_*(S)$ as

$$S \xrightarrow{f_1} V \xrightarrow{v_2^t} V \xrightarrow{f_2} S, \quad \text{with } t \geq 0.$$

For this family to be of interest, one must also show that the elements are nonzero, for example by identifying their Adams–Novikov representatives.

Let $S/3$ denote the mod 3 Moore space, and for $m \geq 1$ let $S/(3, v_1^m)$ denote the cofiber of the m -fold iterate of Adams’ v_1 self-map $S/3 \xrightarrow{v_1} S/3$; see Toda [19]. Behrens and Pemmaraju [3] show there is a v_2^9 self-map on $S/(3, v_1)$ and use this to prove the existence of nonzero homotopy classes represented by β_{9t+s} for $s = 1, 2, 5, 6, 9$ and $t \geq 0$. The second author [18] proves the existence of β_{9t+3} . By comparison to L_2 -local homotopy, he shows in [15] that the elements

$$\beta_{9t+4}, \quad \beta_{9t+7}, \quad \beta_{9t+8}, \quad \beta_{9t+3/3}, \quad \beta_{9t/3,2}, \quad \beta_{3^i s/3^i}$$

are not permanent cycles for $t \geq 1$, $s \not\equiv 0 \pmod{3}$, and $i > 1$. The main goal of this paper is to construct a v_2^9 self-map on $S/(3, v_1^8)$ and show that the remaining β elements in $\pi_s(S)$ for $s \leq |v_2^9| = 144$ also give rise to infinite families.

Theorem 5.1 *For all $t \geq 0$, the classes*

$$\begin{aligned} \beta_{9t+3/j} & \text{ for } j = 1, 2, & \beta_{9t+6/j} & \text{ for } j = 1, 2, 3, \\ \beta_{9t+9/j} & \text{ for } j = 1, \dots, 8, & \alpha_1 \beta_{9t+3/3} & \text{ and } \alpha_1 \beta_{9t+7} \end{aligned}$$

are permanent cycles in the Adams–Novikov spectral sequence for the sphere.

These families are interesting in part because the Hurewicz map $\pi_*(S) \rightarrow \pi_*(\text{tmf})$ detects β_{9t+1} , $\alpha_1\beta_{9t+3/3}$, $\beta_{9t+6/3}$ and $\alpha_1\beta_{9t+7}$, as we show in [Theorem 6.5](#). Together with the well-known behavior in the 0- and 1-lines, this completely determines the Hurewicz image of tmf at $p = 3$. Behrens, Mahowald and Quigley [2] calculate the Hurewicz image of tmf at $p = 2$. Since $\pi_*(\text{tmf}[\frac{1}{6}]) = \mathbb{Z}[\frac{1}{6}, a_4, a_6]$ is concentrated on the Adams–Novikov 0-line, our work together with the $p = 2$ case forms the complete determination of the Hurewicz image of tmf at all primes.

Following the strategy outlined above, much of the work involves showing that v_2^9 is a permanent cycle in the Adams–Novikov spectral sequence computing $\pi_*(S/(3, v_1^8))$; this is [Theorem 4.6](#). All of our explicit calculations are in the Adams–Novikov spectral sequence for $S/3$. To relate this to $S/(3, v_1^8)$ we use a lemma due to the second author ([Lemma 4.1](#)) that relates v_1^m -extensions in the Adams–Novikov spectral sequence for $S/3$ to differentials in the Adams–Novikov spectral sequence for $S/(3, v_1^m)$. Combined with Oka’s result [10] that $S/(3, v_1^m)$ is a ring spectrum for $m \geq 2$, this implies the existence of a v_2^9 self-map.

Corollary 4.7 *For $2 \leq m \leq 8$, there is a nonzero self-map*

$$v_2^9: \Sigma^{144} S/(3, v_1^m) \rightarrow S/(3, v_1^m).$$

There is also a similar result for $m = 9$, but correction terms for v_2^9 are needed; see [Remark 4.8](#).

Our proof that v_2^9 is a permanent cycle in the Adams–Novikov spectral sequence computing $\pi_*(S/(3, v_1^8))$ relies on analysis of the 143 stem in the Adams–Novikov spectral sequence for the sphere. This is greatly aided by software written by Wang [21; 20], which computes the E_2 page of the Adams–Novikov spectral sequence for the sphere using the algebraic Novikov spectral sequence. In addition, the software computes multiplication by p , α_1 and arbitrary $\beta_{i/j}$ elements. Wang’s software was originally written for use at $p = 2$; the minor modifications we used to change the prime are available at [5] and data, charts and more documentation are available at [4]. The calculations that make use of computer data occur solely in [Section 3](#).

We now comment on the overlap between this work and the preprint [16] by the second author: both works construct v_2^9 and the families $\beta_{9t+3/3}$, $\beta_{9t+6/3}$, and $\beta_{9t+9/8}$, but we find the methods here to be more straightforward. The earlier preprint uses the machinery of infinite descent to control the complexity of the Adams–Novikov spectral

sequence, while we opt to work directly with the Adams–Novikov E_2 page, controlling the complexity using Wang’s program. In particular, our analysis in Section 3, which is the crucial input for the construction of v_2^9 , follows from the β_1 –multiplication structure given by the computer calculations, as most of the elements in play are highly β_1 –divisible.

We conclude this section by giving an outline of the rest of the paper. In Section 2 we state notational conventions for the rest of the paper and write down some easy facts about the Adams–Novikov spectral sequence that will be used extensively in the remaining sections. Most of the work for proving Theorem 4.6 occurs in Section 3, which makes use of computer calculations to determine the Adams–Novikov spectral sequence for $S/3$ near the 143 stem. Theorem 4.6, which constructs v_2^9 , is proved in Section 4. In Section 5 we prove Theorem 5.1, which constructs the promised β families. This involves explicit calculations in a tractable range of stems to prove that $v_1^2 v_2^3$, $v_1 v_2^6$ and $\alpha_1 v_1 v_2^3$ in $E_2(S/(3, v_1^4))$, and $\alpha_1 v_1 v_2^7$ in $E_2(S/(3, v_1^2))$, are permanent cycles. In Section 6 we determine the 3–primary Hurewicz image of tmf (Theorem 6.5).

Acknowledgements Belmont would like to thank Paul Goerss for suggesting this project, and for many helpful conversations along the way. She would also like to thank Guozhen Wang for explaining some aspects of his program, and for some code changes.

2 Notation and preliminaries

At a fixed prime p , the Brown–Peterson spectrum BP has coefficient ring $\text{BP}_* = \mathbb{Z}_{(p)}[v_1, v_2, \dots]$ with $|v_i| = 2p^i - 2$, and ring of co-operations $\text{BP}_*\text{BP} = \text{BP}_*[t_1, t_2, \dots]$ with $|t_i| = 2p^i - 2$. Given a finite p –local spectrum X , the Adams–Novikov spectral sequence

$$(2) \quad E_2 = \text{Ext}_{\text{BP}_*\text{BP}}^{*,*}(\text{BP}_*, \text{BP}_* X) \Rightarrow \pi_*(X)_{(p)}$$

converges. Henceforth everything will be implicitly localized at the prime $p = 3$. The E_2 page of (2) can be calculated as the cohomology of the normalized cobar complex

$$(3) \quad \text{BP}_* \xrightarrow{\eta_R - \eta_L} \overline{\text{BP}_*\text{BP}} \rightarrow \overline{\text{BP}_*\text{BP}}^{\otimes 2} \rightarrow \dots,$$

although this is not an efficient means of computation. For further background on the Adams–Novikov spectral sequence, see [6] and [14, Sections 4.3–4.4].

Let $E_r^{s,f}(X)$ denote the E_r page of (2), restricted to stem s and Adams–Novikov filtration f . We say that an element in $\pi_s X$ is detected in filtration f if it is represented by a nonzero class in $E_\infty^{s,f}(X)$. Throughout, any equality of homotopy or E_2 page elements should be understood to be true up to units (that is, up to signs).

We will make frequent use of the cofiber sequence

$$S \xrightarrow{3} S \xrightarrow{i} S/3 \xrightarrow{j} \Sigma S.$$

We will also consider the cofiber sequences

$$\Sigma^{4m} S/3 \xrightarrow{v_1^m} S/3 \xrightarrow{i_m} S/(3, v_1^m) \xrightarrow{j_m} \Sigma^{4m+1} S/3 \quad \text{for } m \geq 1,$$

eventually focusing primarily on $m = 8$. Henceforth, degree shifts in cofiber and long exact sequences will usually not be shown. The maps i, j, i_m and j_m induce maps of Adams–Novikov spectral sequences, which we will denote with the same letters. We note the effect on degrees: given $x \in E_2^{s,f}(S/3)$, we have $j(x) \in E_2^{s-1, f+1}(S)$; given $x \in E_2^{s,f}(S/(3, v_1^m))$, we have $j_m(x) \in E_2^{s-4m-1, f+1}(S/3)$. The maps i and i_m preserve degrees. Sometimes we omit applications of i or i_m in the notation for brevity; for example, we write $\beta_1 \in E_2^{10,2}(S/3)$ to refer to $i(\beta_1)$. This is justified by regarding $E_2(S/3)$ as a module over $E_2(S)$.

By the geometric boundary theorem (see [14, Theorem 2.3.4]), the maps induced by j and j_m on E_2 pages coincide with the boundary maps in the long exact sequences of Ext groups

$$\begin{aligned} \text{Ext}_{\text{BP}_* \text{BP}}^{*,*}(\text{BP}_*, \text{BP}_*/3) &\rightarrow \text{Ext}_{\text{BP}_* \text{BP}}^{*+1,*}(\text{BP}_*, \text{BP}_*), \\ \text{Ext}_{\text{BP}_* \text{BP}}^{*,*}(\text{BP}_*, \text{BP}_*/(3, v_1^m)) &\rightarrow \text{Ext}_{\text{BP}_* \text{BP}}^{*+1,*}(\text{BP}_*, \text{BP}_*/3). \end{aligned}$$

Definition 2.1 We will say that an element $x \in E_2^{*,*}(S/3)$ is a *bottom cell element* if it is in the image of $i : E_2^{*,*}(S) \rightarrow E_2^{*,*}(S/3)$. An element $x \in E_2^{*,*}(S/3)$ is a *top cell element* if its image under the boundary map $j : E_2^{*,*}(S/3) \rightarrow E_2^{*-1,*}(S)$ is nonzero.

Notation 2.2 (i) If $x \in E_2^{s,f}(S)$ is 3–torsion, we will let $\bar{x} \in E_2^{s+1, f-1}(S/3)$ denote a class such that $j(\bar{x}) = x$. Note that \bar{x} may not always be uniquely determined.

(ii) If $x \in E_2(X)$ is a permanent cycle converging to $y \in \pi_*(X)$, write $y = \{x\}$.

In the rest of this section we present some preliminaries that are important for working with the Adams–Novikov spectral sequences for $S, S/3$ and $S/(3, v_1^m)$. All of these

facts are well known, and the rest of this section can be skipped by a knowledgeable reader. First we recall some frequently encountered permanent cycles in the 3–primary Adams–Novikov spectral sequence for the sphere. The comparisons below to the Adams spectral sequence are not needed in the rest of the paper, and are just presented for those readers who are more familiar with the Adams elements; a reference for computational facts about the Adams spectral sequence E_2 page is [14, Section 3.4], and for the corresponding Adams–Novikov elements is [9] or [6, Section 6] for the Greek letter construction and [14, Theorem 4.4.20] for low stems.

- $\alpha_1 \in E_2^{3,1}(S)$ is represented by $[t_1]$ in the cobar complex (3), and is called h_0 in the Adams spectral sequence.
- $\beta_1 \in E_2^{10,2}(S)$ equals the Massey product $\langle \alpha_1, \alpha_1, \alpha_1 \rangle$, and is called $b_0 = b_{10}$ in the Adams spectral sequence.
- $\beta_2 \in E_2^{26,2}(S)$ is called $k = k_0 = \langle h_0, h_1, h_1 \rangle$ in the Adams spectral sequence. (This does not correspond to an Adams–Novikov Massey product since h_1 does not exist in the Adams–Novikov E_2 page.)

The 0–line (generated by just $1 \in E_2^{*,0}(S)$) and the 1–line $E_2^{*,1}(S)$ consist of the image of the J homomorphism. These classes are all permanent cycles; the image of the 1–line under i is $\alpha_1 v_1^m$ for $m \geq 0$. The 0–line of $E_2(S/3)$ is the polynomial algebra on v_1 .

Fact 2.3 *Let $X = S, S/3,$ or $S/(3, v_1^m)$ for $m \geq 1$.*

- (i) $E_2^{s,f}(X) = 0$ if $s + f \not\equiv 0 \pmod{4}$.
- (ii) $E_2^{s,f}(X) = E_5^{s,f}(X)$.

Proof See [14, Proposition 4.4.2] for the statement about $X = S$. This sparseness for the sphere also implies the first statement for $X = S/3$ and $S/(3, v_1^m)$, as can be seen by looking at the degrees of the long exact sequences in Ext groups corresponding to the short exact sequences

$$\text{BP}_* \xrightarrow{3} \text{BP}_* \rightarrow \text{BP}_*/3 \quad \text{and} \quad \text{BP}_*/3 \xrightarrow{v_1^m} \text{BP}_*/3 \rightarrow \text{BP}_*/(3, v_1^m).$$

The second statement follows from the first. □

Most of our calculations in the Adams–Novikov spectral sequence for $S/(3, v_1^m)$ for $m \geq 2$ implicitly use the following fact.

Theorem 2.4 [10] For $m \geq 2$, $S/(3, v_1^m)$ is a ring spectrum.

It is also well known that $S/3$ is a ring spectrum.

Lemma 2.5 (i) If $x \in E_{10}(S)$, then $\beta_1^6 x$ is zero in $E_{10}(S)$. If $x \in E_{10}(S/3)$, then $\beta_1^6 x$ is zero in $E_{10}(S/3)$.

(ii) If $x \in E_6(S)$, then $\alpha_1 \beta_1^3 x$ is zero in $E_6(S)$. If $x \in E_6(S/3)$, then $\alpha_1 \beta_1^3 x$ is zero in $E_6(S/3)$.

(iii) We have $v_1^2 \cdot \beta_1 = 0$ in $E_2(S/3)$.

Proof For (i), the classical differential $d_9(\alpha_1 \beta_4) = \beta_1^6$ (see Table 1) implies that $\beta_1^6 = 0$ in $E_{10}(S)$, and hence $\beta_1^6 = 0$ in $E_{10}(S/3)$. Part (ii) is an analogous consequence of the Toda differential $d_5(\beta_{3/3}) = \alpha_1 \beta_1^3$. Part (iii) is [17, Lemma 2.13]. \square

Next, we record some basic facts about transferring differentials by naturality across various Adams–Novikov spectral sequences.

Lemma 2.6 Let $m \geq 1$.

(i) If there is a nontrivial differential $d_5(x) = y$ in $E_5(S)$ where y is not 3–divisible, then there is a nontrivial differential $d_5(i(x)) = i(y)$ in $E_5(S/3)$.

(ii) If there is a nontrivial differential $d_5(j(x)) = j(y)$ in $E_5(S)$ for $x, y \in E_5(S/3)$, then $d_5(x) \neq 0$, and $d_5(x) \equiv y \pmod{\text{Im}(i)}$.

Proof For (i), by naturality of i , there is a differential $d_5(i(x)) = i(y)$. We just need to check that $i(y)$ is nonzero in $E_5(S/3)$. This follows from the fact that $E_2(S/3) = E_5(S/3)$ and the assumption that $i(y)$ is nonzero in $E_2(S/3)$. For (ii), since j commutes with the differential and $E_2(S/3) = E_5(S/3)$, we have that $d_5(x) \equiv y \pmod{\ker(j)}$. The long exact sequence

$$\dots \xrightarrow{3} E_2(S) \xrightarrow{i} E_2(S/3) \xrightarrow{j} E_2(S) \xrightarrow{3} \dots$$

implies that $\ker(j) = \text{Im}(i)$. \square

We use the following lemma without further mention when working with β elements, applying it to the case $x = v_2^i$.

Lemma 2.7 *Let $m \geq 1$. For any $x \in E_2(S/(3, v_1^m))$, we have $j_m(x) = j_{m+k}(v_1^k x)$.*

Proof The map of short exact sequences

$$\begin{array}{ccccc}
 \mathrm{BP}_*/3 & \xrightarrow{v_1^m} & \mathrm{BP}_*/3 & \longrightarrow & \mathrm{BP}_*/(3, v_1^m) \\
 = \downarrow & & \downarrow v_1^k & & \downarrow v_1^k \\
 \mathrm{BP}_*/3 & \xrightarrow{v_1^{m+k}} & \mathrm{BP}_*/3 & \longrightarrow & \mathrm{BP}_*/(3, v_1^{m+k})
 \end{array}$$

induces a map of long exact sequences after applying $\mathrm{Ext}_{\mathrm{BP}_*\mathrm{BP}}(\mathrm{BP}_*, -)$. In particular, we have a commutative diagram

$$\begin{array}{ccc}
 \overbrace{\mathrm{Ext}_{\mathrm{BP}_*\mathrm{BP}}(\mathrm{BP}_*, \mathrm{BP}_*/(3, v_1^m))}^{E_2(S/(3, v_1^m))} & \xrightarrow{j_m} & \overbrace{\mathrm{Ext}_{\mathrm{BP}_*\mathrm{BP}}(\mathrm{BP}_*, \mathrm{BP}_*/3)}^{E_2(S/3)} \\
 \downarrow v_1^k & & \downarrow = \\
 \mathrm{Ext}_{\mathrm{BP}_*\mathrm{BP}}(\mathrm{BP}_*, \mathrm{BP}_*/(3, v_1^{m+k})) & \xrightarrow{j_{m+k}} & \mathrm{Ext}_{\mathrm{BP}_*\mathrm{BP}}(\mathrm{BP}_*, \mathrm{BP}_*/3)
 \end{array} \quad \square$$

3 Computer-assisted calculations in the 143 stem

In this section we study the Adams–Novikov spectral sequence for $S/3$ in the 143 stem and nearby stems; this is the main technical input needed for [Theorem 4.6](#). We make use of computer calculations of the Adams–Novikov E_2 page for the sphere; the specific facts from the computer data we use are given in [Lemma 3.1](#). The results from this section that are used later are [Lemma 3.2](#) and [Proposition 3.7](#). The former follows immediately from the \mathbb{F}_3 -vector space structure of $E_2^{*,*}(S)$. The rest of the section is devoted to proving the latter, which says that every permanent cycle in $\pi_{143}(S/3)$ is detected in filtration ≤ 5 . This requires more careful analysis using the multiplicative structure of the E_2 page. [Lemmas 3.5](#) and [3.6](#) give the differentials responsible for killing higher filtration elements in $E_2^{143,*}(S/3)$.

We encourage the reader to refer to the Adams–Novikov chart in [\[4\]](#) while reading this section. [Table 1](#) is a summary of this data: all of the differentials in the chart in [\[4\]](#) are derived from α_1 , β_1 and β_2 -multiples of the classes in [Table 1](#). Here x_{57} is the generator of $E_2^{57,3}(S)$, x_{75} is the generator of $E_2^{75,5}(S)$, and x_{96} is the generator of $E_2^{96,4}(S)$. Moreover, the differentials are complete through stem 108.

source		source	d_r	target	reason
s	f				
34	2	$\beta_{3/3}$	d_5	$\alpha_1\beta_1^3$	Toda differential
57	3	x_{57}	d_5	$\beta_1^3\beta_2$	forced by [14, Table A3.4]
58	2	β_4	d_5	$\alpha_1\beta_1^2\beta_{3/3}$	forced by [14, Table A3.4]
61	6	$\alpha_1\beta_4$	d_9	β_1^6	forced by [14, Table A3.4]
89	3	nonzero class	d_5	nonzero class	forced by [14, Table A3.4]
96	4	x_{96}	d_5	$\beta_1^2x_{75}$	see Lemma 3.4

Table 1: Some classical Adams–Novikov differentials.

- Lemma 3.1** [4; 5] (i) We have $\dim(E_2^{81,3}(S)) = 2$, $E_2^{81,7}(S) = \mathbb{F}_3\{\alpha_1\beta_1^2\beta_4\}$ and $\dim(E_2^{81,f}(S)) = 0$ if $f \neq 3, 7$.
- (ii) $E_2^{95,9}(S) = \mathbb{F}_3\{\beta_1^2x_{75}\}$, where x_{75} is the generator of $E_2^{75,5}(S)$, and we have $\alpha_1\beta_1^2x_{75} \neq 0$. The only other generator in $E_2^{95,\geq 9}(S)$ is $\alpha_1\beta_1^4\beta_2^2 \in E_2^{95,13}(S)$.
- (iii) We have that $\dim(E_2^{99,5}(S)) = 2$ and $\dim(\alpha_1 E_2^{99,5}) = 1$, and one of the generators of $E_2^{99,5}(S)$ is α_1x_{96} , where x_{96} is the generator of $E_2^{96,4}(S)$. Moreover, $E_2^{99,17}(S) = \mathbb{F}_3\{\alpha_1\beta_1^7\beta_2\}$, and $E_2^{99,f}(S) = 0$ for $f \neq 5, 17$.
- (iv) $E_2^{135,5}(S) = 0 = E_2^{134,6}(S)$.
- (v) $E_2^{141,15}(S) = \beta_1^6 E_2^{81,3}(S)$, and this group has dimension 2.
- (vi) Figure 2, left, displays the vector space structure of $E_2^{s,f}(S)$ for $140 \leq s \leq 144$, as well as selected multiplicative structure.

In Figure 2, left, the names in $E_2^{141,15}(S)$ follow from the proof of Lemma 3.3; other names are multiplications computed using Wang’s program.

Lemma 3.2 If $x \in E_2^{143,5}(S/3)$ is v_1^2 -divisible, then $x = 0$.

Proof Lemma 3.1(iv) implies $E_2^{135,5}(S/3) = 0$. □

Lemma 3.3 We have that $E_2^{81,3}(S)$ is 2-dimensional, and both generators are permanent cycles.

Proof By [14, Table A3.4], $\pi_{81}(S)_3^\wedge$ is 2-dimensional, and is generated by γ_2 and $\langle \alpha_1, \alpha_1, \beta_5 \rangle$. Lemma 3.1(i) gives the structure of $E_2^{81,*}(S)$. It suffices to show that

$\alpha_1\beta_1^2\beta_4$ supports a nontrivial differential; Table 1 implies $d_9(\alpha_1\beta_1^2\beta_4) = \beta_1^8$. Moreover, it is clear from an $E_2(S)$ chart (see [4]) that β_1^8 cannot be the target of a shorter differential. □

Lemma 3.4 *There is a differential*

$$d_5(x_{96}) = \beta_1^2x_{75}.$$

The \mathbb{F}_3 -vector space $\alpha_1 E_2^{99,5}(S)$ is 1-dimensional and is generated by a class α_1x_{99} , where x_{99} is a permanent cycle.

See Lemma 3.1 for element definitions. The second sentence is used implicitly when identifying one of the generators of $E_2^{142,14}(S)$ as $\alpha_1\beta_1^4x_{99}$ (as seen in Figure 2, left): Wang’s program only shows that there is a nonzero element in $E_2^{142,14}(S)$ that is $\alpha_1\beta_1^4$ times an element of $E_2^{99,5}(S)$.

Proof By [14, Table A3.4], $\pi_{96}(S)_3^\wedge = 0$, and the generator $x_{96} \in E_2^{96,4}(S)$ must support a nontrivial differential as it cannot be a target for degree reasons. We claim this implies a differential $d_5(x_{96}) = \beta_1^2x_{75}$: by Lemma 3.1(ii) the only other possible target is $\alpha_1\beta_1^4\beta_2^2 \in E_2^{95,13}(S)$ (a possibility for $d_9(x_{96})$), but this is zero in E_6 by Lemma 2.5(ii) as it is $\alpha_1\beta_1^3$ times the permanent cycle $\beta_1\beta_2^2$.

From Lemma 3.1(ii), we have that $\alpha_1d_5(x_{96}) = d_5(\alpha_1x_{96}) = \alpha_1\beta_1^2x_{75}$ is nonzero. By [14, Table A3.4], we have $\pi_{99}(S)_3^\wedge / \text{Im } J \cong \mathbb{F}_3$. We claim this permanent cycle

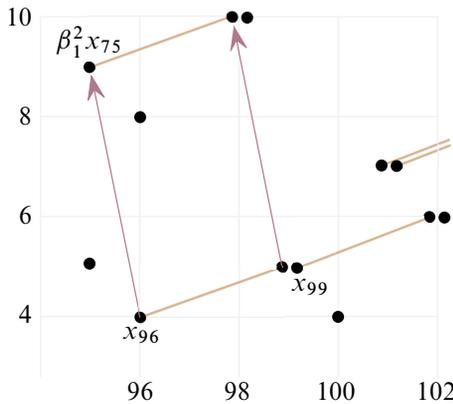


Figure 1: $E_2^{s,f}(S)$ in degrees $95 \leq s \leq 102$ and $4 \leq f \leq 10$. Brown lines represent α_1 -multiplication. Each dot represents a copy of \mathbb{F}_3 . The information in this chart used in the proof of Lemma 3.4 is summarized in Lemma 3.1(ii) and (iii).

is detected in filtration 5. By Lemma 3.1(iii), the only other possibility is $\alpha_1 \beta_1^7 \beta_2 \in E_2^{99,17}(S)$, which is the target of a d_5 differential by Lemma 2.5(ii). Let x_{99} denote the permanent cycle in $E_2^{99,5}(S)$. Since $\alpha_1^2 = 0$ and $\dim(\alpha_1 E_2^{99,5}(S)) = 1$ by Lemma 3.1(iii), we have that $\alpha_1 E_2^{99,5}(S)$ is generated by $\alpha_1 x_{99}$. \square

Lemma 3.5 *The generator of $E_2^{142,10}(S)$ supports a nontrivial Adams–Novikov d_5 differential. The generator of $E_2^{142,6}(S)$ supports a nontrivial Adams–Novikov d_9 differential.*

Proof Combining Lemma 3.1(v) with Lemma 3.3, we have that the 2–dimensional vector space $E_2^{141,15}(S)$ is generated by β_1^6 –divisible permanent cycles. By Lemma 2.5(i), both classes in $E_2^{141,15}(S)$ are hit by some differential. By the vector space structure of $E_2^{142,*}(S)$ displayed in Figure 2, left, the only possibilities are the indicated d_5 and d_9 . \square

Lemma 3.6 *The generator of $E_2^{143,9}(S)$ supports a nontrivial Adams–Novikov d_5 differential hitting $\alpha_1 \beta_1^4 x_{99}$, where x_{99} is the permanent cycle introduced in Lemma 3.4. One of the two generators of $E_2^{143,5}(S)$ supports a nontrivial Adams–Novikov d_9 differential hitting $\beta_1^6 \beta_{6/3}$.*

Proof This proof relies on Figure 2, left, in particular the fact that the elements mentioned are all nonzero. For the first statement, we have $d_5(\beta_{3/3} \cdot \beta_1 x_{99}) = \alpha_1 \beta_1^3 \cdot \beta_1 x_{99}$ since x_{99} (and hence $\beta_1 x_{99}$) is a permanent cycle. Since $\beta_{6/3} \in E_2^{82,2}(S)$ is a permanent cycle by [14, Table A3.4], we may apply Lemma 2.5(i) to show that $\beta_1^6 \beta_{6/3} \in E_2^{142,14}(S)$ is the target of a differential d_r for $r \leq 9$. Since the group $E_2^{143,9}(S)$ is one-dimensional and we proved above that the generator supported a nontrivial d_5 , the element $\beta_1^6 \beta_{6/3}$ must be hit by a d_9 . \square

Proposition 3.7 *Every element in $\pi_{143}(S/3)$ is detected in Adams–Novikov filtration ≤ 5 .*

Proof We list the elements in $E_2^{143,f}(S/3)$ for $f > 5$ in Table 2.

We encourage the reader to refer to Figure 2 alongside the rest of the proof: the diagram on the right is derived from that on the left.

Filtration 9 We claim that both classes $E_2^{143,9}(S/3)$ support d_5 differentials. The bottom cell class does so because of Lemmas 3.6 and 2.6(i), and the top cell class does so because of Lemmas 3.5 and 2.6(ii).

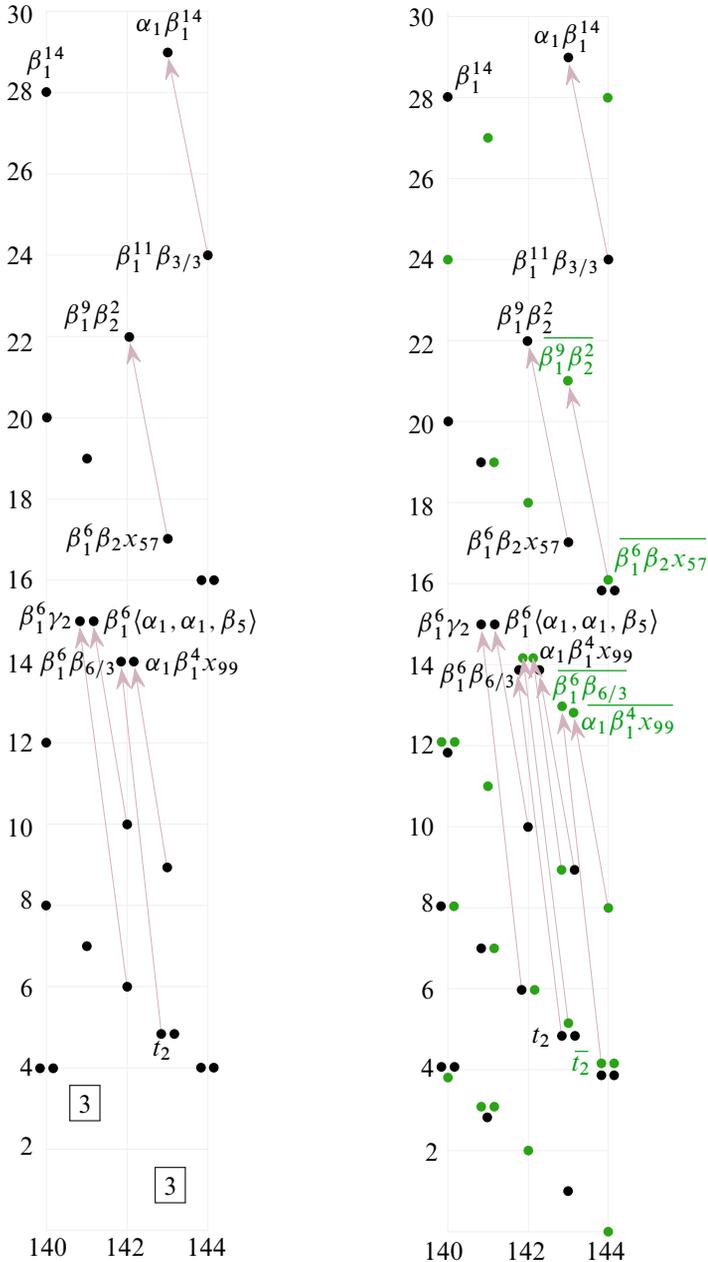


Figure 2: Left: $E_2^{s,f}(S)$ in degrees $140 \leq s \leq 144$, along with some Adams–Novikov differentials. A box containing “3” denotes a copy of $\mathbb{Z}/27$. Multiplications by α_1 are not shown. Right: $E_2^{s,f}(S/3)$ in degrees $140 \leq s \leq 144$, along with some Adams–Novikov differentials. Green dots denote top cell classes. Multiplications by α_1 are not shown.

filtration f	# bottom cell generators	# top cell generators
9	1	1
13	0	2
17	1	0
21	0	1
29	1	0

Table 2: Classes in $E_2^{143,f}(S/3)$ for $f \geq 2$.

Filtration 13 We may take the two generators of $E_2^{143,13}(S/3)$ to be classes $\alpha_1\beta_1^4x_{99}$ and $\beta_1^6\beta_{6/3}$ defined such that their image under j is $\alpha_1\beta_1^4x_{99}$ and $\beta_1^6\beta_{6/3}$, respectively. By Lemma 2.6(ii), the d_5 in Lemma 3.6 induces a d_5 differential hitting $\alpha_1\beta_1^4x_{99}$; note that $\text{Im}(i) = 0$ in this degree.

By Lemma 3.6 we have a class $t_2 \in E_2^{143,5}(S)$ such that $d_9(t_2) = \beta_1^6\beta_{6/3}$. Let \bar{t}_2 be the top cell class in $E_2^{144,4}(S/3)$ associated to the 3-torsion element t_2 . We wish to show that there is a differential $d_9(\bar{t}_2) = \beta_1^6\beta_{6/3}$. First we check that \bar{t}_2 survives to the E_9 page. The only possible targets for such a shorter differential are in $E_2^{143,9}(S/3)$, and we showed above that these both support nontrivial d_5 differentials. The map induced by j on E_2 pages shows that $d_9(\bar{t}_2) \equiv \beta_1^6\beta_{6/3}$ modulo $\ker(j)$. We have $E_9^{143,13}(S/3) = \mathbb{F}_3\{\beta_1^6\beta_{6/3}\}$, and $j(\beta_1^6\beta_{6/3}) = \beta_1^6\beta_{6/3}$, which is nonzero in $E_9(S)$. Thus there is a nonzero d_9 differential as claimed.

Filtration 17 The generator of $E_2^{143,17}(S)$ is $\beta_1^6\beta_2x_{57}$, where x_{57} is the generator of $E_2^{57,3}(S)$. Using a differential in Table 1, we have a differential $d_5(\beta_1^6\beta_2x_{57}) = \beta_1^9\beta_2^2$. By Lemma 2.6(i) we have a differential $d_5(i(\beta_1^6\beta_2x_{57})) = i(\beta_1^9\beta_2^2)$.

Filtration 21 By Lemma 2.6(ii), the d_5 differential on $\beta_1^6\beta_2x_{57}$ discussed in the filtration 17 case above gives rise to a differential $d_5(\beta_1^6\beta_2x_{57}) = \beta_1^9\beta_2^2$ over $S/3$.

Filtration 29 The generator of $E_2^{143,29}(S/3)$ is $i(\alpha_1\beta_1^{14})$; this class is zero in $E_6(S/3)$ by Lemma 2.5(ii). □

Remark 3.8 The dependence of Proposition 3.7 on computer calculations would be reduced if we could make precise the observation that much of the Adams–Novikov E_2 -page is β_1 -periodic, and classes in high filtrations are highly β_1 -divisible. Using [12, Theorem 2.3.1, Remark 2.3.5(c)], one can prove that multiplication by β_1 is an

isomorphism on the Adams E_2 page restricted to Adams filtration f_A , stem s , and filtration ν in the algebraic Novikov spectral sequence $\text{Ext}_A^{*,*}(\mathbb{F}_3, \mathbb{F}_3) \Rightarrow E_2^{*,*}(S)$ if

$$f_A > \frac{1}{23}s + \frac{24}{23}\nu + \frac{159}{23}.$$

By keeping track of the effect on the algebraic Novikov spectral sequence, one can derive that β_1 acts injectively (up to higher algebraic Novikov filtration) on the subspace of $E_2^{s,f}(S)$ in algebraic Novikov filtration ν if

$$(4) \quad f > \frac{1}{23}s + \frac{1}{23}\nu + \frac{169}{23}.$$

Surjectivity is harder to prove. Even if we knew that β_1 acted isomorphically on the region (4) (which is often true), this is not enough to prove the β_1 -divisibility results we need. For example, in Proposition 3.7 we use the fact that the generator x of $E_2^{143,17}(S)$ is divisible by β_1^6 . This element has $\nu = 0$, and $\beta_1^{-1}x$ and $\beta_1^{-2}x$ lie in the region (4) but $\beta_1^{-3}x$ does not. Improving this bound would also be of use more generally to the study of the 3-primary Adams and Adams–Novikov spectral sequences.

4 Survival of v_2^9

In this section, we prove Theorem 4.6, which says that v_2^9 is a permanent cycle in $E_2(S/(3, v_1^8))$. We first explain the choice of exponent of v_1 . Since $\eta_R(v_2) \equiv v_2 + v_1 t_1^3 - v_1^3 t_1 \pmod{3}$ in the Hopf algebra (BP_*, BP_*BP) (see eg [14, (6.4.16)]), we have that v_2^3 is an element of $E_2(S/(3, v_1^m))$ for $m \leq 3$, and v_2^9 is an element of $E_2(S/(3, v_1^m))$ for $m \leq 9$. On the other hand, we would like to work with $m \geq 8$, since those are the values of m for which $\beta_{9/8}$ is in the image of the composition of Adams–Novikov E_2 page boundary maps $E_2(S/(3, v_1^m)) \rightarrow E_2(S/3) \rightarrow E_2(S)$. Trivial modifications to the work in this section show that $v_2^9 \pm v_1^8 v_2^7$ is a self-map on $S/(3, v_1^9)$; see Remark 4.8. However, this slight strengthening is not necessary for our purposes, and we write down our results for $v_2^9 \in \pi_*(S/(3, v_1^8))$ essentially for cosmetic reasons, avoiding the correction term. To obtain the families in Theorem 5.1 other than $\beta_{9t+9/j}$, it suffices to work with $S/(3, v_1^4)$.

The main ingredients for proving Theorem 4.6 are Lemma 3.2 and Proposition 3.7 from the previous section, and the following lemma (below, specialized to our setting) due to the second author. It draws a connection between hidden v_1^8 -extensions in $\pi_*(S/3)$, and differentials of the minimum length (ie d_5 differentials) in the Adams–Novikov spectral sequence for $S/(3, v_1^8)$.

Lemma 4.1 [15, Lemma 1.4] *Let $m \geq 1$. Suppose we have $y \in E_5(S/(3, v_1^m))$ such that $j_m(y)$ is a nontrivial permanent cycle in $E_5^{s,f}(S/3)$, and let w denote an element in $E_5(S/3)$ detecting the product $v_1^m \cdot \{j_m(y)\} \in \pi_*(S/3)$. Then there is a differential*

$$d_5(y) = i_m(w)$$

in $E_5(S/(3, v_1^m))$.

In the next lemma we separate out the general strategy used to prove that v_2^9 and other elements in Section 5 are permanent cycles.

Lemma 4.2 (i) *Let $x \in E_2^{s,f}(S/3)$ for $f \leq 3$ be such that $j(x) \in E_2^{s-1,f+1}(S)$ is a permanent cycle and $\{j(x)\} \in \pi_{s-1}(S)$ is an essential element of order 3. Furthermore, suppose that $\text{Im}(i: E_2^{s,f}(S) \rightarrow E_2^{s,f}(S/3))$ consists of permanent cycles. Then $x \in E_2^{s,f}(S/3)$ is a permanent cycle.*

(ii) *Let $x \in E_2^{s,f}(S/(3, v_1^m))$ for $f \leq 3$ be such that $j_m(x) \in E_2^{s-4m-1,f+1}(S/3)$ is a permanent cycle and $\{j_m(x)\}$ is an essential element with $v_1^m \cdot \{j_m(x)\} = 0$ in $\pi_*(S/3)$. Furthermore, suppose that $\text{Im}(i_m: E_2^{s,f}(S/3) \rightarrow E_2^{s,f}(S/(3, v_1^m)))$ consists of permanent cycles. Then $x \in E_2^{s,f}(S/(3, v_1^m))$ is a permanent cycle.*

Proof We just prove (i), as (ii) is analogous. Consider the exact sequences

$$(5) \quad \begin{array}{ccccccc} E_2^{s,f}(S) & \xrightarrow{3} & E_2^{s,f}(S) & \xrightarrow{i} & E_2^{s,f}(S/3) & \xrightarrow{j} & E_2^{s-1,f+1}(S), \\ \pi_s(S) & \xrightarrow{3} & \pi_s(S) & \xrightarrow{i} & \pi_s(S/3) & \xrightarrow{j} & \pi_{s-1}(S) \xrightarrow{3} \pi_{s-1}(S), \end{array}$$

associated to the cofiber sequence $S \xrightarrow{3} S \xrightarrow{i} S/3 \xrightarrow{j} S$. (For the first long exact sequence, we are using the fact that j induces the zero map in BP-homology.) Suppose $x \in E_2^{s,f}(S/3)$ is an element such that $j(x)$ is a permanent cycle with $3 \cdot \{j(x)\} = 0$. Then there exists an element $\xi \in \pi_s(S/3)$ such that $j(\xi) = \{j(x)\}$. Since $j: S/3 \rightarrow \Sigma S$ induces a map of Adams–Novikov spectral sequences, the induced map on homotopy $j: \pi_*(S/3) \rightarrow \pi_*(\Sigma S) = \pi_{*-1}(S)$ respects Adams–Novikov filtration; thus $j(x)$ being detected in filtration $f + 1$ implies ξ is detected in filtration $\leq f$. The assumption $f \leq 3$ combined with Fact 2.3 implies that ξ is detected in filtration f . We may write the detecting element as $x + y$ for some $y \in E_2^{s,f}(S/3)$. By the geometric boundary theorem, $j(x + y)$ converges to $j(\xi)$, and we also have that $j(x)$ converges to $j(\xi)$. So $j(y)$ is a boundary. But $j(y)$ has filtration ≤ 4 , so Fact 2.3 implies $j(y) = 0$ in $E_2(S)$. By (5), we have that y is in the image of i . By the assumption about $\text{Im}(i)$, y is a permanent cycle, and we have from above that $x + y$ is a permanent cycle. Therefore, x is a permanent cycle. □

Lemma 4.3 The element $\overline{\beta_{9/8}} = j_8(v_2^9) \in E_2^{111,1}(S/3)$ is a permanent cycle.

Proof By [14, Table A3.4], there exists $c \in \{\pm 1\}$ such that $x_{106} = \beta_{9/9} + c\beta_7$ in $E_2^{106,2}(S)$ is a 3-torsion permanent cycle. Lemma 4.2(i) applies since $E_2^{*,1}(S)$ consists of permanent cycles; it implies that $\overline{\beta_{9/9}} + c\overline{\beta_7} = j_9(v_2^9) + cj_1(v_2^7) \in E_2^{107,1}(S/3)$ is a permanent cycle. Hence $v_1 \cdot (j_9(v_2^9) + cj_1(v_2^7)) = v_1 \cdot j_9(v_2^9) = j_8(v_2^9)$ is a permanent cycle. □

Lemma 4.4 We have that $d_5(v_2^9) = 0$ in $E_5^{143,5}(S/(3, v_1^8))$.

Proof Let $x = d_5(v_2^9)$. We first consider the image of this differential along the natural map induced by $i'_3: S/(3, v_1^8) \rightarrow S/(3, v_1^3)$. Since v_2^3 is an element of $E_2^{48,0}(S/(3, v_1^3))$, by the Leibniz rule and Theorem 2.4, we have $i'_3(x) = i'_3(d_5(v_2^9)) = d_5((v_2^3)^3) = 3v_2^6 d_5(v_2^3) = 0$.

Using Lemma 4.3, we have

$$j_8(x) = j_8(d_5(v_2^9)) = d_5(j_8(v_2^9)) = 0$$

in $E_5(S/3) = E_2(S/3)$. Thus the exact sequence

$$E_2^{143,5}(S/3) \xrightarrow{i_8} E_2^{143,5}(S/(3, v_1^8)) \xrightarrow{j_8} E_2^{111,1}(S/3)$$

gives $x = i_8(y)$ for some $y \in E_2^{143,5}(S/3)$.

Consider the commutative diagram of cofiber sequences obtained using Verdier’s axiom:

$$(6) \quad \begin{array}{ccccc} S/3 & \xrightarrow{v_1^3} & S/3 & \xrightarrow{i_3} & S/(3, v_1^3) \\ i_5 \downarrow & & \downarrow i_8 & & \parallel \\ S/(3, v_1^5) & \xrightarrow{v_1^3} & S/(3, v_1^8) & \xrightarrow{i'_3} & S/(3, v_1^3) \\ j_5 \downarrow & & \downarrow j_8 & & \downarrow \\ S/3 & \xlongequal{\quad} & S/3 & \longrightarrow & * \end{array}$$

This implies that $i_3 = i'_3 \circ i_8$; in particular, $i_3(y) = i'_3(i_8(y)) = i'_3(x) = 0$. By the long exact sequence corresponding to the top row of (6), we have that y is v_1^3 -divisible. By Lemma 3.2, $y = 0$. □

Lemma 4.5 The product $v_1^8 \cdot \{j_8(v_2^9)\}$ is zero in $\pi_*(S/3)$.

Proof Let $w \in E_5^{143, f}(S/3)$ be a representative of $v_1^8 \cdot \{j_8(v_2^9)\} \in \pi_{143}(S/3)$. By [Proposition 3.7](#), we have $f \leq 5$, and by [Fact 2.3](#), the only possibilities are $f = 1, 5$. The product $v_1^8 \cdot j_8(v_2^9)$ is zero on the E_2 page, so we must have $f = 5$.

By [Lemma 4.3](#), [Lemma 4.1](#) applies to v_2^9 ; combining this with [Lemma 4.4](#), we have

$$0 = d_5(v_2^9) = i_8(w) \quad \text{in } E_5^{143, 5}(S/(3, v_1^8)) = E_2^{143, 5}(S/(3, v_1^8)).$$

Thus w is divisible by v_1^8 in $E_2^{143, 5}(S/3)$. By [Lemma 3.2](#), we must have $w = 0$ in $E_5^{143, 5}(S/3)$. Since w was defined to be an element detecting $v_1^8 \cdot \{j_8(v_2^9)\}$, this product is zero in homotopy. □

Theorem 4.6 *The element $v_2^9 \in E_2^{144, 0}(S/(3, v_1^8))$ is a permanent cycle in the Adams–Novikov spectral sequence computing $\pi_*(S/(3, v_1^8))$.*

Proof This will follow from applying [Lemma 4.2\(ii\)](#) to v_2^9 . The first two hypotheses of that lemma are satisfied due to [Lemmas 4.3](#) and [4.5](#). For the last hypothesis, note that $E_2^{144, 0}(S/3)$ is generated by v_1^{36} , which is a permanent cycle. □

Corollary 4.7 *For $2 \leq m \leq 8$, the class v_2^9 in $\pi_{144}(S/(3, v_1^m))$ lifts to a class v_2^9 in $[S/(3, v_1^m), S/(3, v_1^m)]_{144}$.*

Proof Naturality of the map $S/(3, v_1^8) \rightarrow S/(3, v_1^m)$ for $m \leq 8$ means that [Theorem 4.6](#) directly implies $v_2^9 \in E_2^{144, 0}(S/(3, v_1^m))$ is a permanent cycle. By [\[10, Theorem 6.1\]](#), $R = S/(3, v_1^m)$ is a (homotopy) ring spectrum for $m \geq 2$. Thus the desired self-map may be obtained as

$$R \rightarrow S \wedge R \xrightarrow{v_2^9 \wedge I} R \wedge R \xrightarrow{\mu} R. \quad \square$$

Remark 4.8 Essentially the same argument shows that $v_2^9 \pm v_1^8 v_2^7$ is a permanent cycle in $E_2^{144, 0}(S/(3, v_1^9))$. In the proof of [Lemma 4.3](#) we show that $j_9(v_2^9) \pm j_1(v_2^7) = j_9(v_2^9 \pm v_1^8 v_2^7)$ is a permanent cycle. The proofs of [Lemmas 4.4](#) and [4.5](#) go through without modification to show that $d_5(v_2^9 \pm v_1^8 v_2^7) = 0$ in $E_5^{143, 5}(S/(3, v_1^9))$ and $v_1^9 \cdot \{j_9(v_2^9 \pm v_1^8 v_2^7)\} = 0$. In the proof of [Theorem 4.6](#), we have $x = c_1(v_2^9 \pm v_1^8 v_2^7) + c_2 v_2^9 + c_3 v_1^{36}$. This time, v_2^9 and $v_1^8 v_2^7$ are not permanent cycles since $\beta_{9/9} = j(j_9(v_2^9))$ and $\beta_7 = j(j_9(v_1^8 v_2^7))$ are not permanent cycles, and v_1^{36} is a permanent cycle. Thus $v_2^9 \pm v_1^8 v_2^7$ is a permanent cycle for some choice of sign.

5 Survival of beta elements

Our goal in this section is to prove that several infinite families of $\beta_{a/b}$ elements are permanent cycles in the Adams–Novikov spectral sequence for the sphere. For indices a and b satisfying the conditions in [9, Theorem 2.6], Miller, Ravenel, and Wilson define cycles $\beta_{a/b}$ in $E_2^{*,2}(S)$ as the image of certain classes in $E_2^{*,0}(S/(3, v_1^b))$ under the composition $j \circ j_b$. In this section we will only consider classes $\beta_{sp^n/b}$ (with $p \nmid s$) such that $b \leq p^n$, which enables us to use the equivalent, but simpler, definition

$$\beta_{a/b} = j(j_b(v_2^a)) \in E_2^{16a-4b-2,2}(S)$$

(at $p = 3$). These elements are defined using the boundary maps j and j_b on Ext associated to the short exact sequences

$$BP_* \xrightarrow{3} BP_* \rightarrow BP_*/3 \quad \text{and} \quad BP_*/3 \xrightarrow{v_1^b} BP_*/3 \rightarrow BP_*/(3, v_1^b);$$

by the geometric boundary theorem these coincide with the maps induced on Adams–Novikov spectral sequences by the maps j and j_b of spectra that we have been considering in this paper. Recall the convention that $\beta_a := \beta_{a/1}$.

Suppose that $\beta_{a/b}$ is a permanent cycle with $b \leq 8$ and, in addition, suppose that the corresponding element in homotopy $\beta_{a/b}^h \in \pi_*(S)$ factors as

$$(7) \quad \beta_{a/b}^h : S \xrightarrow{B_{a/b}} S/(3, v_1^b) \xrightarrow{jj_b} S \quad \text{for some } B_{a/b} \in \pi_*(S/(3, v_1^b)).$$

In this case, for $t \geq 1$, Corollary 4.7 allows us to define elements in $\pi_*(S)$:

$$\beta_{9t+a/b}^h : S \xrightarrow{B_{a/b}} S/(3, v_1^b) \xrightarrow{(v_2^9)^t} S/(3, v_1^b) \xrightarrow{jj_b} S.$$

We warn that existence of a factorization (7) is not automatic, even if $\beta_{a/b}$ is a permanent cycle and such a factorization exists on the level of Adams–Novikov E_2 pages. Our goal is to show the following, proved at the end of the section.

Theorem 5.1 *For all $t \geq 0$, the classes*

$$\begin{aligned} &\beta_{9t+3/j} \quad \text{for } j = 1, 2, && \beta_{9t+6/j} \quad \text{for } j = 1, 2, 3, \\ &\beta_{9t+9/j} \quad \text{for } j = 1, \dots, 8, && \alpha_1 \beta_{9t+3/3} \quad \text{and} \quad \alpha_1 \beta_{9t+7} \end{aligned}$$

are permanent cycles in the Adams–Novikov spectral sequence for the sphere.

Since $\beta_{3/3}$ and β_7 support Adams–Novikov differentials, none of the families in [Theorem 5.1](#) are trivially multiplicative consequences of a different family. Instead, we have $\alpha_1\beta_{3/3} \in \langle \alpha_1, \alpha_1, \beta_1^3 \rangle$ and $\alpha_1\beta_7 \in \langle \alpha_1, \alpha_1, \beta_1^2\beta_{6/3} \rangle$. As we will see in [Section 6](#), the families $\alpha_1\beta_{9t+3/3}$, $\beta_{9t+6/3}$ and $\alpha_1\beta_{9t+7}$ have nontrivial image in $\pi_* \text{tmf}$, along with the family β_{9t+1} constructed in [\[3, Corollary 1.2\]](#).

Lemma 5.2 *The class $j_4(v_1^2v_2^3) \in E_2^{39,1}(S/3)$ is a permanent cycle such that*

$$j(j_4(v_1^2v_2^3)) = \beta_{3/2} \in E_2^{38,2}(S) \quad \text{and} \quad v_1^4 \cdot \{j_4(v_1^2v_2^3)\} = 0 \in \pi_*(S/3).$$

Proof We have $j(j_4(v_1^2v_2^3)) = j(j_2(v_2^3)) = \beta_{3/2}$ in $E_2(S)$ by the definition of the β elements along with [Lemma 2.7](#). By classical computations of the Adams–Novikov E_2 page (see eg [\[14, Figure 1.2.19\]](#)), $E_2^{38,f}(S) = 0 = E_2^{37,f+1}(S)$ for $f \geq 3$, so $E_2^{38,f}(S/3) = 0$ for $f \geq 3$. Thus $j_4(v_1^2v_2^3) \in E_2^{39,1}(S/3)$ cannot support a differential of any length.

As $v_1^4 \cdot j_4 = 0$ as a map $E_2(S/(3, v_1^4)) \rightarrow E_2(S/3)$, it remains to rule out hidden v_1^4 -extensions on $\beta'_{3/2} := \{j_4(v_1^2v_2^3)\} \in \pi_{39}(S/3)$. Using [\[14, Table A3.4\]](#) we have $\pi_{51}(S/3) = \mathbb{F}_3\{\alpha_{13}, \beta_1^5\}$, and so $v_1^3 \cdot \beta'_{3/2} = c\beta_1^5 = c\beta_1^2 \cdot \beta_1^3$ for some $c \in \mathbb{F}_3$. (If there were an α_{13} component, then the extension would not be hidden.) We have $v_1 \cdot \beta_1^2 = 0$ for degree reasons, as Ravenel’s table implies $\pi_{25}(S/3) = 0$. Thus $v_1 \cdot v_1^3 \cdot \beta'_{3/2} = 0$. \square

Lemma 5.3 *The class $j_4(v_1v_2^6) \in E_2^{83,1}(S/3)$ is a permanent cycle such that*

$$j(j_4(v_1v_2^6)) = \beta_{6/3} \in E_2^{82,2}(S) \quad \text{and} \quad v_1^4 \cdot \{j_4(v_1v_2^6)\} = 0 \in \pi_*(S/3).$$

Proof By Ravenel’s table [\[14, Table A3.4\]](#), $\beta_{6/3}$ and β_6 are 3–torsion permanent cycles. Since $j(j_4(v_1v_2^6)) = \beta_{6/3}$ and $j(j_4(v_1^3v_2^6)) = \beta_6$, we apply [Lemma 4.2\(i\)](#) to $j_4(v_1v_2^6) \in E_2^{83,1}(S/3)$ and $j_4(v_1^3v_2^6) \in E_2^{91,1}(S/3)$, noting that $E_2^{*,1}(S)$ consists of permanent cycles. This shows that $j_4(v_1v_2^6)$ and $j_4(v_1^3v_2^6)$ are permanent cycles.

To determine $v_1^4 \cdot \{j_4(v_1v_2^6)\}$, we first consider the possibilities for $v_1^2 \cdot \{j_4(v_1v_2^6)\} \in \pi_{91}(S/3)$: from Ravenel’s tables, we have

$$\pi_{91}(S/3) = \mathbb{F}_3\{\alpha_{23}, \beta_1\gamma_2, \beta_1x_{81}, \beta_6'\},$$

where $j(\beta_6') = \beta_6$. Since $v_1^2 \cdot \beta_1 = 0$ in homotopy by [Lemma 2.5\(iii\)](#), we have $v_1^2 \cdot \{j_4(v_1v_2^6)\} = c_1\alpha_{23} + c_2\beta_6'$ for $c_i \in \mathbb{F}_3$. If $c_1 \neq 0$, then $v_1^4 \cdot \{j_4(v_1v_2^6)\}$ would be detected in filtration 1, contradicting the fact that $v_1^4 \cdot j_4(v_1v_2^6) = 0$ in $E_2(S/3)$. So

it suffices to show that $v_1^2 \cdot \beta'_6 = 0$. From above, we may write $\beta'_6 = \{j_4(v_1^3 v_2^6)\} = \{j_2(v_1 v_2^6)\}$. By [11, Lemma 3], $v_1 v_2^6 \in E_2^{100,0}(S/(3, v_1^2))$ is a permanent cycle, and hence so is $j_2(v_1 v_2^6)$. We have $\{j_2(v_1 v_2^6)\} = j_2(\{v_1 v_2^6\})$ by the geometric boundary theorem, and $v_1^2 \cdot j_2(\{v_1 v_2^6\}) = 0$ by definition of j_2 as a map on homotopy groups. \square

Lemma 5.4 *The classes $v_1^2 v_2^3 \in E_2^{56,0}(S/(3, v_1^4))$ and $v_1 v_2^6 \in E_2^{100,0}(S/(3, v_1^4))$ are permanent cycles in the Adams–Novikov spectral sequence computing $\pi_*(S/(3, v_1^4))$.*

Proof Use Lemma 4.2(ii), with Lemmas 5.2 and 5.3 as input. To check the condition about the image of $i_4 : E_2(S/3) \rightarrow E_2(S/(3, v_1^4))$ in these degrees, note that $E_2^{*,1}(S/3)$ is generated by the image of $i : E_2^{*,1}(S) \rightarrow E_2^{*,1}(S/3)$, which consists of permanent cycles (these are all image of J classes), along with elements that map to β elements under j . Standard theory about the β elements [9] implies that $\overline{\beta_{3/2}}$ and $\overline{\beta_{6/3}}$ are the only such elements in the relevant degrees; these are both permanent cycles by Lemmas 5.2 and 5.3. \square

Lemma 5.5 *The class $\alpha_1 v_1 v_2^3 \in E_2^{55,1}(S/(3, v_1^4))$ is a permanent cycle in the Adams–Novikov spectral sequence.*

Proof There is a Toda bracket $\langle \alpha_1, \alpha_1, \beta_3^1 \rangle \in \pi_{37}(S)$ detected by $\alpha_1 \beta_{3/3}$ in filtration 3, and this class is 3–torsion; see [14, Table A3.4]. In order to apply Lemma 4.2(i) to $j_4(\alpha_1 v_1 v_2^3) \in E_2^{38,2}(S/3)$, we must check that $E_2^{38,2}(S)$ consists of permanent cycles. It follows from standard facts about the Adams–Novikov 2–line [9] that $E_2^{38,2}(S) = \mathbb{F}_3\{\beta_{3/2}\}$. So we may conclude that $j_4(\alpha_1 v_1 v_2^3)$ is a permanent cycle.

Moreover, $v_1^4 \cdot \{j_4(\alpha_1 v_1 v_2^3)\}$ is zero in homotopy: since $\pi_{53}(S) = 0 = \pi_{54}(S)$ by [14, Table A3.4], we have $\pi_{54}(S/3) = 0$. In order to apply Lemma 4.2(ii) to $\alpha_1 v_1 v_2^3 \in E_2^{55,1}(S/(3, v_1^4))$, we must check that $E_2^{55,1}(S/3)$ consists of permanent cycles. This is true because the image of $E_2^{*,1}(S)$ consists of permanent cycles, and analysis of the 2–line reveals that there cannot be a class with nontrivial image in $E_2^{54,2}(S)$. Thus we have that $\alpha_1 v_1 v_2^3$ is a permanent cycle. \square

Lemma 5.6 *The class $\alpha_1 v_1 v_2^7 \in E_2^{119,1}(S/(3, v_1^2))$ is a permanent cycle in the Adams–Novikov spectral sequence.*

Proof Since $S/(3, v_1^2)$ is a ring spectrum (Theorem 2.4), we may consider this element as a product $v_1 v_2^5 \cdot \alpha_1 v_2^2$. Oka [11, Lemma 2] showed that v_2^5 is a permanent cycle in $E_2^{80,0}(S/(3, v_1))$. This implies that $v_1 v_2^5$ is a permanent cycle in $E_2^{84,0}(S/(3, v_1^2))$.

Next we consider possible differentials on $\alpha_1 v_2^2 \in E_2^{35,1}(S/(3, v_1^2))$. An element in $E_2^{34,f}(S/(3, v_1^2))$ either has nonzero image under j_2 in $E_2^{25,f+1}(S/3)$ or is the image under i_2 of an element of $E_2^{34,f}(S/3)$. From classically known computations of the Adams–Novikov E_2 page (see eg [14, Figure 1.2.19]), we deduce that

$$E_2^{25,*}(S/3) = 0 \quad \text{and} \quad E_2^{35,\geq 3}(S/3) = \mathbb{F}_3\{\overline{\alpha_1\beta_1^3}\}.$$

This implies $E_2^{34,\geq 3}(S/(3, v_1^2))$ is generated by $i_2(\overline{\alpha_1\beta_1^3})$. Observe that $\overline{\alpha_1\beta_1^3} = \overline{\alpha_1}\beta_1^3 = v_1 \cdot \beta_1^3$. Thus the only possible nonzero differential on $v_1 v_2^5 \cdot \alpha_1 v_2^2$ is a d_5 with target $v_1 v_2^5 \cdot v_1 \beta_1^3$. But the target is divisible by v_1^2 , hence zero in $E_5(S/(3, v_1^2))$. \square

Proof of Theorem 5.1 We show that $\beta_{9t+3/2}$ and $\beta_{9t+3/1}$ are permanent cycles for $t \geq 0$. Since v_2^9 is a permanent cycle in $E_2(S/(3, v_1^8))$ by Theorem 4.6, its image in $E_2(S/(3, v_1^4))$ is a permanent cycle. Lemma 5.4 says that $v_1^2 v_2^3$ is a permanent cycle in $E_2(S/(3, v_1^4))$, so the product $v_1^2 v_2^3 \cdot v_2^{9t} \in E_2(S/(3, v_1^4))$ is a permanent cycle. Recall that $\beta_{9t+3/2} \in E_2(S)$ is defined as $j(j_2(v_2^{9t+3})) = j(j_4(v_1^2 v_2^{9t+3}))$ in $E_2(S/3)$. Since $j_4(v_1^2 v_2^{9t+3})$ is a permanent cycle, so is $j(j_4(v_1^2 v_2^{9t+3}))$. Since $v_1^2 v_2^3$ is a permanent cycle in $E_2(S/(3, v_1^4))$, so is $v_1^3 v_2^3$, so $\beta_{9t+3/1} = j(j_1(v_2^{9t+3})) = j(j_4(v_1^3 v_2^{9t+3}))$ is a permanent cycle in $E_2(S)$.

The family $\beta_{9t+9/8}$ (and hence $\beta_{9t+9/j}$ for $j < 8$) follows directly from the fact that v_2^9 is a permanent cycle in $E_2(S/(3, v_1^8))$. The other families of permanent cycles follow analogously, using Lemma 5.4 again as the input for $\beta_{9t+6/3} = j(j_4(v_1 v_2^{9t+6}))$, Lemma 5.5 as the input for $\alpha_1 \beta_{9t+3/3} = j(j_4(\alpha_1 v_1 v_2^{9t+3}))$, and Lemma 5.6 as the input for $\alpha_1 \beta_{9t+7} = j(j_2(\alpha_1 v_1 v_2^{9t+7}))$. \square

6 3–Primary Hurewicz image of tmf

In this section we determine the image of the Hurewicz map $h: \pi_* S \rightarrow \pi_* \text{tmf}$ induced by the unit map $S \rightarrow \text{tmf}$. The target $\pi_* \text{tmf}$ has been computed via the elliptic spectral sequence (see [1, Section 3]); this is the $Y(4)$ –based Adams spectral sequence for tmf, where $Y(4)$ is the Thom spectrum of $\Omega U(4) \rightarrow \mathbb{Z} \times BU$. We will denote this spectral sequence by $E_r^{\text{ell}}(\text{tmf})$.

Theorem 6.1 (Hopkins–Mahowald, Bauer [1, Section 6]) *At $p = 3$, $\pi_* \text{tmf}$ is generated by $c_4, c_6, \Delta, \alpha, \beta$ and b , subject to the relation $c_4^3 - c_6^2 = 1728\Delta$ and the relations on the other generators displayed in Figure 3. Multiplication by $\Delta^3 \in \pi_{72}(\text{tmf})$ is injective.*

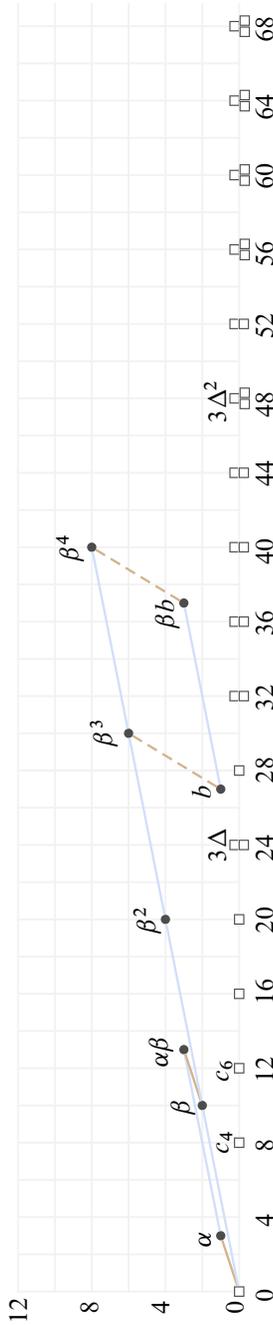


Figure 3: The E_∞ page of the elliptic spectral sequence computing $\pi_s \text{tmf}$ for $0 \leq s \leq 76$. Dashed brown lines represent hidden α -multiples. Squares indicate copies of $\mathbb{Z}_{(3)}$ and dots indicate copies of \mathbb{F}_3 .

We will show (Theorem 6.5) that all classes in filtration ≥ 2 are in the Hurewicz image, and the only classes in filtrations 0 and 1 in the image are the summands generated by 1 and α . Instead of directly mapping to the elliptic spectral sequence, we use the $K(2)$ -local E -based Adams spectral sequence

$$E_2^E(\text{TMF}) = H^*(G_{24}; E_*) \Rightarrow \pi_*(L_{K(2)}\text{TMF})$$

where $E = E_2$ is height 2 Morava E -theory and TMF is the periodic version of tmf . There is a map of spectral sequences $E_r(S) \rightarrow E_r^E(\text{TMF})$ induced by the natural maps $\text{BP} \rightarrow E$ and $S \rightarrow \text{TMF}$. Henn, Karamanov and Mahowald [7, Theorem 1.1] completely determine $E_2^E(\text{TMF}/3)$ and provide formulas that we use to compute the map on E_2 pages $E_2(S) \rightarrow E_2^E(\text{TMF})$ in cases of interest; see Lemmas 6.2 and 6.3. For each class in $E_2^E(\text{TMF})$ in filtration ≥ 2 , we identify a preimage in $E_2(S)$ that is among the classes proved to be permanent cycles in Theorem 5.1 or [3]; see Proposition 6.4. As we explain in the proof of Theorem 6.5, it suffices to understand the Hurewicz image in $\pi_*(L_{K(2)}\text{TMF})$ because there is an injection $\pi_*(\text{tmf}) \rightarrow \pi_*(L_{K(2)}\text{TMF})$; see Lemma 6.6.

First we review some notation and basic facts. We have $E_*/3 = \mathbb{F}_9\llbracket u_1 \rrbracket\llbracket u^{\pm 1} \rrbracket$, and there is a natural map $\text{BP}_* \rightarrow E_*$ that sends $v_1 \mapsto u_1 u^{-2}$, $v_2 \mapsto u^{-8}$ and $v_i \mapsto 0$ for $i > 2$. Abusing notation, we will let v_i denote its image in $E_*/3$.

Recall $j : S/3 \rightarrow \Sigma S$ denotes the boundary map in the cofiber sequence $S \xrightarrow{3} S \rightarrow S/3$. We will also use j to refer to the map $j \wedge \text{TMF} : \text{TMF}/3 \rightarrow \Sigma \text{TMF}$. Similarly, j_m will denote both boundary maps $S/(3, v_1^m) \rightarrow S/3$ and $\text{TMF}/(3, v_1^m) \rightarrow \text{TMF}/3$, depending on context.

Lemma 6.2 *In E_* we have*

$$\begin{aligned} v_2^3 &\equiv -\Delta^2 - v_1^2 v_2 \Delta && \text{mod } (3, v_1^6), \\ v_2^6 &\equiv \Delta^4 - v_1^2 v_2 \Delta^3 && \text{mod } (3, v_1^3), \\ v_2^{3^n} &\equiv -\Delta^{2 \cdot 3^{n-1}} - v_1^{2 \cdot 3^{n-1}} v_2^{3^{n-1}} \Delta^{3^{n-1}} && \text{mod } (3, v_1^{2 \cdot 3^n}). \end{aligned}$$

Proof The formula $\Delta \equiv (1 - \omega^2 u_1^2 + u_1^4) \omega^2 u^{-12} \text{ mod } (3, u_1^6)$ from [7, Proposition 5.1] implies

$$\begin{aligned} \Delta^2 &\equiv (1 - 2\omega^2 u_1^2 + u_1^4)(-v_2^3) && \text{mod } (3, v_1^6), \\ v_1^2 v_2 \Delta &\equiv v_2^3 (\omega^2 u_1^2 + u_1^4) && \text{mod } (3, v_1^6). \end{aligned}$$

where ω denotes an 8th root of unity in \mathbb{F}_9 . Combining these facts, we obtain the formula for v_2^3 ; the formulas for v_2^6 and $v_2^{3^n}$ follow from it by squaring and successive cubing, respectively. □

Let

$$\begin{aligned} H &: E_2(S) \rightarrow E_2^E(\text{TMF}), \\ H' &: E_2(S/3) \rightarrow E_2^E(\text{TMF}/3), \\ H'_m &: E_2(S/(3, v_1^m)) \rightarrow E_2^E(\text{TMF}/(3, v_1^m)), \end{aligned}$$

denote the natural maps of spectral sequences.

Lemma 6.3 *We have*

$$\begin{aligned} H(\alpha_1) &= \alpha, & H'(j_3(v_2^3)) &\doteq \Delta \tilde{\alpha}, & H'(j_3(v_1^2 v_2^7)) &\doteq \Delta^4 \tilde{\alpha}, \\ H(\beta_1) &\doteq \beta, & H(\beta_{3/3}) &\doteq \Delta \beta, & H(\beta_7) &\doteq \Delta^4 \beta, \\ H'(j_3(v_1^2 v_2)) &\doteq \tilde{\alpha}, & H'(j_3(v_2^6)) &\doteq \Delta^3 \tilde{\alpha}, \\ & & H(\beta_{6/3}) &\doteq \Delta^3 \beta, \end{aligned}$$

where $j(\tilde{\alpha}) = \beta$. (Here \doteq denotes equality up to multiplication by a unit.)

Proof Following Bauer [1, Section 6], we have $H(\alpha_1) = \alpha$ since they both come from the cobar class $[t_1]$, and $H(\beta_1) = \beta$ because of the Massey products $\beta_1 = \langle \alpha_1, \alpha_1, \alpha_1 \rangle$ and $\beta = \langle \alpha, \alpha, \alpha \rangle$. We have $j(\tilde{\alpha}) = \beta$ and $j(j_3(v_1^2 v_2)) = \beta_1$, so

$$j(H'(j_3(v_1^2 v_2))) = H(j(j_3(v_1^2 v_2))) = \beta.$$

This specifies $H'(j_3(v_1^2 v_2))$ up to the image of $E_2^E(\text{TMF})$, but since $E_2^E(\text{TMF}/3)$ is 1-dimensional in the degree of $\tilde{\alpha}$, there is no ambiguity.

For the next column, we have in $E_2^E(\text{TMF}/(3, v_1^3))$ that

$$H'(j_3(v_2^3)) = j_3(H'_3(v_2^3)) = j_3(-\Delta^2 - v_1^2 v_2 \Delta) = -\Delta j_3(v_1^2 v_2) = -\Delta \cdot \tilde{\alpha},$$

using Lemma 6.2 and the earlier fact about $H'(j_3(v_1^2 v_2))$. Note that $j_3(\Delta^n) = 0$ since Δ^n is in the image of $E_2^E(\text{TMF}/3)$. Now apply j to get the statement about $H(\beta_{3/3})$. The remaining facts in this column are analogous, using the fact that $\beta_{6/3} = j(j_3(v_2^6))$. The last column is also proved similarly, using the fact that $\beta_7 = j(j_3(v_1^2 v_2^7))$. □

By our convention about naming elements in the image of the map $\text{BP}_* \rightarrow E_*$, we have $H'_3(v_2) = v_2$.

Proposition 6.4 For $t \geq 0$ the map $H: E_2(S) \rightarrow E_2^E(\text{TMF})$ satisfies

- (i) $H(\beta_{9t+1}) \doteq \Delta^{6t} \beta,$
- (ii) $H(\beta_{9t+3/3}) \doteq \Delta^{6t+1} \beta,$
- (iii) $H(\beta_{9t+6/3}) \doteq \Delta^{6t+3} \beta,$
- (iv) $H(\beta_{9t+7}) \doteq \Delta^{6t+4} \beta.$

Proof These statements are all proved the same way; we show (ii). First observe that Lemma 6.2 implies $v_2^9 \equiv -\Delta^6 \pmod{(3, v_1^6)}$. Using Lemmas 6.2 and 6.3 we have

$$\begin{aligned} H(\beta_{9t+3/3}) &= H(j(j_3(v_2^{9t+3}))) = j(j_3(H'_3(v_2^{9t+3}))) \\ &= j(j_3(H'_3(v_2^3) \cdot H'_3(v_2^{9t}))) = j(j_3((-\Delta^2 - v_1^2 v_2 \Delta) \cdot (-1)^t \Delta^{6t})) \\ &= j(j_3((-1)^{t+1} \Delta^{6t+2})) + j(j_3((-1)^{t+1} v_1^2 v_2 \Delta^{6t+1})) \\ &= 0 + (-1)^{t+1} \Delta^{6t+1} j(j_3(v_1^2 v_2)) = (-1)^{t+1} \Delta^{6t+1} \beta. \end{aligned}$$

The last line uses the fact that $j_3(v_1^2 v_2) = \tilde{\alpha}$ in $E_2^E(\text{TMF}/3)$ from Lemma 6.3. □

In the next theorem, we show that every element in $\pi_* \text{tmf}$ detected in filtration ≥ 2 is in the Hurewicz image. This result is stated without proof in [8, Section 1], but we do not know of any prior proof in the literature.

Theorem 6.5 The image of the map $h: \pi_* S \rightarrow \pi_* \text{tmf}$ at $p = 3$ consists of the $\mathbb{Z}_{(3)}$ summand generated by 1 and the \mathbb{F}_3 summands generated by

$$\alpha, \quad \Delta^{3t} \beta^i, \quad \Delta^{3t} \alpha \beta, \quad \Delta^{3t} \beta b \quad \text{for } 1 \leq i \leq 4 \text{ and } t \geq 0.$$

More precisely, we have

$$\begin{aligned} h(\alpha_1) &= \alpha, & h(\beta_1^{i-1} \beta_{9t+1}) &= \Delta^{6t} \beta^i & \text{for } 1 \leq i \leq 4, \\ h(\alpha_1 \beta_{9t+3/3}) &= \Delta^{6t} \beta b, & h(\beta_1^{i-1} \beta_{9t+6/3}) &= \Delta^{6t+3} \beta^i & \text{for } 1 \leq i \leq 4, \\ h(\alpha_1 \beta_{9t+7}) &= \Delta^{6t+3} \beta b. \end{aligned}$$

Proof Let $E_2^{\text{ell}}(\text{tmf})$ denote the elliptic spectral sequence for tmf (see [1, Section 6]); recall this is the $Y(4)$ -based Adams spectral sequence for tmf . There is a map of spectral sequences $L: E_r^{\text{ell}}(\text{tmf}) \rightarrow E_r^E(\text{TMF})$ that comes from the map on Adams

towers induced by the maps $Y(4) \rightarrow MU_P \rightarrow E$ (where MU_P denotes periodic MU) and $\text{tmf} \rightarrow \text{TMF}$. These maps assemble into a diagram of spectral sequences

$$(8) \quad \begin{array}{ccccc} E_2^{\text{ell}}(\text{tmf}) & \xrightarrow{L} & E_2^E(\text{TMF}) & \xleftarrow{H} & E_2(S) \\ \Downarrow & & \Downarrow & & \Downarrow \\ \pi_* \text{tmf} & \xleftarrow{L} & \pi_* L_{K(2)}\text{TMF} & \xleftarrow{H} & \pi_* S \end{array}$$

\xleftarrow{h}

No element $x \in E_\infty^{\text{ell},s,0}(\text{tmf})$ for $s \neq 0$ is in the image of h : **Lemma 6.6(ii)** implies x would be detected in filtration 0 of $E_\infty^E(\text{TMF})$, and $H: E_\infty(S) \rightarrow E_\infty^E(\text{TMF})$ is zero in filtration 0 for nonzero stems.

Next we turn to elements detected in filtration 1. We have $H(\alpha_1) = \alpha$ by **Lemma 6.3**; since we have $H = L \circ h$ as maps $\pi_* S \rightarrow \pi_* L_{K(2)}\text{TMF}$ and L is injective by **Lemma 6.6**, this implies $h(\alpha_1) = \alpha \in \pi_* \text{tmf}$. The other elements of $E_\infty^{\text{ell}}(\text{tmf})$ in filtration 1 are $\Delta^{3^t}\alpha$ for $t \geq 1$ and $\Delta^{3^t}b$ for $t \geq 0$; we will show that the permanent cycles they represent are not in the Hurewicz image. By **Lemma 6.6(i)** they are in the image of h if and only if their images in $\pi_* L_{K(2)}\text{TMF}$ are in the image of H . By **Lemma 6.6(ii)** they are also detected in filtration 1 in $E_\infty^E(\text{TMF})$, so if they were in the image of H , they would be the image of a class in $E_2(S)$ in filtration 0 or 1. We have $E_2^{s,0}(S) = 0$ for $s > 1$, so it suffices to show that the elements in $E_2^{s,1}(S)$ except for α_1 are in the kernel of h . If $x \in E_2^{s,1}(S)$ with $s > 3$ then $i(x) = \alpha_1 v_1^k$ for some $k \geq 1$. If h' denotes the map $\pi_*(S/3) \rightarrow \pi_*(\text{tmf}/3)$ induced by h , we have $h'(\alpha_1 v_1) = 0$ since $\pi_7(\text{tmf}/3) = 0$. Thus $i(h(x)) = h'(i(x)) = 0$ in $\pi_*(S/3)$, which implies that $h(x)$ is 3-divisible. But **Figure 3** shows that there are no 3-divisible nonzero targets in Adams–Novikov filtration 1.

We will now show how to use **Proposition 6.4** to derive the remaining claims about h ; for multiplicative reasons, it suffices to show $i = 1$ in those statements. We will illustrate this with the element $\alpha_1 \beta_{9t+3/3}$; the other elements are analogous, using **Theorem 5.1** for $\beta_{9t+6/3}$ or [3, Corollary 1.2] for β_{9t+1} in place of **Theorem 5.1** below as necessary. By **Proposition 6.4**, $H(\alpha_1 \beta_{9t+3/3}) = \Delta^{6t+1}\alpha\beta$ in $E_2^E(\text{TMF})$. Since $\Delta^{6t+1}\alpha\beta$ is a permanent cycle in $E_2^{\text{ell}}(\text{tmf})$ converging to $\Delta^{6t}\beta b$, we have that $\Delta^{6t+1}\alpha\beta$ is a permanent cycle in $E_2^E(\text{TMF})$ converging to $\Delta^{6t}\beta b$. **Theorem 5.1** shows that $\alpha_1 \beta_{9t+3/3}$ is a permanent cycle in the Adams–Novikov spectral sequence; write $\alpha_1 \beta_{9t+3/3}$ for the (non- α_1 -divisible) element in homotopy it converges to. The following diagram summarizes these statements by illustrating (8) applied to these

elements:

$$\begin{array}{ccccc}
 \Delta^{6t+1}\alpha\beta & \xrightarrow{L} & \Delta^{6t+1}\alpha\beta & \xleftarrow{H} & \alpha_1\beta_{9t+3/3} \\
 \vdots & & \vdots & & \vdots \\
 \Delta^{6t}\beta b & \xrightarrow{L} & \Delta^{6t}\beta b & \xleftarrow{H} & \alpha_1\beta_{9t+3/3} \\
 & & \xleftarrow{h} & &
 \end{array}$$

Thus $H: \pi_* S \rightarrow \pi_* L_{K(2)}\text{TMF}$ satisfies $H(\alpha_1\beta_{9t+3/3}) = \Delta^{6t}\beta b$. Since H factors through h and $L: \pi_* \text{tmf} \rightarrow \pi_* L_{K(2)}\text{TMF}$ is injective by Lemma 6.6(i), we have that $h(\alpha_1\beta_{9t+3/3}) = \Delta^{6t}\beta b$. □

Lemma 6.6 (i) *The map $L: \pi_* \text{tmf} \rightarrow \pi_* L_{K(2)}\text{TMF}$ is injective on the classes in Theorem 6.5.*

(ii) *The map $L: E_\infty^{\text{ell}}(\text{tmf}) \rightarrow E_\infty^E(\text{TMF})$ is injective in filtrations 0 and 1.*

In fact, L is injective on E_∞ pages in all filtrations, but we do not need this fact.

Proof (i) We have

$$\pi_*(L_{K(2)}\text{TMF}) = (\pi_*(\text{tmf})[(\Delta^3)^{-1}])_I^\wedge,$$

where $I = (3, c_4)$; see [8, Section 2]. It is clear from the calculation of $\pi_* \text{tmf}$ that the localization map $\pi_*(\text{tmf}) \rightarrow (\Delta^3)^{-1}\pi_*(\text{tmf})$ is an injection. It suffices to show that completion at I is injective on the specified classes. This holds because $0 = c_4 \cdot \alpha = c_4 \cdot \beta = c_4 \cdot b$ in $(\Delta^{24})^{-1}\pi_* \text{tmf}$ for degree reasons (and these classes are also all 3-torsion).

(ii) Consider an element of $\ker(L: E_\infty^{\text{ell}}(\text{tmf}) \rightarrow E_\infty^E(\text{TMF}))$ represented by $x \in E_2^{\text{ell}}(\text{tmf})$ in filtration 0 or 1. We claim that x is in $\ker(L_2: E_2^{\text{ell}}(\text{tmf}) \rightarrow E_2^E(\text{TMF}))$: since $L_2(x)$ is in filtration 0 or 1, it cannot be the target of a d_r differential for $r \geq 2$. By comparing the calculations of $E_2^{\text{ell}}(\text{tmf})$ and $E_2^E(\text{TMF}/3)$ in [1, Section 5] and [7, Theorem 1.1], respectively, it is clear that $L'_2: E_2^{\text{ell}}(\text{tmf}/3) \rightarrow E_2^E(\text{TMF}/3)$ is an injection, so the image of x in $E_2^{\text{ell}}(\text{tmf}/3)$ is zero, which implies (using exactness of the top row in the diagram) $x \in E_2^{\text{ell}}(\text{tmf})$ is 3-divisible:

$$\begin{array}{ccccc}
 E_2^{\text{ell}}(\text{tmf}) & \xrightarrow{3} & E_2^{\text{ell}}(\text{tmf}) & \xrightarrow{i} & E_2^{\text{ell}}(\text{tmf}/3) \\
 & & \downarrow L_2 & & \downarrow L'_2 \\
 & & E_2^E(\text{TMF}) & \xrightarrow{i} & E_2^E(\text{TMF}/3)
 \end{array}$$

Since $E_2^{\text{ell}}(\text{tmf})$ has no 3–divisible classes in filtration 1, we now focus on the filtration 0 case. Let $y = x/3^n \in E_2^{\text{ell}}(\text{tmf})$ be the non-3–divisible generator, which then has nonzero image $i(y)$ in $E_2^{\text{ell}}(\text{tmf}/3)$. Since L'_2 is an injection, $L'_2(i(y)) = i(L_2(y)) \neq 0$. We claim that the (nonzero) group generated by $L_2(y)$ is torsion-free: if not, then the corresponding top cell class would be a nonzero class in $E_2^E(\text{TMF}/3)$ in filtration -1 , contradicting [7, Theorem 1.1]. So $L_2(x) = 3^n L_2(y) \neq 0$, contradicting the fact above that $x \in \ker(L_2)$. \square

Remark 6.7 Our methods are not sufficient to completely determine the image of the map $h': \pi_*(S/3) \rightarrow \pi_*(\text{tmf}/3)$. The remaining nontrivial part of this question is to determine which elements $\Delta^n \alpha$ are in the image. Arguments similar to those we have given in this section show that $h'(\overline{\beta_{9t+2}}) = \Delta^{6t+1} \alpha$ and $h'(\overline{\beta_{9t+5}}) = \Delta^{6t+3} \alpha$. However, the families $\Delta^{6t} \alpha$ for $t \geq 1$ and $\Delta^{6t+4} \alpha$ for $t \geq 0$ fit into patterns that are not described by our work in this paper. For example, $\Delta^4 \alpha \in \pi_{99}(\text{tmf}/3)$ is not in the image of h' for degree reasons. On the other hand, using the more precise definitions of the β elements in [9, (2.4)] and calculating analogously to Lemma 6.3, we find that the map $E_2(S/3) \rightarrow E_2^E(\text{TMF}/3)$ sends $\overline{\beta_{18/11}}$ to $\Delta^{10} \alpha$. As we do not know if $\overline{\beta_{18/11}}$ is a permanent cycle, we are unable to conclude whether $\Delta^{10} \alpha$ is in the image of $\pi_*(S/3)$.

References

- [1] **T Bauer**, *Computation of the homotopy of the spectrum tmf*, from “Groups, homotopy and configuration spaces” (N Iwase, T Kohno, R Levi, D Tamaki, J Wu, editors), Geom. Topol. Monogr. 13, Geom. Topol. Publ., Coventry (2008) 11–40 [MR](#) [Zbl](#)
- [2] **M Behrens, M Mahowald, J D Quigley**, *The 2–primary Hurewicz image of tmf*, Geom. Topol. 27 (2023) 2763–2831 [MR](#)
- [3] **M Behrens, S Pemmaraju**, *On the existence of the self map v_2^9 on the Smith–Toda complex $V(1)$ at the prime 3*, from “Homotopy theory: relations with algebraic geometry, group cohomology, and algebraic K –theory” (P Goerss, S Priddy, editors), Contemp. Math. 346, Amer. Math. Soc., Providence, RI (2004) 9–49 [MR](#) [Zbl](#)
- [4] **E Belmont, G Wang**, *Adams–Novikov data*, online data set (2023) Available at https://github.com/ebelmont/ANSS_data/raw/master/anss_E2_158.pdf
- [5] **E Belmont, G Wang**, *MinimalResolution (code for computing algebraic Novikov spectral sequence, $p = 3$ version)*, source code (2023) Available at <https://github.com/ebelmont/MinimalResolution>

- [6] **P G Goerss**, *The Adams–Novikov spectral sequence and the homotopy groups of spheres*, lecture notes (2008) [arXiv 0802.1006](#)
- [7] **H-W Henn, N Karamanov, M Mahowald**, *The homotopy of the $K(2)$ -local Moore spectrum at the prime 3 revisited*, *Math. Z.* 275 (2013) 953–1004 [MR](#) [Zbl](#)
- [8] **A G Henriques**, *The homotopy groups of tmf and of its localizations*, *Mathematical Surveys and Monographs* 201, Amer. Math. Soc., Providence, RI (2014) [MR](#) [Zbl](#)
- [9] **H R Miller, D C Ravenel, W S Wilson**, *Periodic phenomena in the Adams–Novikov spectral sequence*, *Ann. of Math.* 106 (1977) 469–516 [MR](#) [Zbl](#)
- [10] **S Oka**, *Ring spectra with few cells*, *Japan. J. Math.* 5 (1979) 81–100 [MR](#) [Zbl](#)
- [11] **S Oka**, *Note on the β -family in stable homotopy of spheres at the prime 3*, *Mem. Fac. Sci. Kyushu Univ. Ser. A* 35 (1981) 367–373 [MR](#) [Zbl](#)
- [12] **J H Palmieri**, *Stable homotopy over the Steenrod algebra*, *Mem. Amer. Math. Soc.* 716, Amer. Math. Soc., Providence, RI (2001) [MR](#) [Zbl](#)
- [13] **D C Ravenel**, *The non-existence of odd primary Arf invariant elements in stable homotopy*, *Math. Proc. Cambridge Philos. Soc.* 83 (1978) 429–443 [MR](#) [Zbl](#)
- [14] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, *Pure and Applied Mathematics* 121, Academic, Orlando, FL (1986) [MR](#) [Zbl](#)
- [15] **K Shimomura**, *The homotopy groups of the L_2 -localized Toda–Smith spectrum $V(1)$ at the prime 3*, *Trans. Amer. Math. Soc.* 349 (1997) 1821–1850 [MR](#) [Zbl](#)
- [16] **K Shimomura**, *The existence of β_{9r+3} in the stable homotopy of spheres at the prime three*, preprint (2006) Available at <http://www.math.kochi-u.ac.jp/katsumi/paper/beta9-shimomura.pdf>
- [17] **K Shimomura**, *The beta elements $\beta_{tp2/r}$ in the homotopy of spheres*, *Algebr. Geom. Topol.* 10 (2010) 2079–2090 [MR](#) [Zbl](#)
- [18] **K Shimomura**, *Note on beta elements in homotopy, and an application to the prime three case*, *Proc. Amer. Math. Soc.* 138 (2010) 1495–1499 [MR](#) [Zbl](#)
- [19] **H Toda**, *On spectra realizing exterior parts of the Steenrod algebra*, *Topology* 10 (1971) 53–65 [MR](#) [Zbl](#)
- [20] **G Wang**, *Computations of the Adams–Novikov E_2 -term*, *Chinese Ann. Math. Ser. B* 42 (2021) 551–560 [MR](#) [Zbl](#)
- [21] **G Wang**, *MinimalResolution (code for computing algebraic Novikov spectral sequence)*, source code (2023) Available at <https://github.com/ebelmont/MinimalResolution>

*Department of Mathematics, Applied Mathematics, and Statistics
Case Western Reserve University
Cleveland, OH, United States*

*Department of Mathematics, Faculty of Science, Kochi University
Kochi, Japan*

eva.belmont@case.edu, katsumi@kochi-u.ac.jp

Received: 3 September 2021 Revised: 25 June 2022

Uniform foliations with Reeb components

JOAQUÍN LEMA

A foliation on a compact manifold is uniform if each pair of leaves of the induced foliation on the universal cover are at finite Hausdorff distance from each other. We study uniform foliations with Reeb components. We give examples of such foliations on a family of closed 3–manifolds with infinite fundamental group. Furthermore, we prove some results concerning the behavior of a uniform foliation with Reeb components on general 3–manifolds.

57R30

1 Introduction

Consider a foliation \mathcal{F} on a compact Riemannian 3–manifold M . This foliation lifts to a foliation $\tilde{\mathcal{F}}$ on the universal cover \tilde{M} . We will say that \mathcal{F} is *uniform* if any pair of leaves of $\tilde{\mathcal{F}}$ are at finite Hausdorff distance from each other.

A lot can be said about a uniform foliation if we further assume \mathcal{F} to be Reebless; see for example Fenley and Potrie [6] and Thurston [17]. In this paper, we will focus on the opposite case. More precisely, we will study the following question posed by Fenley and Potrie [6, Question 1]:

Question 1.1 If \mathcal{F} is uniform in M with infinite fundamental group, does it follow that \mathcal{F} is also Reebless?

Our first result will be to give a negative answer to the question (see Section 4):

Theorem 1.2 For every $l, m \in \mathbb{N}$ and every choice M_1, \dots, M_l of 3–manifolds with finite fundamental group, there exists a uniform foliation with Reeb components on $M = (\#_{i=1}^m S^1 \times S^2) \# (\#_{i=1}^l M_i)$.

One can arrange the foliations given by Theorem 1.2 to be C^∞ –smooth; see Remark 4.3. We will also give an example of uniform foliation with Reeb components on the solid torus, which is trivial on the boundary. However, these 3–manifolds are “small” in the

sense that they are the only ones not admitting an immersed essential closed surface of genus $g \geq 1$; see [Proposition 2.5](#). In this paper, we will say that an immersed surface is *essential* if the immersion induces an injective morphism from the fundamental group of the surface to $\pi_1(M)$.

We can say the following for the rest of the 3–manifolds:

Theorem 1.3 *Let \mathcal{F} be a uniform foliation on a compact 3–manifold M admitting an immersion $i: \Sigma \rightarrow M$ from a closed surface Σ of genus $g \geq 1$ such that $i_*: \pi_1(\Sigma) \rightarrow \pi_1(M)$ is injective. Then $i(\Sigma)$ must intersect the set of Reeb components.*

The idea of the proof is to put the immersion in general position with respect to the foliation, and then apply Poincaré–Bendixson theory on the induced foliation on $\tilde{\Sigma} \cong \mathbb{R}^2$.

[Theorem 1.3](#) and some remarks in [Section 3](#) motivate us to modify [Question 1.1](#): does an *irreducible* 3–manifold with infinite fundamental group admit a uniform foliation with Reeb components?

Acknowledgements I am profoundly grateful to my advisor Rafael Potrie for his patience and suggestions during every step of my thesis. This work could not be done without him. I would also like to thank Sergio Fenley for his comments on earlier versions of this paper, which motivated the statement of [Theorem 1.3](#).

The author was supported by a CAP grant for Master’s students at UdelaR.

2 Preliminaries

2.1 Foliations on 3–manifolds

We will work on a compact 3–manifold M endowed with a $C^{\infty,0+}$ codimension-one foliation \mathcal{F} . From now on, by foliation on a 3–manifold we refer to a foliation of codimension one. We will assume some familiarity with foliation theory; the reader may find a comprehensive treatment of the subject in [\[3; 4\]](#) and [\[9; 10\]](#).

A Reeb component is a foliation of the solid torus such that the boundary is a leaf, and there is a circle’s worth of planar leaves in the interior spiraling towards the boundary torus. We will also call some quotient of this foliation a Reeb component. These are crucial in the study of foliations on 3–manifolds by Novikov’s celebrated theorem [\[11\]](#).

It says (among other things) that if \mathcal{F} is orientable and transversely orientable, then it is Reebless if and only if every leaf L is essential (as defined in the introduction). In this sense, for Reebless foliations, the topology of a leaf is tied to the topology of the manifold.

Considering this fact, it is natural to ask if there exists some property of a foliation \mathcal{F} with Reeb components “reading” the topology of the 3–manifold. This does not seem plausible, as every 3–manifold admits a foliation with Reeb components. Furthermore, Thurston shows in [16] that every plane field on M is homotopic to the tangent space of a foliation. His construction is local in nature, so we do not care about the global topology of M . The Reeb components play a key role because holes can easily be filled using them. See [4, Section 8.5] for a detailed treatment of this construction.

However, we believe that if a compact 3–manifold M is “big enough”, then it does not admit a *uniform* foliation with Reeb components; see Section 3. In particular, Lemma 3.1 would tell us that every foliation on these manifolds must have leaves lifting to unbounded sets on the universal cover. This would be interesting because not much is known about the behavior of a foliation with Reeb components on a general 3–manifold.

Suppose that M is equipped with a Riemannian metric, and let us denote the universal cover of M by $p: \tilde{M} \rightarrow M$, equipping \tilde{M} with the pullback metric.

Definition 2.1 Let \mathcal{F} be a foliation on a compact manifold M , and denote the lifted foliation on \tilde{M} by $\tilde{\mathcal{F}}$. We will say that \mathcal{F} is *uniform* if any pair of leaves of $\tilde{\mathcal{F}}$ are at finite Hausdorff distance from each other.

This notion was introduced by Thurston in [17], who further required \mathcal{F} to be Reebless; see also [2, Section 9.3]. We stick to the definition given by S Fenley and R Potrie in [6].

2.2 Turbulization

To prove Theorem 1.2, we will rely on a method for modifying foliations along a transverse curve known as *turbulization*.

Suppose that a 3–manifold M is endowed with an oriented and cooriented foliation \mathcal{F} admitting an embedded closed curve τ transverse to the foliation. Let $\overline{N(\tau)}$ be the closure of an embedded tubular neighborhood of τ . Then the orientability and coorientability of \mathcal{F} implies that if $N(\tau)$ is small enough, $\mathcal{F}|_{\overline{N(\tau)}}$ is homeomorphic to the product foliation by disks on the solid torus $\overline{N(\tau)}$.

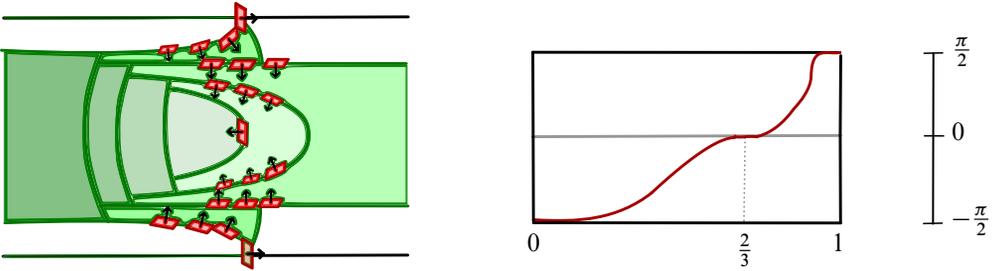


Figure 1: Left: some leaves of the foliation \mathcal{T} . Right: the function λ controlling the vector field normal to $\ker \omega$.

Fix an identification of $\overline{N(\tau)}$ with $D^2 \times S^1$ sending the leaves of $\mathcal{F}|_{\overline{N(\tau)}}$ to $D^2 \times \{\cdot\}$. We will denote points in $D^2 \times S^1$ using cylindrical coordinates, by which we mean that (r, θ, z) represents a point whose projection to D^2 has polar coordinates (r, θ) and which projects to $z \in S^1$. Notice that the plane field tangent to the foliation by disks is given by the kernel of the 1-form $\alpha_0 = dz$.

Now we will construct a new foliation \mathcal{T} on $D^2 \times S^1$ with a Reeb component on the interior and coinciding with the foliation by disks on a neighborhood of the boundary torus. This will give us a new foliation \mathcal{F}' on M , defined as \mathcal{F} in the complement of $\overline{N(\tau)}$ and as \mathcal{T} on $\overline{N(\tau)}$.

Take $\lambda: [0, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$ to be a smooth function which is a strictly increasing bijection restricted to the interval $[0, \frac{3}{4}]$, and satisfies $\lambda(\frac{2}{3}) = 0$ and $\lambda|_{[3/4, 1]} = \frac{\pi}{2}$; see Figure 1, right. Now define the 1-form

$$\omega_{(r, \theta, z)} = \cos(\lambda(r)) dr + \sin(\lambda(r)) dz.$$

This is well-defined on $D^2 \times S^1$ if $\lambda^{(k)}(0) = 0$ for every $k \geq 1$. Using Frobenius' theorem, one can check that $\pi = \ker \omega$ is an integrable plane field. The foliation tangent to π is the desired foliation \mathcal{T} depicted at the left of Figure 1.

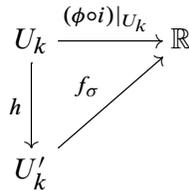
If we are careful with the choice of λ , it can be shown that the resulting foliation \mathcal{F}' on M is as regular as \mathcal{F} . The reader can find a detailed description of this construction in [3, Example 3.3.11].

2.3 General position

Let \mathcal{F} be a foliation on a 3-manifold M and $i: \Sigma \rightarrow M$ an immersion. We can always take a small perturbation of i in order to suppose that $i(\Sigma)$ is “as transverse as possible” to the foliation.

Definition 2.2 Let Σ be a closed surface and let $i : \Sigma \rightarrow M$ be an immersion into a 3-manifold M endowed with a foliation \mathcal{F} . We will say that *the immersion is in general position with respect to \mathcal{F}* if the following happens:

- (1) Except at a finite set of points $\{p_1, \dots, p_l\} \subset \Sigma$, i is transverse to \mathcal{F} .
- (2) The points $\{i(p_1), \dots, i(p_l)\}$ lie on distinct leaves of \mathcal{F} .
- (3) For every p_k and every submersion ϕ defined on a neighborhood around $i(p_k)$ locally defining the foliation¹ sending $i(p_k)$ to zero, there exists a neighborhood U_k of p_k such that $\phi \circ i|_{U_k}$ is topologically conjugated to f_c or f_s , where $f_c(x, y) = x^2 + y^2$ and $f_s(x, y) = x^2 - y^2$. By this we mean that there exists a homeomorphism between U_k and some neighborhood U'_k of $0 \in \mathbb{R}^2$ such that the diagram



commutes, where σ may be c or s .

Geometrically the last condition says that the tangencies between Σ and the foliation look like a critical point of a Morse function on Σ (up to homeomorphism); see Figure 2.

The following theorem dates back to Haefliger when the foliation \mathcal{F} is sufficiently regular. It was generalized to the case of C^0 foliations by Solodov in [14].

Theorem 2.3 *Let $i : \Sigma \rightarrow M$ be an immersion of a closed surface Σ on a 3-manifold M endowed with a oriented and cooriented foliation \mathcal{F} . Then for every $\varepsilon > 0$, there exists an immersion $j : \Sigma \rightarrow M$ in general position with respect to \mathcal{F} and ε -close to i in the C^0 -topology.*

Suppose that \mathcal{F} is oriented and cooriented and $i : \Sigma \rightarrow M$ is an immersion in general position with respect to \mathcal{F} . The first condition of Definition 2.2 tells us that \mathcal{F} induces a foliation on $\Sigma \setminus \{p_1, \dots, p_l\}$ whose leaves are intersections of leaves of \mathcal{F} and $i(\Sigma)$. The third condition gives us a model neighborhood around a singularity; more precisely, they look like saddles in the case of f_s , or centers in the case of f_c ; see Figure 2.

¹By this we mean a map $\phi : U \rightarrow \mathbb{R}$ such that the preimages of the regular values are disks contained in a leaf.

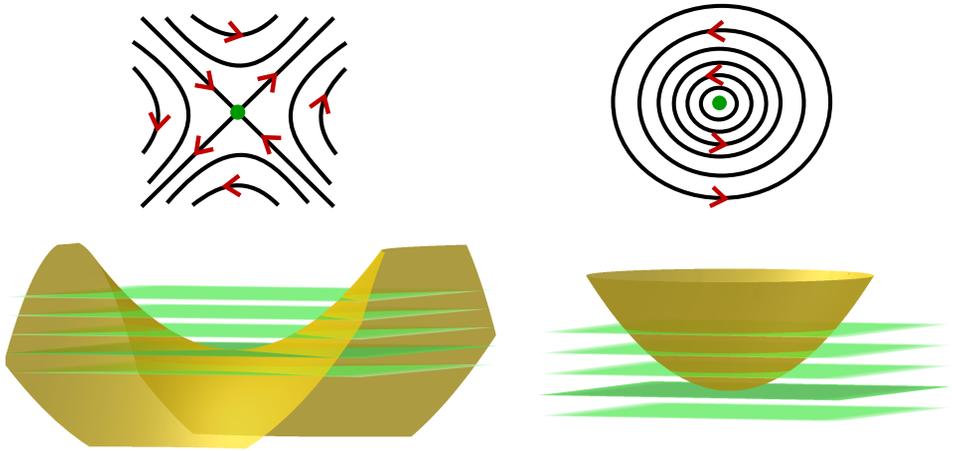


Figure 2: Bottom: tangencies of a surface in general position (in yellow) and the foliation (in green). Top: the induced singular foliation on the surface. The orientation and coorientation of \mathcal{F} induces an orientation on the singular foliation.

These local models give us what is called a *singular foliation with Morse singularities* on Σ . The orientation of \mathcal{F} induces an orientation on this singular foliation (this is an orientation outside its singular points).

The second condition tells us that a separatrix cannot join two distinct singularities; see below for a definition of separatrix.

2.4 Singular foliations of Morse type

In this section, we will fix some notation that we are going to use throughout this text. Let Σ be a surface endowed with an oriented singular foliation with Morse singularities \mathcal{G} . Suppose that p is some nonsingular point; we will denote the leaf passing through p by L_p . If L_p is noncompact (ie not a circle) then p separates L_p into two components L_p^+ and L_p^- , where L_p^+ is composed of points greater than p with respect to the order imposed by the orientation, and L_p^- of points smaller than p .

If L is some noncompact leaf, we will say that the ω -limit of the leaf is the set $\omega(L) = \bigcap_{p \in L} \overline{L_p^+}$. Analogously we define the α -limit of the leaf as $\alpha(L) = \bigcap_{p \in L} \overline{L_p^-}$. A leaf L is a *separatrix* if the α - or ω -limit of L is a singularity.

Let C be a union of singularities and separatrices S_i such that the α -limit and ω -limit of S_i are singularities in C . This set defines a directed graph with one vertex for each

singularity in C and one edge for each S_i , with the orientation induced by the foliation. We will say that C is a *closed graph* if there exists a closed path on the graph which travels through every edge only once (respecting the orientation).

We will end up studying Morse-type singular foliations on \mathbb{R}^2 . In this context, we can apply the classical Poincaré–Bendixson theorem (see for instance [12, Theorem 1.8]):

Theorem 2.4 (Poincaré–Bendixson) *Let \mathcal{G} be a Morse type singular foliation on \mathbb{R}^2 and L a leaf with compact closure. Then $\alpha(L)$ and $\omega(L)$ can be a saddle singularity, a closed leaf or a closed graph.*

2.5 Essential immersions

Let M be some closed 3–manifold. We will say that a 2–sided embedded closed surface Σ is *compressible* if there exists some embedded disk D on M such that $D \cap \Sigma = \partial D$ and ∂D is not homotopically trivial in Σ . We will call D a *compressing disk* for Σ . A 2–sided embedded surface Σ of genus $g \geq 1$ is *incompressible* if there are no compressing disks.

If a 2–sided embedded surface Σ is compressible, we can do surgery on a compressing disk D in order to obtain a simpler surface [8, Section 1.2]. This operation preserves the homology class $[\Sigma] \in H_2(M, \mathbb{Q})$. Doing finitely many surgeries on Σ , we obtain an embedded 2–sided surface, whose connected components are incompressible surfaces or spheres. The sum of the homology class of the connected components is the homology class of Σ .

The disk theorem [8, Theorem 3.1] tells us that a 2–sided embedded surface Σ is incompressible if and only if the embedding $i : \Sigma \rightarrow M$ induces an injective morphism $i_* : \pi_1(\Sigma) \rightarrow \pi_1(M)$, ie the surface is essentially embedded.

Thanks to the virtual Haken conjecture proved by Agol in [1], we can say exactly which closed 3–manifolds admit essentially embedded closed surfaces of genus $g \geq 1$:

Proposition 2.5 *A closed 3–manifold M admits an essentially immersed surface Σ of genus $g \geq 1$ if and only if some factor of the prime decomposition of M is irreducible with an infinite fundamental group.*

Proof The “if” part is a direct consequence of the virtual Haken conjecture. Suppose P is a factor of the prime decomposition, which is irreducible with an infinite fundamental

group. Then it has a finite cover \hat{P} which admits an embedded incompressible surface of genus $g \geq 1$. Projecting this surface to P and avoiding the balls of the connected sums, we obtain an essential immersion on M .

To see the “only if” part, it is enough to show that the fundamental group of the remaining 3–manifolds does not admit a subgroup isomorphic to the fundamental group of a closed surface. No factor of the prime decomposition of M is irreducible with an infinite fundamental group. Therefore, factors can be irreducible with a finite fundamental group or $S^1 \times S^2$ (prime but not irreducible). The fundamental group of such a 3–manifold must be a free product of infinite cyclic and finite groups.

If the fundamental group of a closed surface Γ were isomorphic to a subgroup of such a group, then the Kurosh subgroup theorem [13, Section 1.5.5] tells us that Γ is isomorphic to a free product of infinite cyclic groups and finite groups. However, the fundamental group of a closed surface is freely indecomposable; this means that it is not isomorphic to a free product of two nontrivial groups (this follows from Stallings’ end theorem [15, Section 4.A.6]). Therefore it would be isomorphic to a cyclic group or a finite group. Both options are impossible. \square

We remark that the last proof shows that the only closed 3–manifolds which do not admit an essential immersion are those where we can use [Theorem 1.2](#).

3 Some remarks

We begin with some general facts about uniform foliations *with Reeb components* on 3–manifolds. We will assume that M is equipped with some Riemannian metric; this induces a metric on the universal cover $p: \tilde{M} \rightarrow M$.

The following was observed by Fenley and Potrie in [6]; we include the proof for completeness.

Lemma 3.1 *Let \mathcal{F} be a uniform foliation with Reeb components on a compact manifold M . Then every lift of a leaf to the universal cover \tilde{M} has compact closure. In particular, the inclusion of a leaf $i: L \rightarrow M$ induces a morphism $i_*: \pi_1(L) \rightarrow \pi_1(M)$ with finite image.*

Proof It is enough to show that a boundary leaf of some Reeb component lifts to a compact leaf because any pair of leaves of the foliation $\tilde{\mathcal{F}}$ are at finite Hausdorff

distance from each other. So assume for the sake of contradiction that the lift of a boundary leaf of some Reeb component R is noncompact. Let $q \in R$ and $\gamma \subset R$ be a loop at q homotopic to the core of the Reeb component. The homotopy class of this curve is nontrivial; otherwise, R would lift to a compact set. Choose some lift of γ starting at $\tilde{q} \in p^{-1}(q)$, we claim that $d_{\tilde{M}}(\tilde{q}, \gamma^n(\tilde{q})) \rightarrow_n \infty$.

One way to see this is to use that the map $\alpha \in \pi_1(M) \rightarrow \alpha(\tilde{p})$ is a quasi-isometry (by the Milnor–Švarc lemma). This fact implies that the distance between a lift of a plane on the interior of the Reeb component and $\gamma^n(\tilde{q})$ goes to infinity with n , contradicting the uniform condition.

To see that the inclusion of a leaf $i: L \rightarrow M$ induces a morphism with finite image, suppose for the sake of contradiction that $\#(i_*\pi_1(L)) = \infty$. Then the cardinality of the stabilizer of a connected component \tilde{L} of $p^{-1}(L)$ is infinite. In particular, the orbit of every $p \in \tilde{L}$ is infinite. But the action is proper, so \tilde{L} escapes every compact set, contradicting the last paragraph. \square

This lemma motivates us to look for counterexamples in foliations such that every leaf of $\tilde{\mathcal{F}}$ has compact closure. The following gives us a criterion for a foliation to verify this condition.

Lemma 3.2 *Let \mathcal{F} be a foliation on a compact manifold M . Suppose that there exist compact leaves A_1, \dots, A_k such that every connected component of the complement of $\bigcup_{i=1}^k p^{-1}(A_i)$ has compact closure. Then \mathcal{F} is uniform.*

Proof Take a leaf L of $\tilde{\mathcal{F}}$ different from a lift of some A_i . Then some connected component of the complement of $\bigcup_{i=1}^k p^{-1}(A_i)$ contains L because the boundaries of these regions are composed of leaves. Therefore, the leaf has compact closure because each one of these regions is compact. \square

This remark motivated our first counterexamples to [Question 1.1](#); see [Constructions 4.2](#) and [4.10](#). However, the reader may notice that the existence of these compact leaves imposes conditions on the topology of the 3-manifold M . [Lemma 3.1](#) allows us to use compression disks on compact leaves until we get spheres on M . These spheres must bound some topology in order for us to use [Lemma 3.2](#).

The following is a consequence of this idea.

Lemma 3.3 *Let \mathcal{F} be an oriented and cooriented uniform foliation with Reeb components on a closed, irreducible 3–manifold M with an infinite fundamental group. Then every compact leaf is a torus, bounding a solid torus on one of its sides.*

Proof The inclusion i of a compact leaf L induces the zero morphism $i_*: \pi_1(L) \rightarrow \pi_1(M)$, because the fundamental group of a closed, irreducible 3–manifold is torsion-free. We can compress the surface L using the disk theorem until we get a union of spheres S . The compression operation preserves the class $[L] \in H_2(M; \mathbb{Q})$. However, S is composed of spheres and M is irreducible, therefore $[L] = [S] = 0$.

This implies that the leaf L cannot admit a closed transversal τ , because if it did then the intersection product between $[\tau] \in H_1(M; \mathbb{Q})$ and $[L] \in H_2(M; \mathbb{Q})$ would be nontrivial and $[L] \neq 0$. A theorem of Goodman (see [3, Theorem 6.3.5]) implies that L must be a torus. To see that it bounds a solid torus, do surgery on L with some compressing disk to get a sphere. This sphere bounds a ball on one side, so we obtain a solid torus bounded by L by reversing the compression operation. \square

Lemmas 3.1 and 3.3 paint a strange picture. For instance, if the universal cover is \mathbb{R}^3 then the foliation fills all the space, but every compact leaf cannot bound topology. An exceptional minimal set is not much worse because they must have compact closure. On the other hand, our main theorem, Theorem 1.3, tells us that the set of Reeb components must intersect every essentially immersed surface. Figure 3 shows a possible picture for the set of Reeb components on a uniform foliation on \mathbb{T}^3 .

Question 3.4 *If \mathcal{F} is uniform in an irreducible 3–manifold M with an infinite fundamental group, does it follow that \mathcal{F} is also Reebless?*

4 Constructions

We begin with an explicit counterexample to Question 1.1 on $S^1 \times S^2$. To do so, consider the characterization of $S^1 \times S^2$ given by the following remark.

Remark 4.1 The manifold $S^1 \times S^2$ is homeomorphic to M , which is constructed as follows. Start with the solid torus $S^1 \times D^2$ and a circle K , which is the boundary of an embedded disk D . Identify a tubular neighborhood of the knot K with $S^1 \times D^2$ in such a way that the longitude $S^1 \times \{\cdot\}$ bounds an embedded disk in the complement of the neighborhood. Drilling this tubular neighborhood, we obtain a compact 3–manifold N with two boundary components homeomorphic to $S^1 \times S^1$. Identifying both boundary components via the identity map, we obtain the closed 3–manifold M .

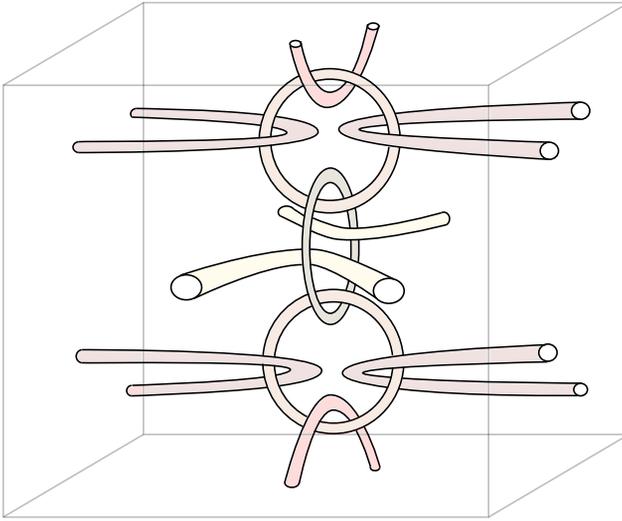


Figure 3: Can we have a uniform foliation on \mathbb{T}^3 so that the Reeb components are the ones depicted in this image?

To prove this, notice that $S^1 \times S^2$ is homeomorphic to the quotient of $\mathbb{R}^3 \setminus \{0\}$ induced by the properly discontinuous \mathbb{Z} -action generated by the map $f(x) = \frac{1}{2}x$. If S^2 is the standard sphere, the set bounded by S^2 and $f(S^2)$ is a fundamental domain. Now consider a torus S resulting from S^2 by adding a small handle. Notice that $f(S)$ is contained in the solid torus bounded by S and containing 0. Furthermore, the set N bounded by S and $f(S)$ is also a fundamental domain for the action. Therefore, $S^1 \times S^2$ is homeomorphic to N , where we identify both boundary components via the map f . This identification coincides with the identification defining M .

We will be referring to a curve K on a 3-manifold which is the boundary of an embedded disk as an *unknot*. We will always equip an unknot with a frame such that the longitude of the tubular neighborhood is the boundary of some embedded disk in the complement of the neighborhood.

The following construction is motivated by [Lemma 3.2](#).

Construction 4.2 We will construct a uniform foliation with Reeb components \mathcal{F} on $S^1 \times S^2$ which admits a transversal unknot.

Thanks to [Remark 4.1](#), it is enough to find a foliation tangential to the boundary of the manifold N obtained from $S^1 \times D^2$ by drilling a tubular neighborhood of an unknot.

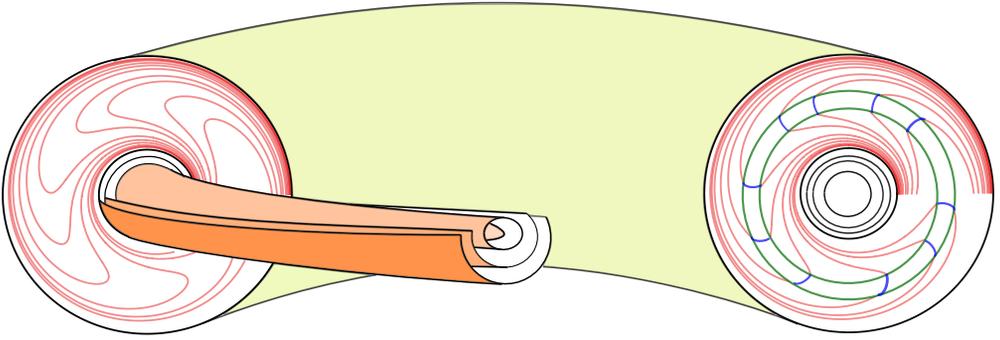


Figure 4: The foliation on N before turbulizing along τ .

To see this, notice that $S^1 \times S^2$ is obtained from N by identifying the boundary components. The foliation on N projects to a foliation of $S^1 \times S^2$ with a toric leaf S coming from the boundary components of N . We saw in [Remark 4.1](#) that the lift of S to the universal cover $\mathbb{R}^3 \setminus \{0\}$ bounds a compact fundamental domain. Therefore this foliation falls under the hypothesis of [Lemma 3.2](#). In particular, it is uniform.

We may construct a foliation tangential to the boundary on N as follows: let D^2 be the unit disk on \mathbb{R}^2 and $A \subset D^2$ be the annuli with inner and outer radii $\frac{1}{2}$ and 1, respectively. Equip A with a Reeb component and consider the product foliation induced on $S^1 \times A \subset S^1 \times D^2$. We can complete this foliation to $S^1 \times D^2$ by adding a Reeb component on $S^1 \times (D^2 \setminus A)$. Notice that the core of a Reeb annulus is transversal to the foliation; this induces a transversal τ to the foliation on the solid torus bounding a disk. Turbulizing along τ and drilling the newly generated Reeb component, we obtain the desired foliation on N ; see [Figure 4](#).

Remark 4.3 The constructed foliation can be made C^∞ . We just have to choose the Reeb annulus and the turbulization process in such a way that the resulting foliation on N has C^∞ -trivial holonomy on the boundary; see [\[3, Definition 3.4.1\]](#). The reader may check that each construction on this section can be made C^∞ . We will not pay attention to this.

Now we describe a procedure that will allow us to start with a pair of foliated 3-manifolds M and N and construct an explicit foliation on its connected sum. Furthermore, if we look at M and N with balls removed inside $M \# N$, the foliation on the connected sum coincides with the original pair of foliations outside the boundary of the drilled balls. The idea comes from the following remark.

Remark 4.4 Let M_1 and M_2 be a pair of 3-manifolds. For $i = 1, 2$, consider an unknot K_i on each M_i contained in a certain 3-ball $B_i \subset M_i$ — for example, a tubular neighborhood of the spanned disk. Drilling a tubular neighborhood of K_i contained on B_i from M_i , we obtain N_i . Identifying this tubular neighborhood with $S^1 \times D^2$ via the standard framing, we get an identification between the boundary component of N_i coming from K_i and $S^1 \times S^1$.

We claim that if we start from $N_1 \cup N_2$ and we identify the boundaries of the drilled tubular neighborhoods of K_1 and K_2 via the map $f: S^1 \times S^1 \rightarrow S^1 \times S^1$ defined as $f(x, y) = (y, x)$, we obtain a 3-manifold M homeomorphic to the connected sum $M_1 \# M_2$. To prove this claim, notice that if we start with the balls B_i , drill an unknot K_i from each of them, and sew the resulting sets along the boundary tori according to the map f , then we obtain S^3 with a pair of 3-balls removed, ie $S^2 \times [0, 1]$. Therefore $M \cong M_1 \# S^3 \# M_2$, as desired.

Lemma 4.5 *Let M_1 and M_2 be compact 3-manifolds endowed with foliations \mathcal{F}_1 and \mathcal{F}_2 . Suppose that there exists an unknot K_i in M_i , transverse to \mathcal{F}_i and contained in a ball B_i . Then there exists a foliation \mathcal{F} on $M_1 \# M_2$ admitting a transversal unknot and coinciding with \mathcal{F}_i on $M_i \setminus B_i$. Furthermore if \mathcal{F}_1 and \mathcal{F}_2 are uniform, then \mathcal{F} is also uniform.*

Proof We start by turbulizing \mathcal{F}_i along the transversal K_i . This process can be done by letting the foliation \mathcal{F}_i fixed outside the ball B_i . Drilling the newly generated Reeb component, we obtain a 3-manifold N_i equipped with a foliation tangential to the component S_i of ∂N_i coming from the Reeb component. Identifying $S_i \subset \partial M_i$ as in Remark 4.4, we obtain the desired foliation \mathcal{F} on $M_1 \# M_2$. Notice that the leaf obtained by identifying S_i is a torus S such that the inclusion $i: S \rightarrow M_1 \# M_2$ induces the zero morphism on the fundamental group. Near this leaf, there are a lot of transversal unknots resulting from the turbulization process.

Now we will see that this construction preserves the uniform condition. Notice that as \mathcal{F}_i admits a homotopically trivial transversal, it must have Reeb components by Novikov's theorem; see, for instance, [4, Theorem 9.1.4]. Therefore, Lemma 3.1 tells us that every leaf of the induced foliation $\tilde{\mathcal{F}}_i$ on the universal cover \tilde{M}_i has compact closure.

Let N_i be as above and consider the cover $q_i: \hat{N}_i \rightarrow N_i$ defined as \tilde{M}_i with the solid tori bounded by lifts of S_i drilled, and the projection $q_i = p_i|_{p_i^{-1}(N_i)}$, where $p_i: \tilde{M}_i \rightarrow M_i$

is the universal cover of M_i . We claim that if $p: \tilde{M} \rightarrow M_1 \# M_2$ is the universal cover of the connected sum, then each connected component of $p^{-1}(N_i)$ is homeomorphic to \hat{N}_i (notice that each N_i is naturally embedded in $M_1 \# M_2$).

To see this, it suffices to notice that if $j_i: N_i \subset M_i \hookrightarrow M_1 \# M_2$ is the inclusion on the connected sum, then $\ker(j_i)_*$ is isomorphic to $\pi_1(\hat{N}_i)$, by Galois correspondence. By Van Kampen's theorem $\pi_1(N_i) \cong \pi_1(M_i) \# \langle m_i \rangle$, where m_i is the homotopy class of a meridian of the drilled solid torus on M_i . The morphism $(j_i)_*$ sends each word γ of $\pi_1(N_i)$ to the word obtained by deleting all appearances of $\langle m_i \rangle$ from γ , because $i_*: \pi_1(S) \rightarrow \pi_1(M_1 \# M_2)$ is zero. Therefore, $\ker(j_i)_*$ is the normalizer of $\langle m_i \rangle$, which is exactly the image of $\pi_1(\hat{N}_i) \subset \pi_1(N_i)$.

To conclude the lemma, notice that every leaf apart from S is either contained in N_1 or N_2 . In any case, the lift of each of those leaves cannot leave a connected component of $p^{-1}(N_i)$ because its boundary is composed of leaves. The restricted foliation on this connected component is the foliation on $\hat{N}_i \subset \tilde{M}_i$. Therefore every leaf has compact closure, and the foliation is uniform. \square

Lemma 4.5 and **Construction 4.2** gives us the following.

Corollary 4.6 *Given $k \in \mathbb{N}$, there exists a uniform foliation \mathcal{F} with Reeb components on $\#_{i=1}^k S^1 \times S^2$. Furthermore, this foliation admits a transverse unknot.*

Every foliation on a compact 3-manifold with a finite fundamental group is uniform because the universal cover is compact. Furthermore, it has Reeb components by Novikov's theorem. We will see that we can construct such foliations in a way that we fall under the hypothesis of **Lemma 4.5**.

For the next construction, we will use the notion of *open book decomposition*. An open book decomposition on a closed 3-manifold M is a pair (B, π) , where B is an oriented link on M (called the *binding*) and $\pi: M \setminus B \rightarrow S^1$ is a fibration. We ask the fibers $\pi^{-1}(\cdot)$ to be the interior of compact surfaces whose boundary is B ; these are called the *pages* of the open book. There is a classical theorem (due to Alexander) saying that every closed 3-manifold has an open book decomposition; see [5] for a nice introduction to these objects.

Construction 4.7 We will construct a (uniform) foliation \mathcal{F} on every closed 3-manifold M with a finite fundamental group, such that there exists a torus transverse to \mathcal{F} which bounds a solid torus T . In particular, every meridian of ∂T is an unknot.

To see this, we will briefly recall the argument which shows that every closed 3-manifold admits a foliation. Choose an open book decomposition (B, π) on M . Let T be a tubular neighborhood of B and identify every connected component of T with $S^1 \times D^2$. With this identification, define a tubular neighborhood $T' \subset T$ whose connected components are $S^1 \times D'^2 \subset S^1 \times D^2$, where D'^2 is an open disk properly contained on D^2 .

On $M \setminus T'$ we have a foliation transverse to the boundary defined by the fibers of the fibration $\pi: M \setminus B \rightarrow S^1$. Spinning this foliation along the boundary and filling the boundary with a Reeb component (see for instance [2, Example 4.10]), we obtain a foliation on M . Notice that ∂T is composed of a torus transversal to the constructed foliation, bounding a solid torus.

Construction 4.7 and Corollary 4.6 immediately give us Theorem 1.2.

Corollary 4.8 *Given $k \in \mathbb{N}$ and a choice M_1, \dots, M_l of 3-manifolds with finite fundamental group, there exists a uniform foliation with Reeb components on*

$$\left(\#_{i=1}^k S^1 \times S^2 \right) \# \left(\#_{i=1}^l M_i \right).$$

We will finish this section by sketching an example of a uniform foliation on the solid torus inducing the trivial foliation by meridians on the boundary. To do so, we will use the following well-known lemma. A visual proof can be found in [7, Section 5.3].

Lemma 4.9 *Identify \mathbb{T}^3 as the quotient $\mathbb{R}^3 / \mathbb{Z}^3$, and let T_x, T_y and T_z be disjoint tubular neighborhoods of curves on \mathbb{T}^3 which are projections of lines parallel to the x -, y - and z -axes, respectively. Then, if L is the Borromean link on S^3 and N is a tubular neighborhood of L , there exists a homeomorphism*

$$h: \mathbb{T}^3 \setminus (T_x \cup T_y \cup T_z) \rightarrow S^3 \setminus N.$$

Furthermore, the homeomorphism takes a meridian of T_x, T_y and T_z to a longitude of a component of the Borromean ring, ie a curve on a component of N bounding a disk on S^3 .

Construction 4.10 We will construct a uniform foliation with Reeb components on the solid torus $S^1 \times D^2$, inducing a foliation by closed curves of meridional slope on the boundary.

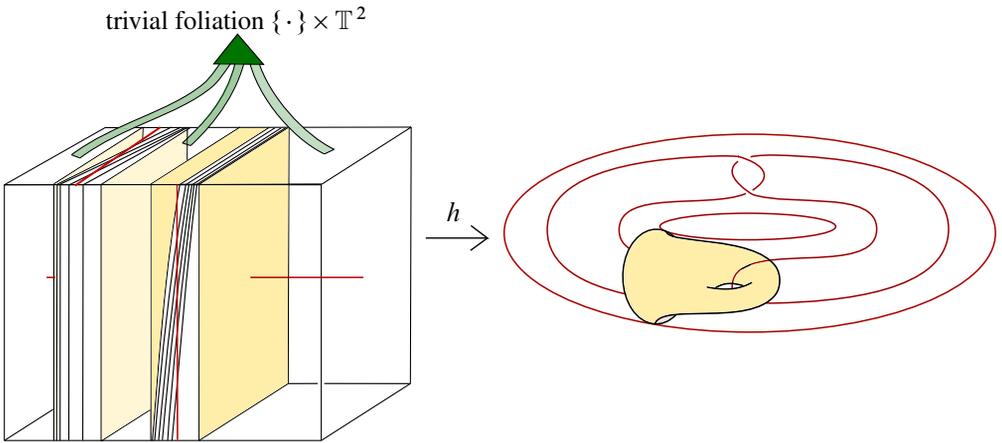


Figure 5: Left: a foliation \mathcal{F}_1 on \mathbb{T}^3 transverse to curves parallel to the x -, y - and z -axes. Right: the image of a compact leaf via h on the complement of the Borromean rings.

Start with \mathbb{T}^3 endowed with a foliation \mathcal{F}_1 , as depicted in Figure 5. This foliation is transversal to three closed curves c_x, c_y and c_z lifting to \mathbb{R}^3 as lines parallel to the x -, y - and z -axes, respectively. Take tubular neighborhoods T_x, T_y and T_z of c_x, c_y and c_z , respectively, where the foliation restricts to a trivial foliation by disks.

By Lemma 4.9, there is a homeomorphism $h: \mathbb{T}^3 \setminus (T_x \cup T_y \cup T_z) \rightarrow S^3 \setminus N$, with N a tubular neighborhood of the Borromean link. Notice that $S^3 \setminus N$ is homeomorphic to a solid torus, with the tubular neighborhood of two linked unknots drilled as to the right of Figure 5. To see this, look at the complement of the tubular neighborhood of one link component.

The homeomorphism h sends the foliation on \mathbb{T}^3 to a foliation \mathcal{F}_2 on the solid torus minus the link. This foliation is transverse to the boundary: it induces on the boundary of the solid torus the trivial foliation by meridians and the trivial foliation by longitudes around the link components. Spiraling around the link in the interior of the solid torus and then adding Reeb components, we obtain a foliation \mathcal{F} on the solid torus, which has the desired behavior on the boundary.

We claim that \mathcal{F} is uniform. To see this, take a compact leaf L of \mathcal{F}_1 in \mathbb{T}^3 intersecting only the curve c_z . Without loss of generality, $h(\partial T_z)$ is identified with the boundary of our solid torus with a link drilled. The leaf L does not separate $\mathbb{T}^3 \setminus (T_x \cup T_y \cup T_z)$. Therefore, its image $h(L)$ is a leaf of the foliation \mathcal{F} which does not separate the solid torus. A nonseparating surface on the solid torus must have a nontrivial intersection

number with the curves $S^1 \times \{\cdot\}$. This implies that when we lift the foliation to the universal cover, the lifts of this compact leaf bound a fundamental domain for $S^1 \times D^2$. Explicitly, this leaf on the solid torus looks like the yellow surface at the right of [Figure 5](#); see the proof of [Lemma 4.9](#) in [7, Section 5.3]. We conclude the uniform condition using [Lemma 3.2](#).

5 Proof of Theorem 1.3

In this section, M will be a compact 3-manifold endowed with a uniform foliation \mathcal{F} with Reeb components. We equip M with a Riemannian metric; this induces on the universal cover $p: \tilde{M} \rightarrow M$ the pullback metric. Lifting the foliation \mathcal{F} to \tilde{M} we obtain a foliation $\tilde{\mathcal{F}}$. By [Lemma 3.1](#), every leaf of this foliation has compact closure.

We will prove [Theorem 1.3](#) by contradiction. Suppose that there exists an essential immersion $i: \Sigma \rightarrow M$ from a closed surface Σ of genus $g \geq 1$ such that $i(\Sigma)$ does not intersect \mathcal{R} , the union of the Reeb components. Up to taking a finite cover, we can assume that \mathcal{F} is oriented and cooriented. Using [Theorem 2.3](#) we can assume that i is in general position with respect to the foliation. This is because there are immersions C^0 -close to i which are in general position, therefore if $i(\Sigma)$ is at a finite distance from \mathcal{R} , a sufficiently small perturbation of i will also be disjoint from \mathcal{R} . The foliation \mathcal{F} induces on Σ a singular foliation of Morse type, which will be denoted by \mathcal{G} .

Denote by $q: \mathbb{R}^2 \cong \tilde{\Sigma} \rightarrow \Sigma$ the universal cover of Σ . Lifting the immersion we obtain a map $\tilde{i}: \mathbb{R}^2 \cong \tilde{\Sigma} \rightarrow \tilde{M}$. As the immersion is essential, \tilde{i} is equivariant with respect to the actions of $\pi_1(\Sigma)$ on \mathbb{R}^2 and \tilde{M} . More precisely, for every $\gamma \in \pi_1(\Sigma)$ we have $\tilde{i} \circ \gamma = i_*(\gamma) \circ \tilde{i}$. The immersed plane $\tilde{i}(\mathbb{R}^2)$ is also in general position with respect to $\tilde{\mathcal{F}}$. In fact the induced singular foliation on \mathbb{R}^2 is exactly the lift of \mathcal{G} to the universal cover; we will denote it by $\tilde{\mathcal{G}}$. Notice that every leaf of $\tilde{\mathcal{G}}$ has compact closure because every leaf of $\tilde{\mathcal{F}}$ has compact closure.

We can classify every closed graph of $\tilde{\mathcal{G}}$:

Remark 5.1 Let S be a separatrix of $\tilde{\mathcal{G}}$ whose α -limit and ω -limit are singularities. Then $\alpha(S) = \omega(S)$. To see this, let $\{x\}$ and $\{y\}$ be the α - and ω -limit, respectively. The separatrices of \mathcal{G} cannot joint two distinct singularities by the second item in the definition of immersions in general position; therefore $y = \gamma x$ for some $\gamma \in \pi_1(\Sigma)$.

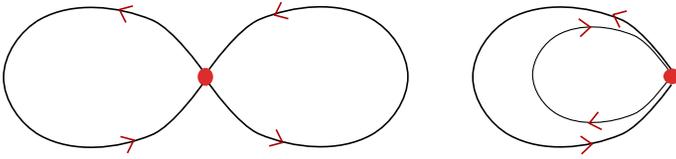


Figure 6: The possible configurations for the saddles in $\tilde{\mathcal{G}}$.

If γ is not the identity, then $q(S \cup \{x\} \cup \{\gamma x\})$ is a homotopically nontrivial curve. As the immersion is essential, this curve is sent to a homotopically nontrivial curve on M contained on certain leaf of \mathcal{F} . This is a contradiction by [Lemma 3.1](#); therefore $x = y$.

We will call a closed leaf or closed graph of $\tilde{\mathcal{G}}$ which is maximal with respect to inclusion (ie which is not strictly contained on another closed graph) a *generalized closed leaf*. The last remark tells us that a separatrix limits on at most one singularity. Therefore, if S is a separatrix of a generalized closed leaf limiting on a singularity x , then $S \cup \{x\}$ bounds a disk (by the Jordan curve theorem). This leaves us with exactly two possible configurations for a generalized closed leaf containing a singularity. Either the disk bounded by one separatrix contains the disk bounded by the other, or both disks are disjoint; both configurations are depicted in [Figure 6](#). These disks project homeomorphically to Σ , because their boundaries project homeomorphically.

Suppose that $D \subset \mathbb{R}^2$ is a disk bounded by a separatrix S of a generalized closed leaf. Then there is a well-defined notion of holonomy transport on the side of D . More precisely, take a half-open transversal τ of a point of S , with $\tau \subset D$. Start with a point x in τ and look at the next point of intersection between the leaf passing through x and τ (with respect to the orientation on \mathcal{G}). This defines a map from τ to itself. Choosing finitely many foliated neighborhoods of S and a model neighborhood of its corresponding singularity, one can see that this map is well-defined up to taking a smaller half-open transversal. Analogously, we can define the holonomy on the unbounded side of a generalized closed leaf.

Lemma 5.2 *Let \mathcal{F} be an oriented and transversely oriented uniform foliation with Reeb components on a compact 3-manifold M , and let $i: \Sigma \rightarrow M$ be an essential immersion of a surface Σ of genus ≥ 1 in general position with respect to \mathcal{F} . Suppose we further assume that $i(\Sigma)$ is disjoint from the set of Reeb components. Then every leaf of the induced singular foliation \mathcal{G} on Σ is either compact or a separatrix such that its α - and ω -limits coincide.*

Proof First, we prove that no generalized closed leaf of $\tilde{\mathcal{G}}$ has holonomy on any of its sides. Let C be a generalized closed leaf. If we had nontrivial holonomy on a certain side of C , then the curve $q(C) \subset \Sigma$ would be sent via i to a homotopically nontrivial curve on a leaf of \mathcal{F} . To see this, if $q(C)$ were sent to a homotopically trivial curve on a leaf of \mathcal{F} then it would have trivial holonomy on \mathcal{F} , and therefore trivial holonomy on $\tilde{\mathcal{G}}$.

Now we can use standard Haefliger-type arguments: by [Remark 5.1](#), $q(C)$ bounds one or two disks which are sent via i to immersed disks on M . The curve $i(q(C))$ is not homotopically trivial on its corresponding leaf of \mathcal{F} ; therefore, the boundary of at least one of the immersed disks is not homotopically trivial. This implies the existence of a vanishing cycle on at least one of the immersed disks; see for instance [\[4, Lemmas 9.2.2 and 9.2.4\]](#). However, Novikov's theorem says that if a leaf admits a vanishing cycle, then this leaf is the boundary of a Reeb component; see [\[4, Theorem 9.4.1\]](#). Therefore, $i(\Sigma)$ intersects the set of Reeb components, which is a contradiction.

Now we prove that either every leaf of \mathcal{G} is compact, or a separatrix limiting on a single singularity. It is enough to see this on $\tilde{\mathcal{G}}$. Suppose that L is a separatrix of $\tilde{\mathcal{G}}$ with $\alpha(S) \neq \omega(S)$, or a noncompact leaf. In any case, $\omega(L)$ or $\alpha(L)$ must be a generalized closed leaf or a closed leaf, by [Remark 5.1](#) and the Poincaré–Bendixson theorem. This generalized closed leaf must have nontrivial holonomy on one of its sides, which is a contradiction by the last paragraphs. \square

It is not hard to see that a singular foliation \mathcal{G} on Σ without holonomy must have a generalized closed leaf which is homotopically nontrivial:

Lemma 5.3 *Let Σ be a closed Riemannian surface of genus $g \geq 1$ and \mathcal{G} an oriented singular foliation of Morse type. Furthermore, suppose that every leaf is compact or a separatrix with the same α - and ω -limit. Then some leaf of the lifted singular foliation $\tilde{\mathcal{G}}$ on $\mathbb{R}^2 \cong \tilde{\Sigma}$ is unbounded.*

Proof Assume for the sake of contradiction that every leaf of $\tilde{\mathcal{G}}$ is bounded, and therefore has compact closure. Let \mathcal{C} be the set of generalized closed leaves of $\tilde{\mathcal{G}}$, and define the function $A: \mathcal{C} \rightarrow \mathbb{R}_{\geq 0}$ such that $A(C)$ is the area of the bounded component of the complement of $C \in \mathcal{C}$. Each connected component of the bounded component of an element of \mathcal{C} is a disk that projects homeomorphically to Σ , therefore its area cannot be greater than the area of Σ . This shows that the function A is bounded.

We claim that we can achieve the supremum $\sup_{C \in \mathcal{C}} A(C)$. Choose $x_n \in \mathbb{R}^2$ such that there exists a generalized closed leaf C_n with $x_n \in C_n$ and $A(C_n) \rightarrow \sup_{C \in \mathcal{C}} A(C)$ an increasing sequence. As the surface Σ is compact, up to translating $\{x_n\}$ by $\pi_1(\Sigma)$ we can assume that the sequence $\{x_n\}_{n \in \mathbb{N}}$ converges to $x \in \mathbb{R}^2$.

Let C be the generalized closed leaf passing through x ; we claim that $A(C)$ is the desired supremum. To see this, notice that if $m > n$ then $A(C_n) < A(C_m)$ and therefore, C_n is contained on the bounded component of C_m . This implies that for n sufficiently big, C_n is contained on the bounded component of C . Otherwise, there would exist some C_k bounding a region containing C_n with $n > k$, which is a contradiction.

Now choose C as in the last paragraph. This leaf cannot have holonomy because every leaf is compact. Therefore we can find a foliated neighborhood of C composed of closed orbits. In particular, some closed leaf bounding C exists, which contradicts the fact that C bounds the largest area. □

Therefore, Lemma 5.2 tells us that the induced foliation on the surface Σ satisfies the hypothesis of Lemma 5.3. However, this lemma contradicts the uniformity of the foliation \mathcal{F} . This contradiction concludes the proof of Theorem 1.3.

Remark 5.4 Suppose that \mathcal{F} is an oriented and transversely oriented uniform foliation with Reeb components on a compact 3-manifold M , and let $i: \Sigma \rightarrow M$ be an essential immersion in general position as before. We just proved that $i(\Sigma)$ must intersect the

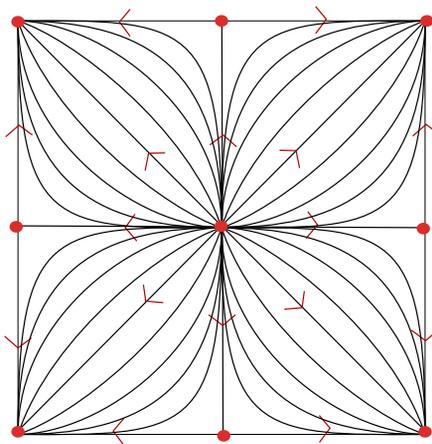


Figure 7: Turbulizing around the attracting/repelling fixed points we obtain a uniform foliation on \mathbb{T}^2 .

set of Reeb components. Thus, the induced singular foliation on Σ is uniform and necessarily contains vanishing cycles.

There exist singular foliations on surfaces satisfying these conditions; see for example [Figure 7.2](#). It is not obvious how the singular foliation induced on Σ would help in answering [Question 3.4](#). We believe that an extra argument is needed.

References

- [1] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 [MR](#) [Zbl](#)
- [2] **D Calegari**, *Foliations and the geometry of 3–manifolds*, Oxford Univ. Press (2007) [MR](#) [Zbl](#)
- [3] **A Candel**, **L Conlon**, *Foliations, I*, Graduate Studies in Mathematics 23, Amer. Math. Soc., Providence, RI (2000) [MR](#) [Zbl](#)
- [4] **A Candel**, **L Conlon**, *Foliations, II*, Graduate Studies in Mathematics 60, Amer. Math. Soc., Providence, RI (2003) [MR](#) [Zbl](#)
- [5] **JB Etnyre**, *Lectures on open book decompositions and contact structures*, from “Floer homology, gauge theory, and low-dimensional topology” (D A Ellwood, P S Ozsváth, A I Stipsicz, Z Szabó, editors), Clay Math. Proc. 5, Amer. Math. Soc., Providence, RI (2006) 103–141 [MR](#) [Zbl](#)
- [6] **S R Fenley**, **R Potrie**, *Minimality of the action on the universal circle of uniform foliations*, Groups Geom. Dyn. 15 (2021) 1489–1521 [MR](#) [Zbl](#)
- [7] **D Hardorp**, *All compact orientable three dimensional manifolds admit total foliations*, Mem. Amer. Math. Soc. 233, Amer. Math. Soc., Providence, RI (1980) [MR](#) [Zbl](#)
- [8] **A E Hatcher**, *Notes on basic 3–manifold topology*, course notes (2007) Available at <http://pi.math.cornell.edu/~hatcher/3M/3Mfds.pdf>
- [9] **G Hector**, **U Hirsch**, *Introduction to the geometry of foliations, A: Foliations on compact surfaces, fundamentals for arbitrary codimension, and holonomy*, Aspects of Mathematics 1, Vieweg & Sohn, Braunschweig (1981) [MR](#) [Zbl](#)
- [10] **G Hector**, **U Hirsch**, *Introduction to the geometry of foliations, B: Foliations of codimension one*, 2nd edition, Aspects of Mathematics E3, Vieweg & Sohn, Braunschweig (1987) [MR](#) [Zbl](#)
- [11] **S P Novikov**, *The topology of foliations*, Trudy Moskov. Mat. Obšč. 14 (1965) 248–278 [MR](#) [Zbl](#) In Russian; translated in Trans. Mosc. Math. Soc. 14 (1965) 268–304
- [12] **J Palis, Jr**, **W de Melo**, *Geometric theory of dynamical systems: an introduction*, Springer (1982) [MR](#) [Zbl](#)

²We thank Sergio Fenley for pointing out this example.

- [13] **J-P Serre**, *Trees*, Springer (1980) [MR](#) [Zbl](#)
- [14] **V V Solodov**, *Components of topological foliations*, *Mat. Sb.* 119(161) (1982) 340–354, 447 [MR](#) [Zbl](#) In Russian; translated in *Math. USSR Sb.* 47 (1984) 329–343
- [15] **J Stallings**, *Group theory and three-dimensional manifolds*, Yale Mathematical Monographs 4, Yale Univ. Press, New Haven, CT (1971) [MR](#) [Zbl](#)
- [16] **W P Thurston**, *Existence of codimension-one foliations*, *Ann. of Math.* 104 (1976) 249–268 [MR](#) [Zbl](#)
- [17] **W P Thurston**, *Three-manifolds, foliations and circles, I*, preprint (1997) [arXiv math/9712268](#)

Centro de Matemática, Udelar
Montevideo, Uruguay

joalema@cmat.edu.uy

Received: 8 December 2021 Revised: 25 April 2022

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the [AGT website](#). You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in \LaTeX , preferably using the journal's class file. More information on preparing articles in \LaTeX for publication in AGT is available on the [AGT website](#).

arXiv papers

If your paper has previously been deposited on the arXiv, we will need its arXiv number at acceptance time. This allows us to deposit the DOI of the published version on the paper's arXiv page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of Bib \TeX 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@msh.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 23

Issue 9 (pages 3909–4400)

2023

Two-dimensional extended homotopy field theories	3909
KÜRŞAT SÖZER	
Efficient multisections of odd-dimensional tori	3997
THOMAS KINDRED	
Bigrading the symplectic Khovanov cohomology	4057
ZHECHI CHENG	
Fibrations of 3–manifolds and asymptotic translation length in the arc complex	4087
BALÁZS STRENNER	
A uniformizable spherical CR structure on a two-cusped hyperbolic 3–manifold	4143
YUEPING JIANG, JIEYAN WANG and BAOHUA XIE	
A connection between cut locus, Thom space and Morse–Bott functions	4185
SOMNATH BASU and SACHCHIDANAND PRASAD	
Staircase symmetries in Hirzebruch surfaces	4235
NICKI MAGILL and DUSA MCDUFF	
Geometric triangulations of a family of hyperbolic 3–braids	4309
BARBARA NIMERSHIEM	
Beta families arising from a v_2^9 self-map on $S/(3, v_1^8)$	4349
EVA BELMONT and KATSUMI SHIMOMURA	
Uniform foliations with Reeb components	4379
JOAQUÍN LEMA	